SCIENTIFIC COMMITTEE
TWENTIETH REGULAR SESSION

Manila, Philippines
14–21 August 2024

Retrospective Forecasting of the 2022 WCPO Bigeye Tuna and Yellowfin Tuna Stock Assessment.

WCPFC-SC20-2024/MI-IP-03
1st August 2024

N. Yao[1], R. Scott, and F. Scott

[1]Oceanic Fisheries Programme, The Pacific Community

Contents

Executive Summary

Retrospective forecasting, also known as hindcasting or backtesting, is a method for evaluating the prediction performance of an assessment model using historical data. This approach involves a retrospective analysis with the additional step of projecting each assessment through to the end of the original time series.

To support the workplan for the adoption of harvest strategies for the Western and Central Pacific Ocean (WCPO) skipjack, bigeye, yellowfin, and South Pacific albacore tuna, we evaluated the 2023 bigeye and yellowfin tuna assessment models using retrospective forecasting. This evaluation aimed to determine the robustness of the stock assessments with varying quantities of data and to assess the quality of the projections in terms of their ability to provide consistent estimates of stock status.

Our results indicate no significant retrospective bias in the assessment of both species. The bigeye projection performs well within 5 peels from the terminal year, while the yellowfin projection performs well within 4 peels from the terminal year, supporting a management period of 3 years based on projections. The models' ability to predict catch per unit effort (CPUE) varied by region, with certain regions showing more accurate predictions (regions 1, 2, and 8 for bigeye; regions 3 and 5 for yellowfin).

Therefore, we conclude:

- The 2023 bigeye tuna and yellowfin tuna stock assessment model is not subject to significant retrospective bias.
- Short-term catch-based projections from the bigeye and yellowfin assessment model grid can provide consistent and accurate indications of stock status in the short term (less than 4 years), although with a tendency towards slight underestimation.
- When using projected CPUE, certain regions should be prioritized based on the hindcast results.

We invited the SC to:

- Acknowledge the ongoing evaluation of the 2023 bigeye and yellowfin tuna stock assessments for the future development of the Management Strategy Evaluation (MSE).
- Recognise that the results of our work support a management period of 3 years for both bigeye and yellowfin tuna stocks.
- Note the advice on the future usage of the projected CPUE time series.

1 Introduction

Operating models (OMs) are fundamental to the harvest strategy approach, representing the "true" state of stock and fisheries. They simulate future stock status and fisheries data, which can be used to evaluate the performance of management procedures (MPs). A good set of OMs should accurately reflect the future dynamics of fisheries while incorporating critical uncertainties. Within the Western and Central Pacific Fisheries Commission (WCPFC), OMs for harvest strategy approaches are often derived from the stock assessment. Therefore, a thorough examination of the stock assessments has become crucial for adopting them as the basis for OMs. Notably, models that appear adequate for estimating current stock status may not necessarily excel at forecasting future stock status. In the context of management strategy evaluation (MSE), this oversight can lead to the selection of inappropriate MPs. Therefore, assessing the suitability of a set of OMs requires examining both model diagnostics and predictive skills.

Retrospective analysis is an essential tool in stock assessment. It systematically examines the impact of sequentially removing historical data on model performance, providing insights into model reliability and impacts on management actions. While stock assessments are not expected to be entirely accurate, they should provide consistent estimates over time without persistent trends of under- or over-estimation. Persistent trends in updated parameter estimates suggest potential model misspecification, commonly referred to as retrospective bias (Sinclair et al., 1991). Therefore, retrospective analysis is conducted for each new assessment, refitting the final model to progressively truncated data series, iteratively moving the terminal years in backwards steps. Mohn (1999) proposed Mohn's ρ, a metric to measure retrospective bias and to facilitate comparisons of retrospective bias between different assessments.

Hindcast analysis is commonly employed to evaluate the prediction skill of simulation frameworks (Walters and Punt, 1994; Kell and Mosqueira, 2017). This method involves projecting quantities from retrospective models and comparing the simulated values to the observed data. In recent years, hindcast analysis has been widely applied in fisheries science to understand the implications of projections from assessments with systematic bias (Cadigan and Farrell, 2005; Brooks and Legault, 2016) and for model validation (Hurtado-Ferro et al., 2015; Kell et al., 2021).

In this paper, we use retrospective and hindcast approaches to examine the 2023 assessments of bigeye tuna (*Thunnus obesus*) and yellowfin tuna (*Thunnus albacares*). First, we conduct a retrospective analysis to evaluate the performance of the bigeye and yellowfin assessment models. Then, we apply the hindcast approach to project depletion (SBrecent/SBF=0) and catch per unit effort (CPUE) time-series, comparing them to corresponding observations to evaluate each model's prediction skill. Finally, we make recommendations for selecting models capable of projecting plausible future conditions within the management strategy evaluation (MSE) framework.

2 Methods

2.1 Retrospective Analysis

The 2023 stock assessment for bigeye and yellowfin tuna in the WCPO is a size-based, age- and spatially-structured population integrated model. The assessment model partitions the bigeye population into nine spatial regions (Figure.7a), while the yellowfin population is divided into five regions (Figure.7b). One index of relative abundance, referred to as an "index fishery", is introduced for each region to provide standardised catch per unit effort (CPUE) indices of abundance. To address the uncertainty in key model assumptions, both bigeye and yellowfin assessments contain 54 models in the grid.

For each of these 54 full time-series assessment models, a retrospective analysis was performed by sequentially removing all observations for a year from the terminal year backwards (i.e., peeled). The model was then refitted to the truncated series, and the estimated depletion values were compared to the corresponding values from the full stock assessment to identify any systematic patterns. This procedure was repeated for seven peels for each model. Retrospective bias in the terminal estimates of depletion (Mohn's ρ) was calculated in accordance with Equation. 1.

$$\rho = \frac{1}{r}\sum_{i=1}^{r}\frac{\theta_{(Y-1)^*} - \theta_{(Y-1)}}{\theta_{(Y-1)}} \tag{1}$$

Where r is the number of years over which the retrospective analysis is conducted, $\theta_{(Y-1)^*}$ is the depletion estimated from each retrospective model and $\theta_{(Y-1)}$ is the corresponding value of depletion from the stock assessment model using the full time-series of input data (Mohn, 1999). The larger the value of ρ, the greater the retrospective bias. A positive ρ indicates a tendency for the retrospective model to overestimate depletion in the final year compared to the same year of the full assessment, while a negative ρ indicates a tendency for underestimation.

To understand which variable(s) from the model assumptions significantly contributed to the retrospective bias, a fixed-effects analysis of variance (ANOVA) was used to evaluate the proportionate contributions of each variable to the values of Mohn's ρ.

2.2 Hindcast Analysis

The retrospective models described above served as the starting point for the hindcasting step. Each model was projected to 2021 (the terminal year of the stock assessment) based on observed catch data. No observation error was added during the projection. Two recruitment assumptions were considered: the first assumption fixed the recruitment at the assessment estimated level; for the second assumption, stochastic recruitment was applied, with deviations from the predictions of the stock-recruitment relationship in each projection run. For the fixed recruitment scenario, one iteration was run for each model per peel. For the stochastic scenarios, 100 iterations were run for each model per peel.

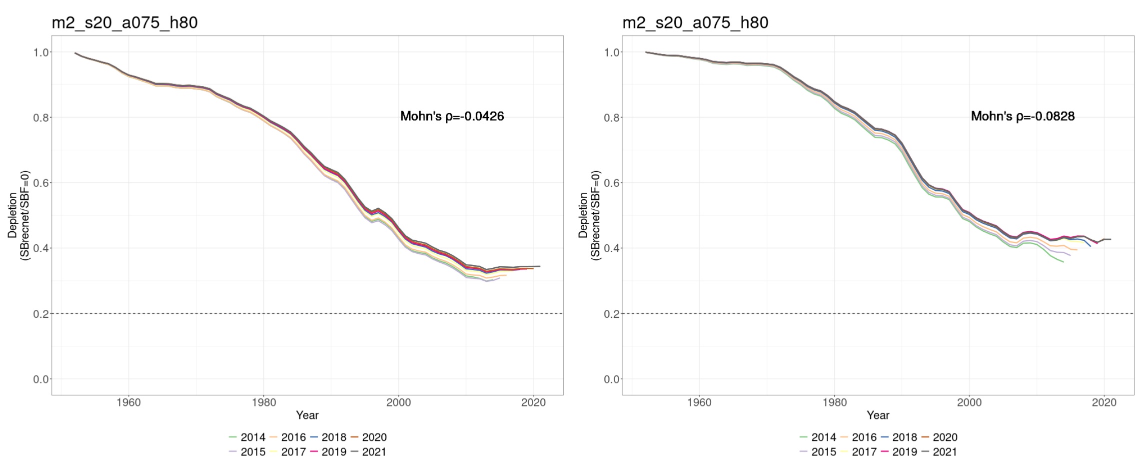
The predictability of each model was evaluated based on the accuracy of the simulated depletion (SBrecent/SBF=0) and the CPUE of each model region. Accuracy was measured using the mean absolute scaled error(MASE) with Equation.2, which scales forecast errors by the in-sample mean absolute error of a naive forecasting method, usually the "no-change" forecast (where the forecast for each period is simply the value of the previous period). A MASE value less than 1 indicates that the forecast method outperforms the naive method, while a value greater than 1 suggests that the naive method performs better than the forecast method.

$$\text{MASE} = \frac{\frac{1}{n}\sum_{t=1}^{n}|\hat{y}_t - y_t|}{\frac{1}{n-1}\sum_{t=2}^{n}|y_t - y_{t-1}|} \tag{2}$$

3 Results

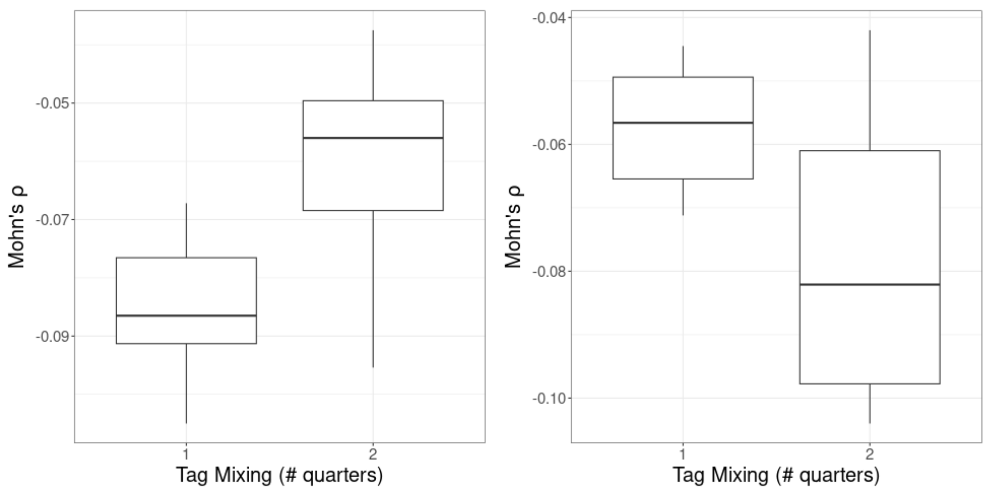
3.1 Retrospective Analysis

The estimated depletion (SBrecent/SBF=0) from each retrospective model exhibits similar trends across the time-series, but is rescaled relative to the full assessment model for both bigeye and yellowfin tuna (e.g., Figure.1a). Systematic underestimation with successive refitted models was observed for both species. However, the retrospective bias (Mohn's ρ) ranges from -0.0375 (minimum bias) to -0.105 (maximum bias) for bigeye, and -0.042 (minimum bias) to -0.104 (maximum bias) for yellowfin (Table. 1), indicating almost no systematic bias over the retrospective period considered (2014 to 2020). The ANOVA results indicate that the tag mixing assumption is the most significant factor impacting the retrospective bias (Mohn's ρ) for both bigeye and yellowfin (Figure.2).



(a) (b)

Figure 1: Two examples of the retrospective results. The annual estimates of depletion determined from the full time-series assessment (gray line) and retrospective assessment runs (terminal years 2014 to 2020, color lines). (a) Bigeye diagnostic model. (b) Yellowfin diagnostic model.

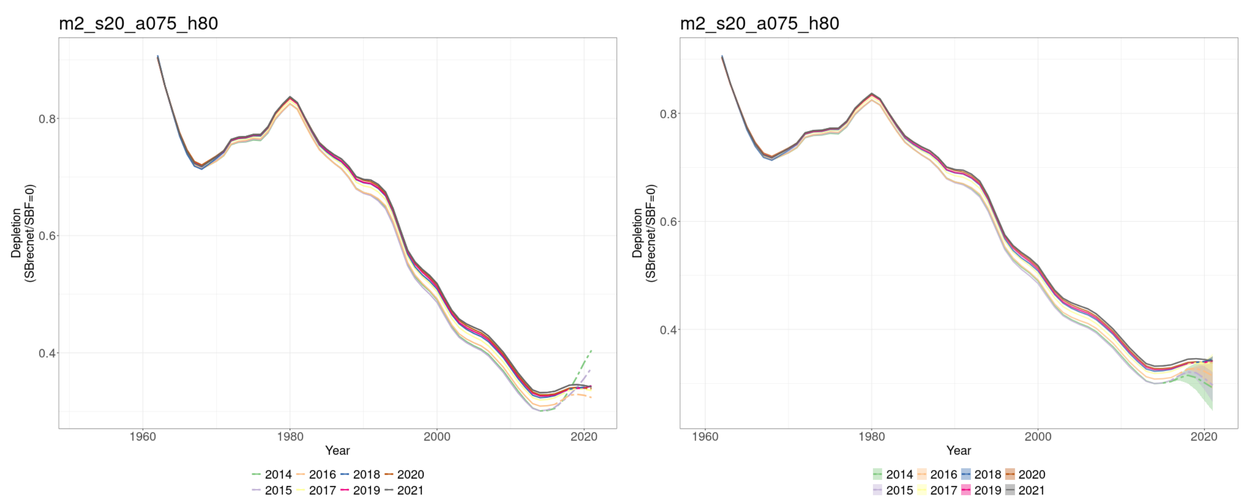


(a) (b)

Figure 2: ANOVA results of the tag mixing assumptions. (a) Bigeye. (b) Yellowfin.

3.2 Hindcast Analysis

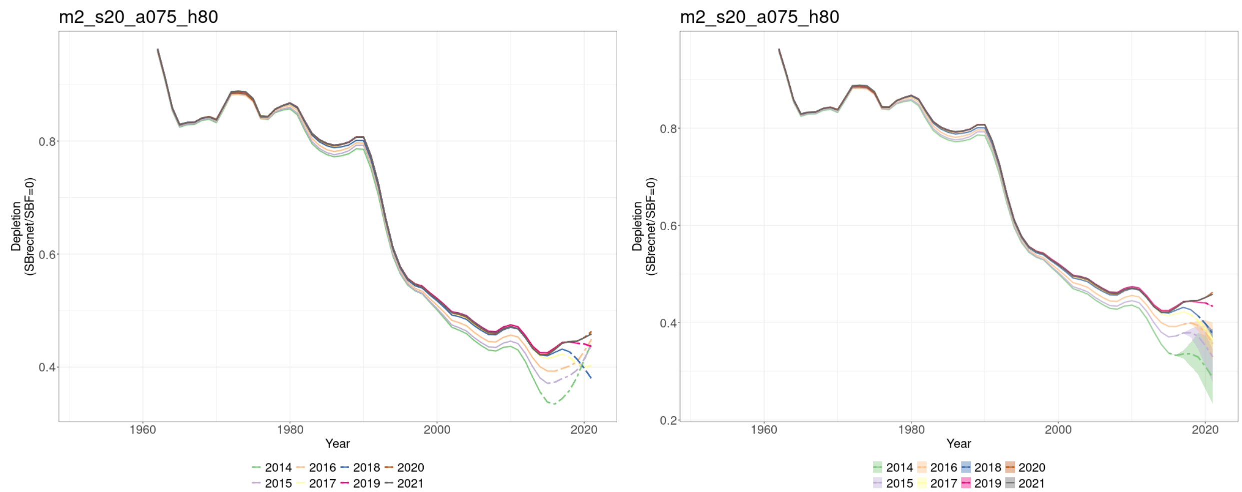
For bigeye tuna, the deterministic catch-based projection displayed a similar depletion(SBrecent/SBF=0) trend to the full time-series for the first five peels (2016 to 2020). Significantly different trends were observed from the results of the last two peels (2014 to 2015; Figure.3a). The stochastic catch-based projections showed variability in estimates of projected depletion, with significant variabilities at the end of the time-series observed from the results of the 2014 to 2016 runs (Figure.3b).



(a) (b)

Figure 3: Hindcast results of the bigeye diagnostic case for depletion (SBrecent/SBF=0). The annual estimates of depletion determined from the full time-series assessment (gray line) and hindcast runs (starting years 2014 to 2020, color lines). (a) Deterministic recruitment assumption. (b) Stochastic recruitment assumption.

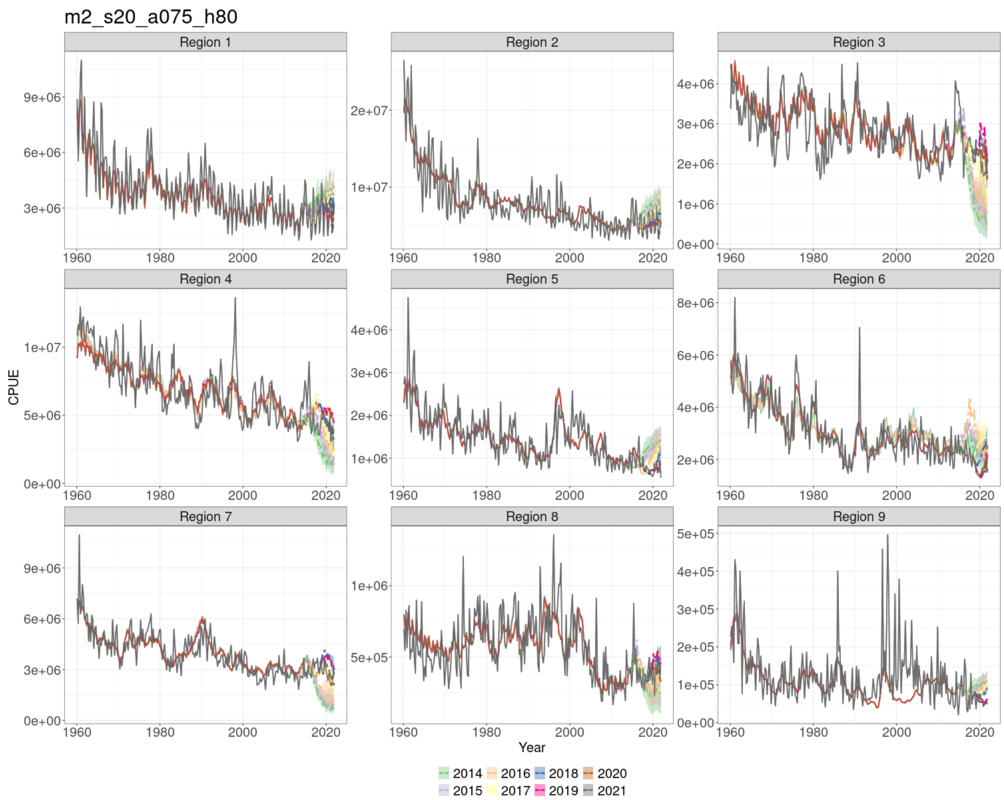
For yellowfin tuna, the estimated depletion in 2021 from deterministic catch-based projections closely matched the value of the stock assessment model, except for two runs starting from years 2017 and 2018 (Figure.4a). In contrast, the stochastic projections exhibited wider variability in the results from last three peels (2014 to 2016;Figure.4b).



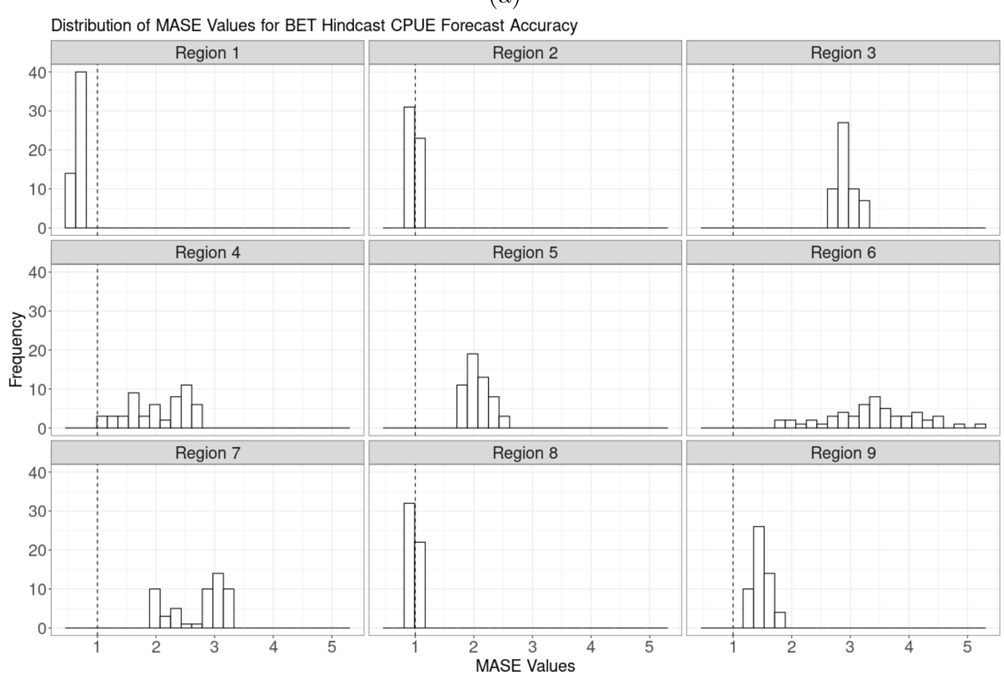
(a) (b)

Figure 4: Hindcast results of the yellowfin diagnostic case for depletion (SBrecent/SBF=0). The annual estimates of depletion determined from the full time-series assessment (gray line) and hindcast runs (starting years 2014 to 2020, color lines). (a) Deterministic recruitment assumption. (b) Stochastic recruitment assumption.

It was evidence that the models' ability to predict CPUE also varies across different regions, as confirmed by the MASE values (Figure.5,6). Using a cut-off value of 1, index fisheries in regions 1, 2, and 8 demonstrated more accurate predictions compared to the others for bigeye. For yellowfin, index fisheries in regions 3 and 5 performed better in predicting CPUE than the rest.

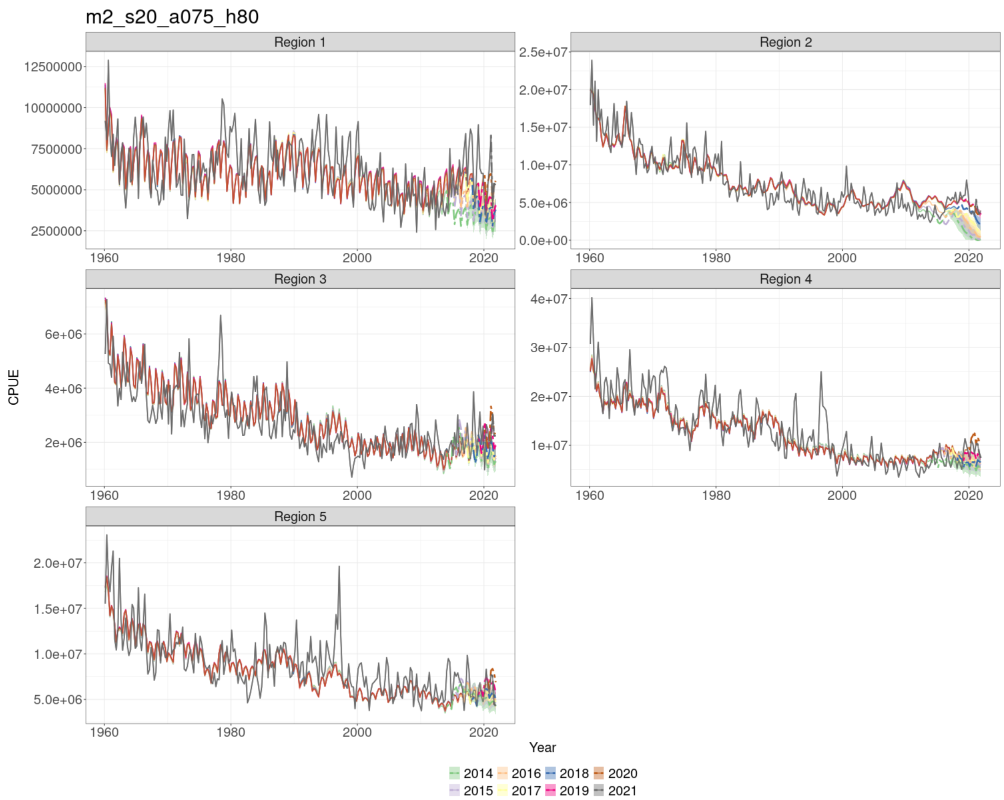


(a)

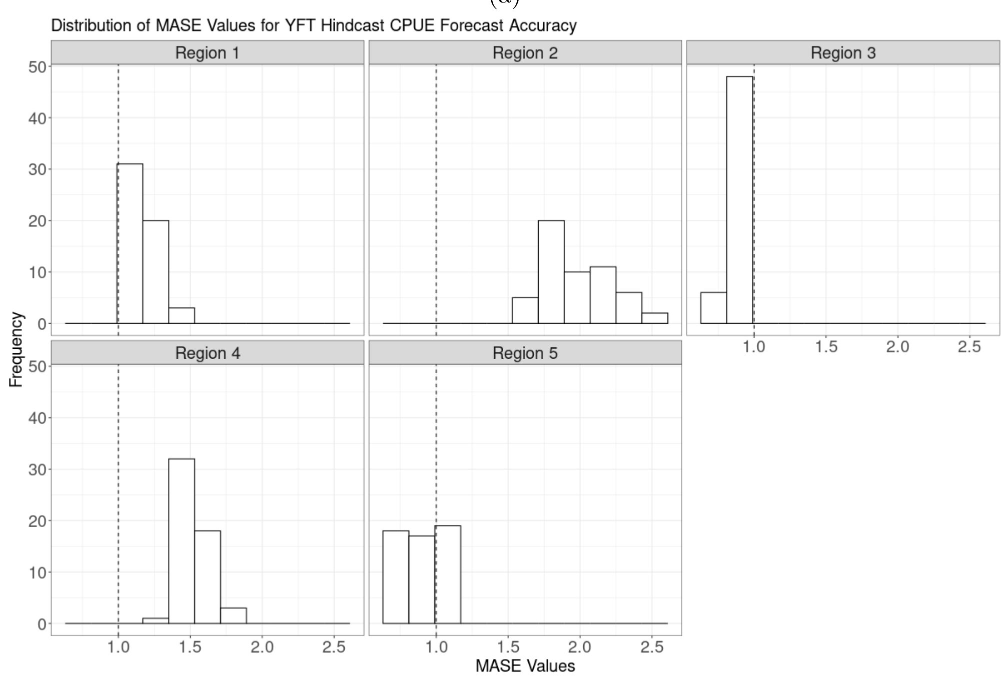


(b)

Figure 5: Hindcast results of the bigeye diagnostic case on CPUE. The annual estimates of CPUE determined from the full time-series assessment (gray line) and hindcast runs (starting years 2014 to 2020, color lines). (a) Stochastic recruitment assumption. (b) Histogram of the MASE values of each model calculated by Equation. 2.



(a)



(b)

Figure 6: Hindcast results of the yellowfin diagnostic case on CPUE. The annual estimates of CPUE determined from the full time-series assessment (gray line) and hindcast runs (starting years 2014 to 2020, color lines). (a) Stochastic recruitment assumption. (b) Histogram of the MASE values of each model calculated by Equation. 2.

4 Discussion

Retrospective analysis are valuable tools for understanding the robustness of model estimates, while hindcast analysis evaluates a model's abilities to predict future quantities. Through these analyses, the MSE approach can be used more confidently to simulate future stock conditions and evaluate the potential outcomes of different management actions, leading to more sustainable fisheries management.

The results of the retrospective analysis showed consistent underestimations of depletion for both bigeye and yellowfin tuna. This indicates that data from the last couple of years influence model estimations. Stronger deviations in estimated depletion values were observed from retrospective models with more peels. This is expected, as excluding more data from the model decreases the likelihood of consistent estimations. However, according to the rule of thumb proposed by Hurtado-Ferro et al. (2015), Mohn's ρ values did not indicate any significant issues.

While the catch-based projections on depletion showed some differences, the overall trends remained consistent for model runs within 4 peels (back to 2017) for both species. The performance of both models significantly decreased with 6 and 7 peels (2014 to 2015), resulting in larger variations towards the end. This variation is primarily due to recruitment fluctuations and is significant enough to impact management decisions in some cases. Therefore, it is advisable to base decisions on projections of less than 4 years, supporting the recommendation of a management period of 3 years.

CPUE projections revealed that the models' performance varies by region, likely due to differences in regional recruitment distribution and population structure in the terminal year. Based on the current model, developing management decisions in MSE framework using predicted CPUE data from regions 1, 2, and 8 for bigeye tuna, and from fisheries 3 and 5 for yellowfin tuna, could result in better performance.

5 Conclusion

Our results suggested the level or retrospective bias was acceptable for both bigeye and yellowfin assessment, although consistent underestimation of depletion was revealed by the retrospective analysis. The hindcast projections highlighted the variability introduced by different recruitment assumptions. Projections with retrospective models with fewer peels generally perform better than models with more peels. Based on the results, the bigeye assessment performs well within 5 peels while the yellowfin assessment performs well within 4 peels, supporting a management period of 3 years based on projections.

The performance of CPUE projections varied by region. Consequently, basing management decisions on CPUE data from specific regions (1, 2, and 8 for bigeye and 3 and 5 for yellowfin) is recommended.

Therefore, we conclude:

- The 2023 bigeye tuna and yellowfin tuna stock assessment model is not subject to significant retrospective bias.
- Short-term catch-based projections from the bigeye and yellowfin assessment model grid can provide consistent and accurate indications of stock status in the short term (less than 4 years), although with a tendency towards slight underestimation.
- When using projected CPUE, certain regions should be prioritized based on the hindcast results.

We invited WCPFC to:

- Acknowledge the ongoing evaluation of the 2023 bigeye and yellowfin tuna stock assessments for the future development of the Management Strategy Evaluation (MSE).
- Recognise that the results of our work support a management period of 3 years for both bigeye and yellowfin tuna stocks.
- Note the advice on the future usage of the projected CPUE time series.

Acknowledgments

We gratefully acknowledge funding for this work from the New Zealand Ministry of Foreign Affairs and Trade (MFAT) funded project “Pacific Tuna Management Strategy Evaluation”.

6 References

Brooks, E. N. and Legault, C. M. (2016). Retrospective forecasting—evaluating performance of stock projections for New England groundfish stocks. *Canadian Journal of Fisheries and Aquatic Sciences*, 73(6):935–950. Publisher: NRC Research Press.

Cadigan, N. G. and Farrell, P. J. (2005). Local influence diagnostics for the retrospective problem in sequential population analysis. *ICES Journal of Marine Science: Journal du Conseil*, 62(2):256–265. _eprint: http://icesjms.oxfordjournals.org/content/62/2/256.full.pdf+html.

Hurtado-Ferro, F., Szuwalski, C. S., Valero, J. L., Anderson, S. C., Cunningham, C. J., Johnson, K. F., Licandeo, R., McGilliard, C. R., Monnahan, C. C., Muradian, M. L., and others (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. Publisher: ICES/CIEM.

Kell, L. and Mosqueira, I. (2017). Conditioning operating models on data and knowledge and rejecting and weighting of hypotheses. *Collec. Vol. Sci. Pap. ICCAT*, (SCRS/2016/014):3000–3008.

Kell, L. T., Sharma, R., Kitakado, T., Winker, H., Mosqueira, I., Cardinale, M., and Fu, D. (2021). Validation of stock assessment methods: is it me or my model talking? *ICES Journal of Marine Science*, page fsab104.

Mohn, R. (1999). The Retrospective problem in sequential population Analyses: An investigation using cod fishery and simulated data. *ICES Journal of Marine Science*, 56:473–488.

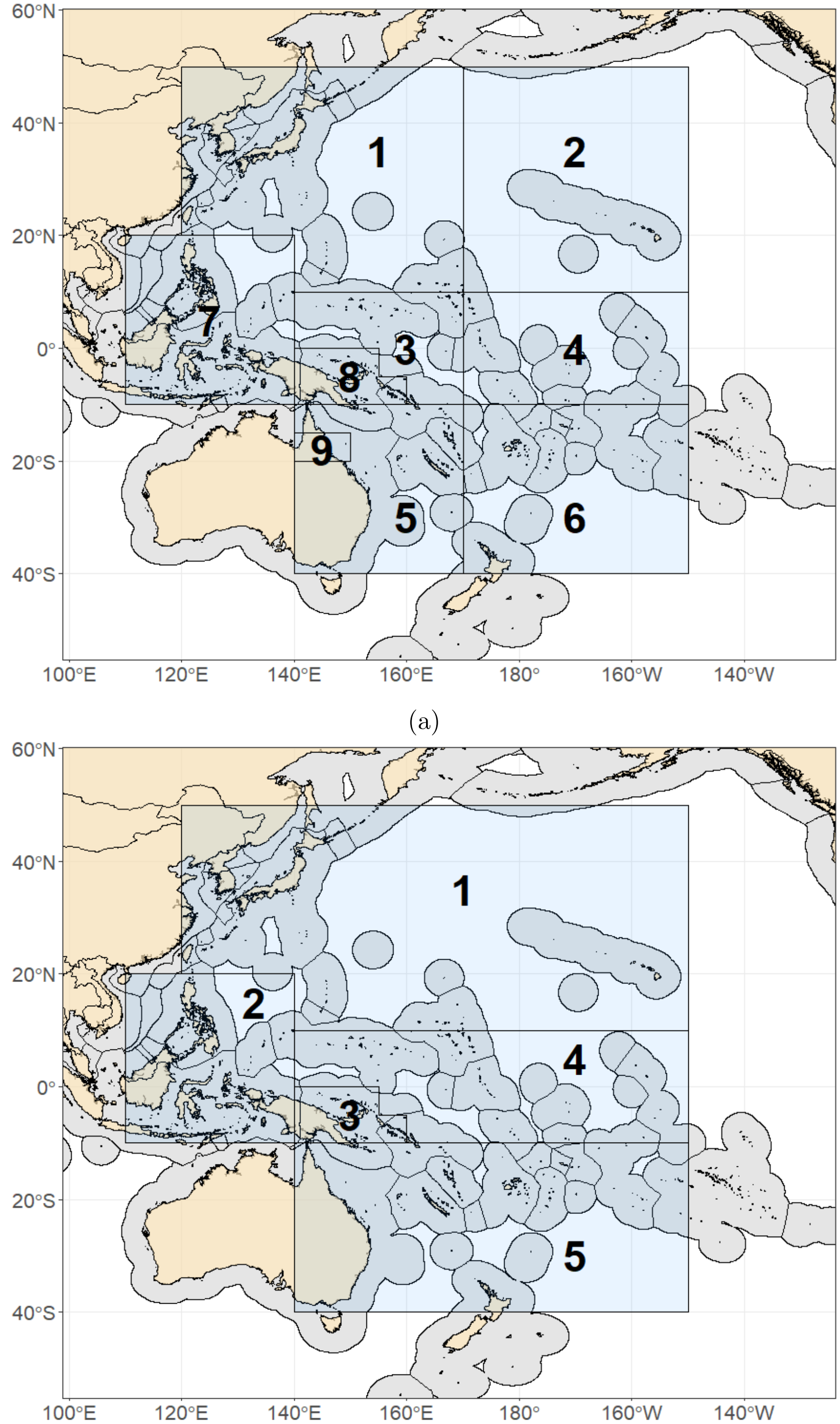
Sinclair, A., Gascon, D., Rivard, D., and Garvaris, S. (1991). Consistency of some northwest Atlantic groundfish stock assessments. *NAFO Sci. Counc. Stud.*, 16:59–77.

Walters, C. and Punt, A. (1994). Placing odds on sustainable catch using virtual population analysis and survey data. *Canadian Journal of Fisheries and Aquatic Sciences*, 51(4):946–958.

7 Appendix

Model	Bigeye Mohn's ρ values	Yellowfin Mohn's ρ values
m1_s10_a050_h65	-0.0844	-0.0561
m1_s10_a050_h80	-0.0935	-0.0566
m1_s10_a050_h95	-0.0885	-0.0567
m1_s10_a075_h65	-0.105	-0.0712
m1_s10_a075_h80	-0.0902	-0.0705
m1_s10_a075_h95	-0.0865	-0.0661
m1_s10_a100_h65	-0.0995	-0.0492
m1_s10_a100_h80	-0.0773	-0.0496
m1_s10_a100_h95	-0.074	-0.0493
m1_s20_a050_h65	-0.0966	-0.067
m1_s20_a050_h80	-0.087	-0.0651
m1_s20_a050_h95	-0.0722	-0.0635
m1_s20_a075_h65	-0.0915	-0.0679
m1_s20_a075_h80	-0.0797	-0.0658
m1_s20_a075_h95	-0.0716	-0.0641
m1_s20_a100_h65	-0.0937	-0.066
m1_s20_a100_h80	-0.0817	-0.064
m1_s20_a100_h95	-0.0672	-0.0624
m1_s40_a050_h65	-0.091	-0.0495
m1_s40_a050_h80	-0.079	-0.0481
m1_s40_a050_h95	-0.0743	-0.0502
m1_s40_a075_h65	-0.105	-0.0509
m1_s40_a075_h80	-0.0911	-0.0526
m1_s40_a075_h95	-0.0758	-0.0481
m1_s40_a100_h65	-0.0876	-0.0477
m1_s40_a100_h80	-0.0793	-0.0493
m1_s40_a100_h95	-0.0681	-0.0445
m2_s10_a050_h65	-0.0684	-0.101
m2_s10_a050_h80	-0.0954	-0.1
m2_s10_a050_h95	-0.0609	-0.0988
m2_s10_a075_h65	-0.0743	-0.104
m2_s10_a075_h80	-0.0557	-0.102
m2_s10_a075_h95	-0.073	-0.101
m2_s10_a100_h65	-0.0815	-0.0983
m2_s10_a100_h80	-0.0712	-0.0972
m2_s10_a100_h95	-0.0685	-0.096
m2_s20_a050_h65	-0.0433	-0.0836
m2_s20_a050_h80	-0.0409	-0.0821
m2_s20_a050_h95	-0.0389	-0.0807
m2_s20_a075_h65	-0.0474	-0.0847
m2_s20_a075_h80	-0.0426	-0.0828
m2_s20_a075_h95	-0.0382	-0.0812
m2_s20_a100_h65	-0.0548	-0.0839
m2_s20_a100_h80	-0.0518	-0.0819
m2_s20_a100_h95	-0.0375	-0.0803
m2_s40_a050_h65	-0.0726	-0.0632
m2_s40_a050_h80	-0.0598	-0.0611
m2_s40_a050_h95	-0.0533	-0.042
m2_s40_a075_h65	-0.0623	-0.0609
m2_s40_a075_h80	-0.056	-0.0588
m2_s40_a075_h95	-0.0584	-0.0546
m2_s40_a100_h65	-0.0592	-0.0558
m2_s40_a100_h80	-0.0521	-0.0569
m2_s40_a100_h95	-0.0551	-0.0555

Table 1: Retrospective bias (Mohn's ρ) of the estimated depletion of the bigeye and yellowfin stock models from assessment grid.



(a)

(b)

Figure 7: Spatial structure of the 2023 bigeye and yellowfin assessments. (a) nine model regions for bigeye. (b) five model regions for yellowfin.