Western and Central Pacific Fisheries Commission

SCIENTIFIC COMMITTEE
TWENTY-SECOND REGULAR SESSION
11-19 August 2026
Apia, Samoa (Hybrid)

SUMMARY OF SC22 ONLINE DISCUSSION FORUM

WCPFC-SC22-2026-ODF01

INTRODUCTION

1. The twenty-second meeting of the Scientific Committee (SC22) was held in Apia, Samoa. SC22 made use of an online discussion forum (ODF) to facilitate consideration of discussions on 2026 SC projects and other items. The ODF was closed on 16 August 2026, during SC22, to allow the outcomes of the ODF discussions to be considered by CCMs at SC22.

2. For reference during the SC22 Work Programme and Budget discussions, the table below summarizes the input provided by SC participants on ODF Topics related to WCPFC projects. The full comments are presented in this paper under each Topic.

SUMMARY OF INPUT FROM SC22 ON SC PROJECTS

Topic #	Subject	Comments
11	**Project 100d:** Application of CKMR methods to SP albacore and cost-benefit evaluation and prioritization of CKMR to WCPFC stocks (tuna, billfish, sharks)	**Sharks Pacific:** supports prioritizing sharks for future CKMR applications, noting the need for additional investment in ageing, reproductive biology, population structure, and genomic assay development.
18	**Project 130:** Joint bycatch assessment workshop for billfish and sharks	**Sharks Pacific:** supports the proposed workshop on bycatch assessment for billfish and sharks and recommends that Project 130 consider monitoring strategies for these non-target species and their potential application to CKMR-based abundance estimation.
20	**Project 110 and 110a:** Non-entangling and Biodegradable FAD Trial in the Western and Central Pacific Ocean	**PNA and Vanuatu:** support Project 110 and its recommendations, particularly work on monitoring biodegradable FAD materials and clarifying biodegradability standards, while proposing revisions concerning catch comparisons and the availability of biodegradable materials in atoll ports.
22	**Project 121:** Ecosystem and Climate Indicators	**USA:** supports the continued work on the indicators and recommends replacing ONI with the RONI as the official NOAA ENSO index, and updating the archived climate.gov reference in the ENSO section.

TOPIC 1. WCPFC Project127/127a - review and reconciliation of size data collected in the WCPFC-CA for stock assessment purposes: Phase 2 progress report

No questions or comments

TOPIC 2. Streamlining scientific data reporting to the Commission: progress, challenges, and next steps

Related papers: WCPFC-SC22-2026-ST-WP03: J. Leiria Schattschneider, T. Jones, T. Vidal. Streamlining scientific data reporting to the Commission: progress, challenges, and next steps

Questions and Comments

Chinese Taipei

Is it possible to develop a list of filter criteria for mandatory data field of standardized data reporting? It will be very helpful.

Reply (SPC)

Hi Fan, thanks a lot for your comment. Are you thinking about a list summarising all data checks applied to mandatory fields? For example, data checks in the *latitude* field in the logbook activity (longline) are:

- Mandatory field (nulls/NAs are not allowed)
- Data type: Numeric
- Minimum value: -90
- Maximum value: +90

If this is what you are thinking of, it is captured in the third item in Areas for Improvement and we can definitely provide the list of information requested. We thought about adding it to the documentation page, to make sure we concentrate all the relevant information for scidata submissions into a single place. Let us know how that sounds to you and to your team.

Response to SPC (Chinese Taipei)

Hi Jessica,

Thanks for your reply. Yes, that’s what I am thinking about. Because my colleague had some data filtering issue while uploading SciData this year, probably there’s some data fields violated the criteria set by SPC.

Ren-Fen

Response to Chinese Taipei (SPC)

Perfect, I will make sure we implement this ahead of time of the next scidata submissions and share the updates with all members to give all opportunity to review and provide additional feedback.

TOPIC 3. Proposal for Conservation and Management Measure to Implement A WCPFC Electronic Monitoring

Related papers: WCPFC-SC22-2026-ST-WP10: United States. Proposal for Conservation and Management Measure to Implement A WCPFC Electronic Monitoring

Questions and Comments

Sharks Pacific

Sharks Pacific welcomes and supports the initiative and ambition of the US to present this proposal advancing observer coverage through the use of EM. Particularly, we’d like to highlight the inclusion of transshipment monitoring as a discrete EM objective, which we believe is important as a means of verifying and validating data collection on high seas operations, but admit there is a lot of work remaining to ensure that transshipment EM coverage meets defined data collection objectives.

However, we'd like to raise concerns regarding the minimum EM Analysis Rate as it potentially counts toward coverage targets. Specifically, we note that under the proposal CCMs may satisfy coverage requirements through ROP observers, EM, or a mix of the two. Similar to the unnecessary selective use of coverage metrics other than "number of hooks," which represents the best available science, because CCMs will be able to satisfy coverage requirements through EM or human observers interchangeably it becomes imperative to ensure that coverage is carefully defined under both scenarios in a way that is consistent across both approaches so that creative accounting isn't allowed. While we appreciate the US' attempt to find a solution that appeals to some members, a coverage figure that blends installation with an unspecified analysis/review rate can look compliant on paper while delivering weaker observer coverage in practice and, more importantly, potentially allows for manipulation of coverage rate figures. For example, a nominal 20% figure could mask a much lower rate of footage actually reviewed if analysis rates are not independently floored. If a similar standard were applied to human observers, an observer could just go for a pleasure cruise on a fishing vessel and, at the end, submit a minimal amount of data collected on one random day out of ten such that the vessel meets the 10% threshold. We simply cannot create a lower bar for EM than currently exists for the ROP.

Nonetheless, we would also like to note the positive synergy of this proposal with elasmobranch conservation objectives. SA-IP24 confirms that deep-set longline gear drives most elasmobranch fishing mortality, so since EM targets this fishery we welcome its application. We again note that improved, genuinely independent EM coverage in longline could materially improve bycatch/discard verification and mitigation-measure compliance for sharks and rays, provided independence in the execution and analysis of those EM systems are ensured. There are, of course, still a lot of devils in the details to be discussed and we look forward to further robust discussions.

TOPIC 4. Progressing the use of cannery receipt data for scientific work (P114) – a new approach

No questions or comments

TOPIC 5. Project 123: Scoping the next generation of tuna stock assessment software

No questions or comments

TOPIC 6. Scientific outputs from the Pacific Tuna Tagging Programme, and addressing requests from FAC19

No questions or comments

TOPIC 7. Yellowfin Tuna Stock assessment

Related papers: WCPFC-SC22-2026-SA-WP05: A. Magnusson et al. Stock assessment of yellowfin tuna in the western and central Pacific Ocean: 2026

Questions and Comments

New Zealand

New Zealand thanks the authors for a substantial piece of work delivered under real time pressure, much of it in response to the 2023 peer review, SC19 comments and the PAW. We recognise these comments come late in the ODF period and we do not expect a response before it closes. We offer them as a contribution to future work rather than as requests for additional analysis now.

Model files

1. Providing the input and output files for the diagnostic model would help to support review, as was done for bigeye.

Tag weighting and τ

2. Section 5.5.3 states that τ is not a penalty weight and is implicit in the likelihood, but sections 6.1.1, 7.1 and 7.6 describe it as downweighting the tag data. This distinction matters, since a variance parameter should be estimated while a weight needs justification. No justification is provided for τ = 2 beyond noting that τ = 1 implies a Poisson distribution — could the authors clarify whether there is a source or empirical basis for this value, or whether it is a convention carried forward?
3. Only τ = 1 and τ = 2 were tested, and τ is not among the ensemble factors, so all 48 models share τ = 2 on an axis identified as a major scale driver. Our provisional analysis indicates the plausible range could be wider than has been explored. We suggest τ be estimated outside the model from fits to the tag data and also query whether the authors could consider adding τ to the ensemble?
4. We also note that steps 1 to 15 were developed under τ near 1, with τ = 2 only added once the diagnostic model was settled, and that Section 7.6 flags an interaction with the mixing period setting. It would be worth re-testing across a range of τ values.

Direction of the τ effect, and growth

5. Fifteen reporting rates are at the 0.99 bound and the model wants them higher, so it compensates through F, which caps biomass. Figure 30 shows the tag component preferring lower biomass than CPUE. Downweighting the tags should therefore raise biomass, but step 16 does the opposite. From the app, we estimate L1 falls from about 20 cm at step 15 to 12 cm at step 16, with τ the only change. Because the Lorenzen curve derives from growth with M fixed at age 40, mortality at age shifts at the same step. Given that M scale is the most influential ensemble factor, much of the step 16 biomass change may come through growth and M rather than tag weighting. A run at τ = 2 with growth fixed at the step 15 values would separate the two. We note this runs against Section 8.2.3, which attributes the low L1 to the tag likelihood and so predicts L1 rising when tags are downweighted, and that the step 16 L1 sits below effectively all age 1 CAAL observations while the step 15 value sits within them.
6. Raising τ also increases the influence of the size data, which may not be internally consistent since growth likely varies spatially — as SPC document for bigeye in SA-IP-11 — while the assessment applies a single curve across all regions. The model then settles on a compromise that fits no region well, and because M at age derives from growth, that compromise propagates into mortality and scale; this may explain much of the large growth change at step 16. Combining the fleets-as-areas approach in SA-IP-09 with regionally converted length-at-age data, as in SA-IP-11, would let regional growth differences be represented — is this tractable for the next assessment?

Tag data

7. Section 7.1 states that the two mixing period **reporting rate** options bias in opposite directions, and that conclusion carries the argument, since this is the second most influential ensemble factor. If both err in the same direction, the ensemble displaces rather than brackets. Could the authors set out the basis for the claim? Our understanding is that option 0 reduces the F estimate somewhat and option 1 reduces it further. Have the authors considered **using option 0 only, and** removing the release groups that reach the threshold at which problems arise under option 0?
8. Fifteen reporting rates are at 0.99, of 49 parameters on bound, and Figure 23 suggests a bimodal pattern, either at the bound or near 0.3. Is this a lack of identifiability between reporting rate, F and scale? The estimates also depart from the seeding priors consistently in one direction, a

conflict with independent experimental data that is not treated as such in Section 7.4.1. Because the rates cannot rise further, the model compensates through F, so limited movement in the sensitivity results may reflect the constraint rather than robustness. How should SC interpret the Hessian based percentile ranges with 49 parameters at bounds?

9. Section 7.2.3 describes underestimation of tag returns at quarters 4 to 6; our reading of Figure 21 is that quarters 4 through 8 all sit above the predicted line, implying **modelled** attrition is too rapid after mixing. The authors may wish to consider revising this description, and noting that the close fit at quarters 1 and 2 carries little information, since the mixing period desensitises the likelihood there. Residuals against time at liberty are in the app but not the paper — would it be possible to include versions disaggregated by release region and tagging programme; and plotted on the log scale.

Convergence

10. Our reading of Figure 26 is that no jitter run recovered the diagnostic solution, with completed runs 36 to 56 units worse, clustering into around five groups. This may suggest a rugged likelihood surface with distinct local optima, rather than providing reassurance of a well-defined minimum — we would welcome the authors' comment, and the basis for the 200-unit acceptance criterion. We also note that Section 8.1.2 states OPR was the only way to obtain a PDH, so the PDH cannot be treated as independent evidence that the model is well behaved. Are regional recruitment estimates without OPR available for comparison?

Retrospective

11. Mohn's rho is within conventional bounds, but Figure 28 shows the peels sitting below the full-data run across the whole series, back to 1952. This appears to reflect a scale shift associated with one year of data, rather than a terminal-year pattern alone. If the cause is the 2024 Indonesian handline reallocation, could the authors indicate its estimated effect on historical spawning potential, and how SC should treat the 2026 results while the correction is outstanding?

Spatial structure

12. Movement rates are low and 29 movement coefficients are at bounds. Movement and the regional distribution of vulnerable biomass together determine regional depletion and how the tags inform scale. How well determined are the movement parameters?
13. SA-IP-02 reports that in the bigeye test the penalty did not reach its targets in three of nine regions, that heavier weights came closer, and that some freed index fishery selectivities were spurious; it also found the penalty raised mean total biomass by 28–36%, left the stock about 8% less depleted, and shifted growth. Step 8 of the yellowfin stepwise moved in the opposite direction, toward lower SB and greater depletion, and the yellowfin paper does not state the penalty weight used. Could the authors report the weight and the estimated regional shares against the targets, confirm whether the index selectivities were checked, and comment on why the direction differs and whether growth also changed at step 8? This matters because freeing the regional catchabilities removes regional scaling information from the indices, so the penalty now carries essentially all of it — and since this was the largest single stepwise change, using the full time series rather than the PAW-recommended 1965–1969 period, is a sensitivity on the time period available?
14. Recruitment is concentrated in region 2, attributed to catches of small yellowfin in the Indonesian and Philippine fisheries. How sensitive is the regional recruitment distribution to those catch and size inputs?

Fits to data

15. Section 7.2.2 reports that the model does not follow observed declines in mean length in several longline fisheries, and overpredicts median length in all five main industrial purse seine fisheries. Are standardised residual plots by fishery and year available, with input against implied effective sample sizes? Size weighting uses divisors of 10, 20 and 40, the Dirichlet multinomial having been rejected because it degraded the index fits, which suggests the tension with the indices is being managed by choice of divisor.
16. Figure 13 presents CPUE residuals on the observation scale rather than standardised by the assumed error. With time varying CVs, the funnel could reflect either fit or weighting, so we suggest standardised residuals, a runs test, and RMSE by index.
17. What does the PEN component in Figure 30 comprise? It is minimised at low biomass and rises monotonically, so it appears to contribute to scale. The figure is also truncated before the CAAL component reaches a minimum, so the size of that conflict cannot be read. Could a finer, wider version be provided?

Further work

18. New Zealand supports the recommendations in Section 8.2, particularly the tag modelling workshop and the observation that the MFCL-specific workarounds currently sit separately from the DTU external tagging analysis. We also support tracking the estimated regional biomass proportions against the prior and across stepwise models, as several comments above would draw on. We note two gaps: τ does not appear among the recommendations despite being identified as a major driver of scale, and we suggest adding it; and while Section 8.2.3 identifies the hard link between the Lorenzen curve and growth as the underlying concern, the recommendation is limited to fixing or estimating L1, which would not address that link itself.
19. More broadly, and as we noted for bigeye, simplification is now a scientific priority as well as a resourcing one. Many of the difficulties above stem from the interaction of spatial structure, movement estimation and tagging data within a single model, which lets settings interact in ways that are hard to anticipate or diagnose. The paper's own recommendations already point toward three complementary directions — analysing tagging data outside the model (Section 8.2.4), a fleets-as-areas structure (Section 8.2.2, explored in SA-IP-09), and converting length compositions to ages externally using region-specific age-at-length (as applied to bigeye in SA-IP-11) — each of which would remove parameters that are currently poorly determined or at bounds, and together would be stronger than any one alone.

Reply (SPC)

We acknowledge the commentary and a number of constructive recommendations provided by New Zealand on the yellowfin assessment report. We appreciate the time taken by NZ to review the assessment and develop this detailed ODF submission, especially considering the late delivery of the paper. We may not have time to review and respond to these comments in suitable detail before the ODF closure. They will be reviewed in detail and may be considered in future work for this stock assessment.

TOPIC 8.	**Bigeye Tuna Stock Assessment**

Related papers: WCPFC-SC22-2026-SA-WP06: K. Kim et al. Stock assessment of bigeye tuna in the western and central Pacific Ocean: 2026

Questions and Comments

New Zealand

We would like to thank the assessment team for a very substantial amount of work. Several aspects stand out.

The diagnostics are a major step forward — phase-aware jitter, the first formal self-test, the retrospective, and parameters-on-bounds reporting are all substantial improvements, as is the transparent reporting of assessment results. Publishing the model files, jitter runs and executable also deserves mention: it allows us to repeat parts of the analysis independently.

Allowing each index fishery its own selectivity, rather than one curve shared across regions, is a real improvement and should make size-composition fits far more meaningful. We also acknowledge the work on tag over-dispersion, including correcting how it was specified in the MULTIFAN-CL documentation, and the software changes in SA-IP02.

> **Reply (SPC)**
> The assessment team thanks New Zealand for recognising the substantial work undertaken for this assessment.

Several comments below exist only because the diagnostics were good enough to reveal the issues.

A general observation
Much of what follows traces back to complexity. The assessment has 1,997 estimated parameters, 33 fisheries and five regions. The combination of multiple regions and tagging data is among the biggest single sources, since it also requires movement between regions to be estimated. And the assessment sits on top of other models — CPUE standardisations, tag mixing estimated externally, growth weighting from a spatial density model — each with its own assumptions, and each needing a different data type assembled and cleaned first.

This complexity carries two consequences. It creates room for settings to interact in ways that may not have been intended, and we believe this is where several of the issues below originate. It also makes assessments slower to build, check and document, which may contribute to papers arriving close to the meeting with limited time for review.

We therefore support the work under Project 123 (SA-WP01) and would encourage simplification to be treated as a scientific priority. A simpler model that can be understood, tested and developed faster may serve management better than a more elaborate one that cannot.

> **Reply (SPC)**
> We thank New Zealand for this observation and agree that model simplification should be considered carefully. Simplification has been a substantial component of this assessment: the spatial structure was reduced from nine regions to five, a single-region fleets-as-areas model was examined in SC22-SA-IP09, and its results were compared with the diagnostic and alternative models in the assessment appendix.
>
> Further simplification would involve trade-offs, including the loss of regional, movement, tagging, and fishery-specific information. It should therefore be discussed across the Scientific Committee, together with requests for additional regional analyses. If there is a collective agreement on a simpler assessment structure and acceptance of the associated information loss, the assessment team would be happy to follow that direction in future assessments.

We also note that the additional CPUE work requested by PAW, that is supported in the general observations above, took considerable additional time of about 6 weeks, that was unplanned time. This was on top of the additional work to improve the size data preparation and developing the additional resources for the open science approach, enhance diagnostic analyses, and additional age length data - done since PAW. The difficulty with timing this year is not only related to complexity, but the reality is also that the time availability is problematic with the current timing of the SC and when data are finalised, and the increase in work that is being requested in each assessment. This issue has been ongoing and has not been resolved.

Comments / Questions

1. Tag over-dispersion (τ)

The diagnostic model uses τ = 2.0, but every ensemble member uses 1.2, 1.3 or 1.4 — all giving the tagging data more influence than the diagnostic model, which the paper notes pushes results toward a more optimistic answer. We have done some provisional work to estimate tau from the assessment's own data which gives a value of about 5. We would be happy to share the methodology we have used to do this.

Could the ensemble be centred on the estimated values, rather than sitting entirely below even the diagnostic model?

Reply (SPC)

The assessment team has discussed tau estimation with the New Zealand delegation, including the proposed methodology and its interpretation. These discussions identified areas of agreement as well as unresolved methodological disagreements. We agree that further research is warranted, but there is currently no shared basis for treating the provisional estimate of about 5 as sufficiently reliable to centre the ensemble. We therefore do not support re-centering the current ensemble on that value.

The values tau = 1.2, 1.3, and 1.4 were selected deliberately. Sensitivity analyses showed that key management quantities changed below 1.4 but were very similar at 1.4 and above, including at the diagnostic value of 2. The selected range represents the transition from stronger to weaker tag influence while retaining uncertainty about that influence, an alternate could possibly be 2, 1.3 and 1.2. Centering the ensemble on a provisional value of 5 would instead place substantial tag downweighting before the estimation method and interpretation have been agreed.

Further analyses by the assessment lead indicated that a common tau was not reliably estimable and that its estimation could increase numerical instability. A single common value is also unlikely to be fully realistic, because over-dispersion may differ among release groups, fisheries, assumed mixing periods and periods at liberty. Estimating multiple tau values would require further methodological development and broader technical discussion.

Within the retained ensemble, the marginal results for tau = 1.2, 1.3, and 1.4 were very similar (Figure 59), while natural mortality and tag-mixing assumptions were more influential. Because all ensemble axes vary jointly, these comparisons should not be interpreted as independent one-factor effects. Any future change to the tau treatment should follow further research and broader agreement of the SC. Re-centering the current ensemble on a provisional estimate from a single CCM.

2. Tags disappear faster than the model expects

Figure 23 shows this, and the text notes over-prediction between 10 and 15 quarters at liberty. We would suggest a slightly stronger reading: the shortfall appears to grow steadily with time at liberty rather than being confined to that window, which suggests a systematic pattern. Ongoing tag loss seems a less likely explanation — double-tagging studies have found shedding rates low, and what is allowed for is a reduction in release numbers, which cannot create a trend through time. An underestimate of total mortality, whether natural or fishing mortality, would appear to fit the pattern better. Higher M reduced the trend when we ran the model, and a biomass scale that is too high would make F too low, which points in the same direction as comments 1 and 7.

Could this be used as a diagnostic on total mortality rather than read as a local feature of the tag fit?

> **Reply (SPC)**
> We agree that the time-at-liberty pattern is potentially informative about total tag attrition, but it is not a unique diagnostic of natural or fishing mortality. Predicted recaptures also depend on tag mixing, movement, reporting rates, fishery selectivity and effort, release-group heterogeneity, and population scaling. Increasing M will mechanically reduce predicted recaptures, but an improved fit does not independently demonstrate that M is underestimated or identify which component of total mortality is responsible.
>
> The diagnostic model identified this pattern and informed the broader sensitivity analysis. Alternative M values are represented in the uncertainty ensemble, drawing on the Hamel and Cope longevity approach, the independent tag-based analysis in SC22-SA-IP14, and values considered in previous assessments. Natural mortality was also identified as one of the dominant driver s of ensemble uncertainty.
>
> The pattern can therefore be recognised as a diagnostic of the combined tag-survival and recapture process, but it should not be interpreted as direct evidence for a higher M. The ensemble provides a broader treatment of uncertainty in M than calibrating it to this single residual pattern would provide.
>
> This point also raises the issue of tagged bigeye, and likely the population in general, gradually and possibly permanently leaving the WCPO model region for the EPO. Currently, we only treat the tag data by omitting extra-WCPO recaptures and lowering the release numbers to preserve the original recapture rate of the remaining included recaptures. Perhaps a better way is, as alluded in this comment, to increase the M to include a permanent emigration component. Then the adjustment to the tag releases would not be required and it would also treat the untagged population equivalently (which the current model does not). Another way would be to add an EPO region to the model, which would include a movement parameter WCPO to EPO, although this would add complexity needing an additional spatial region and fishery(s). A good topic for the proposed tagging data workshop.

3. Natural mortality is the largest single influence
We welcome its inclusion as an uncertainty axis. In our model runs, raising M moved recent depletion from 0.18 to 0.44, and again to 0.72 — a larger effect than anything else we investigated, τ included. It also reduced the time-at-liberty trend in comment 2, so two independent lines point the same way. The paper reports the response to M is inconsistent in direction and calls the low-M result unexpected.

Has that been resolved, and does the ensemble's range cover what SA-IP14 supports?

Reply (SPC)

SA-IP14 is explicitly represented in the natural-mortality uncertainty range. The design range for M at age 40 quarters is 0.050–0.165 per quarter. The SA-IP14 estimate of 0.0624, including its 90% interval of 0.0500–0.0749, is therefore fully covered. The distribution also considers the previous-assessment value and the Hamel and Cope longevity analysis.

The response of depletion to M should not be interpreted as a simple one-factor relationship. Changing M simultaneously affects survival, recruitment, population scale, fishing mortality, and the fit to tagging and other data. Large or non-monotonic changes should also be checked against convergence diagnostics because changing a fixed M value can move the optimisation between local optima. A direction observed in a small number of individual runs is therefore not necessarily general across model configurations.

The depletion response and the tag time-at-liberty pattern are not independent lines of evidence. Both arise within the same fitted model after changing M, and increasing M mechanically reduces predicted recaptures at longer times at liberty. An improved tag-attrition fit therefore does not independently validate a higher M.

For these reasons, M was treated as an uncertainty axis and varied jointly with the other ensemble assumptions. The ensemble confirms that M is highly influential, but it does not support assigning a single, model-independent direction to its effect. The range supported by SA-IP14 is represented in the ensemble.

4. Tag reporting rates

One group is estimated at its 0.99 upper limit and others sit where the prior considers very unlikely, which may suggest the model is struggling to reconcile predicted and observed recoveries. There is also an issue in how reporting rates are treated during the mixing period, which affects 44% of all returns; we expect others will comment on that in more detail.

Could the influence of the boundary and low-prior estimates on the biomass scale be reported?

Reply (SPC)

The biomass-scale effect of the reporting-rate estimates cannot be isolated from the existing model fit. In the catch-conditioned MFCL model, catches are fixed, tightly coupling biomass and fishing mortality. Reporting rates therefore interact with fishing mortality, movement, mixing, and tau. Boundary estimates indicate this joint tension, not a standalone effect on biomass scale.

The boundary issue has nevertheless decreased: two groups reached the upper bound in 2020, three in 2023, and one in 2026. A catch-errors model could provide more flexibility, but would require substantially more parameters and computation, and the 2020 catch-er rors model still had two boundary estimates.

Reporting-rate treatment during mixing was tested and included in the ensemble. Following discussion with the US delegation at SC22, we agree that further improvement requires broader methodological work and cannot be resolved by a single additional analysis in the current assessment. This work could be considered for the next assessment in 2029.

5. Growth, and spatial variation in it

One curve is used for the whole region. Against the ageing data it grows fish too fast, reaching a given size more than a year too early for the largest fish, in every region. The size data agree: in 14 of 16 longline fisheries the model predicts too few fish at 130–150 cm and too many above 150 cm. So this is not two data sources disagreeing — both point the same way, and the ageing data that would settle it carry about one percent of the effective weight. The underlying cause does not yet appear to be firmly established; a likelihood profile on the growth rate may offer a useful and low-cost first step.

Reply (SPC)

We agree that these residual patterns warrant attention. The may not necessarily distinguish spatial variation in growth from a common growth-model mismatch or other model processes. Length compositions reflect not only growth, but also selectivity, recruitment, mortality, movement, and spatial and temporal sampling. The ageing residuals provide relevant information, but a similar mismatch across all regions may not by itself demonstrate variation among regions.

Recoverability and practical identifiability of the common growth parameters have been assessed through the parametric-bootstrap self-test. A likelihood profile would provide complementary information about support for the common growth rate in the observed data, but it would remain conditional on MFCL's single-curve structure. It therefore could not test spatial growth variation or establish the cause of the residual pattern.

A defensible profile analysis would require model setup, convergence checks, validation, interpretation, and reporting; it is not simply an additional model run. Given the substantial new analyses incorporated and the timeline for completing the current assessment, this additional work cannot be accommodated. Spatial growth variation is more appropriately considered through future methodological development informed by SA-IP08 and SA-IP11, and the next assessment, also pending more grow th data becoming available from age-length sampling.

6. What sets the biomass scale

The paper identifies population scale as the dominant data conflict, tags favouring a higher scale and CPUE and size data a lower one. In our model runs what resisted a lower scale appeared to involve the regional scaling constraint and a persistent shortfall in region 1, but we do not understand the mechanism.

We believe this would benefit from further diagnostic work: would it be possible to examine the region 1 behaviour directly, and to report the assessment with the constraint relaxed?

Reply (SPC)

We agree that the regional scaling constraint and region 1 behaviour may warrant further investigation. However, examining this mechanism and presenting an assessment with the constraint relaxed would constitute substantial new diagnostic and sensitivity work rather than clarification of the existing results. A relaxed-constraint model would require full rerunning, validation, and interpretation before it could be presented as an assessment alternative.

The current assessment includes substantial new analyses and developments beyond the previous SC accepted assessments, and no single assessment can resolve every diagnostic question or model alternative. Given the timeline for completing the current assessment, this may be considered in a future assessment if it remains relevant.

7. How wide is the real uncertainty

The ensemble gives a median depletion of 0.277, with 15 of 80 models below the limit reference point. The diagnostic model is among the most pessimistic in the ensemble, and on the τ axis one reason is visible: it uses a value closer to what the tag data support than any ensemble member does.

Could the ensemble's position relative to the diagnostic model be reported axis by axis? If it is systematically displaced, the probability of being below the limit may be displaced as well, and the operating models could inherit a similar gap.

> **Reply (SPC)**
> The diagnostic model is a reference for model development, it is not necessary that it should be the centre of the ensemble, and it was not included in the 80-model grid. Its tau = 2 setting was applied for consistency with the value reported in the 2017, 2020, and 2023 assessments, that on face value, was accepted as suitable by multiple SCs before this assessment, not because it was reliably estimated from the tag data, or a value the 'tag data preferred'. While a diagnostic model in the centre of the ensemble might look better, distance from the diagnostic model does not necessarily measure ensemble bias.
>
> Further exploration by the lead assessment scientist indicated that tau is not reliably estimable in the current model. These findings have been shared and discussed with the New Zealand delegation and FFA member delegations. The rationale for using tau = 1.2, 1.3, and 1.4 is reported. Because all ensemble assumptions vary jointly, an axis-by-axis comparison with the diagnostic model would not isolate causal effects or demonstrate that the probability of falling below the limit reference point is displaced. The existing axis-based ensemble results provide an appropriate comparison.

Priorities for future work

1. Spatially varying growth

For a model fitting mainly to size data, growth converts what is observed into what is modelled, so errors there affect everything downstream. We think this is the key structural issue to address. SA-IP08 on spatial sampling bias in growth estimation and SA-IP11 on regional growth variation both address it, and should feed into the bigeye tuna assessment rather than remain background papers.

> **Reply (SPC)**
> We thank New Zealand for this recommendation and agree that SA-IP08 and SA-IP11 provide relevant information for future model development. The assessment team will consider how spatial sampling bias and regional growth variation can be evaluated and, subject to methodological, improved data on growth from more age-length samples, and software feasibility, incorporated into future bigeye assessments.

2. Simpler spatial structure

SA-IP09 on single-region fleets-as-areas models is promising. If a simpler structure reproduces the main results it would be easier to test, faster to run and easier to explain, and would reduce reliance on the regional constraints in comment 6.

> **Reply (SPC)**

We agree that SA-IP09 provides a promising basis for further simplification. Subject to further validation and consideration of the associated loss of spatial and tagging information, the single-region fleets-as-areas model could be considered as a structural sensitivity and potentially as a component of a future ensemble.

3. How tagging data are used inside the model

Tags are one of the main things setting the scale, but sit awkwardly in a five-region model: tagged fish move at much smaller scales than the regions they are assigned to. Estimating mixing periods outside the assessment with SEAPODYM and Ikamoana (SA-IP07, SA-IP10) is a sensible response, but it is another model stacked on the assessment, and the small-scale movement that matters most may still not be captured. If not, the tag data are biased in ways that are hard to see from inside. This is one of the strongest arguments for analysing tagging data outside of the stock assessment model.

Reply (SPC)

We agree, and this issue is recognised in the assessment report. The assessment team supports further analysis of tagging data outside the integrated assessment model, particularly for movement, mixing, mortality, and reporting processes. The current assessment explored externally estimated mixing periods and examines models with tagging data removed and movement fixed from MFCL or SEAPODYM. This recommendation is therefore consistent with the explorations that were conducted by the assessment team. The SEAPODYM movement option maybe a viable alternative if population scaling is thought to be badly biased due to the tagging data not conforming to model assumptions, but in that case the other data needs to be sufficiently informative on population scaling.

4. Simulation testing

Several issues above concern whether a method behaves as intended rather than the data, and can only be answered by testing on data with known properties. We would encourage this in the development cycle — for the tag likelihood, and for the spatial growth methods above, where the whole question is whether a method recovers growth correctly when it varies by region. It applies equally to the modelling platform as a whole. We support the paper's recommendation to extend self-tests to structural cross-tests and would suggest making this a priority.

Reply (SPC)

We agree that simulation testing and structural cross-tests are valuable, and this is recommended in the assessment report. The current assessment introduced phase-aware jittering and the first formal model self-test, alongside extensive sensitivity analyses and a large uncertainty ensemble.

Extending simulation testing would require substantial model development, computation, validation, and interpretation. Its scope and priority therefore need to be balanced against the existing diagnostic workload and the need to provide sufficient review time. Future work should focus on targeted tests of the most important structural questions rather than exhaustive simulation testing of every model configuration.

Concluding remarks

We would like to close by again thanking the SPC assessment team for the depth and rigour of this year's work.

We wish to stress that none of these issues are critical in isolation, and many are already acknowledged in the paper. Our main concern is that they appear to push mostly in the same direction and that, on the settings we have been able to check, the ensemble appears to sit on the optimistic side of the diagnostic model. We would like to understand how much of the reported status depends on that combination, and we think the medium-term answer lies in simpler, better-tested models.

Reply (SPC)

We thank New Zealand for its constructive engagement and acknowledge that the identified issues warrant continued investigation. However, their apparent directional consistency should not be interpreted as independent evidence of a common bias, because many arise from the same interacting tag, mortality, population-scaling, and spatial assumptions. The diagnostic model is also not necessarily intended to define the centre or the preferred outcome of the uncertainty ensemble.

We agree that simpler and better-tested models are an appropriate medium-term objective. Achieving that objective will require Scientific Committee agreement on the information and spatial detail that may be lost, together with clear priorities for future methodological and simulation work. The current assessment reports the identified l imitations transparently and represents the principal uncertainties through the ensemble.

TOPIC 9.	**North Pacific Albacore Stock Assessment**

No questions or comments

TOPIC 10.	**Pacific Blue Marlin (Makaira nigricans) Stock Assessment**

Related papers: WCPFC-SC22-2026-SA-WP10: ISC. Stock Assessment Report for Pacific Blue Marlin (*Makaira nigricans*) through 2024

Questions and Comments

Vanuatu

PNA and Vanuatu reiterate previous statement on billfish that Pacific blue marlin makes up a small but important component of our domestic longline fisheries they are also important for our sport fisheries. As such this stock is important to us and we want to ensure that it is sustainably managed.

We believe that this assessment is an improvement on the previous assessments. We would like to ask the assessment team

1. Why does the projected biomass decline very rapidly under the FMSY scenario? We would not expect this to happen faster than the FHigh scenario.

2. Is this issue a result of the assumptions in the assessment or with the recruitment scenarios used in the projections?

We note that the paper suggests that additional catch would be possible, however, we would like to ask. How could that happen given that this is almost entirely a bycatch fishery?

Lastly, noting the issues with the projections if a catch increase is proposed PNA and Vanuatu believe that any catch increases should be undertaken with caution.

Reply (USA)

Thanks to PNA and Vanuatu for your comments and questions. The FHigh projection corresponded to the highest F estimated across the time series. This occurred in 2003 and was estimated to be F=0.23. The estimated FMSY was F=0.26, so the Fishing mortality on the stock was never above FMSY, which meant that the removals estimated for the FMSY projection were actually higher than the FHigh projection. As a result, the FMSY projected biomass decreased faster than the FHigh projected biomass.

The WG would also like to note that there are currently no catch limits for BUM, as you noted it is a bycatch species and only targeted occasionally. Our suggestion that the stock could withstand somewhat higher catches was intended to indicate that should the catch increase in the future, it is unlikely to lead to overfishing, and that no actions to reduce catch at this time would likely be necessary.

TOPIC 11. Project 100d: Application of CKMR methods to SP albacore and cost-benefit evaluation and prioritization of CKMR to WCPFC stocks (tuna, billfish, sharks)

Related papers: WCPFC-SC22-2026-SA-WP11: SPC-OFP and CSIRO. Report on Project 100d: Application of CKMR methods to SP albacore and cost-benefit evaluation and prioritisation of CKMR to WCPFC stocks (tuna, billfish, sharks)

Questions and Comments

Sharks Pacific

Because CKMR uses genetics to estimate absolute abundance and mortality independent of a conventional stock assessment model, it could have real implications for future data-generation capacity for shark stocks, particularly those that are currently data deficient. Currently, sharks are triaged as the least CKMR-ready group of all candidate stocks in this paper noting that Pacific blue shark, oceanic whitetip, silky shark, and both southwest and northwest Pacific shortfin mako all need "additional investment" across every readiness category, including ageing, reproductive biology, population structure, and genomic assay development, in contrast to all the other target species that are rated as Validated/Established/Available. That recognition alone should be enough to elevate sharks in the consideration and development of future CKMR application.

We note the challenge non-retention of some species presents and support the suggestion of non-lethal sampling that is proposed. We would like to highlight our side event that will be exploring some of these potential methodologies and approaches on Tuesday, August 11. However, we also emphasise, again, that any sampling must be facilitated and supported by adequate observer coverage onboard vessels. Specifically, we acknowledge the passage in the paper stating "Sampling for these species will likely also present logistical challenges given the low longline observer coverage and the incidental nature of their catch…" Like with so many other issues before the SC, any sampling needed for any CKMR application relies on increased observer coverage, which must be considered an imperative and absolute priority to improve scientific data collection across the board.

Most importantly, we note that this paper highlights the need for sustained investment in the Shark Research Plan and Project 130 because it explicitly makes shark life-history knowledge gaps (ageing, reproductive biology, population structure) a precondition for future shark abundance science.

Lastly, we propose that existing CKMR sampler training infrastructure (241 samplers trained across 17 Pacific countries since 2023) be leveraged to build non-lethal shark tissue-sampling and molecular monitoring application capacity across the regional observer cadre as part of the ROP and National observer programme training platforms, noting also the recent publication of a paper on non-lethal sampling technologies: Boube, T., Guilbert, A., Winkler, Y. et al. A new underwater probe for simultaneous tagging and biopsy of marine megafauna: a case study on the Critically Endangered Great hammerhead (Sphyrna mokarran). Mar Biol 173, 150 (2026).

Reply (SPC)

Thanks Sharks Pacific and Bubba for your comments. The broader CKMR aspect of the paper only provides a desktop triage on information available for tuna, billfish and sharks that the SC may be considering for CKMR application. The result suggests that there is a split between species where logistics and gaps in life history knowledge would be unlikely to compromise CKMR application, whereas for other species the consideration of CKMR application should be matched with investment in improving logistic and life-history knowledge to reduce any risk of compromise during implementation. The result in particular supports the implementation of both the billfish research plan and shark research plan.

SC22 is also advised to read SA-IP26. This paper notes that the genotyping completed to date and the number of half sibling pairs and parent offspring pairs observed indicates that these numbers are more closely aligned with the 2024 stock assessment population scenario used to inform the sampling design than other scenarios.

Both papers for P100d note that the delay in development of a cost effective epigenetic age methodology necessitates the collection of otoliths in the remainder of 2026 to provide age information for the CKMR model.

Both papers also note that progress towards provision of CKMR estimates for south Pacific albacore at SC23 is on track. The budget requested from WCPFC to complete this work in 2027 is USD115k.

TOPIC 12.	**Project 129 update: Building an age-length data stream for tuna assessments**
No questions or comments	

TOPIC 13.	**Estimation of tag mixing periods for the 2026 WCPO tuna stock assessments**
Related papers: WCPFC-SC22-2026-SA-IP10: J. Scutt Phillips, I. Senina and L. Bonin. Estimation of tag mixing periods for the 2026 WCPO tuna stock assessments	

Questions and Comments

New Zealand

Thank you to the authors for SA-IP-10, a substantial and useful piece of work. The move from a binary mixed/not-mixed judgement to a continuous measure is a sensible reframing, and the observed tag density plots (e.g. Figures 9 and 10) are a welcome addition.

Why these plots matter: Ikamoana now does more than describe mixing – it determines which tag recaptures enter the assessment likelihood, so the SEAPODYM movement model has become an assessment input as well as an ecosystem-modelling result, and now needs validation against observations to support this role. If the simulated movement is wrong, K will be wrong in a specific way: tags will be admitted when the fishing pressure they experience is not actually representative of the fishing pressure on the population.

This could happen: a) if dispersal is simulated too fast, so that dissimilarity falls faster than it should and mixing periods are set too short (the 2026 PAW noted SEAPODYM predicts greater diffusion than observed in the tag data); or b) because a real process, such as entrainment around islands/reefs or association with FADs, is missing from the simulation. Tag data enter the SEAPODYM likelihood aggregated to 6-degree cells, so such fine-scale processes cannot inform the fitted parameters regardless of dynamics resolution, and fish held near land or their release point by such processes would stay there longer than simulated, whatever the diffusion coefficient. Both mechanisms produce the same signature: tag density that stays elevated close to land or the release location for longer than simulated – exactly what the observed tag density plots could show. If the pooled plots show no such pattern, that would be a useful result too. This matters because no comparison currently tests the movement model against anything independent: the dissimilarity statistic compares two simulations, and comparing SEAPODYM with MFCL compares two models, neither of which is the truth. Tag recapture data are the only independent observation available, and tag density plots are the natural way to summarise them.

A few requests would help get more out of these plots:

1. Cell-inclusion criteria. Coloured cells are often widely separated, and it looks like a cell is shown only if it contains at least one recaptured tag – could you confirm the criteria? Cells with catch and no tags are also informative: they show fishing occurred without catching a tagged fish, which is exactly the contrast that indicates uneven tag distribution. Otherwise we would expect more coloured cells, not fewer, over longer periods (e.g. 270–1000 days). We suggest including all cells with catch in the time window (perhaps above a minimum catch threshold), not just cells with a tag.

2. Pooled plots. Individual release groups are too sparse to assess, but pooled plots, along the lines of SC9-SA-IP-06, could address whether tag density is systematically elevated near land or the release location – directly relevant to mixing, independent of the Ikamoana simulations, and a question about the systematic behaviour of release groups rather than any individual group, so pooling is the natural approach. To avoid confounding by the shared response of fish and fishing to a moving ocean, density should be estimated separately for each release group and time window and then averaged, not calculated as total recoveries over total catch. Because tag density is tags per tonne of catch, the ratio is largely unaffected when fish and fishing concentrate together on a feature. And because moving features lose influence under averaging, pooling is conservative. A pattern that survives it is likely real, and one that recurs in the same place across years is a persistent feature the assessment needs to account for.

We suggest pooled plots by assessment region: absolute coordinates for archipelagic regions, where proximity to land is fixed in space and releases are clustered; and, as a subsidiary, more demanding option, positions relative to release location for non-archipelagic regions, where releases are on TAO buoys/FADs and locations vary between cruises. Showing the number of contributing release groups per cell would help readers judge sampling coverage.

The most direct version of these plots would use only the release groups classified as sufficiently mixed under each mixing-period cutoff, since those are the tags that actually enter the assessment. Additional modifications are possible: (a) plot simulated Ikamoana density and observed tag density with each expressed relative to its own spatial mean, pooled over the same release groups and time windows, so predicted and realised distributions can be compared directly; and (b) summarise dispersal with the TART statistics of Kolody and Hoyle (2015), which give a formal, region-by-region comparison of simulated and observed spread. Both use data and methods that already exist.

Finally, observed recapture density is also affected by spatial variation in reporting rates. The SA-IP-05 estimates should make it possible to adjust for this – could you comment on whether this could account for any of the pattern seen?

Reply (SPC)

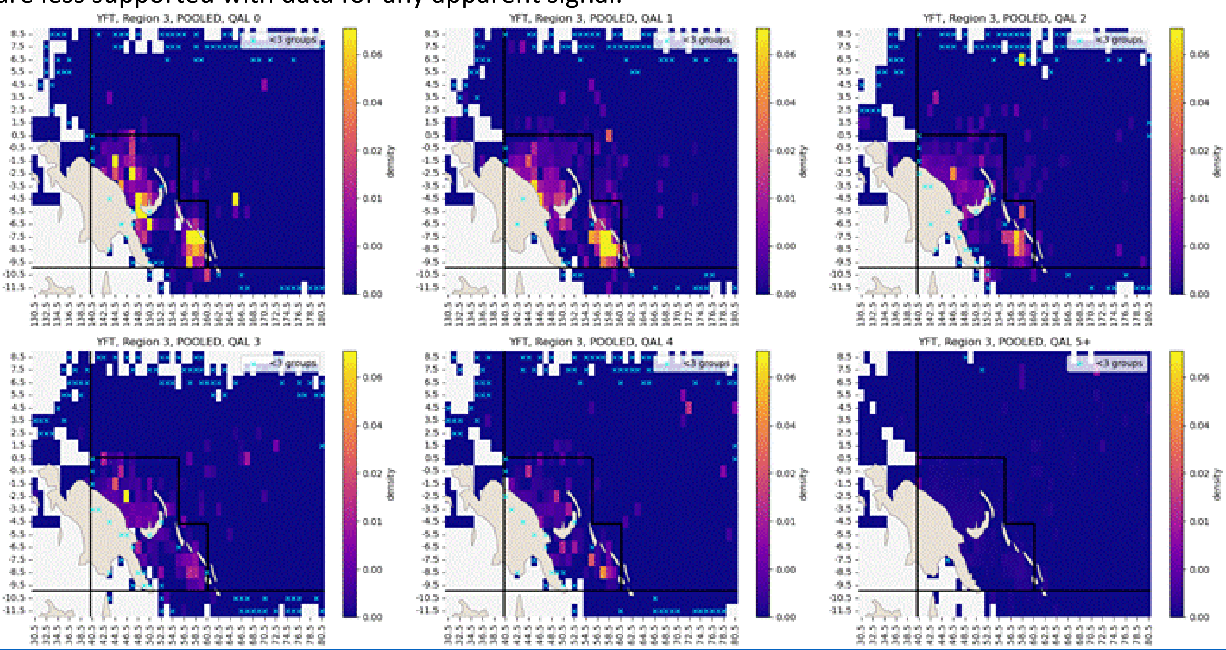
We thank New Zealand for the detailed technical comments and requests, and we have tried to collate and answer them here.

1. Regarding the validation of the SEAPODYM movement parameters, we direct those interested to the papers SC22-EB-WP-01, SC22-EB-IP-04 and SC21-IP-20, where diagnostic methods and validation scores are provided- noting that there remain areas for improvement, depending on the particular solution in question. It is incorrect to state that no comparison currently tests SEAPODYM movement models against anything independent, and such validation of the movement model against tagging data is presented in these papers, including held-out cross-validation for skipjack and out-of-sample tag validation for bigeye. We note that the authors of these papers have addressed related questions in more detail in the ODF responses to EB-WP-01. The suggestion that SEAPODYM and MFCL are somehow compared within this analysis is also untrue: the dissimilarity metric compares two recapture-probability distributions generated by the Ikamoana IBM under a single SEAPODYM movement solution. It does not compare SEAPODYM against MFCL.

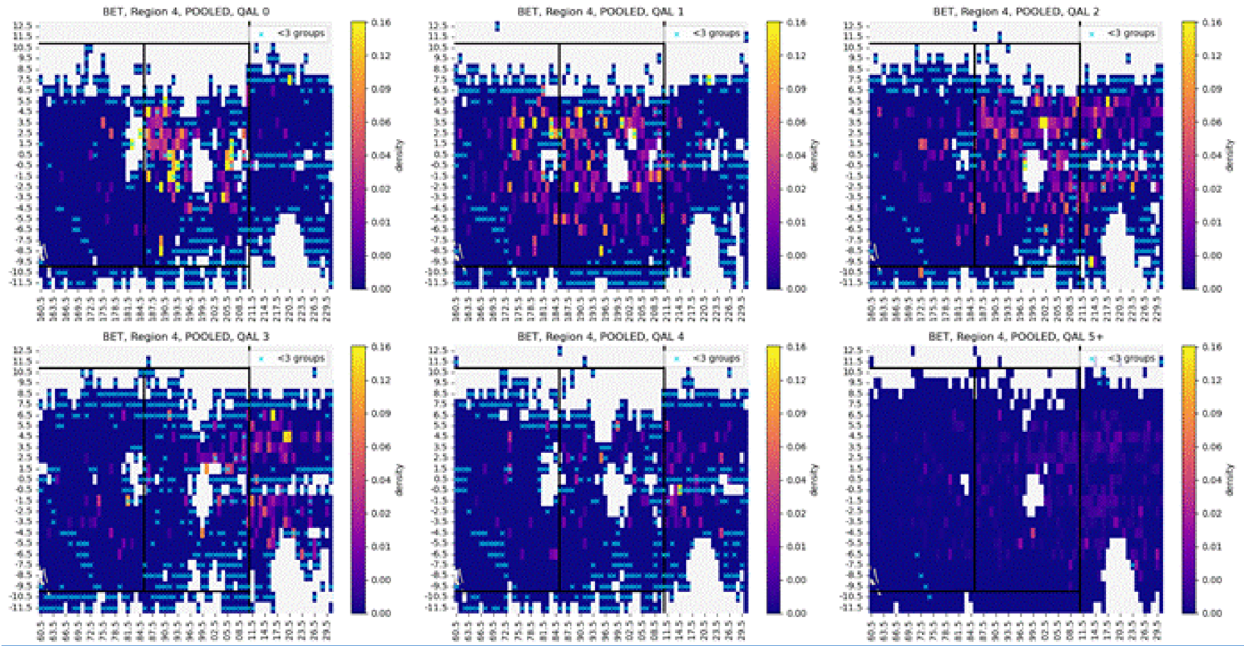
2. Regarding the impact of local effects on movement parameters, such as islands or FADs, we acknowledge the potential for their influence on tuna movement at larger scales, and that this may influence the appropriate admission of tags into the MF-CL likelihood. It is important to note that this indirect influence of SEAPODYM on MF-CL assessments remains just one of a number of auxiliary analyses which inform the number and nature of tagging data, alongside many other data sources, that are included into the assessment (e.g. tagging condition effects SA-IP-04). However, we remain doubtful in the ability of any currently available movement model, estimated on real data, to model such influences at the assessment scale. The density of coastal regions, as well as anchored FADs, in the archipelagic region is such that at one degree density, descriptive statistics or plots of recapture tag densities are unlikely to reveal consistent patterns in recapture because this is confounded with the vast majority of release locations that coincide with these same areas. This confounding of release location with coastal/FAD proximity affects any method applied to these data, observational or simulation-based, and not only the one that we present here. Furthermore, given that the archipelagic area exhibits a unique combination of environmental variables, disentangling the impacts from the "island" or any dense FAD network with environmental drivers of movement will be challenging when using any estimation of diffusion/advection/distance attrition. There are no comparable areas where environmental data differ, but similar high density of islands and anchored FAD networks exist, with which to compare such results. Reporting rate also introduces further problems when interpreting data, as in local coastal areas of the archipelagic are reporting rate estimations are highly uncertain due to the increased small-scale and artisanal catch, further confounding the effects of any possible bias in a movement exhibited in these areas from tag recaptures.

3. The justifications for improved tag density plots are well taken, particularly the inclusion of catch-positive, recapture negative cells. However, the data still remain extremely sparse for almost, if not all, individual release groups, and drawing conclusions from recapture density plots remains challenging. Pooling release groups using a mean of each release groups' recapture density (tags recaptured/tonne of catch) does introduce a smoothing of the data, but a strong signal of consistently

higher recapture densities along coastlines may be possible to identify. To this end, we are providing a revised version of SA-IP-10, with pooled mean recapture densities for the species/region combinations that represent the most significant sources of PTTP tagging data. These recapture densities exclude catch from longline, due to differences in data reporting, ignore any contributions from cells that experienced < 1 tonne of catch during a particular quarter strata, as well as apparent recapture densities above the 98th percentile, typically driven by few numbers of recaptures in very low catch cells, which we here consider to be noise. Similarly, cells for which fewer than three release groups contributed data are also highlighted with crosses, to indicate that they are less supported with data for any apparent signal.



The pooled mean recapture-density surfaces for region-3 yellowfin release groups show density elevated near the release area, persisting for roughly one to two quarters at liberty and relaxing thereafter as tags disperse, notably eastward and out of the archipelago into more oceanic waters. This near-release elevation is carried by cells with multiple contributing release groups (i.e. it is systematic across groups, not an artefact of individual groups), and includes some moderately high recapture cells even at four quarters. However, the surfaces do not resolve into a consistent, spatially-fixed pattern of elevated density along particular coastlines or at recurring locations: the elevation tracks release location rather than recurring at fixed coastal cells, and because essentially every fished cell in the archipelagic region is close to land, there is no non-coastal fished area within the region against which a distinct proximity-to-land effect could be contrasted. The pattern is therefore consistent with incomplete near-release mixing that relaxes with time at liberty, but does not provide evidence for, nor can it isolate, a specific island- or FAD-driven stickiness effect.



The pooled density for all bigeye tuna release groups in region four shows a highly patchy surface, even at three quarters time-at-liberty, driven mainly by more heterogeneous fishing effort as evidenced by the increased instances of cells for which less than three release groups contained catch data. Note the significant, consistent recapture density in the EPO, which may influence the assessment results of SA-WP-06 over estimating tag recaptures through time, as observed EPO recaptures are removed from the release group in the assessment. As with the results for yellowfin tuna in region 3, some consistently high density cells exist at 4, and to some degree 5 quarters plus time-at-liberty, despite this area having no archipelagic waters nor dense network of anchored FADs.

4. Calculation of the proposed distance-from-release gradient is similarly challenging for the same reasons. Particularly for the species explored here, individual release group recapture densities are sparse, but pooling releases groups confounds the many and spread-out release locations across these groups against any clear signal that may be related to landmasses or dense networks of anchored FADs. Given resources available, and the confounding nature of release location, reporting rate, and variable environment with the local impacts being examined, we feel that these analyses will not yield insightful results for the two species being assessed at SC22 and hence are not seen as priorities within the existing workload.

The results of the simulation study presented here remain just that, a simulation study. They are based on an independently estimated and validated model, but which nevertheless exhibits its own uncertainties and assumptions. There is no doubt that the requirement of a mixing period assumption within MF-CL stock assessments appears highly influential on the stock status indicators for WCPO tunas. The purpose of this work is to quantify the degree of difference between potential mixing across tag release groups, at the dimensionality that they are aggregated within the assessment, and to provide a sliding scale that can be used by stock assessment scientists to bracket this uncertainty in MF-CL results. The additional tag recapture density calculations we have provided here form a further, useful contextual layer for the question of tag mixing, despite having no dissimilarity K estimates for yellowfin tuna for comparison, and those figures for bigeye being largely uninformative for selecting a mixing period. We hope that New Zealand finds them useful.

TOPIC 14. Project 120: Revisiting the Reproductive Biology of Western and Central Pacific Ocean Yellowfin Tuna (Thunnus albacares)

No questions or comments

TOPIC 15. Project 122a: Global Joint Tuna RFMO Longline CPUE Technical Workshop

No questions or comments

TOPIC 16. Project 125: Biology of South Pacific striped marlin, blue marlin, black marlin, shortbill spearfish, and sailfish in the WCPO from longline fisheries

No questions or comments

TOPIC 17. Projects 117,118 and 126: Biological sampling plans for tuna, billfish and sharks

No questions or comments

TOPIC 18. WCPFC Project 130: Joint bycatch assessment workshop for billfish and sharks

Related papers: WCPFC-SC22-2026-SA-IP21_Rev01: P. Neubauer, P. Hamer.WCPFC Project 130 — Joint bycatch assessment workshop for billfish and sharks: progress and outcomes of the first steering group meeting

Questions and Comments

Sharks Pacific

Sharks Pacific would like to thank Drs Neubauer and Hamer for leading the development of this long overdue workshop on bycatch assessment for billfish and sharks. We note the importance of this workshop not only to generally advance better methodologies to assess bycatch and non-target species, but also to help review and identify potential candidate stocks for CKMR-based abundance estimation. We would like to recommend that Project 130 consider scoping various monitoring strategies for sharks and billfish, so that non-target species are not managed only as a residual assumption within tuna-focused MPs. Lastly, Sharks Pacific looks forward to participating or supporting this workshop in whatever capacity it is able.

TOPIC 19. SEAPODYM modelling framework: methodology, validation, and analyse

Related papers: WCPFC-SC22-2026-EB-WP01: I. Senina, L. Bonnin, R. Forestier, and S. Nicol. Advancing the SEAPODYM modelling framework: methodology, validation, and analyses

WCPFC-SC22-2026-EB-IP04: I. Senina, L. Bonnin, R. Forestier, M. Lengaigne, K. Guihou, J. Temple-Bouyer, and S. Nicol. Updated SEAPODYM reference models for four Pacific tuna species

Questions and Comments

Chinese Taipei

I am just curious about whether SEAPODYM model could apply to sea turtles distribution research or not.

Ren-Fen

Reply (SPC)

The underlying equations can be used to model sea turtle habitat and movement. Previously we did this for juvenile loggerhead turtles in the North Pacific using MOVEMOD, a single-cohort simplified version of SEAPODYM (Abecassis et al., 2013, A Model of Loggerhead Sea Turtle (Caretta caretta) Habitat and Movement in the Oceanic North Pacific). Note, however, that the current framework is specifically designed for modelling tuna population dynamics and has no single-cohort module, though one could be added.

Response to SPC (Chinese Taipei)
Merci beaucoup, Inna. This information is very helpful. Ren-Fen

New Zealand

Thank you to the authors for the substantial work behind EB-WP01 and EB-IP04.

We raise four items on these two papers below; comments on the use of SEAPODYM outputs in the 2026 assessments, including SA-IP-10, are posted separately under the stock assessment item. Given the scope of the validation material, some of the questions below may simply reflect places where we've missed something.

1. Validation against observations (Comment)
New Zealand thanks the authors for the validation framework set out in EB-WP01 Section 8. The out-of-sample tag scores in Section 8.2.3 are equal to or better than the in-sample scores. We would be interested to know whether any of the tag validation is spatially independent, for instance whether the EPO recapture groups were withheld from estimation, since spatial hold-out would test the movement model more strongly than cross-validation by recapture group alone. Section 8.5 also reports a cross-basin transfer test, where the Pacific albacore solution reproduced Atlantic dynamics, without re-fitting parameters, at a level of skill comparable to an Atlantic-fitted baseline. This is a demanding test of structural transferability. Would SPC consider extending it to a tropical tuna?

2. Forward prediction and out-of-window diagnostics (Question)
Section 8.5 states that the full set of out-of-window diagnostics for skipjack, yellowfin and bigeye is reported in EB-IP04, but Table 3 reports scores only over the common 1986 to 2010 window; we couldn't locate scores for the validation periods defined in EB-IP04's Table 1 (e.g. bigeye 1960 to 1994, skipjack 2013 to 2022). Separately, even where reported, within-window scores measure fit rather than forecast skill. Given SEAPODYM's role in forward projection, we'd find a hindcast (truncate, refit, score forward predictions via MASE; Kell et al. 2016, Carvalho et al. 2021) more informative for how heavily the 2050 projections should be caveated. We'd also note Section 8.2.1 confines CPUE validation to EEZ scales within each species' current core area, which the projections expect to shift.

Could the Table 1 scores be reported please, and has hindcast cross-validation been considered (and if so, which quantities would be re-derived within each peel, given Section 5 selected the skipjack ceiling using the 2013 to 2022 period)?

3. Uncertainty in stock scale and recruitment (Comment and question)
Section 7.2 reports a condition number of 3.1 x 10^7 for the full joint skipjack fit, above the paper's own threshold of 10^6 for severe ill-conditioning, traced to a saturated prey coefficient in the spawning habitat and to correlated recruitment-scale, larval-catchability and mortality parameters. We note the movement and habitat parameters carry none of this. Section 5 shows the profile against stock size flattens rather than reaching a minimum, so the stock ceiling (demonstrated for skipjack on the WCPO) must be selected rather than estimated; Section 9 notes this scale is configuration-dependent for all species and must be re-validated when forcing or resolution changes, but we couldn't find the equivalent profile or ceiling reported for bigeye, yellowfin or albacore.

We'd also note the catch-removal formulation enforces biomass at least equal to the reported catch as a one-sided penalty: an overstated catch forces the model to supply biomass that wasn't there, while an understated one costs nothing, and Section 5 identifies the free-school purse-seine fishery as the binding constraint on the skipjack ceiling.

Could standard errors be provided for the EB-IP04 Table 2 reference parameters please, and is the suggested remedy for high uncertainty planned? How sensitive is regional scaling to the choice of ceiling across the range in Figure 4? And what is the form and weight of the catch-removal penalty, noting the bigeye high-latitude overestimation reported in EB-IP04 Section 4.1.3: how does that overestimation affect the utility of the bigeye model for regional scaling and regional recruitment distribution?

4. Validating movement predictions, including in archipelagic waters (Comment and question)
EB-WP01 Section 8.2.3 validates movement through recapture maps, profiles and cross-validated scalar metrics, but aggregate scores can mask directional error. SC21-SA-IP-20 reported that meridional skipjack movement is poorly reproduced, that the model under-predicts the northernmost extent of feeding grounds, and that a zonal bias displaces EPO tags east of where they were observed. A model can disperse tags the right distance in the wrong direction and still score well. We'd expect this to be compounded by SEAPODYM's lack of FAD/island attraction (which SPC noted at SC21 enters only through coarse oceanographic forcing): if retention is absent, predicted dispersal should exceed observed dispersal, with the gap growing in archipelagic waters, where many Pacific EEZs sit. We also note tag recaptures are modelled aggregated to 6-degree squares, even though release position and date must be exact as initial conditions for the advection-diffusion equations for tagged fish.

Does the Section 8 validation include a diagnostic for systematic directional bias, for example spatial residual maps alongside the scalar metrics, and could reliability be reported separately for open-ocean versus archipelagic waters and by quarter since release?

USA

We have a few follow-up questions for the authors, one recommendation and one request for additional work. We note that the authors identified that total population scale is not estimable but is an emergent property of balancing a total population scale cap with producing a large enough population to explain the spatial catches. They also describe that the maximum stock constraint was only applied to the WCPO and not the EPO. Our question is two fold, given that these are basin wide models, how sensitive are outcomes to alternative assumptions made to either WCPO and/or EPO stock constraints? Secondly, how sensitive are key inputs for stock assessment, regional scaling and bulk transfer coefficients, to assumptions made when determining the total population scale?

In the presentation of the Hessian convergence diagnostics for the example SKJ joint model the authors note that while the model appeared to converged to a minimum based on the Newton step diagnostic, their condition number diagnostic indicated that the solution was very poorly defined, and assuming this was the same model as presented in Section 3 had a number of parameters on a bound. Do the authors have any indication of sensitivity in model outputs to alternative parameter values for those that are either poorly determined (highly correlated) or estimated to be on a bound?

Reply (SPC)

Thank you for these questions and recommendations.

Q1. The MLEs reported in SC21/SA-IP-20, SC22/EB-WP-01 and SC22/EB-IP-04 used the WCPO-wide stock constraint. We constrain the WCPO because once applied to the WCPO the stock penalty bounds the stock size used by the assessment directly, applied to the whole Pacific it bounds the basin total and leaves the WCPO share free to adjust. However, this was not the arbitrary choice – we have also tested the Pacific-wide constraint and ran the profile of Section 5 with for the JRA55-forcing models (not reported in the SC22 papers). We found that within a fixed region, the level of the ceiling only sets the absolute scale and does not change the spatial distribution. Thus, the stock constraint was set at 10500, 11000, 11500 and 12000 thousand mt, reached in every optimisation, and the objective decreased monotonically from 5.01 to 4.85 (at these levels, the likelihood starts to saturate hence the change in likelihood is small). The improvement was mainly observed for the catch likelihood, which fell by 10-20%, while the length, tagging and effort components changed by 2% or less, indicating the reduction of the errors was coming from the catch-removal fisheries. Across that 14% change in total biomass the relative spatial distribution remained the same.

The region over which the constraint is applied matters more. Comparing two reference models with the same estimated parameters and the same fixed values, one with a WCPO ceiling of 7300 thousand mt and one with a Pacific-wide ceiling of 11500, we found that the latter configuration led to an increased (by 3.6%) WCPO stock and decreased (by 27%) EPO stock. This resulted in the total likelihood deterioration by only 1%, however the catch component got worse by 8% (attributed to the EPO catches) and tagging likelihood got worse by 1.5%.

Nevertheless, the model with PO constraint was selected as one of the alternative models and added in the climate change impact estimation and therefore we could verify the sensitivity of regional scaling and bulk transfer rates to selected stock constraint. Despite a substantial EPO stock decrease in the PO-wide constrained model, the spatial distribution in the WCPO area did not change, so both, regional scaling and bulk transfer rates in the WCPO are robust to the choice. Shares of the WCPO biomass over the estimation window, for the two configurations above, are:

area WCPO 7300 PO 11500 difference

120–210°E, 20–45°N 12.8% 14.1% +1.3

140–170°E, 20°S–20°N 38.8% 39.9% +1.1

170–210°E, 20°S–20°N 36.1% 33.2% −2.9

120–140°E, 20°S–20°N 12.3% 12.8% +0.5

Residency (diagonals of movement matrices) differs by at most 2.2 percent in any region, and across age. Such effect of the constraint region on the transfer rates is of the same order as the differences between movement rates across six decades of 1960-2022 runs.

Q2. The skipjack model used in the Bridging Analysis (Section 3) has five parameters on the boundary, some close to it, out of 36 estimated. Regarding the ill-conditioning, it is not uncommon in non-linear optimization problems, where the objective function can be very steep in one direction and flat in another, to obtain a large condition number. However, a large condition number can be produced by an exceptionally well-determined direction as well as by a poorly determined one, indicating that the surface is strongly anisotropic rather than poorly determined. In other words, the condition number indicates the presence of a flat direction, but the flatness is always expressed relatively to the stiffest direction, so a Hessian-based severe ill-conditioning does not always warrant that there is an issue. The symptoms of ill-conditioning, such as sensitivity of the MLE to initial parameter values or to small perturbations in the model structure or in the data inputs, are much more important indicators, and we do not observe them for the skipjack model: the optimizations converged to the same MLE from different starting points and under small perturbations in the model structure.

In the skipjack refence model, two parameters are responsible for severe ill-conditioning, and both control recruitment and are informed by the fisheries data: the shape parameter alpha in the normalised Holling type-III functional response of larvae, which has the flattest likelihood surface at MLE, and the parameter defining the range of local densities of predators within "safe" spawning habitat, beta, which has the stiffest likelihood surface and usually well-determined. Removing them from Hessian, moves the condition number from 3.1e7 to 2.3e6, the "moderate" range.

Why not removing flattest-direction parameter from MLE? In case of alpha it is not poor identifiability, but one-sided identifiability and the likelihood becomes flat only at the high end of its range. Skipjack is sensitive to the primary production (as a proxy of prey organisms), as demonstrated by the Bridging Analysis. In optimizations with the full model and joint likelihood, quasi-Newton converged to the upper limit of the feasible range for this parameter in every run of the reference configuration, however from the normalisation of the prey-function, once alpha exceeds alpha=3, the likelihood surface flattens. To measure the flattening we have run a likelihood profile of alpha, fixing it at 3, 5, 8 and 11 and re-estimating all remaining parameters at each node. Relative to the solution with alpha free, the objective function is higher by 0.117%, 0.063%, 0.024% and 0.002% respectively. So, even if the reported value of alpha is near the boundary, it is still estimated, and the boundary is fixed where the likelihood can't gain further. Note, since the parameter transformation into unbounded space changes the likelihood gradient near the boundary, the test was done by moving the upper limit to 15. As a result, the gradient with respect to alpha is unchanged and the Hessian remains positive definite. To resume, data indicate alpha should be large, but how large was chosen by the estimation.

The other four are the epipelagic accessibility coefficient and the variability of diffusion with habitat index, both at their upper bounds, and the slope of the temperature-age relationship together with the coefficient of the thermal habitat function, both at their lower bounds. Each is a shape or scale coefficient of a bounded functional form, and each is one-sided in the same way: the data determine the direction, and the bound marks the limit of the range we considered feasible.

TOPIC 20. Project 110 and 110a: Non-entangling and Biodegradable FAD Trial in the Western and Central Pacific Ocean

Related papers: WCPFC-SC22-2026-EB-WP02: Escalle, G. Moreno, I. Zudaire, J. Uranga, D. David, and P. Hamer. Progress Report of Project 110 and 110a: Non-entangling and Biodegradable FAD Trial in the Western and Central Pacific Ocean

Questions and Comments

<u>Tuvalu</u>

COMMENT: PNA and Vanuatu have the following comments on the paper WCPFC-SC22-2026-EB-WP02**.**

PNA and Vanuatu strongly support the work of Project 110. We thank SPC for leading the Project and thank all who have supported and contributed to the trials. We generally support the recommendations on the top of page 20 and the related Project set out in the Updated FAD Research Plan. We particularly support the focus in the recommendations on establishing arrangements for monitoring the use of biodegradable materials and clarifying the standards for biodegradability.

On the recommendations:

a) We suggest that the recommendation on relative catches in the 3rd dot point on the bottom of page 20 should be reworded for the SC Report to reflect more closely the data in Table 6. One option is to use the median data e.g. “The median catch per set for jelly-FADs was 25.5 mts/set compared to 43.5 mts/set for conventional FADs in the trial and 30.0 to 35.0 mts/set on conventional FADs for the entire WCPO fleet in 2023 to 2025”. Alternatively, the average data could be used.

b) PNA and Vanuatu have a particular concern about the issue of the availability of biodegradable FAD materials. The paper helpfully reports on page 8 that the lack of bamboo availability is a major challenge in using Majuro as a base port for non-entangling and biodegradable dFADs. We note that the same constraint applies to Funafuti, Tarawa and Kiritimati which are all major ports for the PNA purse seine fleets. We propose to add the following text to the 4th dot point in the recommendations on the bottom of page 20 for adoption by the SC.

“… *except that the lack of bamboo availability was identified in the trials as a major challenge in using atoll country ports as base ports for biodegradable dFADs. This may require consideration of alternative designs and some flexibility in the implementation of requirements for biodegradable FAD materials*.”

TOPIC 21. Proposal to Amend the ROP Minimum Standard Data Fields and SciData to Support Sea Turtle Conservation and Management

No questions or comments

TOPIC 22. Ecosystem and Climate Indicators (Project 121)

Related papers: WCPFC-SC22-2026-EB-IP01 : SPC-OFP, CSIRO. Project 121 Update: Ecosystem and Climate Indicators of the western and central Pacific Ocean

Questions and Comments

USA

The U.S. appreciates the continued work on these important indicators and recommends that the authors switch from ONI to the RONI (https://www.cpc.ncep.noaa.gov/products/analysis_monitoring/enso/roni/) next year as that is the new NOAA official ENSO index that replaces the ONI. Additionally, regarding the reference to www.climate.gov/climatedashboard in the ENSO section, the U.S. notes that this is now an archived website that is not updated, suggesting a change in this reference moving forward.

TOPIC 23. Implementation of the Ecosystem Approach to Fisheries Management in tuna RFMOs

No questions or comments

TOPIC 24. Development of early warning tools and expanded climate impact assessments

No questions or comments

TOPIC 25. Updated assessment of Black-footed Albatross overlap with pelagic longline fisheries in the North Pacific

Related papers: WCPFC-SC22-2026-EB-IP16: Thomas A. Clay, Alyssa Piauwasdy, Michelle Hester, Jessie Beck, Scott A. Shaffer, André F. Raine, K. David Hyrenbach. Updated assessment of Black-footed Albatross *Phoebastria nigripes* overlap with pelagic longline fisheries in the North Pacific

Questions and Comments

Birdlife International

Paper IP16 presents new comprehensive tracking data, encompassing 84% of the global breeding population of the Black-footed Albatross illustrates that 97% of its overlap with WCPFC vessels occurs north of 23°N making the effectiveness of this Commission's mitigation measures the primary determinant for reducing mortality.

BirdLife is concerned by Members operating north of 23°N, who have reported very high seabird bycatch rates. The current provisions of CMM 2025-05 for the Northern Hemisphere are unchanged and include measures, such as blue-dyed bait, for which there is limited evidence of effectiveness.

While we appreciate the efforts of the Commission to amend the Seabird CMM last year for the southern hemisphere; however, BirdLife emphasises that the SC must urgently address aligning measures for North of 23N in CMM 2025-05 with ACAP best-practice advice . The analysis has already been completed by New Zealand, and recommendations made for amendments for this fishing area. We encourage the SC to recommend these be adopted to address excessive seabird bycatch in the North Pacific.

TOPIC 26. Single region fleets as areas models for bigeye and yellowfin tuna in the WCPO

No questions or comments

TOPIC 27. Impacts of regional growth variation on the 2023 bigeye tuna stock assessment

No questions or comments

TOPIC 28. Update and workplan on FAD Research

Related papers: WCPFC-SC22-2026-EB-IP06_Rev01: Lauriane Escalle, Jennyfer Mourot and Paul Hamer.Update and workplan on FAD Research

Questions and Comments

USA

The United States appreciates the comprehensive work and SPC's leadership in advancing the viability of an eventual transition to more biodegradable components to the expanding dFAD industry in both the EPO and WCPO.

Regarding future research, the US is supportive of the collaborative work of SPC and ISSF regarding WCPFC Projects 110 and 110a to conduct trials of non-entangling and biodegradable dFADs in the WCPO. Additional data will allow for more robust comparisons of material performance. We also encourage continued collaboration with industry to test biodegradable materials and designs with assistance from SPC to lead their transition through advice, trainings/ workshops and scientific analyses of the performance of dFADs.

The US is interested in the project to assess the feasibility and cost-effectiveness of a range of different dFAD recovery options to reduce the environmental and economic costs of lost, stranded and abandoned dFADs.

As a note, our industry believes there is merit in pursuing all the options presented, however the following two are believed to be most successful: Option3 (yet noting the importance to identify the most vulnerable areas and focus efforts in these areas) and Option 5.

Tuvalu

PNA and Vanuatu thank SPC for the Update and Workplan on FAD research and the value of the ongoing programme of work of FAD Research We broadly support the Plan, including the list of papers for SC23, the proposed workplan for 2027-28, and the TORs in Appendix 2 with a few comments as set out below. We attach high priority to further analysis of FAD buoy transmission data and to further work on biodegradable FADs including clarifying the standards for biodegradable materials. We also support the priority attached to work on stranding and FAD recovery, particularly the work on monitoring and mitigating the impact of FADs outside the main fishing grounds and the valuable support provided by SPC to local FAD recovery programmes.

PNA and Vanuatu have 4 further comments on elements of the Workplan.

1. We suggest the Workplan could include preparation of a set of FAD fishery related indicators. PNA and Vanuatu would welcome SPC providing a proposal for the development of a set of FAD fishery indicators to SC23.
2. Secondly, PNA and Vanuatu are not able to support the work on the legal framework referred to on page 7 being continued within the FAD Research Workplan. We appreciate that this work is being done for SPC Members and not the Commission and is valuable for that purpose. From the preliminary results presented at SC21, it seems that the work is based on legal opinions that are not consistent with applicable national laws and international law. We request that work on the legal framework should not be considered as endorsed by the Commission or part of the Workplan for the Commission.
3. PNA and Vanuatu support the suggested papers for SC23, including the proposed SC23 paper on mitigation of FAD impacts. We appreciate that the proposed work is focused outside the main fishing grounds and support that focus. We consider it may be a bit early to be looking for recommendations on that work and suggest the text on page 12 should refer to "*further analysis of mitigation strategies*" rather than *"recommendations for mitigation strategies"*.
4. On the assessment of FAD use per vessel, we suggest priority on analysis of deployments rather than active FAD buoys monitored per vessel. With the improved coverage of the FAD buoy tracking data, we hope it will be possible to prepare some reasonably reliable estimates of simple trends in the average active FADs monitored per vessel without too much additional analysis.