SCIENTIFIC COMMITTEE
TWELFTH REGULAR SESSION

Bali, Indonesia

3–11 August 2016

Retrospective forecasting of the 2014 WCPO bigeye tuna stock assessment.

SC12-SA-WP-02

R. Scott[1], N. Davies[2], G. M. Pilling[1] and J. Hampton[1]

[1]Oceanic Fisheries Programme, Secretariat of the Pacific Community
[2]Te Takina Ltd. Whangerei, New Zealand

1 Executive Summary

Retrospective forecasting, also known as hindcasting and backtesting, is a method for testing the prediction performance of an assessment model using existing historical data. The approach is based on a retrospective analysis with the additional step that each assessment is then projected through to the end of the original time series. The validity of projections that are made from assessments that are subject to retrospective bias is a significant concern since catch and effort limits that are designed to meet management targets can be systematically under- or over-estimated. Ultimately this can lead to drastic management revisions being required. Such revisions reduce the ability of managers to manage risk because they indicate a source of uncertainty that has not been fully accounted for.

We conducted a retrospective analysis of the 2014 bigeye tuna reference case assessment for the period 2012 to 2007 and a subsequent hindcasting analysis using both deterministic and stochastic projection approaches to determine the robustness of the stock assessment to varying quantities of data and the quality of the projections in terms of their ability to provide consistent estimates of stock status.

Our results show no evidence of any systematically increasing over- or under-estimation of bigeye adult biomass with successive assessments. The ρ adjusted adult biomass value (293909) was within the approximate 68% confidence region (228044, 309339) and therefore well within the 90% region suggesting that bias correction of the assessment estimates was not necessary. Estimates of stock status in 2012 and recent fishing mortality from the deterministic projections consistently indicate the stock to be below the limit reference point and to be exploited at levels in excess of F_{MSY}. The median estimate of adult biomass in 2012 from stochastic projection of the 2007 retrospective lies just outside the 95% confidence interval of the equivalent value estimated by the reference case assessment.

We therefore conclude that:

1. the 2014 bigeye tuna stock assessment model is not subject to significant retrospective bias;
2. short-term deterministic catch based projections, conducted from the 2014 bigeye tuna reference case assessment model, provide consistent and relatively accurate indications of stock status in the short-term;
3. short-term stochastic projections for bigeye tuna will potentially under-estimate overall uncertainty in stock status in the first few years.

We invite the Scientific Committee:

1. to note the results of this analysis and consider them in the context of the provision of short-term management advice;

2. to note the future developments planned for MULTIFAN-CL that will improve estimates of uncertainty and risk in the short-term;

3. to note the importance of retrospective analyses as a diagnostic tool for WCPFC stock assessments; and

4. to note that integrated assessments (such as MULTIFAN-CL) may exhibit different characteristics to virtual population analysis (VPA) based assessments under conditions that generate retrospective bias.

2 Introduction

Estimates of stock status derived from fish stock assessment models are subject to uncertainty arising from a number of sources. Rosenberg and Restrepo (1994) identify the following 5 different types of uncertainty that are present in any stock assessment: (1) Measurement error is the error in observed quantities such as the total catch or length distributions; (2) Process error is the underlying stochasticity in the population dynamics (eg. recruitment variability); (3) Model error arises from the misspecification of model parameter values (eg. steepness or natural mortality) or the structural specifications of the model (eg. spatial configuration); (4) Estimation error is related to the inaccuracy and imprecision in the estimated model parameters; and (5) Implementation error resulting from the inability to fully implement management decisions (implementation error is less important for the stock assessment than for the evaluation of management strategies).

The different types of uncertainty can impact on fisheries management advice in different ways and a variety of approaches have been developed to address them. For example, integrated stock assessment methods (such as MULTIFAN-CL) that are built around formal statistical likelihood functions permit the uncertainty associated with the input data (types 1 and 4) to be propagated to final model outputs (Maunder and Punt, 2013). The "uncertainty grid" of alternative assessment models, that are routinely presented with each new assessment for WCPFC stocks, addresses concerns regarding model mis-specification (type 3) and stochastic projections account for process error (type 2) by using historical variation in for example recruitment as a measure of future variability.

Whilst it is unreasonable to expect that stock assessments will be 100% accurate, it is not unreasonable to expect that they should provide consistent estimates from one year to the next, and specifically that model estimates do not show persistent trends of under- or over-estimation over time. When updated parameter estimates display a persistent trend in relation to previous estimates it suggests that something may be misspecified in the model. Systematic error of this kind is typically referred to as retrospective bias (Sinclair et al., 1991). For this reason a retrospective analysis is typically conducted for each new assessment whereby the final assessment model is refitted to a progressively truncated time series of data (ie. the terminal year of the assessment is iteratively moved backward). Mohn (1999) proposed a metric (Mohn's ρ) to measure the extent of retrospective bias and to facilitate comparisons of retrospective bias between different assessments. More recently a practice has developed for adjusting biomass estimates for retrospective bias (Deroba, 2014).

The validity of projections that are made from assessments that are subject to retrospective bias is a significant concern since catch and effort limits that are designed to meet management targets can be systematically under- or over-estimated ultimately leading to drastic management revisions eventually being required. Such revisions reduce the ability of managers to manage risk because they indicate a source of uncertainty that has not been fully accounted for. Retrospective forecasting (Brooks and Legault, 2015), also known as hindcasting and backtesting, is a method for testing the performance of a predictive model using existing historic data. The approach is based on a retrospective analysis with the additional step that each assessment is then projected through to the end of the original time series.

The most recent reference case assessment of bigeye tuna in the Western and Central Pacific Ocean indicates that spawning biomass has decreased below the limit reference point. The stock is considered to be overfished with overfishing occurring. The design of measures to rebuild the stock to biomass levels above the limit reference point will be informed by analyses that rely on projections (eg. OFP (2015)). It is therefore important that we fully understand the extent of retrospective bias in the WCPO Bigeye Tuna stock assessment and its potential impact on projections of future stock status.

3 Methods

The analysis used the most recent WCPO Bigeye Tuna stock assessment (Harley et al., 2014) as conducted in 2014. The MULTIFAN-CL model was re-fitted to input data that had been successively truncated by one year from 2012, the terminal assessment year, to 2007. This involved the removal of catch and effort data for the terminal years from the *.frq* file, the removal of tag release and recapture data for the terminal years from the *.tag* file and appropriate modification of the *.ini* file for the restructured tag data.

Each assessment was run with the same settings and phases as for the 2014 reference case assessment. The final phase (phase 10) of the fitting procedure ran for 3,000 function evaluations, however, inspection of the results showed that some of the retrospective assessments achieved relatively poor convergence, having high values for the maximum terminal gradient of the fitted model. Each assessment was therefore run for a further 10,000 function evaluations (phase 11). The convergence criterion was set to a maximum gradient of less than 0.01 such that the fitting process would continue for either 10,000 evaluations or until a maximum gradient of less than 0.01 was achieved, whichever was the sooner.

Retrospective bias in the terminal estimates of adult biomass (Mohn's ρ) was calculated in accordance with Equation 1, where r is the number of years over which the retrospective analysis is conducted, $\theta^{*}_{\hat{Y-i}}$ is adult biomass in the terminal year of the retrospective assessment and $\theta_{\hat{Y-i}}$ is the corresponding value of adult biomass from the assessment using the full time series of input data. The larger the value of ρ the greater the retrospective bias. A positive ρ indicates a tendency to over-estimate adult biomass in the final year of the assessment whilst a negative ρ would indicate the tendency for under-estimation.

$$\rho = \frac{1}{r}\sum_{i=1}^{r}\frac{\theta^{*}_{\hat{Y-i}} - \theta_{\hat{Y-i}}}{\theta_{\hat{Y-i}}} \tag{1}$$

Both deterministic and stochastic projections were conducted for the retrospective assessments. The projections were run from the terminal year of each assessment through to 2012. Similar to the short-term stochastic projections presented to SC11 to advise on between-assessment status of tropical tuna stocks (Scott et al., 2015), the projections were based on catch for all fisheries using the actual catches that had been observed in each of the projection years.

Stochastic projections were conducted for each of the assessments that used truncated data (2011 to 2007). In each case, 200 projections were performed with variability in future recruitment implemented by randomly re-sampling from the historical recruitment estimates of the retrospective assessment over the most recent 10 year period, excluding the terminal year (ie. 2002 to 2011 for the 2012 assessment, 2001 to 2010 for the 2011 assessment, etc.). Stochastic recruitments were applied as deviates to the predictions of the stock recruit relationship. Catchability (which can have a trend in the historical component of the model) was assumed to remain constant for the projection at the level estimated in the terminal year of each retrospective assessment.

4 Results

None of the assessment runs achieved the convergence criteria of having a maximum gradient less than 0.01, although the maximum gradient was reduced (sometimes substantially) after running the final phase for an additional 10,000 function evaluations (Table 1). In spite of the improved convergence there was, however, little change to the final model estimates of adult biomass, F, etc. between those estimated in phase 10 and phase 11. The results presented here are those achieved after phase 11.

Table 1: Convergence statistics for the retrospective assessment fits showing for each retrospective assessment the number of parameters estimated by the model and the maximum gradient component after phase 10 (3,000 function evaluations) and phase 11 (an additional 10,000 function evaluations. Estimates of SB_{MSY}, F_{MSY} and $SB_{F=0}$ as estimated in phase 11 are also shown.

Terminal Year	N params	Max Grad (ph.10)	Max Grad (ph.11)	SB_{MSY}	F_{MSY}	$SB_{F=0}$
2012	8467	4.35	0.35	321300	0.035	1616744
2011	8303	8.76	0.13	291700	0.037	1430911
2010	8135	94.23	0.48	326300	0.036	1466691
2009	7967	2.46	1.11	300700	0.036	1416085
2008	7803	3.50	0.14	364500	0.036	1400761
2007	7633	2.16	0.19	310800	0.035	1341844

Estimates of adult biomass determined from each of the retrospective assessments show very similar trends over time (Figure 1) but are re-scaled throughout the time series relative to the reference case assessment. Estimates of adult biomass for most of the retrospectives fall within the 95% confidence interval of the 2012 reference case assessment although estimates from the 2011 retrospective fall

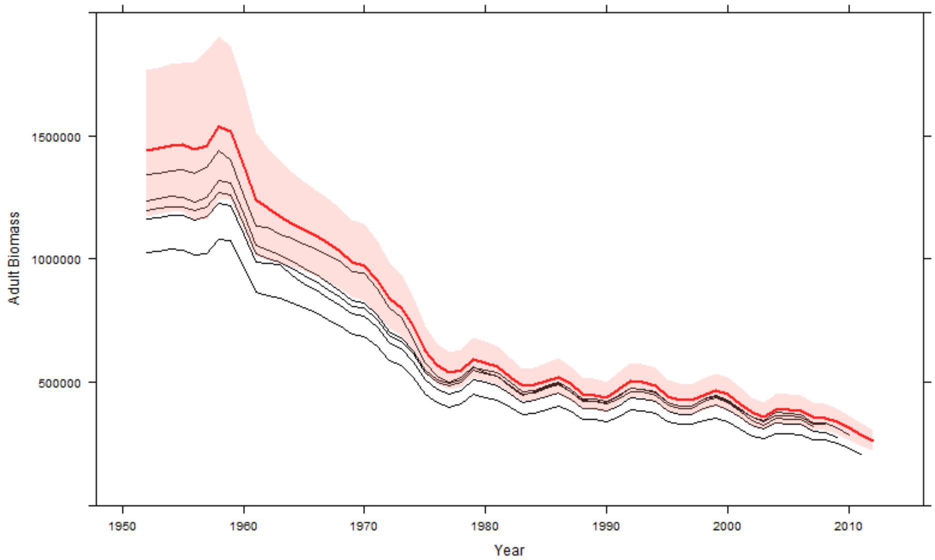

Figure 1: Retrospective results. Annual estimates of adult biomass (summed over areas) determined from the 2014 reference case assessment (1952 to 2012, red line) and retrospective assessment runs (terminal years 2011 to 2007, black lines). Shaded area shows the approximate 95% confidence interval for the reference case assessment estimates.

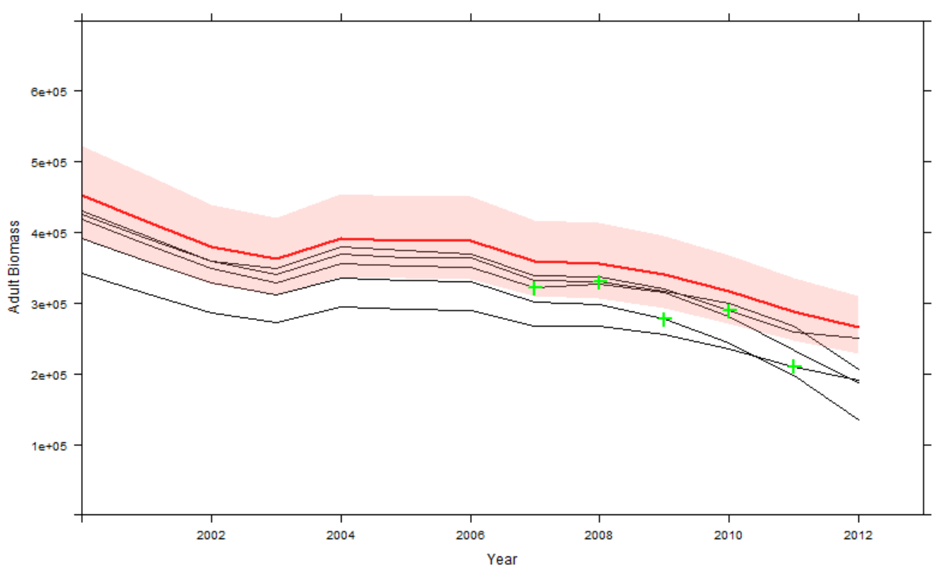

Figure 2: Hindcast results. Annual estimates of adult biomass (summed over areas) determined from the 2014 reference case assessment (2000 to 2012, red line) and retrospective assessment runs (terminal years 2011 to 2007, black lines). Shaded area shows the approximate 95% confidence interval for the reference case assessment estimates. Green crosses show the point at which the historical assessment ends and the deterministic catch based projection begins.

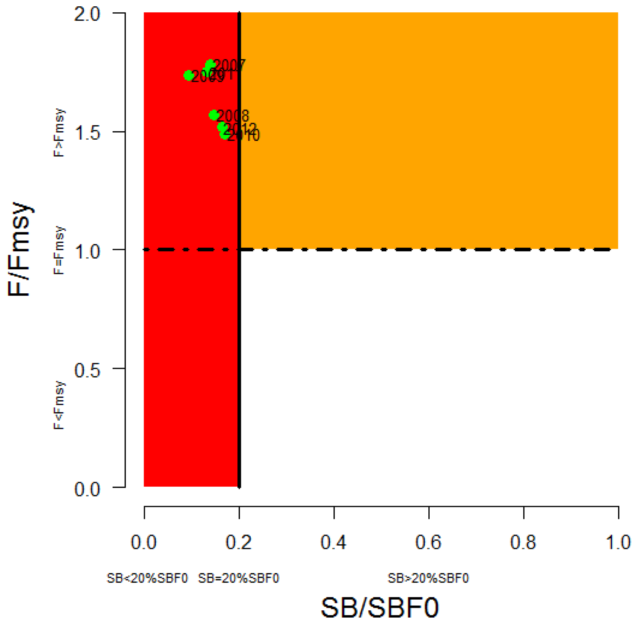

Figure 3: Hindcast results: 2012 estimates of adult biomass relative to $SB_{F=0}$ and fishing mortality (average 2008-2011) relative to F_{MSY} from each of the deterministic projections. Year labels indicate the terminal year of the retrospective assessment for which the projection was conducted.

outside for the full time period. Estimates of adult biomass for all of the retrospective assessments are lower than those estimated by the reference case assessment.

Although there is some variation in the scaling of adult biomass throughout the assessment period, there is no evidence of any systematic over- or under-estimation with successive assessments. The retrospective bias, as calculated from Equation 1, is -0.106 indicating almost no systematic bias over the retrospective period considered (2012 to 2007). Brooks and Legault (2015) propose a diagnostic rule of thumb for when to apply a ρ adjustment to to terminal biomass estimates. If the ρ adjusted point estimate lies within the 90% confidence region no adjustment should be made since the corrected value remains within the overall uncertainty bounds of the assessment. We used the standard deviation of the terminal biomass estimate to compute approximate 68% confidence intervals for comparison with the adjusted value. The ρ adjusted value (293909) was within the approximate 68% confidence region (228044, 309339) and therefore well within the 90% region suggesting that bias correction was not necessary.

The deterministic catch-based projections gave estimates of $SB/SB_{F=0}$ in 2012 that ranged from 0.09 to 0.17 (Figures 2 and 3) with corresponding estimates of F/F_{MSY} in 2012 ranging between 1.48 and 1.78. No persistent trend in over- or under-estimation of either $SB/SB_{F=0}$ or F/F_{MSY} was apparent.

The stochastic catch-based projections show the variability in estimates of projected adult biomass as a consequence of variability in recruitment. Due to the time taken for bigeye tuna to reach maturity (around 4 years to reach 50% maturity) there is very little variation apparent in projected estimates of adult biomass. The longest projection period was for the 2007 retrospective (Figure 4) having a 5 year period over which biomass could evolve with stochastic recruitment. The median estimate of adult biomass in 2012, as determined from stochastic projection of the 2007 retrospective, lies just outside the 95% confidence interval of the equivalent value estimated by the reference case assessment. Figure 4 also shows that the confidence intervals for projected adult biomass, after 5 years, are wider than the confidence intervals for the reference case assessment.

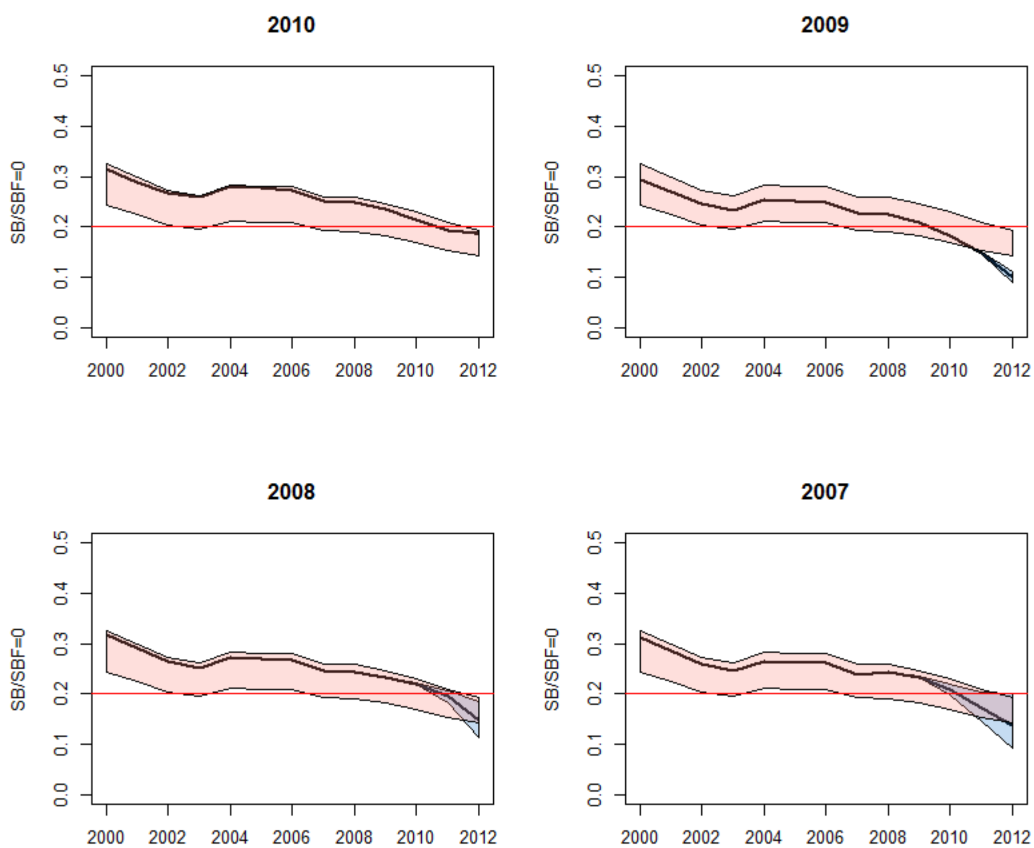

Figure 4: Adult biomass determined from stochastic catch-based hindcasts for 2010 to 2007. Pink shaded area shows the approximate 95% confidence interval for the reference case assessment estimates. Blue shaded area shows the 95% confidence interval of the catch-based retrospective projections. Horizontal red line denotes the LRP of 20% $SB_{F=0}$.

5 Discussion

Retrospective analyses are an important diagnostic tool for determining the robustness of model estimates to varying quantities of data and in recent years have been routinely presented for each new stock assessment. Similarly retrospective forecasts, or hindcasts, provide important information on the quality of projections. Such analyses are of particular importance if we are to make critical use of information about the quality of our assessments and projections to improve the basis of management advice (Brander et al., 2013).

The period of the retrospective analysis considered here is just 5 years (2012 to 2007). As such the hindcasting analysis provides information on short term risk and uncertainty. Brooks and Legault (2015) found the largest source of bias in stock projections in the short term was due to the terminal estimates of population numbers at age. As the projection period increased the results became more sensitive to recruitment variability. We found very similar results in our analysis although at present MULTIFAN-CL can only incorporate uncertainty in future recruitment in its stochastic projections. The inclusion of two other sources of variability (numbers at age in the first year of the projections and fishery effort deviations) is currently under development (Davies et al., 2013) and once implemented will allow for better characterisation of risk and uncertainty in the short term.

Although there is no evidence of a persistent retrospective pattern, estimates of adult biomass for all of the retrospective assessments are lower than those estimated by the reference case assessment (Figure 1). The additional figures shown in Appendix A and specifically Figure 5 show that this is not the case across all regions of the assessment and that for many regions there is close correspondence between the reference case and retrospective results, particularly for the recent period. However, estimates of adult biomass in region 8 are consistently estimated to be higher by the reference case assessment throughout the time series. The reason for this result is currently unclear and will require further investigation.

We note that the investigation of retrospective bias, in terms of the conditions that cause it and the measures to remedy it, has been particularly focussed on VPA based assessments that typically have very strict assumptions about catchability and, importantly, treat the catch data as exact. Indeed, the analyses of Brooks and Legault (2015), Deroba (2014), Mohn (1999) and Sinclair et al. (1991) cited throughout this paper have all considered VPA-based rather than integrated assessments. Integrated assessments (eg. MULTIFAN-CL and Stock Synthesis) are based on different structural assumptions and therefore may exhibit different characteristics in the face of conditions that generate retrospective bias in VPA-based assessments.

It should be noted that Mohn's ρ is just one method to characterise model performance (in fact it is just one of several methods to calculate retrospective bias). Although it cannot identify the cause of the problem, a large value does indicate that there is something mis-specified in the data, the model, or both. Importantly it identifies whether there is uncertainty in the results beyond what can be characterised by a single reference case model. The uncertainty grid of alternative assessment models, commonly constructed for WCPFC assessments, addresses some aspects of potential model mis-specification. The retrospective analysis serves as a further check on the performance and robustness of the final assessment.

Although not considered a problem for MULTIFAN-CL, we highlight a potential problem with post-hoc projections conducted on stock assessment output where the projection settings can assume different values to those used in the initial assessment (Brooks and Deroba, 2015). This can result in future projected quantities that are inconsistent with the parameter values and structural assumptions associated with the assessment that generated the corresponding values for the historical period. We note that projections undertaken with MULTIFAN-CL are run as an extension of the main stock assessment analysis. In effect the projection is integrated into the assessment so that the starting population for the projections and other population and fishery parameters are entirely consistent with the assessment that generated the historical trends.

6 Conclusions

Although adult biomass for all of the retrospective runs is below that of the reference case assessment, our results show no evidence of any systematically increasing over- or under-estimation of adult biomass with successive assessments. The calculated retrospective statistic (ρ) is relatively small and the ρ corrected terminal biomass value remains within the 90% confidence interval of the stock assessment estimate.

Estimates of stock status in 2012 and recent fishing mortality as determined from the deterministic projections are relatively tightly clustered (Figure 3) and consistently indicate the stock to be below the limit reference point and to be exploited at levels in excess of F_{MSY}.

The median estimate of adult biomass in 2012, as determined from stochastic projection of the 2007 retrospective, lies just outside the 95% confidence interval of the equivalent value estimated by the reference case assessment (Figure 4). In addition, the confidence interval for projected adult biomass, after 5 years, envelopes the confidence interval for the reference case assessment.

We therefore conclude that:

1. the 2014 bigeye tuna stock assessment model is not subject to significant retrospective bias;
2. short-term deterministic catch based projections, conducted from the 2014 bigeye tuna reference case assessment model, provide consistent and relatively accurate indication of stock status in the short-term;
3. short-term stochastic projections for bigeye tuna will potentially under-estimate overall uncertainty in stock status in the first few years.

We invite the Scientific Committee:

1. to note the results of this analysis and consider them in the context of the provision of short-term management advice;
2. to note the future developments planned for MULTIFAN-CL that will improve estimates of uncertainty and risk in the short-term;
3. to note the importance of retrospective analyses as a diagnostic tool for WCPFC stock assessments; and
4. to note that integrated assessments may exhibit different characteristics to VPA based assessments under conditions that generate retrospective bias.

Acknowledgments

This study was conducted with financial support from WCPFC and from SPC core and programme funding sources.

References

Brander, K., Neuheimer, A., Andersen, K. H., and Hartvig, M. (2013). Overconfidence in model projections. *ICES Journal of Marine Science*, 70(6):1065–1068.

Brooks, E. N. and Deroba, J. J. (2015). When data are not data: the pitfalls of post-hoc analyses that use stock assessment model output. *Canadian Journal of Fisheries and Aquatic Sciences*, 72(4):634–641.

Brooks, E. N. and Legault, C. M. (2015). Retrospective forecasting - evaluating performance of stock projections for New England groundfish stocks. *Canadian Journal of Fisheries and Aquatic Sciences*, 73:1–16.

Davies, N., Fournier, D. A., Hampton, J., Hoyle, S., Bouye, F., and Harley, S. (2013). Recent developments in the MULTIFAN-CL stock assessment software. WCPFC-SC09-2013/SA-IP-07, Pohnpei, Federated States of Micronesia, 6–14 August 2013.

Deroba, J. J. (2014). Evaluating the consequences of adjusting fish stock assessment estimates of biomass for retrospective patterns useing Mohn's Rho. *North American Journal of Fisheries Management*, 34:380–390.

Harley, S. J., Davies, N., Hampton, J., and McKechnie, S. (2014). Stock assessment of bigeye tuna in the Western and Central Pacific Ocean. WCPFC-SC10-2014/SA-WP-01, Majuro, Republic of the Marshall Islands, 6–14 August 2014.

Maunder, M. N. and Punt, A. E. (2013). A review of integrated analysis in fisheries stock assessment. *Fisheries Research*, 142:61–74.

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.

OFP (2015). Evaluation of CMM2014-01 for bigeye tuna. Technical Report WCPFC12-2015-12 rev.1, Stones Hotel, Kuta, Bali, Indonesia, 3-8 December 2015.

Rosenberg, A. and Restrepo, V. (1994). Uncertainty and risk evaluation in stock assessment advice for U.S. marine fisheries. *Canadian Journal of Fisheries and Aquatic Sciences*, 51:2715–2720.

Scott, R. D., Pilling, G., and Harley, S. (2015). Short-term stochastic projections for skipjack, yellowfin and bigeye tunas. WCPFC-SC11-2015/SA-WP-04, Pohnpei, Federated States of Micronesia, 5–13 August 2015.

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

A Additional Retrospective Figures

We present here, for the dedicated reader, some additional figures that are of interest in terms of the finer inspection of the retrospective analysis but are of less relevance to the central topic of this paper. Specifically we show those results of the retrospective analyses that are typically presented (adult biomass, fishing mortality and recruitment) but at finer spatial and temporal resolutions.

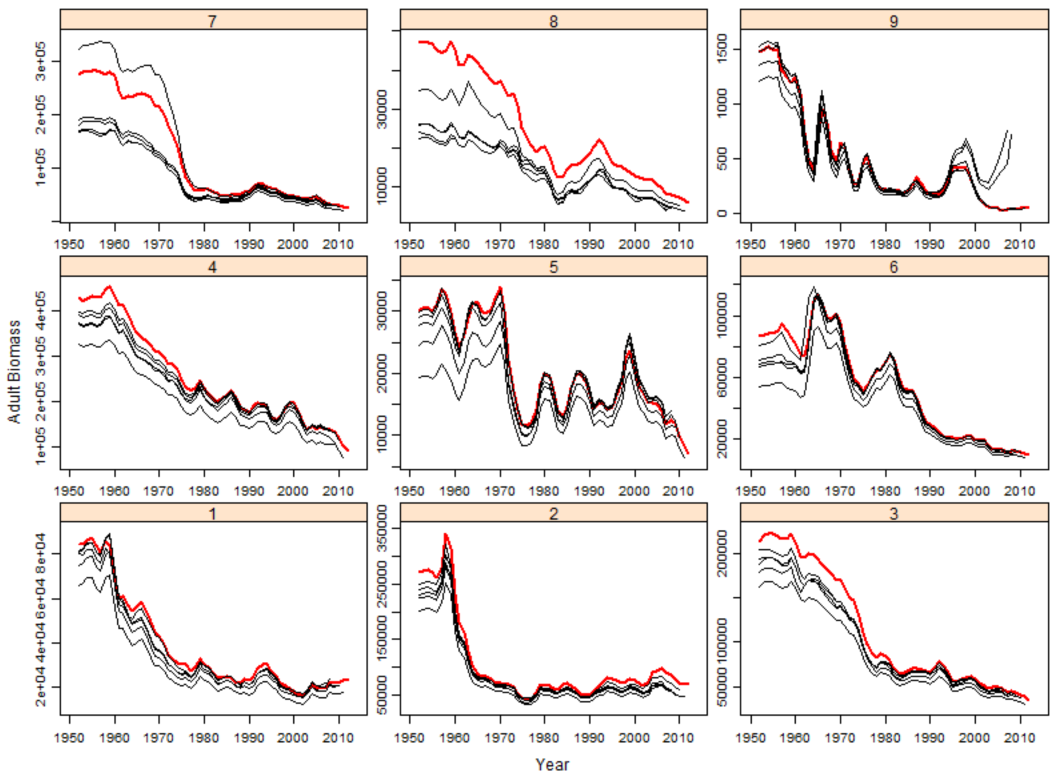

Figure 5: Quarterly estimates of adult biomass by assessment region determined from the 2014 reference case assessment (1952 to 2012, red line) and retrospective assessment runs (terminal years 2011 to 2007, black lines). Note that the y scale differs for each plot.

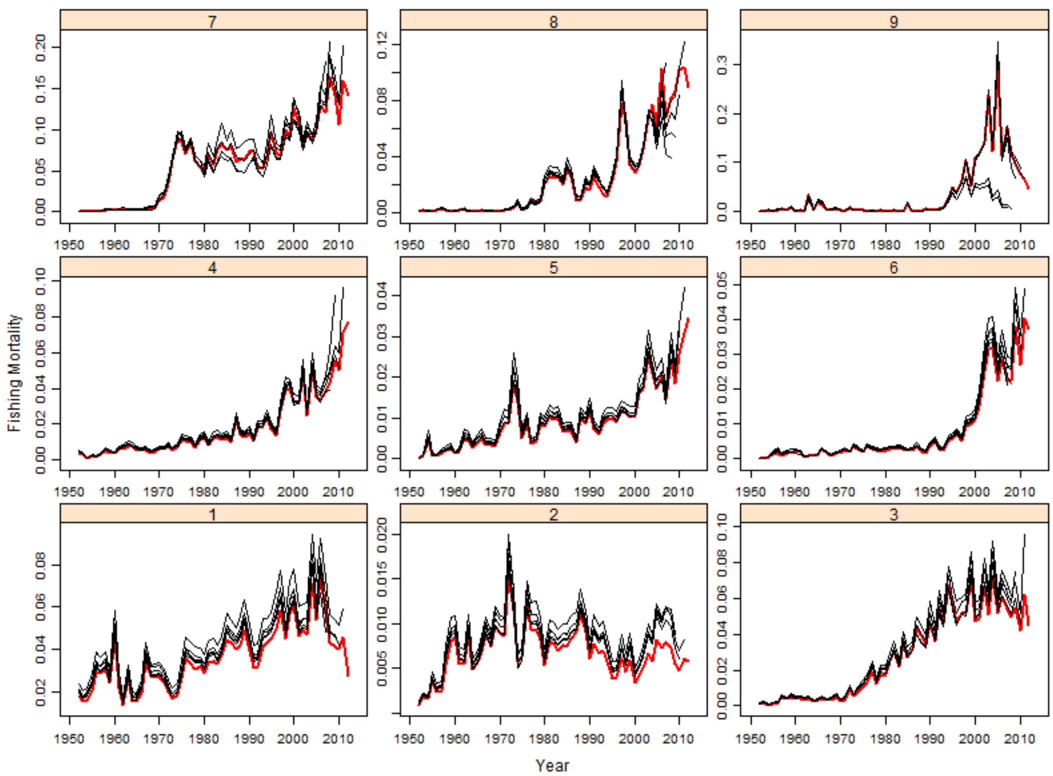

Figure 6: Quarterly estimates of fishing mortality by assessment region determined from the 2014 reference case assessment (1952 to 2012, red line) and retrospective assessment runs (terminal years 2011 to 2007, black lines). Note that the y scale differs for each plot.

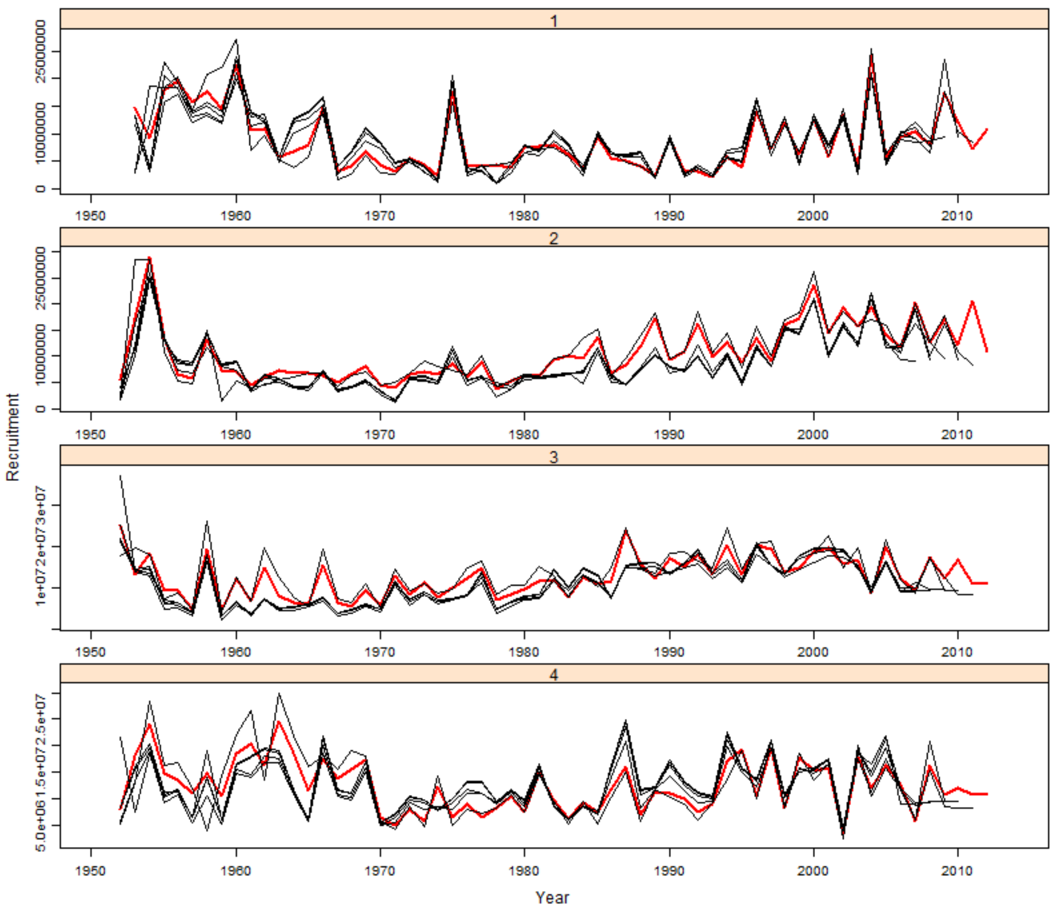

Figure 7: Estimates of recruitment by quarter determined from the 2014 reference case assessment (1952 to 2012, red line) and retrospective assessment runs (terminal years 2011 to 2007, black lines).