SCIENTIFIC COMMITTEE
SEVENTEENTH REGULAR SESSION

Online meeting
11-19 August 2021

Retrospective forecasting of CPUE for South Pacific albacore

WCPFC-SC17-2021/MI-IP-01

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

[1]Oceanic Fisheries Programme, The Pacific Community (SPC), Noumea, New Caledonia

Contents

Executive Summary

Under the current South Pacific albacore Management Strategy Evaluation framework (MSE), future fishing opportunities are "driven" by the CPUE time-series in the harvest strategy. The plausibility of the projected CPUE time-series that is used in the framework is of particular concern when evaluating management procedures (MPs) that rely on CPUE. Retrospective forecasting or hindcasting is a method for testing the performance of a predictive model using existing historic data. Noting the comments of CCMs concerning the stability of the forcasted CPUE, a retrospective forecasting is performed to evaluate the CPUE predictions under the current MSE framework. Our results suggest that the forecasted standardized CPUE time-series in area 3 is the closest to the observed CPUE and shows potential to be used as a stock status indicator in the harvest strategy of South Pacific albacore. Further work is required to test the performance of MPs using CPUE as a stock status indicator to drive future fishing opportunities. We also recommend that alternative approaches to estimating stock status within the MP using simple models (i.e. biomass dynamic model) are explored and compared with the empirical CPUE approach.

We invited SC to:

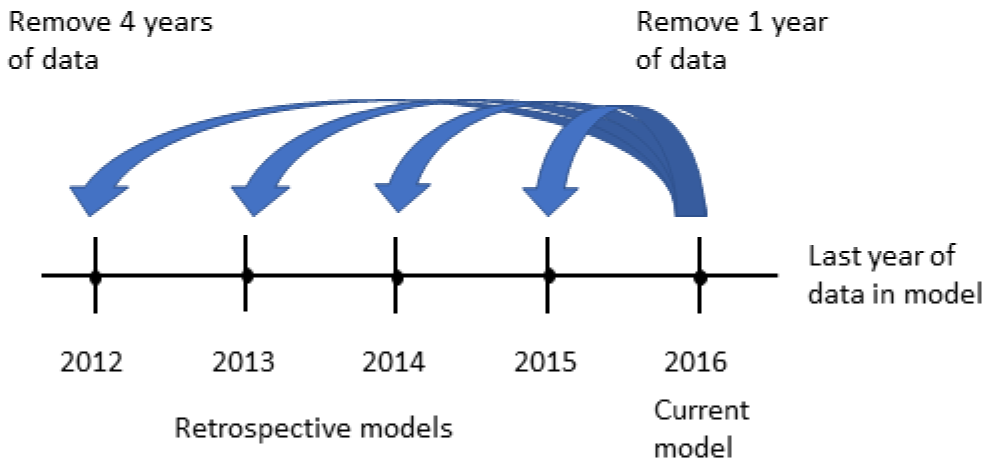
- Note the work done to investigate the reliability of CPUE projections for application in the South Pacific albacore MSE.
- Acknowledge the issues with reliable CPUE projections indicated by this study and support further development of model-based (biomass dynamic model) estimation methods for testing and comparison with CPUE in a MP for South Pacific albacore.

1 Introduction

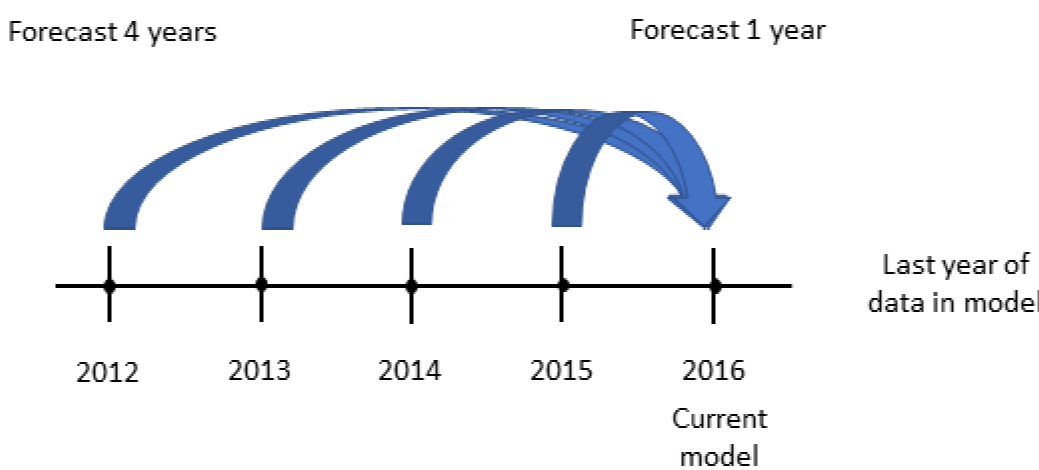
In a harvest strategy framework, management procedures (MPs) set fishing opportunities based on estimates of stock status (Butterworth et al., 1997; Punt et al., 2014). Initial work on harvest strategies for South Pacific albacore has focused on developing empirical MPs that set future fishing opportunities based directly on the observed CPUE (Yao et al., 2020). Future work may seek to develop alternative model-based approaches that employ relatively simple assessment models (e.g. biomass dynamic models). Both of these approaches rely heavily on the use of CPUE data either directly or as part of the estimation method. The use of CPUE as the basis for an empirical MP is consistent with the focus of recent discussions for the southern longline fisheries desire for economically viable catch rates, as reflected in the economic management objectives that were noted at WCPFC14 (see WCPFC14, attachment K). Further, CPUE is the basis for the interim target reference point (TRP) that was agreed at WCPFC15 (see WCPFC15, para 207). Before an MP is adopted, candidate MPs are tested using management strategy evaluation (MSE). In general, future CPUE time-series are commonly generated based on the assumption that future biological parameters and fishery selectivity are similar to the historical period within the MSE framework. However, the observations from reality are usually more noisy and complex than the model simulated data. This approach might fail to reflect the expected dynamics of the observed CPUE. Given the importance of the CPUE signal in setting future fishing opportunities, testing MPs with unrealistic CPUE could lead to the adoption of an inappropriate MP and increase the risks associated with fisheries management. Therefore, the plausibility of the projected CPUE time-series should be investigated before applying in the MPs.

One of the common approaches to evaluate the accuracy of the CPUE projections is to simulate quantities from retrospective models and then compare the simulated values to the observed values. This approach is commonly referred as retrospective forecasting (Brooks and Legault, 2015), also known as hindcasting and backtesting. 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 (Fig. 1). This approach has been used in previous studies to understand the implications of projections from assessments with systematical bias (Brooks and Legault, 2016) and investigate the performance of short term projections for WCPO bigeye tuna (Scott et al., 2016).

Whilst we must acknowledge that no stock assessment will be 100% accurate, it is reasonable 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 overestimation over time. When updated parameter estimates display a persistent trend in relation to previous estimates (i.e. a new assessment with additional data) it suggests that something may be misspecified in the assessment model. Systematic error of this kind is typically referred to as retrospective bias (Sinclair et al., 1991). Retrospective patterns in estimated biomass or depletion could introduce biased perceptions of stock status and poor management advice. In severe cases, the assessment models could be considered as unreliable for management purposes (Cadigan and Farrell, 2005; Valero, 2012). For this reason, a retrospective analysis is typically conducted to evaluate the reliability and internal consistency of stock assessment models (Cadigan and Farrell, 2005; Hurtado-Ferro et al., 2015).



(a)



(b)

Figure 1: Concept diagram of "retrospective forecasting". (a) Retrospective models are created by truncating 1,2,3,4 years of data from the full time-series applied in the current model. (b) Forecasting models are then made from each retrospective model and used to forecast data to the end of 2016. (adopted from (Brooks and Legault, 2016)).

Forecasting from the retrospective models with known characteristics of the stock could provide important information on the performance of the simulations. A significant concern for management is the reliability of the CPUE projections made from assessments that are subject to retrospective bias, since catch and effort limits that are designed to meet management targets can be systematically under or overestimated, ultimately leading to 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. Therefore, it is necessary to examine the implications of any retrospective bias to the performance of the management framework using retrospective forecasting.

In this paper, we conduct a retrospective analysis to examine the performance of the 36 models selected from the South Pacific albacore assessment grid. Then, we apply the retrospective forecasting approach to project CPUE time-series and compare them to the observed CPUE to evaluate the plausibility of projected CPUE under the current MSE framework.

2 Retrospective CPUE standardization

To run the retrospective stock assessments it is necessary to generate standardized retrospective CPUE time-series. The data used in the CPUE standardization is the operational-level longline data described by Tremblay-Boyer et al. (2018a). This data set contains individual records of longline fishing activity, recorded by vessel, day, time, location (at a one degree resolution) and the numbers of fish caught of the four species most likely to be targeted in longline operations: albacore tuna, bigeye tuna, yellowfin tuna, and swordfish. This data set (without the recent updates) has also been used for the CPUE standardization for recent stock assessments (Tremblay-Boyer et al., 2018a). The regional structure used in the standardization is also consistent with the recent stock assessment (Fig. 2) (Tremblay-Boyer et al., 2018a).

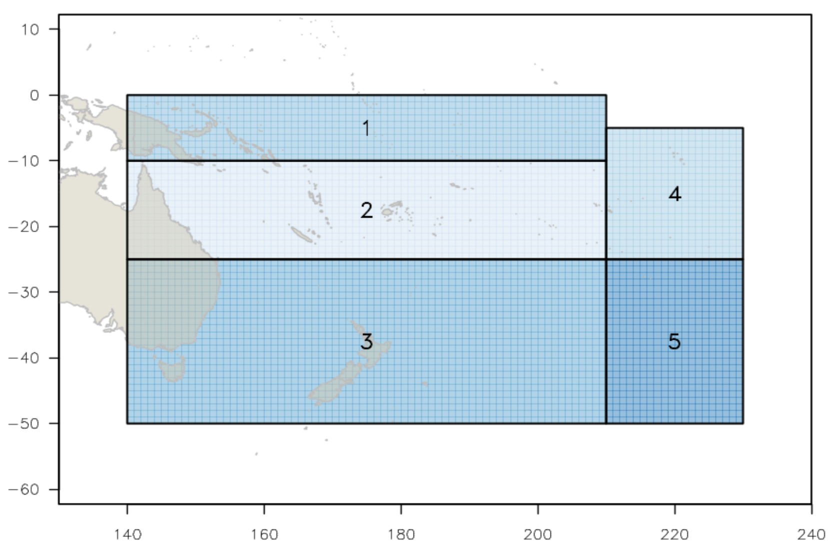


Figure 2: Map of the assessment regional structures used in the 2018 South Pacific albacore stock assessment ("2018 regional structure").

The targeting group of each individual longline set is identified, so that the effect of non-albacore targeting sets could be accounted for within the standardization. The same K-mean clustering analysis described in Tremblay-Boyer et al. (2018a) was applied in this study. This method assumes that sets targeting a specific species should have a higher proportion of that species in their catch. This analysis was applied to the truncated data for each retrospective year.

A geostatistical approach was then used to standardize the nominal CPUE in each area. In this approach, the spatial correlation was assumed to be a random effect and was fitted using a spatial GLMM with a delta lognormal error distribution. The number of knots used in the standardization was 200 and the explanatory variables included the year-quarter and the targeting cluster. The model configuration and assumptions are consistent with CPUE standardization for the previous stock assessment (Tremblay-Boyer et al., 2018a).

The geostatistical CPUE standardization model described above was applied to the operational data, that has been successively truncated by one year from 2016 (the terminal year in the most recent assessment) to 2012. We only considered running the retrospective analysis back to 2012 to eliminate the impacts of removing the age-length data, which is first available to the assessment in 2011.

The fit of the model was examined by the distribution of residuals and the reliability of convergence. All analysis was performed in R and the spatial GLMM models were run with the VAST package (Thorson, 2019).

3 Retrospective Stock assessment analysis

Once the retrospective CPUE time-series were generated the next step was to re-fit the South Pacific albacore stock assessment model with the truncated input data for each retrospective model period.

The 2018 South Pacific albacore stock assessment model grid contains 72 models (Tremblay-Boyer et al., 2018b). These models are included in the stock assessment to accommodate the uncertainty in the model assumptions. However, half of them used the "traditional" CPUE standardization method, which is not expected to be continued in the future. Therefore, these models have not been included for the retrospective analysis. The resulting model grid includes 36 models.

For each of the 36 full time-series (i.e., end year of 2016) assessment model, we created 4 retrospective assessment models (with the end year 2012, 2013, 2014, 2015, referred as "retro models") by successively truncating one year of input data from the end of time-series (Fig. 1a). The corresponding retrospective standardized CPUE described above was applied in each retro model. The regional structure used in the retro models is also consistent with the recent stock assessment (Fig. 2) (Tremblay-Boyer et al., 2018a).

The model configuration and assumptions are identical for all the retro models and the full time-series stock assessment models. The only exception being the period over which the Beverton and Holt stock-recruitment relationship (SRR) is estimated. For the retrospective assessments, this period was necessarily shortened depending on the terminal year.

The performances of the retro models were characterized by the ρ statistic (i.e. Mohn's ρ), which measures the relative difference between an estimated value from an assessment with a reduced time-series and the same value estimated from the full time-series (Mohn, 1999). Here, the retrospective bias (Mohn's ρ) of the adult biomass in the terminal year was estimated using Equation. 1, where r is the number of years over which the retrospective analysis is conducted, θ^*_{Y-i} is adult biomass in the terminal year of each retro model and θ_{Y-i} is the corresponding value of adult biomass from the stock assessment model using the full time-series of input data (Mohn, 1999).

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

The larger the value of ρ the greater the retrospective bias. A positive ρ indicates a tendency for the retro model to overestimate adult biomass in the final year compared to the same year of the full assessment whilst a negative ρ indicates a tendency for underestimation.

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

4 Forecasting simulation framework

The retro models described above served as the starting point in the forecasting step. For each retro model, forecasts were made for each year until 2016 (Fig. 1b). All the fisheries in the stock assessment were projected in the forecasting period except two driftnet fisheries (fishery 15 and 16) which no longer operate (Table. 1). Among the 19 fisheries that were projected in the framework, fisheries 17 to 21 are the index fisheries (shaded in gray), which contain the standardized CPUE time-series of each area (area 1 to 5). The future catches of the projected fisheries were based on the actual catches that had been observed in each of the projection years (2013 to 2016).

Fishery	Nationality	Gear	Area	Longline group
1.DWFN LL 1	Distant-water Fishing Nations	Longline	1	Tropical
1.PICT.AZ LL 1	Pacific Island Countries and Territories	Longline	1	Tropical
3.DWFN LL 2	Distant-water Fishing Nations	Longline	2	Sub-tropical
4.PICT LL 2	Pacific Island Countries and Territories	Longline	2	Sub-tropical
5.AZ LL 2	Australia/New Zealand	Longline	2	Sub-tropical
6.DWFN LL 3	Distant-water Fishing Nations	Longline	3	Temperate
7.PICT LL 3	Pacific Island Countries and Territories	Longline	3	Temperate
8.AZ LL 3	Australia/New Zealand	Longline	3	Temperate
9.DWFN LL 4	Distant-water Fishing Nations	Longline	4	Sub-tropical
10.PICT LL 4	Pacific Island Countries and Territories	Longline	4	Sub-tropical
11.DWFN LL 5	Distant-water Fishing Nations	Longline	5	Temperate
12.PICT LL 5	Pacific Island Countries and Territories	Longline	5	Temperate
13.All TR 3	All nationalities	Troll	3	-
14.All TR 5	All nationalities	Troll	5	-
15.All DR 3	All nationalities	Driftnet	3	-
16.All DR 5	All nationalities	Driftnet	5	-
17.Index LL 1	Index fishery	Longline	1	-
18.Index LL 2	Index fishery	Longline	2	-
19.Index LL 3	Index fishery	Longline	3	-
20.Index LL 4	Index fishery	Longline	4	-
21.Index LL 5	Index fishery	Longline	5	-

Table 1: Definition of fisheries for the MULTIFAN-CL South Pacific albacore tuna retrospective forecasting.

Recruitment deviance for the projections was fixed consistent with estimated recruitment level during the retrospective period. Catchability was assumed to remain constant for the projection at the level estimated for the terminal year of each retro model. In addition, no observation error was added to the forecasting process (i.e. the coefficient of variation (CV) for both the simulated catch and effort were set to 0).

The simulated CPUE in each projection was calculated simply as $CPUE = Catch/Effort$ for each fishery. The simulated CPUE was determined for all projected fisheries. However, subsequent analyses are mostly focused on the five index fisheries (standardized CPUE in each area).

The accuracy of the forecasted CPUE was measured using the mean absolute percent error (MAPE), which estimates the gap in percentage between the simulated values and observation values. Equation. 2 was used to calculate the MAPE value of each of the five index fisheries, where n is the number of observations, $CPUE_i$ is actual observed CPUE value and $CPUE_i^*$ is the corresponding simulated CPUE based on actual catch. The larger the MAPE value, the greater the estimation error of the simulated CPUE, and vice versa. This method is frequently used for hindcasting exercises (e.g. Calvin et al., 2017; Snyder et al., 2017).

$$MAPE = \frac{100}{n} \sum_{i=1}^{n} \frac{|CPUE_i - CPUE_i^*|}{CPUE_i} \tag{2}$$

In summary, the projection assumptions are:

- All the fisheries (except two driftnet fisheries) were projected from the terminal year of each model through to 2016 based on the observed catches for the retrospective projection period.
- Catchability was assumed to remain constant for the projection at the level of the terminal year of each retro model.
- Recruitment deviance was fixed to be consistent with recruitment estimated for the retrospective period.
- No observation error was added to the simulated catch and effort.
- The future catches of the projected fisheries were based on the actual catches that had been observed in each of the projection years.

5 Results

5.1 Retrospective CPUE standardization

The results of the cluster analyses on the retrospective data are similar to the previous cluster analysis carried out for the 2018 South Pacific albacore stock assessment (Tremblay-Boyer et al., 2018a) (Figs. 6 to 10). The results of the cluster analysis for each retrospective year with the truncated data are then used for the retrospective CPUE standardization.

The estimates of the standardized CPUE from each of the retrospective runs show consistent patterns over time (e.g. Fig. 3, complete results are included in the appendix: Figs. 11 to 14). Note that due to the updates of the operational data, the standardized CPUE values of our analysis might be different from those of Tremblay-Boyer et al. (2018a). The retrospective standardized CPUE time-series were then applied to the corresponding retro models.

5.2 Retrospective Stock Assessment Model

All the retro models achieved the convergence criteria of maximum gradients less than 0.01. In the majority of models, estimated adult biomass determined from each of the retro model shows similar trends across the time-series (Fig. 4), but are re-scaled throughout the time-series relative to the full assessment in most cases. The retrospective pattern is detected from several retro models (see appendix: Figs. 15 to 17).

The retrospective bias of each model, as calculated from Equation. 1, are listed in Table. 2. The values of Mohn's ρ ranges from 0.0003 (minimum bias) to -0.213 (maximum bias) across the 36 assessment models selected from the South Pacific albacore assessment grid. Among them, 29 displayed persistent negative bias in the estimation of adult biomass, including for the terminal years, which indicates a general systematic underestimation over the retrospective period (2012 to 2016).

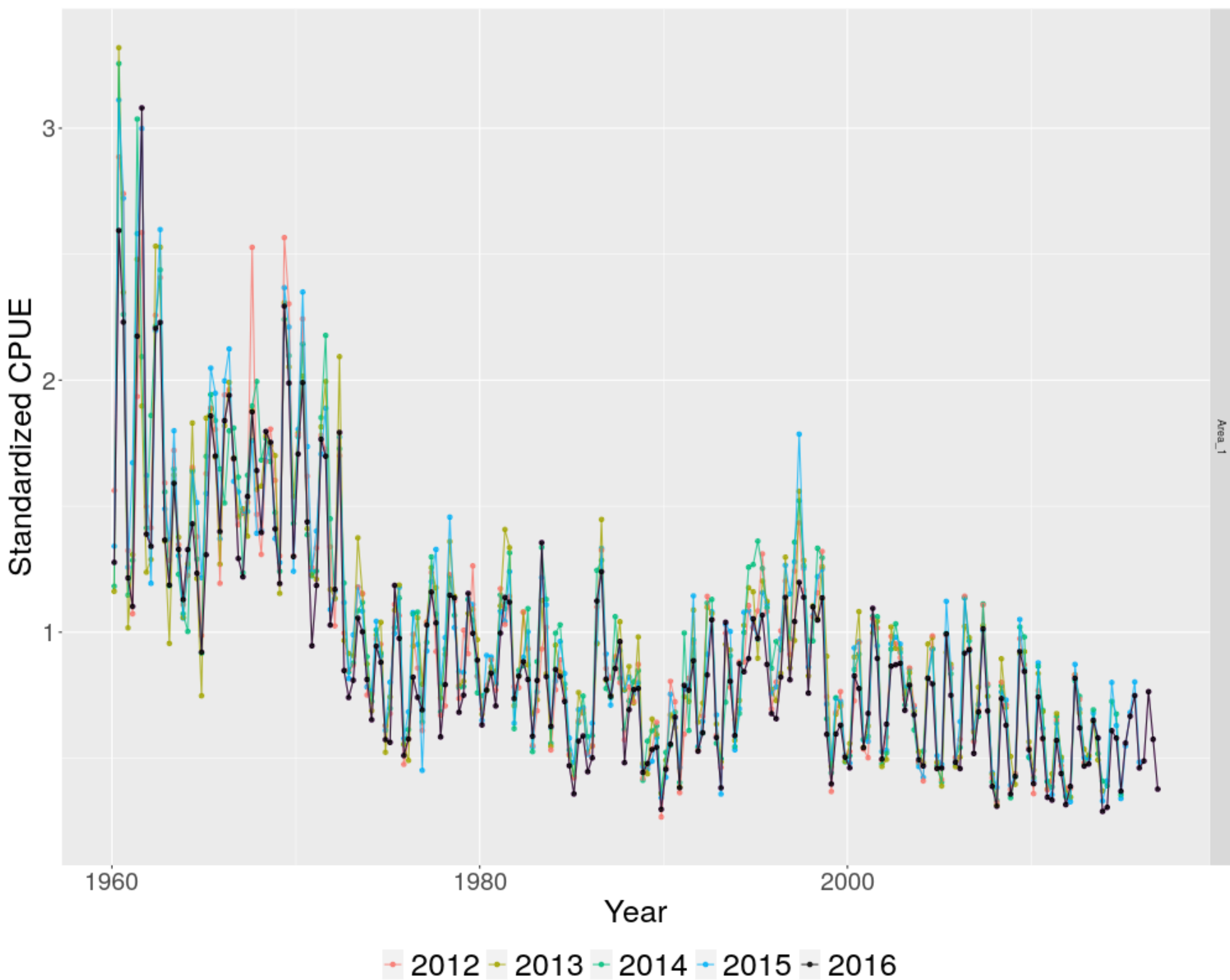
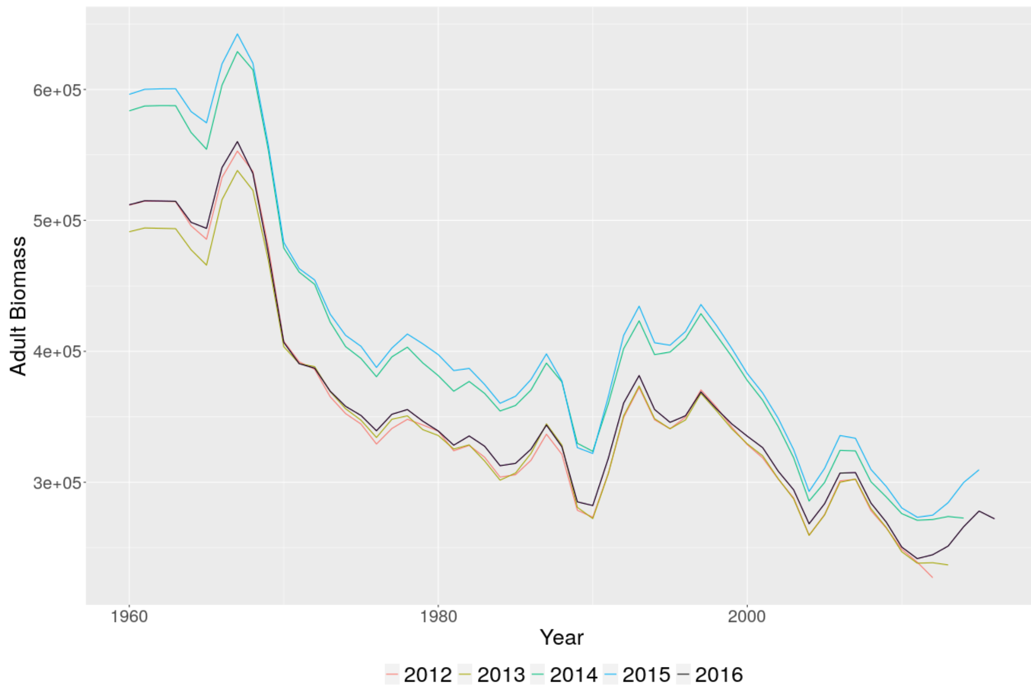
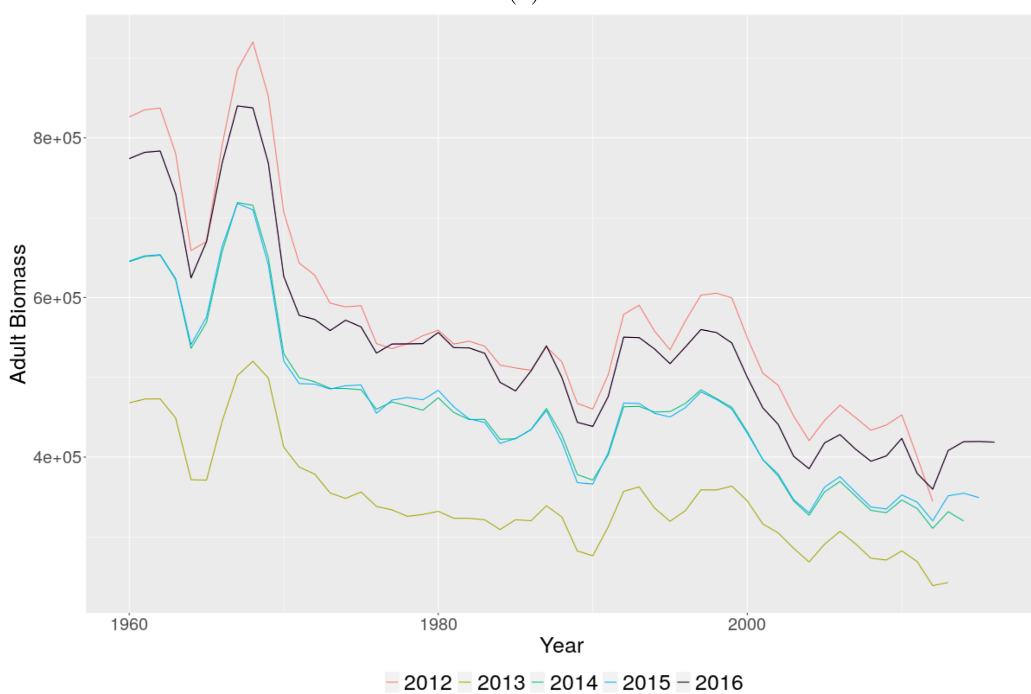


Figure 3: Results of the retrospective CPUE standardization from 2012 to 2016 for stock assessment example plot for model area 1, for other model areas see appendix: Figs. 11 to 14. Note: the black line is for the full assessment period and the color lines are the retrospective periods.



(a)



(b)

Figure 4: Two examples of the retrospective results. The annual estimates of adult biomass (summed over areas) determined from the full time-series assessment (black line) and retrospective assessment runs (terminal years 2015 to 2012, color lines). (a) Model: CPUE:Geo Growth:Chen-Wells Size Weighting:50 Steepness:0.95 Natural Mortality:0.4. This model has the lowest retrospective bias. (b) Model: CPUE:Geo Growth:Estimated Size Weighting:20 Steepness:0.95 Natural Mortality:0.4. This model has the highest retrospective bias.

Model					Mohn's ρ
CPUE	Growth	Size data Weighting	Steepness	Natural Mortality	
Geo	Chen-Wells	20	0.65	0.3	-0.130
Geo	Chen-Wells	20	0.65	0.4	-0.084
Geo	Chen-Wells	20	0.8	0.3	-0.127
Geo	Chen-Wells	20	0.8	0.4	-0.086
Geo	Chen-Wells	20	0.95	0.3	-0.126
Geo	Chen-Wells	20	0.95	0.4	-0.086
Geo	Chen-Wells	50	0.65	0.3	-0.132
Geo	Chen-Wells	50	0.65	0.4	0.002
Geo	Chen-Wells	50	0.8	0.3	-0.130
Geo	Chen-Wells	50	0.8	0.4	0.016
Geo	Chen-Wells	50	0.95	0.3	-0.129
Geo	Chen-Wells	50	0.95	0.4	-0.0004
Geo	Chen-Wells	80	0.65	0.3	-0.111
Geo	Chen-Wells	80	0.65	0.4	0.050
Geo	Chen-Wells	80	0.8	0.3	-0.108
Geo	Chen-Wells	80	0.8	0.4	0.051
Geo	Chen-Wells	80	0.95	0.3	-0.104
Geo	Chen-Wells	80	0.95	0.4	0.095
Geo	Estimated	20	0.65	0.3	-0.011
Geo	Estimated	20	0.65	0.4	-0.175
Geo	Estimated	20	0.8	0.3	-0.160
Geo	Estimated	20	0.8	0.4	-0.086
Geo	Estimated	20	0.95	0.3	-0.213
Geo	Estimated	20	0.95	0.4	-0.153
Geo	Estimated	50	0.65	0.3	-0.069
Geo	Estimated	50	0.65	0.4	-0.116
Geo	Estimated	50	0.8	0.3	-0.063
Geo	Estimated	50	0.8	0.4	-0.118
Geo	Estimated	50	0.95	0.3	-0.067
Geo	Estimated	50	0.95	0.4	-0.122
Geo	Estimated	80	0.65	0.3	-0.064
Geo	Estimated	80	0.65	0.4	0.017
Geo	Estimated	80	0.8	0.3	-0.070
Geo	Estimated	80	0.8	0.4	0.051
Geo	Estimated	80	0.95	0.3	-0.067
Geo	Estimated	80	0.95	0.4	-0.085

Table 2: Mohn's ρ for the retrospective analysis of the 36 models selected from the South Pacific albacore assessment grid.

The ANOVA results suggest that the natural mortality and size data weighting assumptions have the most significant impacts on retrospective bias (Mohn's ρ) (Table. 3).

	Df	$SumSq$	$MeanSq$	$F - value$	$P - value$
Growth	1	0.006	0.006	2.015	0.166
Size Data Weighting	2	0.047	0.023	7.743	0.002*
Steepness	2	0.003	0.001	0.423	0.659
Natural Mortality	1	0.029	0.029	9.508	0.004*
Residuals	29	0.087	0.003		

Table 3: The results of the fixed-effects analysis of variance (ANOVA) on the model assumptions contributing to the retrospective bias (Mohn's ρ).

5.3 Retrospective Forecasting of CPUE

The forecasted CPUE of the five index fisheries mostly show similar trends to the observed CPUE. However, deviations between the forecasted CPUE and the observed CPUE are detected across the forecasting models (Figs. 21 to 32).

The average MAPE value is 17.81%, with a range of 4.67% to 36.39% (Table. 4). The MAPE values of the five index fisheries indicate different levels of ability to forcast accurate CPUE time-series (Fig. 5). Among the five index fisheries, forecasted CPUE from fishery 19 (i.e., the standardized CPUE in area 3) is the closest to the observed CPUE with most models having less than 12% error. The forcasted standardized CPUE in area 2 (fishery 18) is also reasonably close to the observed CPUE (ca. 20% error) and has the shortest confidence interval. Meanwhile, the greatest deviations from the observed CPUE are detected in fishery 20 and 21 (i.e., the standardized CPUE in area 4 and area 5). Overall the results of the forecasted standardized CPUE show that the most internally consistent forecasts are for assessment model areas 3 and 2. Forecasting accuracy for the other assessment model areas is generally poor.

For the retrospective analysis, we investigated the patterns of the estimated biomass for the entire stock assessment area. Whereas in the forecasting component, we mainly focus on testing the plausibility of the forcasted standardized CPUE in each area. However, while our primary focus is the five index fisheries we also explored the potential of forecasted CPUE time-series for the PICT and DWFN fleets separately in each area (Appendix 1: Figs. 33 to 41). The results suggest that the forecasted CPUE from those fleets showed overall larger deviations from the observed CPUE than the index fishery from the same area.

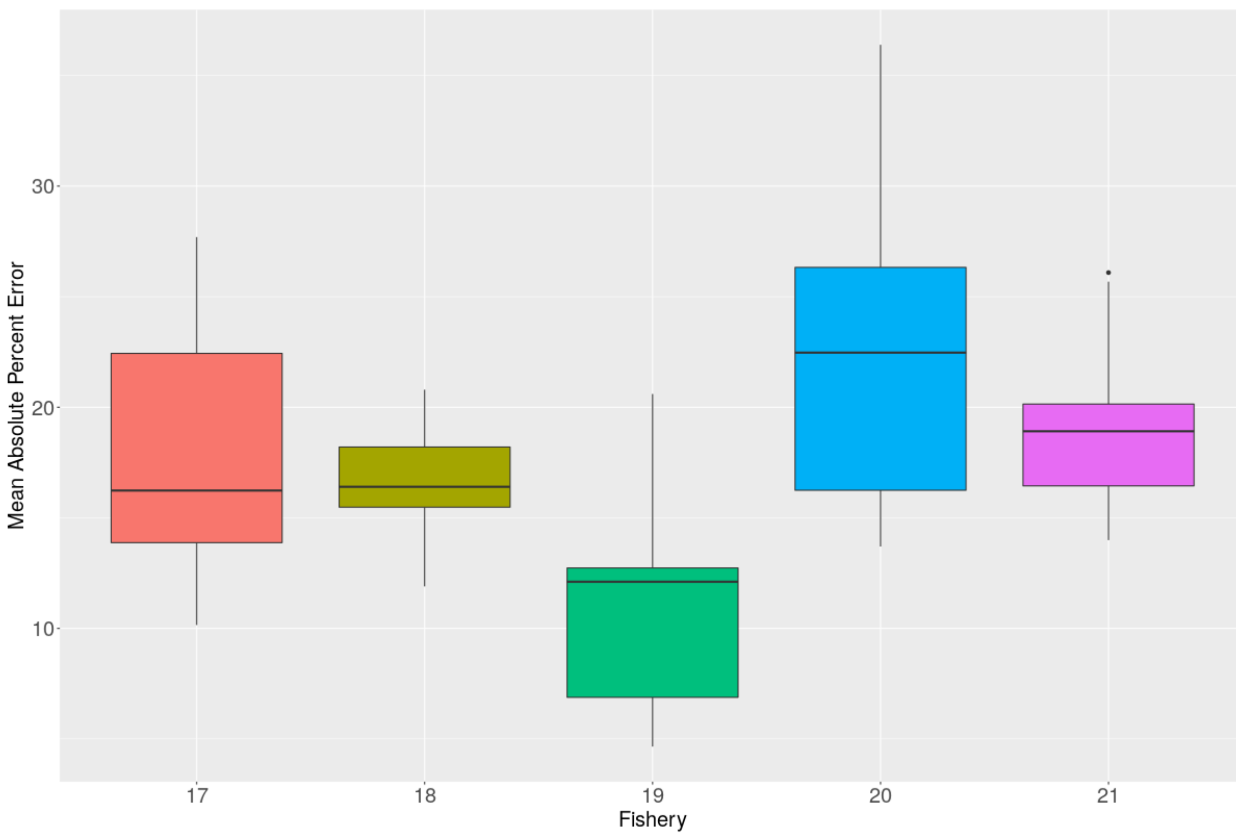


Figure 5: Mean absolute percent error (MAPE) result for the five index fishery (fishery 17 to 21), model area 1 to 5, respectively.

Model					Fishery 17 (area 1)	Fishery 18 (area 2)	Fishery 19 (area 3)	Fishery 20 (area 4)	Fishery 21 (area 5)
CPUE	Growth	Size data Weighting	Steepness	Natural Mortality					
Geo	Chen-Wells	20	0.65	0.3	12.51	20.74	19.45	28.43	17.89
Geo	Chen-Wells	20	0.65	0.4	14.67	20.03	12.02	22.16	19.42
Geo	Chen-Wells	20	0.8	0.3	13.41	20.79	20.23	28.13	17.91
Geo	Chen-Wells	20	0.8	0.4	15.24	20.07	12.59	22.19	18.91
Geo	Chen-Wells	20	0.95	0.3	12.73	20.73	20.60	28.31	18.04
Geo	Chen-Wells	20	0.95	0.4	14.70	20.13	12.69	22.11	18.96
Geo	Chen-Wells	50	0.65	0.3	10.49	15.88	5.29	34.40	21.41
Geo	Chen-Wells	50	0.65	0.4	12.35	14.87	5.03	22.82	20.26
Geo	Chen-Wells	50	0.8	0.3	10.15	15.71	5.99	33.93	20.65
Geo	Chen-Wells	50	0.8	0.4	12.40	14.96	5.65	22.87	19.74
Geo	Chen-Wells	50	0.95	0.3	10.45	16.04	6.07	34.83	20.11
Geo	Chen-Wells	50	0.95	0.4	12.28	14.89	5.85	23.03	19.57
Geo	Chen-Wells	80	0.65	0.3	14.20	16.79	9.83	36.39	25.68
Geo	Chen-Wells	80	0.65	0.4	15.78	13.06	5.31	23.84	25.41
Geo	Chen-Wells	80	0.8	0.3	14.03	16.52	9.05	36.12	24.69
Geo	Chen-Wells	80	0.8	0.4	15.93	13.28	4.66	23.89	24.65
Geo	Chen-Wells	80	0.95	0.3	14.20	16.70	8.31	36.02	23.67
Geo	Chen-Wells	80	0.95	0.4	16.53	11.90	4.94	22.75	26.09
Geo	Estimated	20	0.65	0.3	20.79	14.87	17.49	16.19	18.13
Geo	Estimated	20	0.65	0.4	24.02	18.05	16.65	16.21	15.82
Geo	Estimated	20	0.8	0.3	24.11	18.17	17.12	16.34	15.83
Geo	Estimated	20	0.8	0.4	15.24	20.07	12.59	22.19	18.91
Geo	Estimated	20	0.95	0.3	21.26	15.14	17.55	16.00	17.59
Geo	Estimated	20	0.95	0.4	24.33	18.13	17.14	16.29	16.06
Geo	Estimated	50	0.65	0.3	22.07	17.73	12.23	15.63	16.43
Geo	Estimated	50	0.65	0.4	27.69	15.85	12.67	13.70	16.69
Geo	Estimated	50	0.8	0.3	21.07	16.91	12.59	15.52	15.77
Geo	Estimated	50	0.8	0.4	27.15	15.59	12.26	14.54	17.25
Geo	Estimated	50	0.95	0.3	22.33	17.83	12.86	16.25	16.43
Geo	Estimated	50	0.95	0.4	27.67	15.90	12.59	14.30	16.45
Geo	Estimated	80	0.65	0.3	18.68	18.28	9.46	25.52	15.30
Geo	Estimated	80	0.65	0.4	25.14	16.23	12.20	16.85	19.41
Geo	Estimated	80	0.8	0.3	18.70	18.38	9.25	25.72	15.13

Geo	Estimated	80	0.8	0.4	25.06	16.29	11.54	17.27	19.26
Geo	Estimated	80	0.95	0.3	17.80	16.12	7.15	23.05	13.99
Geo	Estimated	80	0.95	0.4	22.76	15.10	9.95	15.90	19.36
Average across all models					18.00	16.88	11.30	22.77	19.08

Table 4: Mean absolute percentage error (MAPE) calculated between the retrospective forecasted CPUE and the observed CPUE of the five index fisheries (standardized CPUE for each area) from the 36 models selected from the South Pacific albacore tuna assessment grid.

6 Discussion

The development of the South Pacific albacore MSE framework will rely heavily on the use of CPUE data to either directly or indirectly inform on stock status. Therefore, the ability to simulate plausible future CPUE time-series is crucial for testing candidate MPs under the MSE framework.

Retrospective patterns are detected from numerous retro models and the Mohn's ρ values have suggested a common underestimation of the adult biomass across many of the 36 models selected from the assessment grid. Since retrospective patterns in assessment models can lead to severe errors when providing management advice (Hurtado-Ferro et al., 2015), using models that show strong retrospective patterns within MSE frameworks should be treated with caution.

The causes of a specific retrospective pattern are often difficult to determine (Brooks and Legault, 2016). Previous studies have identified model misspecification as a common cause of retrospective pattern (Mohn, 1999; Cadigan and Farrell, 2005). The results of further investigations in our case study have indicated that natural mortality and size data weighting are the two major sources that contributed to the retrospective bias. Natural mortality has also been identified as a major source of uncertainty in the South Pacific albacore assessment. Recommendation has been made to investigate the model misspecification of the key demographic parameters including natural mortality (Tremblay-Boyer et al., 2018b).

When a retrospective pattern is detected in the stock assessment, a common approach is to apply the bias adjustment to correct the retrospective bias (Brooks and Legault, 2016; Deroba, 2014). However, the feasibility of applying this approach to an integrated model (e.g.MFCL) is not clear. Also, there is a risk of introducing further misspecificaiton if these corrections are attempted (Hurtado-Ferro et al., 2015).

The forecasted standardized CPUE time-series display deviations from the observed CPUE time-series among the five areas. The variability in the forecasted CPUE is a consequence of the variability in the terminal biomass of each retro model as suggested by the retrospective analysis. The forecasted standardized CPUE time-series in area 3 shows the most potential to be used in the MP as the stock status indicator.

Besides the standardized CPUE time-series, the nominal CPUE time-series of each longline fishery (Table. 1) was also forecasted. Compared to the standardized CPUE time-series, the performance of forecasted nominal CPUE time-series was worse, with greater prediction error. It is likely due to the fact that the standardized CPUE time-series are used as the principal indices of stock abundance in each area in the stock assessment model (Tremblay-Boyer et al., 2018b), therefore, they are less noisy than the nominal CPUE. Under the current framework, standardized CPUE achieves better prediction accuracy. Conversely, nominal CPUE doesn't require any standardization, it could be viewed as more direct and transparent mechanism for managing the fishery. However, since the most important determining factor for selecting the estimation method is to be able to track the stock status, the standardized CPUE seems a more reasonable option based on the current study.

The caveat of this study is that the period of the retrospective analysis considered is just 5 years (2016 to 2012) and the conclusions may be unrepresentative of a longer time period. However, the focus of this paper is on testing the plausibility of using forecasted CPUE in a MSE framework. Brooks and Legault (2016) recommend to developing a MP that explicitly incorporates buffers in terms of unmodeled, unmeasured and underestimated uncertainty. The empirical-based MP relies solely on the forecasted CPUE time-series as the stock status indicator, which could be problematic when the CPUE time-series doesn't track the biomass well. This situation could be improved if we provide more information to estimate the stock status. For example, a simple biomass dynamic model uses both standardized CPUE and catch information to estimate stock status, which might increase accuracy of the estimation. Therefore, future work on developing the South Pacific albacore MSE framework should involve testing MPs that use standardized CPUE directly along with a simple biomass dynamic model as the estimation model.

7 Conclusions

Our results from the retrospective analysis on the 36 selected models from the 2018 South Pacific albacore grid show evidences of a trend for underestimation of adult biomass with successive removal of years from the data. The forecasted standardized CPUE time-series in area 3 provides the most reliable estimation of future CPUE for informing the MP for MSE testing.

We invited SC to:

- Note the progress on the retrospective forecasting of CPUE for South Pacific albacore.
- Acknowledge the advantage and caveat of the forecasted CPUE time-series.
- Supporting further development of model-based (biomass dynamic model) estimation methods for testing and comparison with CPUE in a MP for South Pacific albacore.

Acknowledgments

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

References

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.

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

Butterworth, D., Cochrane, K. L., and De Oliveira, J. (1997). *Global Trends: Fisheries Management*, chapter Management procedures: a better way to manage fisheries? The South African experience. American Fisheries Society Symposium.

Cadigan, N. G. and Farrell, P. J. (2005). Local influence diagnostics for the retrospective problem in sequential population analysis. *ICES Journal of Marine Science*, 62(2):256–265.

Calvin, K., Wise, M., Kyle, P., Clarke, L., and Edmonds, J. (2017). A hindcast experiment using the gcam 3.0 agriculture and land-use module. *Climate Change Economics*, 8(01):1750005.

Deroba, J. J. (2014). Evaluating the consequences of adjusting fish stock assessment estimates of biomass for retrospective patterns using mohn's rho. *North American Journal of Fisheries Management*, 34(2):380–390.

Hurtado-Ferro, F., Szuwalski, C. S., Valero, J. L., Anderson, S. C., Cunningham, C. J., Johnson, K. F., Licandeo, R., McGilliard, C. R., Monnahan, C. C., Muradian, M. L., et al. (2015). Looking in the rear-view mirror: bias and retrospective patterns in integrated, age-structured stock assessment models. *ICES Journal of Marine Science*, 72(1):99–110.

Mohn, R. (1999). The retrospective problem in sequential population analysis: an investigation using cod fishery and simulated data. *ICES Journal of Marine Science*, 56(4):473–488.

Punt, A. E., Butterworth, D., de Moor, C., De Oliveira, J., and Haddon, M. (2014). Management strategy evaluation: best practices. *Fish and Fisheries*, (DOI:10.111/faf12104).

Scott, R., Davies, N., Pilling, G. M., and Hampton, J. (2016). Retrospective forecasting of the 2014 WCPO bigeye tuna stock assessment. WCPFC-SC12-2016/SA-WP-02, Bali, Indonesia, 3–11 August 2016.

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

Snyder, A. C., Link, R. P., and Calvin, K. V. (2017). Evaluation of integrated assessment model hindcast experiments: a case study of the gcam 3.0 land use module. *Geoscientific Model Development*, 10(12):4307–4319.

Thorson, J. (2019). Guidance for decisions using the vector autoregressive spatio-temporal (vast) package in stock, ecosystem, habitat and climate assessments. *Fisheries Research*.

Tremblay-Boyer, L., J., H., McKechnie, S., and Pilling, G. (2018a). Background analyses for the 2018 stock assessment of south pacific albacore tuna. WCPFC-SC14-2018/SA-IP-07, Busan, Republic of Korea, August 2018.

Tremblay-Boyer, L., J., H., McKechnie, S., and Pilling, G. (2018b). Stock assessment of south pacific albacore tun rev 2 (29 july 2018). WCPFC-SC14-2018/SA-WP-05, Busan, Republic of Korea, August 2018.

Valero, J. L. (2012). Harvest policy considerations on retrospective bias and biomass projections. Technical report.

Yao, N., Singh, F., Scott, F., and Scott, R. (2020). HCR design considerations for South Pacific albacore. techreport WCPFC-SC16-2020/MI-IP-05, SPC, Online.

8 Appendix

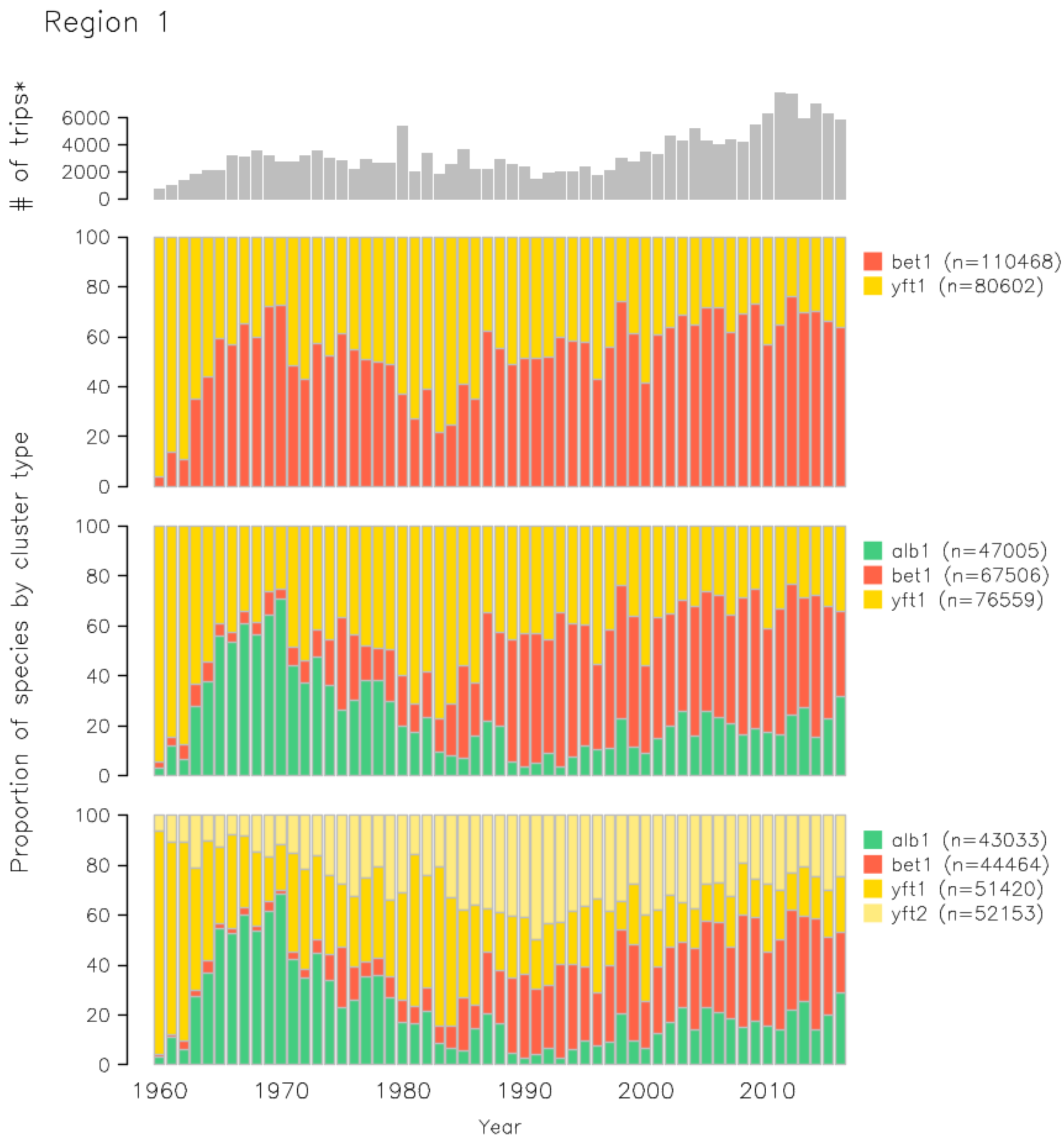


Figure 6: Time-series of cluster membership for the 2, 3 and 4 cluster models in area 1, with the colors matching the most dominant species in the cluster and the top panel indicating the number of trips available in each year.

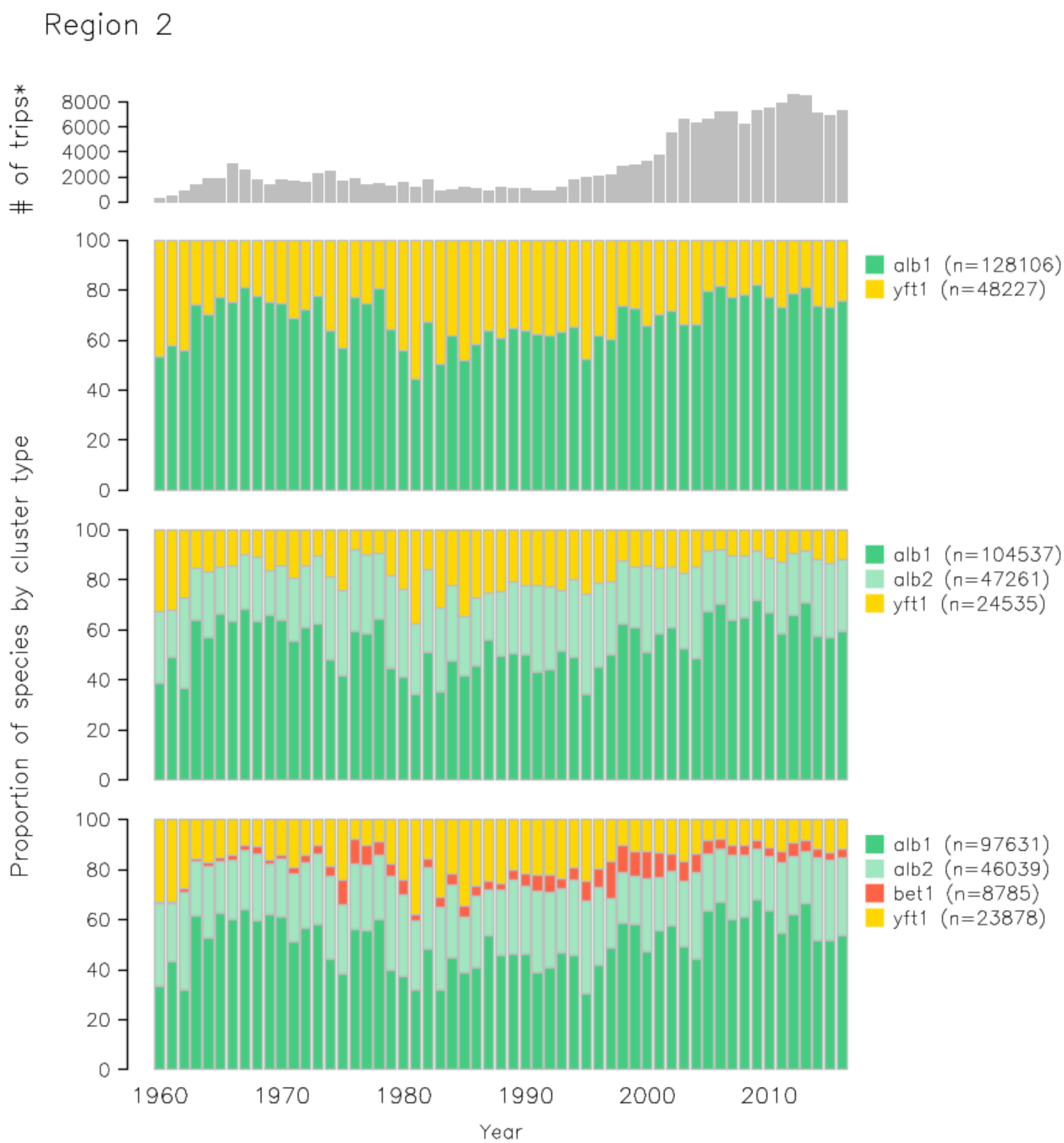


Figure 7: Time-series of cluster membership for the 2, 3 and 4 cluster models in area 2, with the colors matching the most dominant species in the cluster and the top panel indicating the number of trips available in each year.

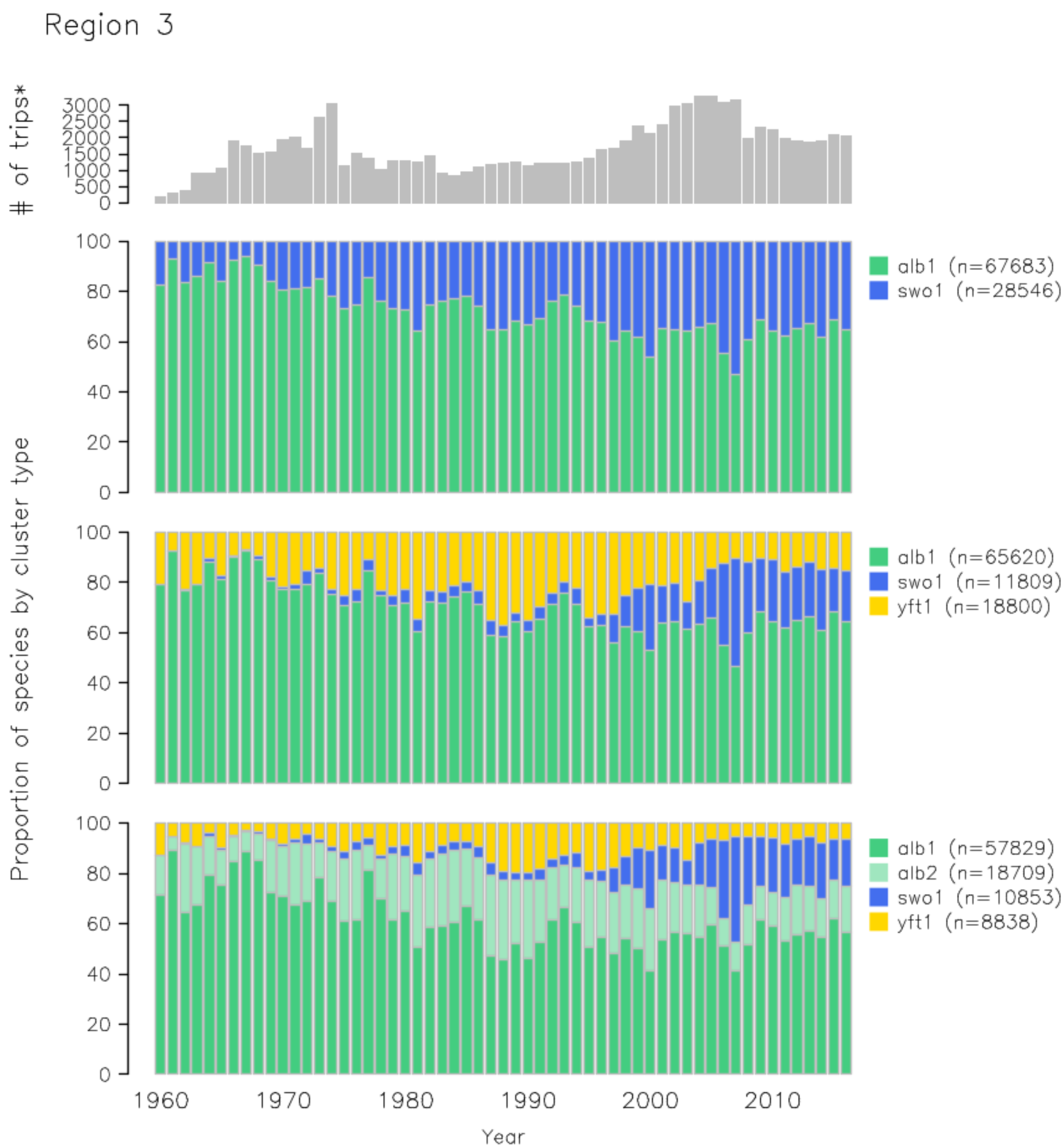


Figure 8: Time-series of cluster membership for the 2, 3 and 4 cluster models in area 3, with the colors matching the most dominant species in the cluster and the top panel indicating the number of trips available in each year.

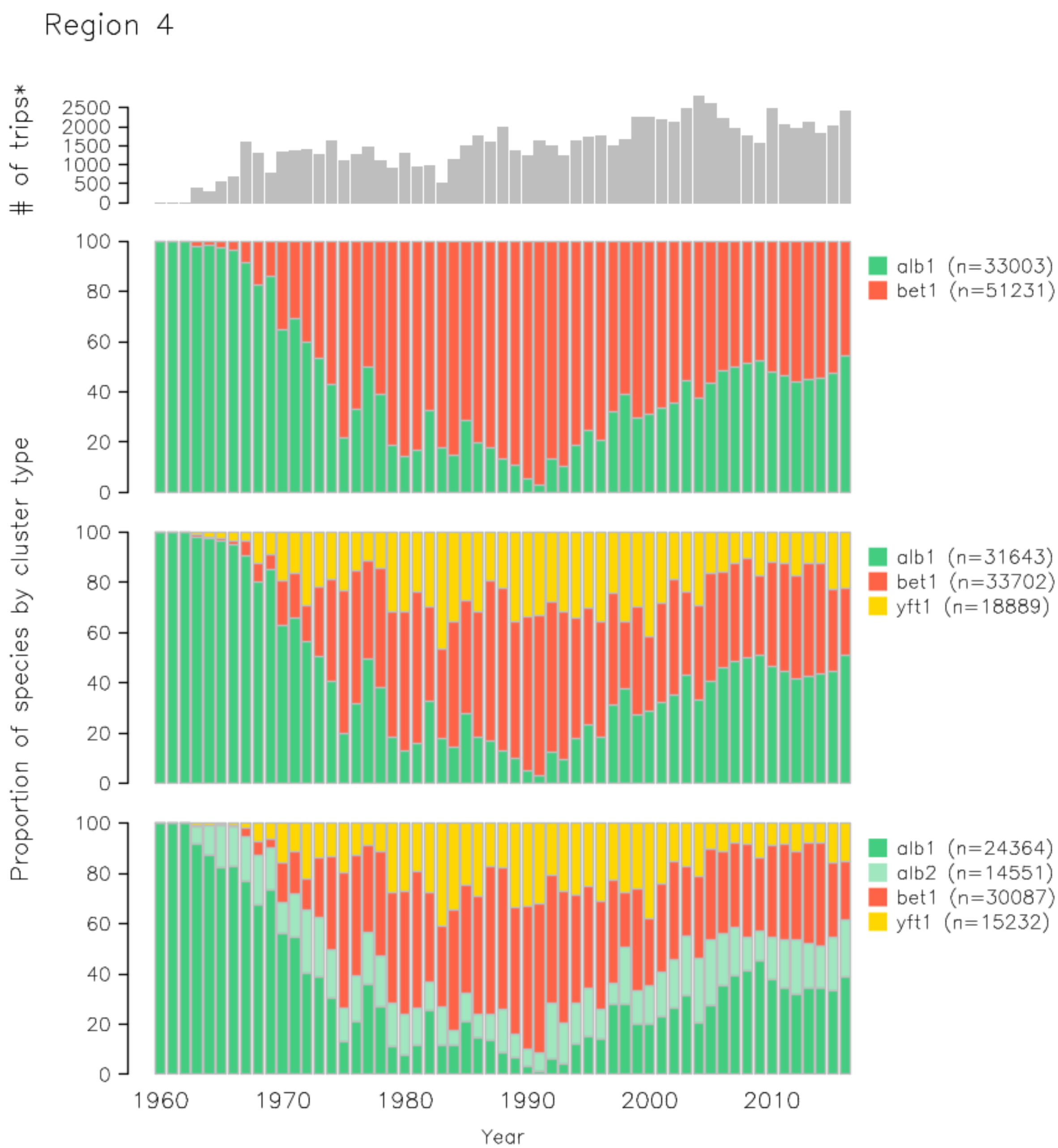


Figure 9: Time-series of cluster membership for the 2, 3 and 4 cluster models in area 4, with the colors matching the most dominant species in the cluster and the top panel indicating the number of trips available in each year.

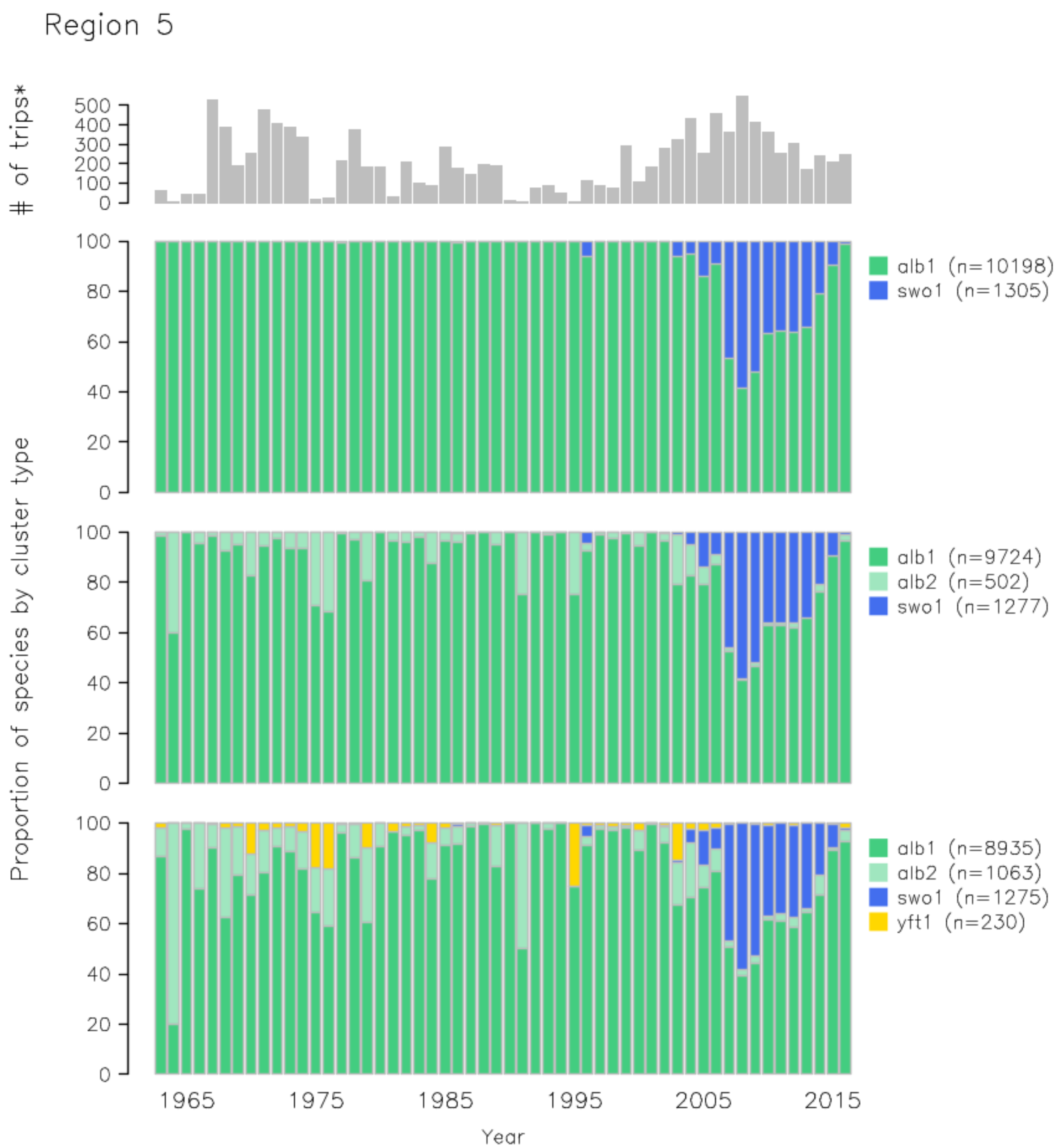


Figure 10: Time-series of cluster membership for the 2, 3 and 4 cluster models in area 5, with the colors matching the most dominant species in the cluster and the top panel indicating the number of trips available in each year.

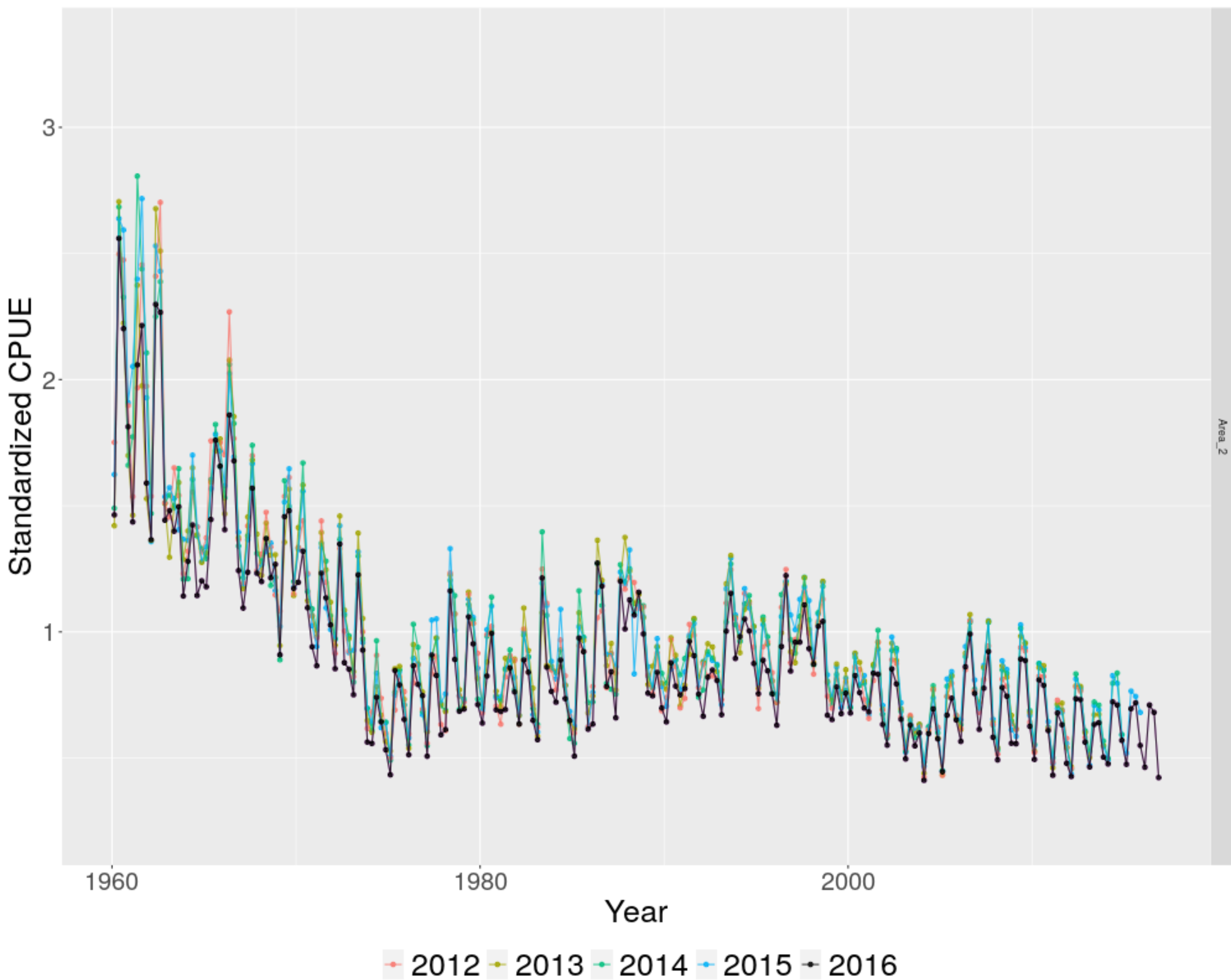


Figure 11: Results of the retrospective CPUE standardization from 2012 to 2016 for stock assessment area 2.

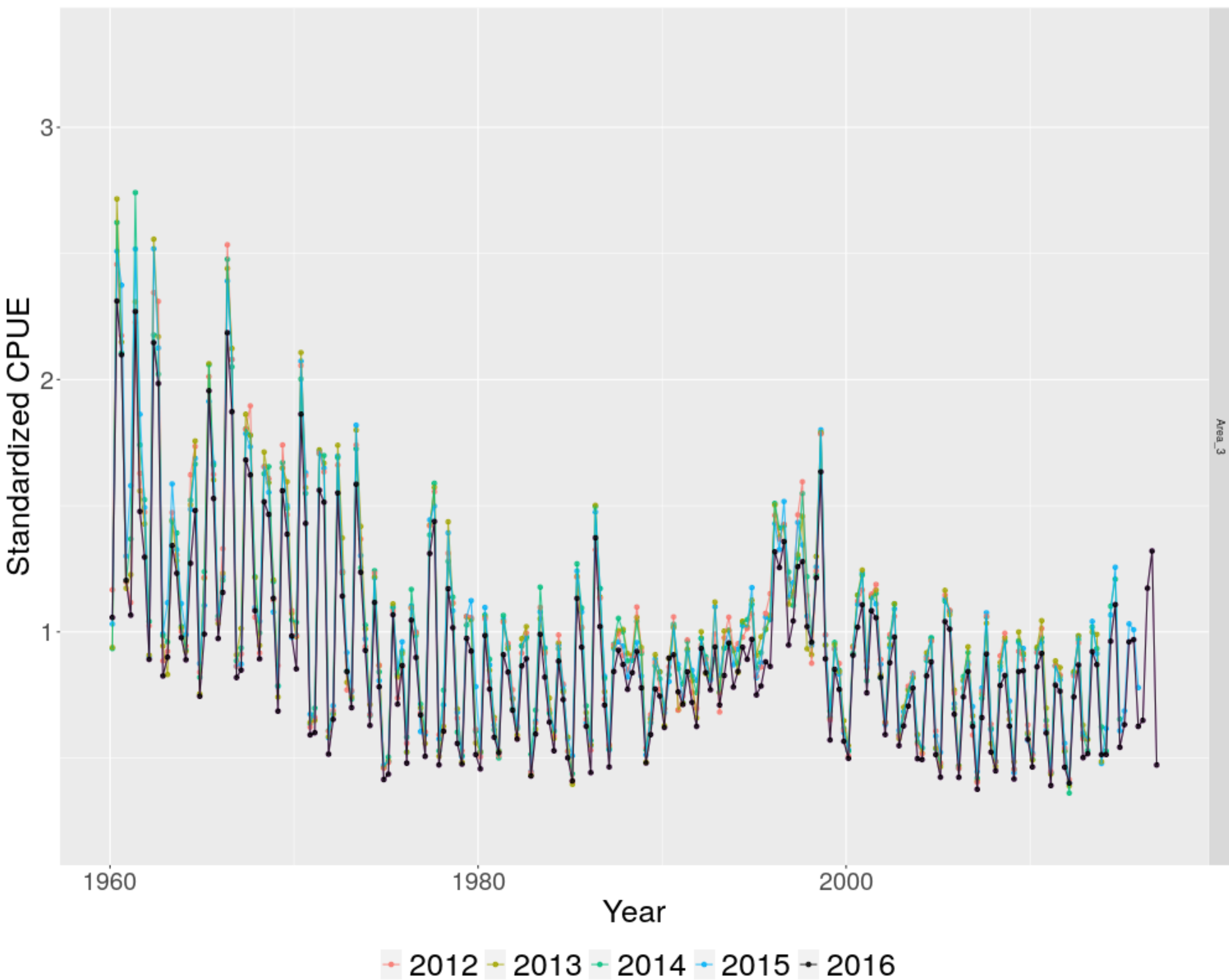


Figure 12: Results of the retrospective CPUE standardization from 2012 to 2016 for stock assessment area 3.

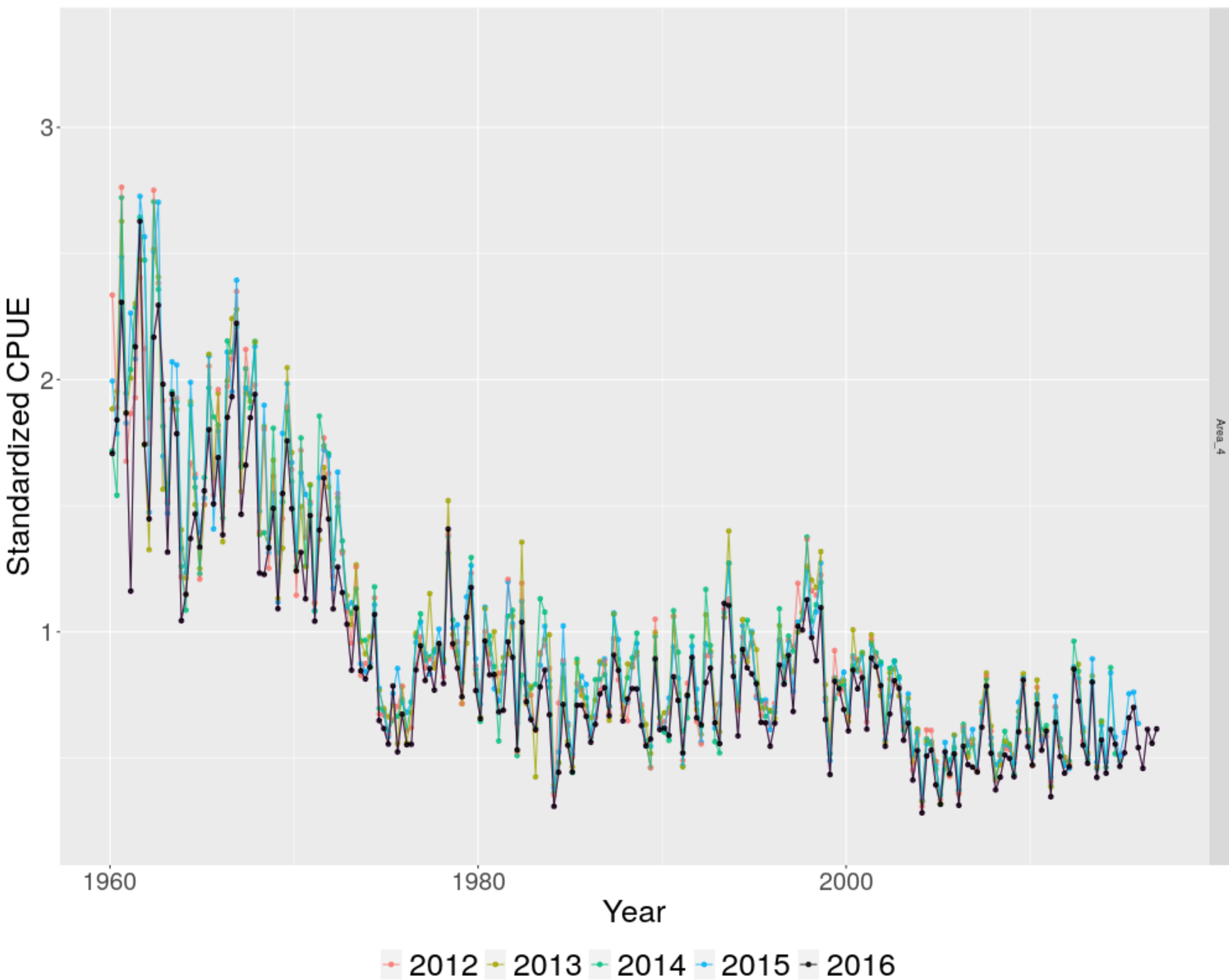


Figure 13: Results of the retrospective CPUE standardization from 2012 to 2016 for stock assessment area 4.

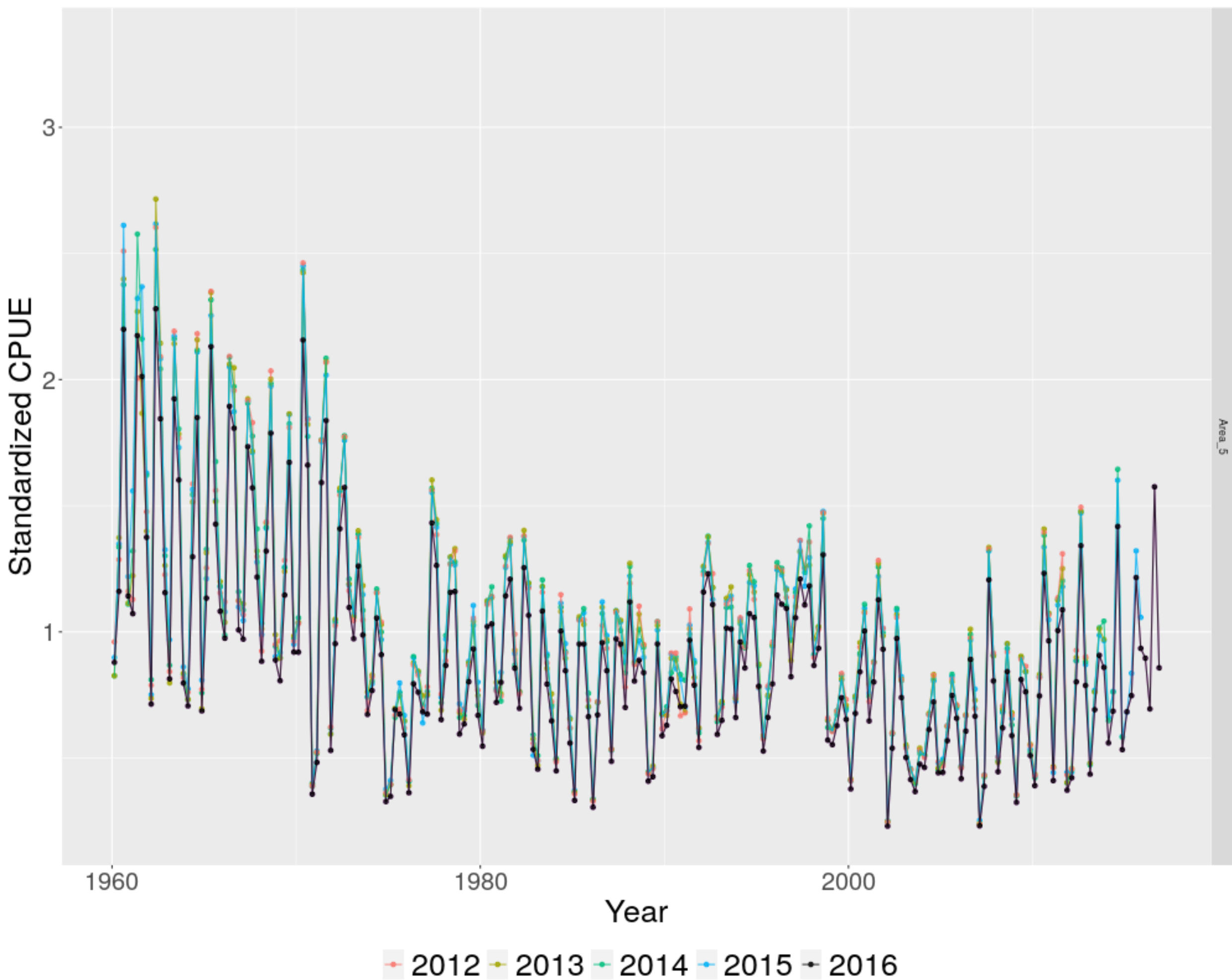


Figure 14: Results of the retrospective CPUE standardization from 2012 to 2016 for stock assessment area 5.

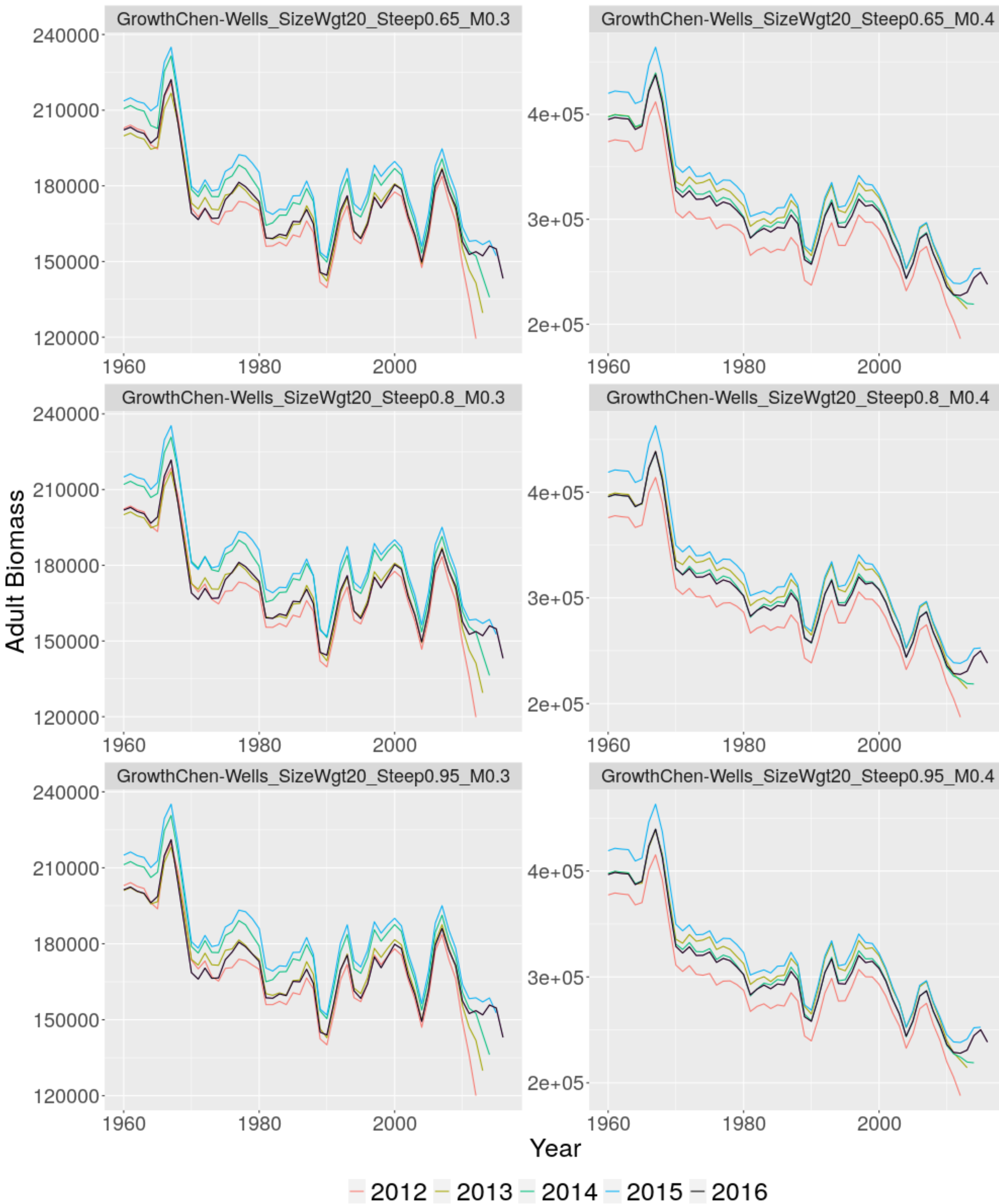


Figure 15: Retrospective results for each individual model. The annual estimates of adult biomass (summed over areas) determined from the full time-series assessment (black line) and retrospective assessment runs (terminal years 2015 to 2012, color lines).

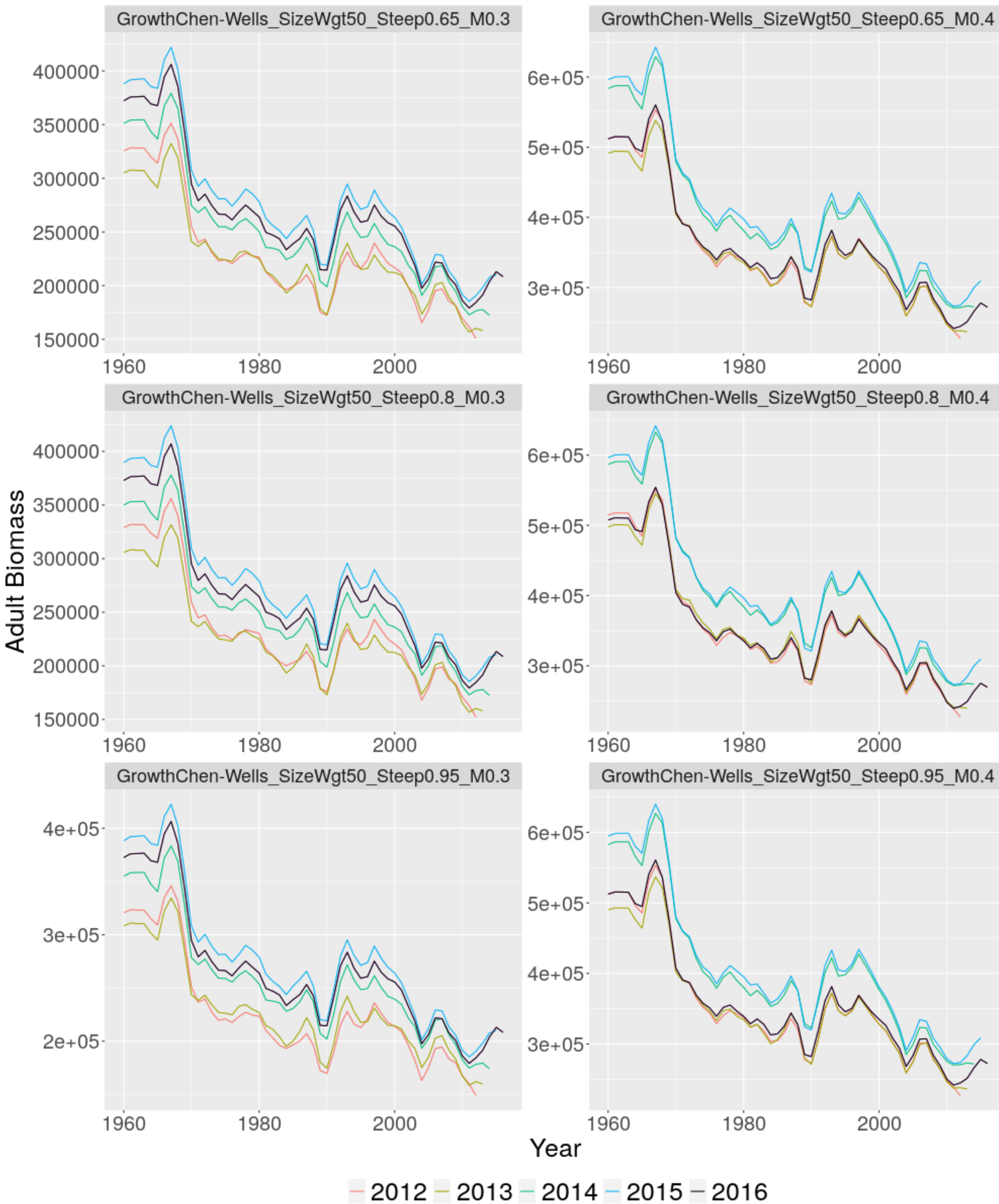


Figure 16: Retrospective results for each individual model. The annual estimates of adult biomass (summed over areas) determined from the full time-series assessment (black line) and retrospective assessment runs (terminal years 2015 to 2012, color lines).

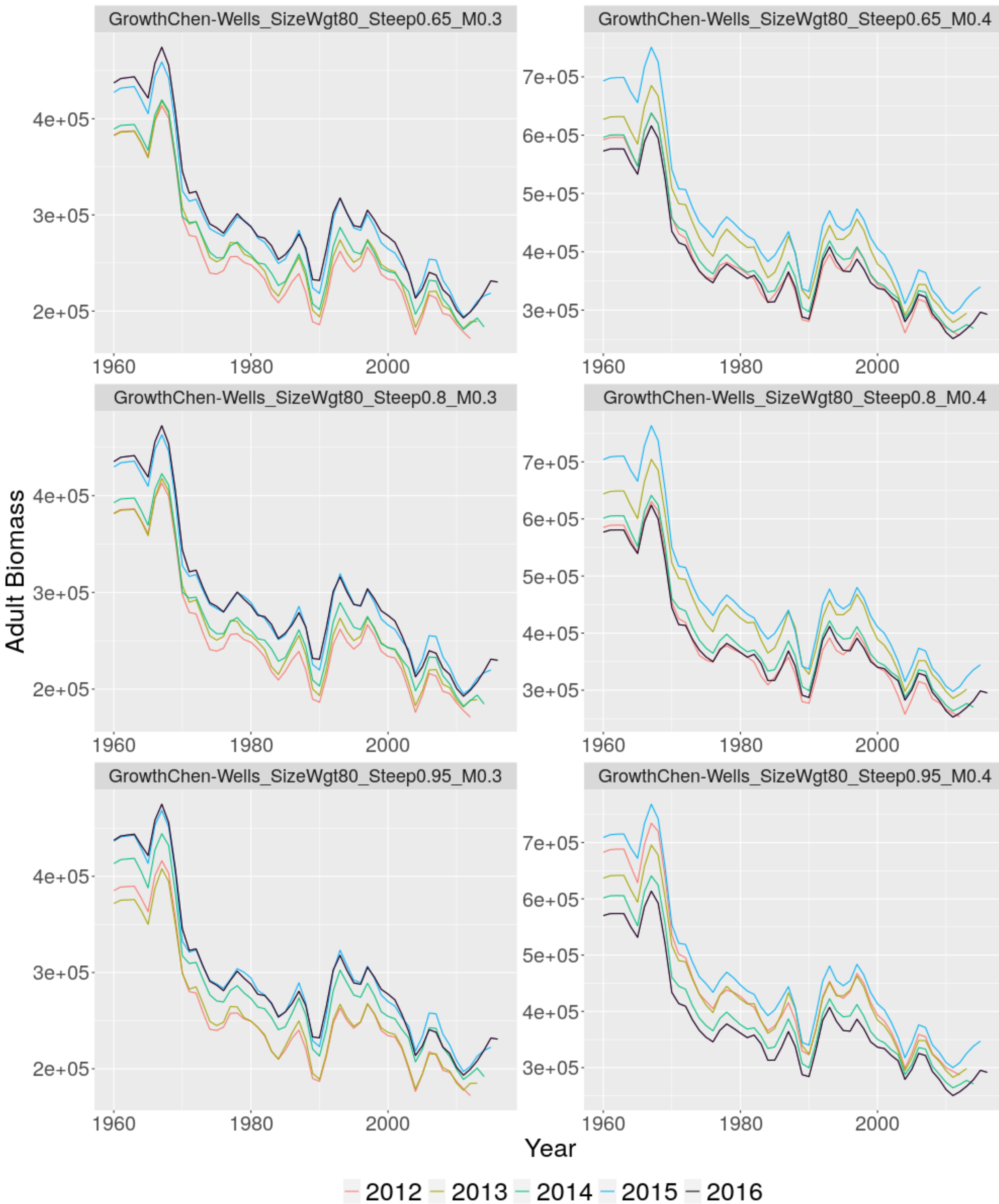


Figure 17: Retrospective results for each individual model. The annual estimates of adult biomass (summed over areas) determined from the full time-series assessment (black line) and retrospective assessment runs (terminal years 2015 to 2012, color lines).

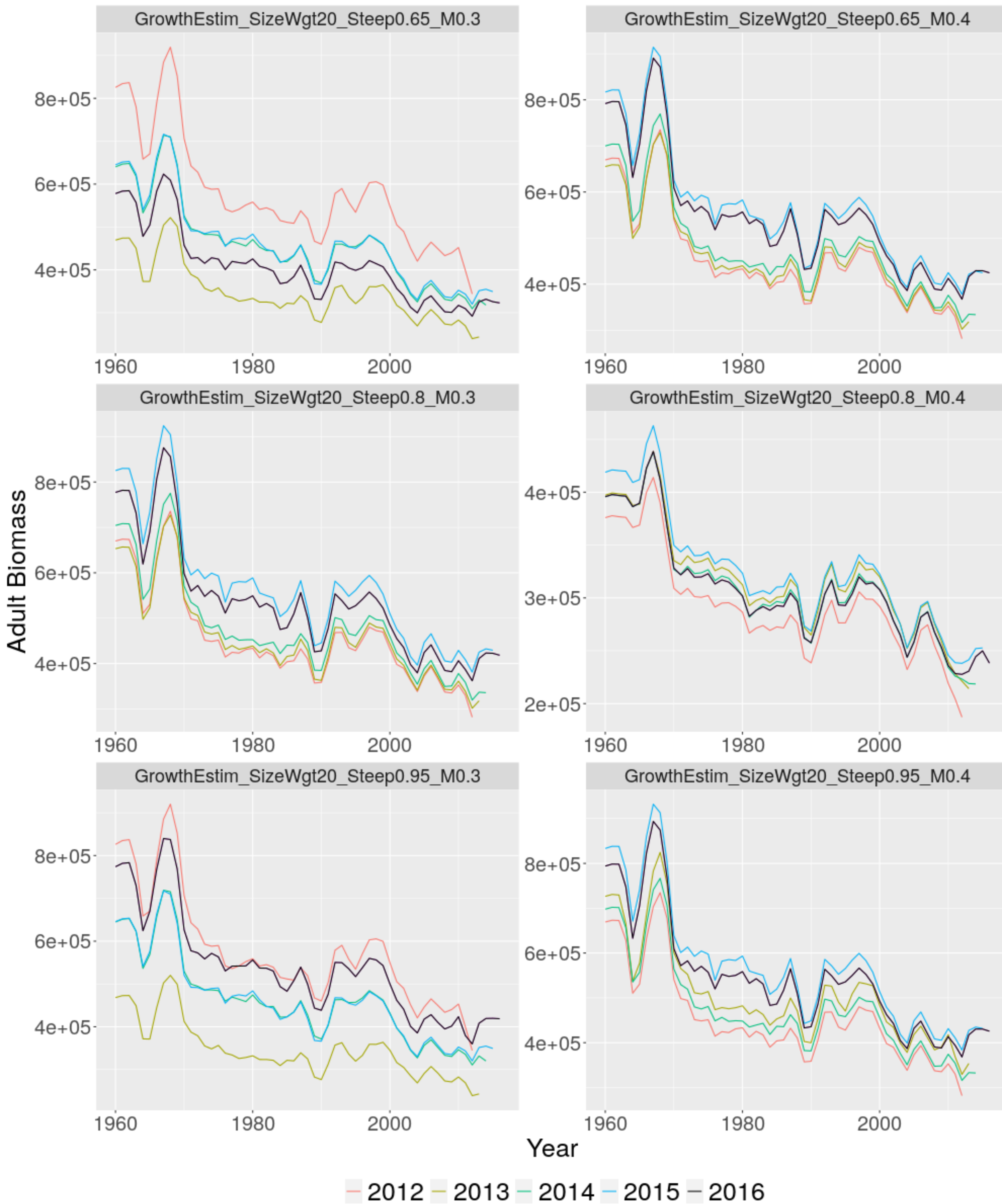


Figure 18: Retrospective results for each individual model. The annual estimates of adult biomass (summed over areas) determined from the full time-series assessment (black line) and retrospective assessment runs (terminal years 2015 to 2012, color lines).

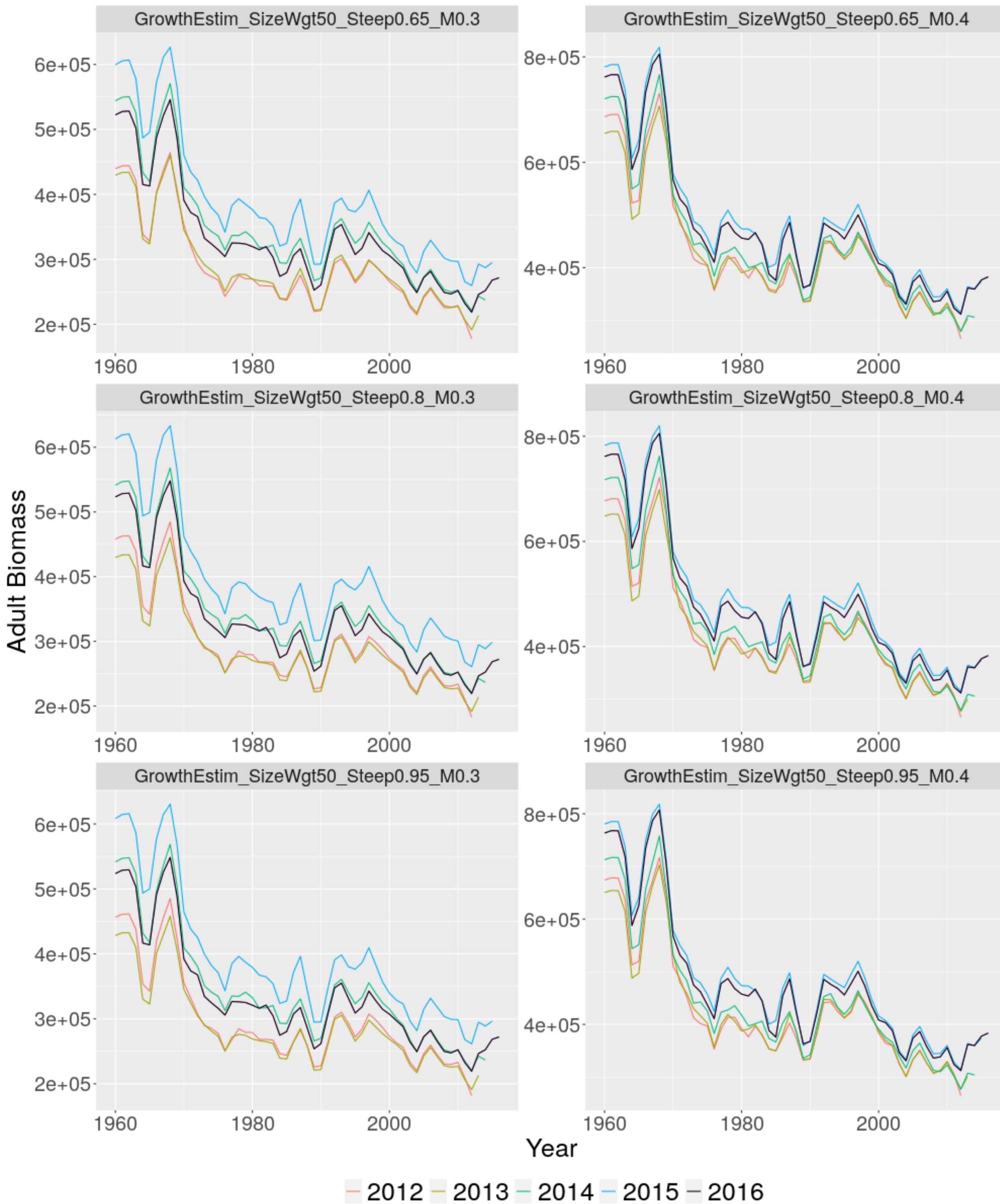


Figure 19: Retrospective results for each individual model. The annual estimates of adult biomass (summed over areas) determined from the full time-series assessment (black line) and retrospective assessment runs (terminal years 2015 to 2012, color lines).

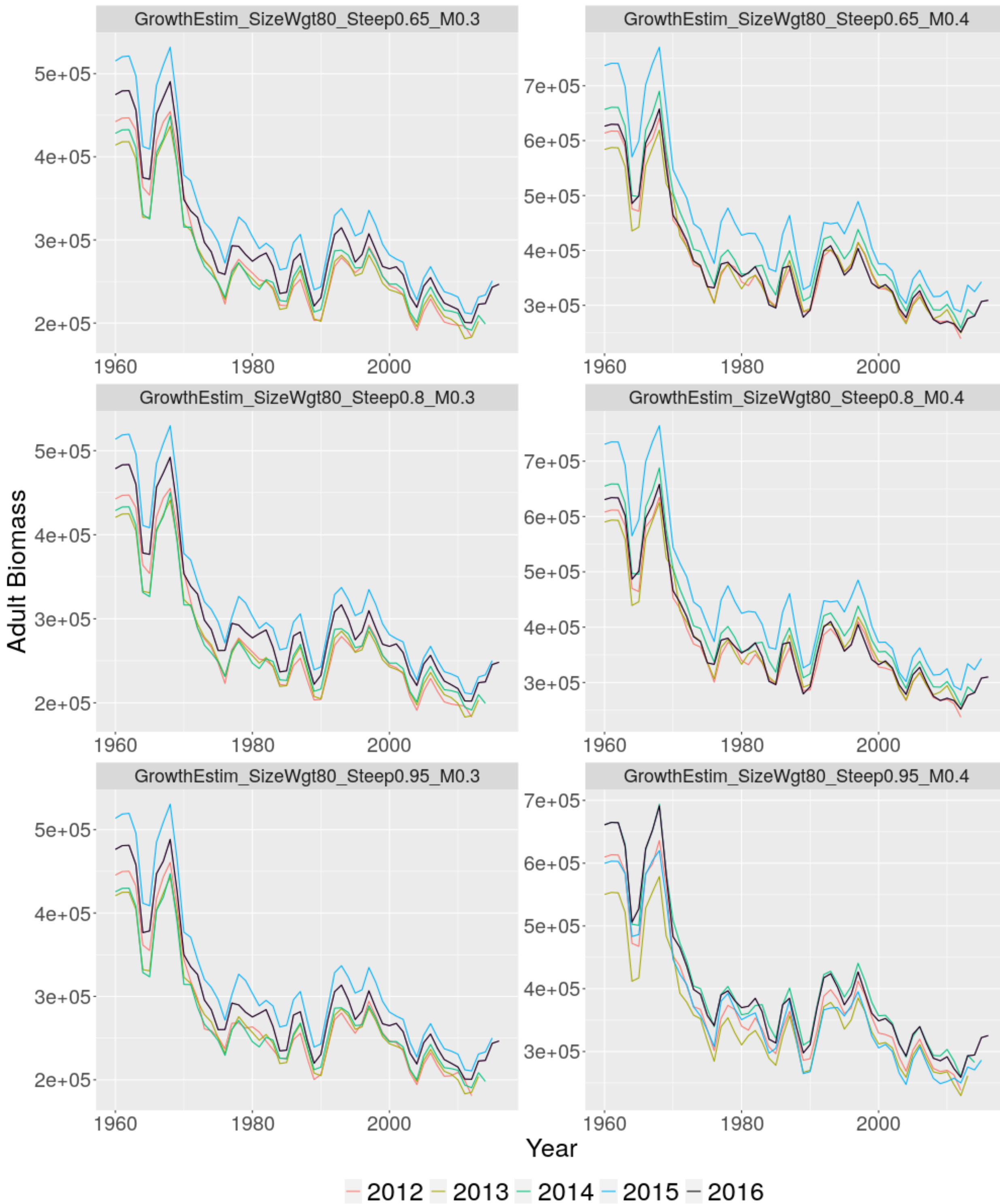


Figure 20: Retrospective results for each individual model. The annual estimates of adult biomass (summed over areas) determined from the full time-series assessment (black line) and retrospective assessment runs (terminal years 2015 to 2012, color lines).

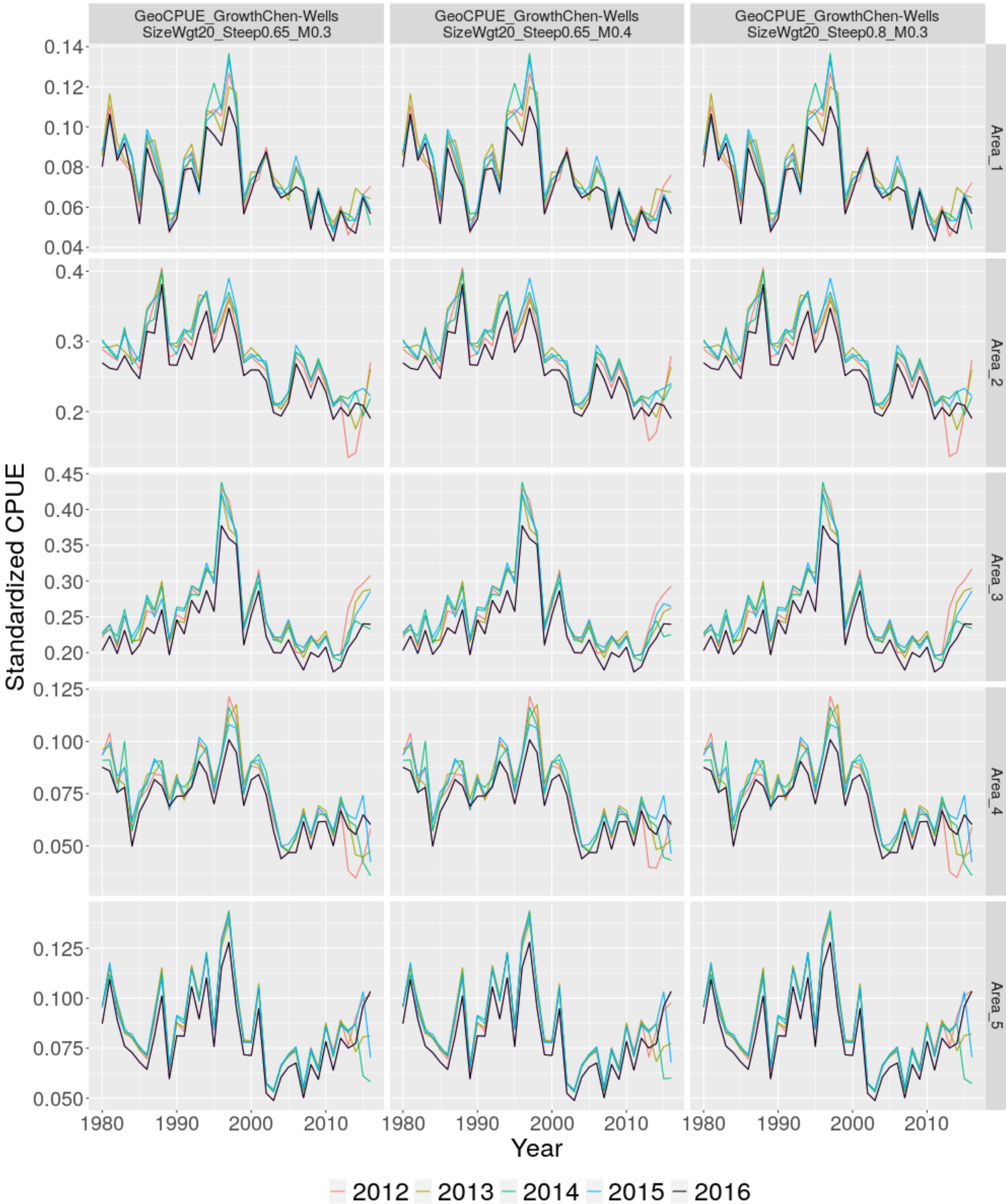


Figure 21: Retrospective forecasting CPUE time-series results for the five index fisheries for each individual model. Annual estimates of the index CPUE (standardized CPUE in each area) determined from the 2018 South Pacific albacore stock assessment (black line) and forecasted CPUE (terminal year 2012 to 2015, color lines).

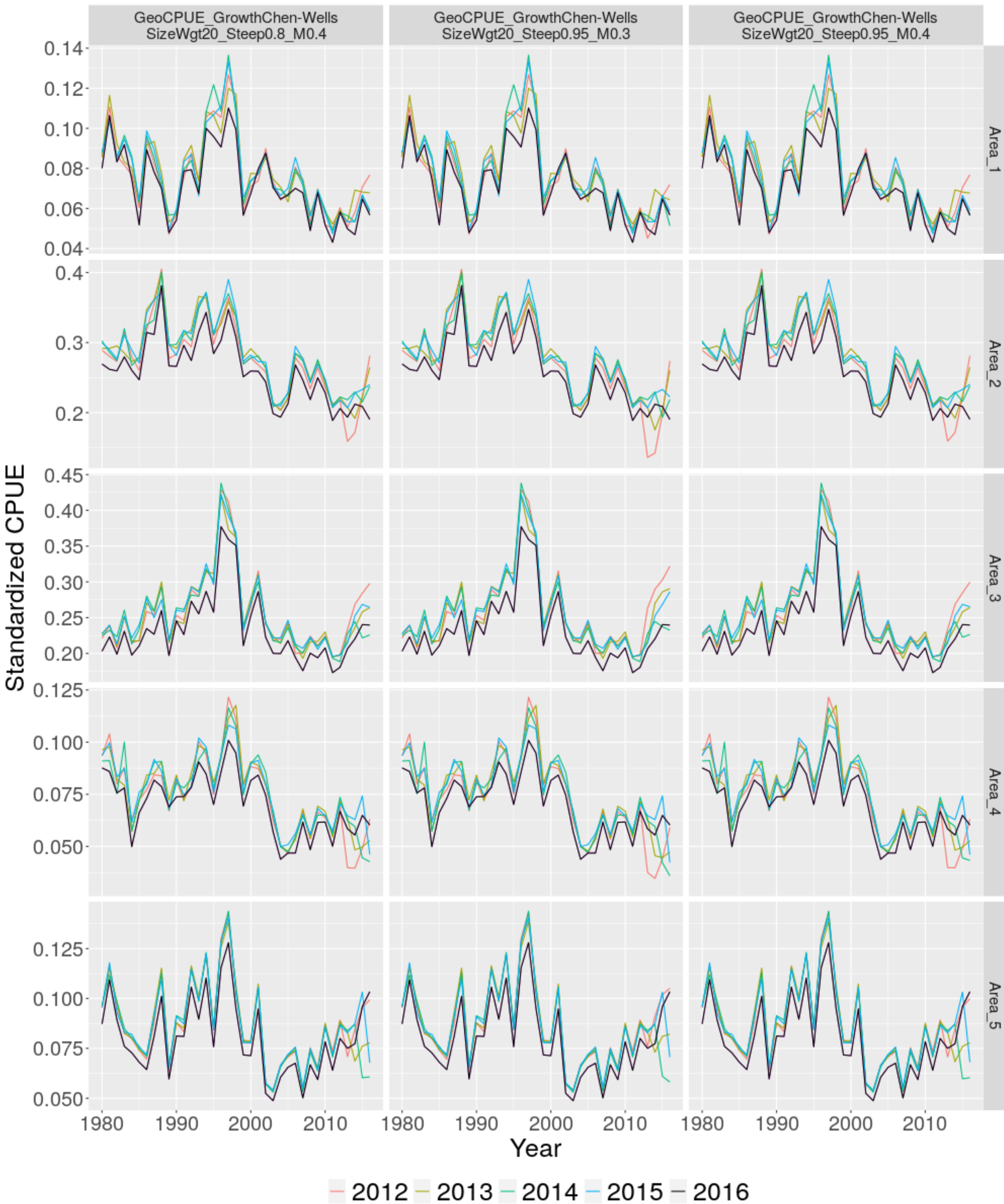


Figure 22: Retrospective forecasting CPUE time-series results for the five index fisheries for each individual model. Annual estimates of the index CPUE (standardized CPUE in each area) determined from the 2018 South Pacific albacore stock assessment (black line) and forecasted CPUE (terminal year 2012 to 2015, color lines).

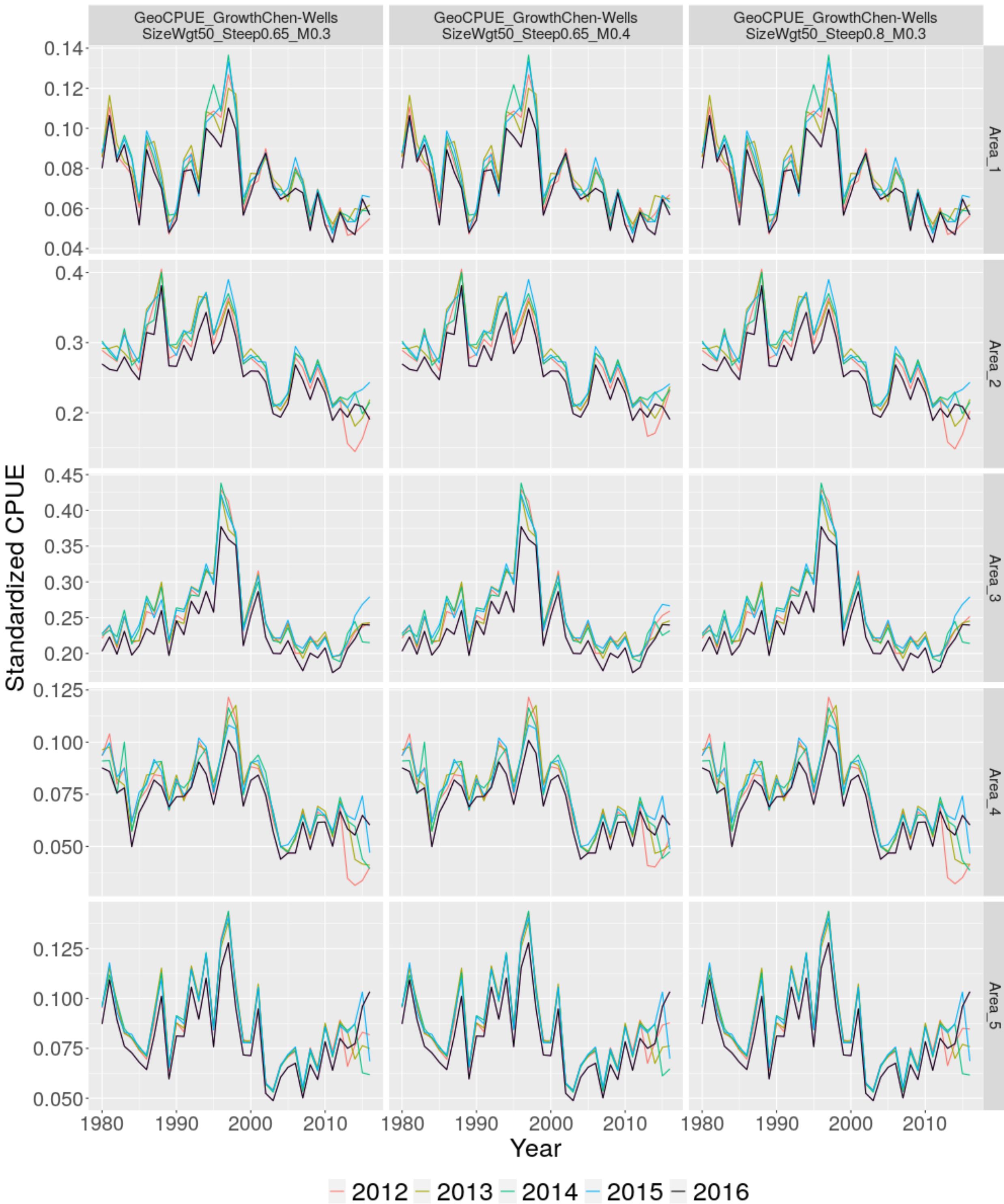


Figure 23: Retrospective forecasting CPUE time-series results for the five index fisheries for each individual model. Annual estimates of the index CPUE (standardized CPUE in each area) determined from the 2018 South Pacific albacore stock assessment (black line) and forecasted CPUE (terminal year 2012 to 2015, color lines).

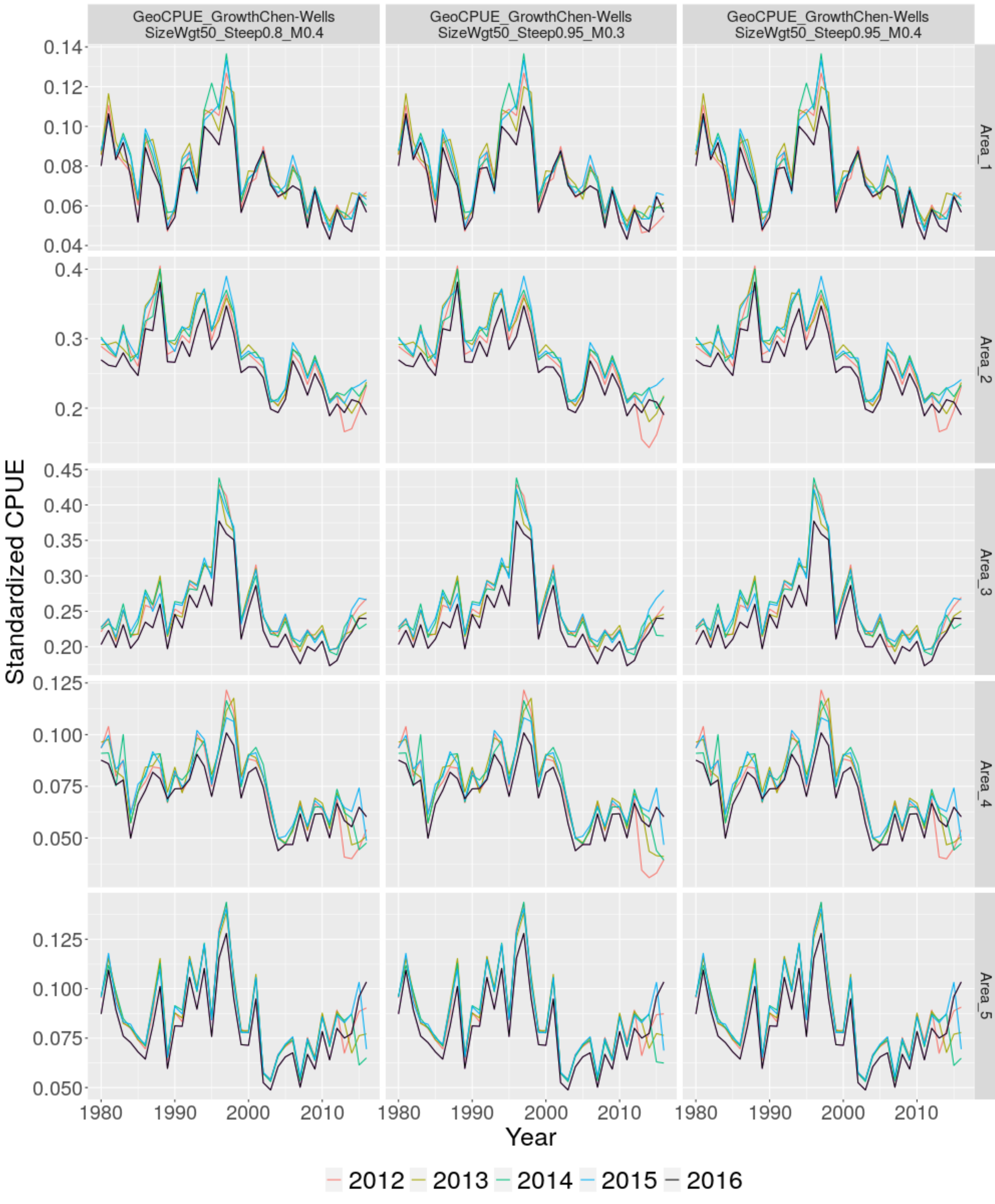


Figure 24: Retrospective forecasting CPUE time-series results for the five index fisheries for each individual model. Annual estimates of the index CPUE (standardized CPUE in each area) determined from the 2018 South Pacific albacore stock assessment (black line) and forecasted CPUE (terminal year 2012 to 2015, color lines).

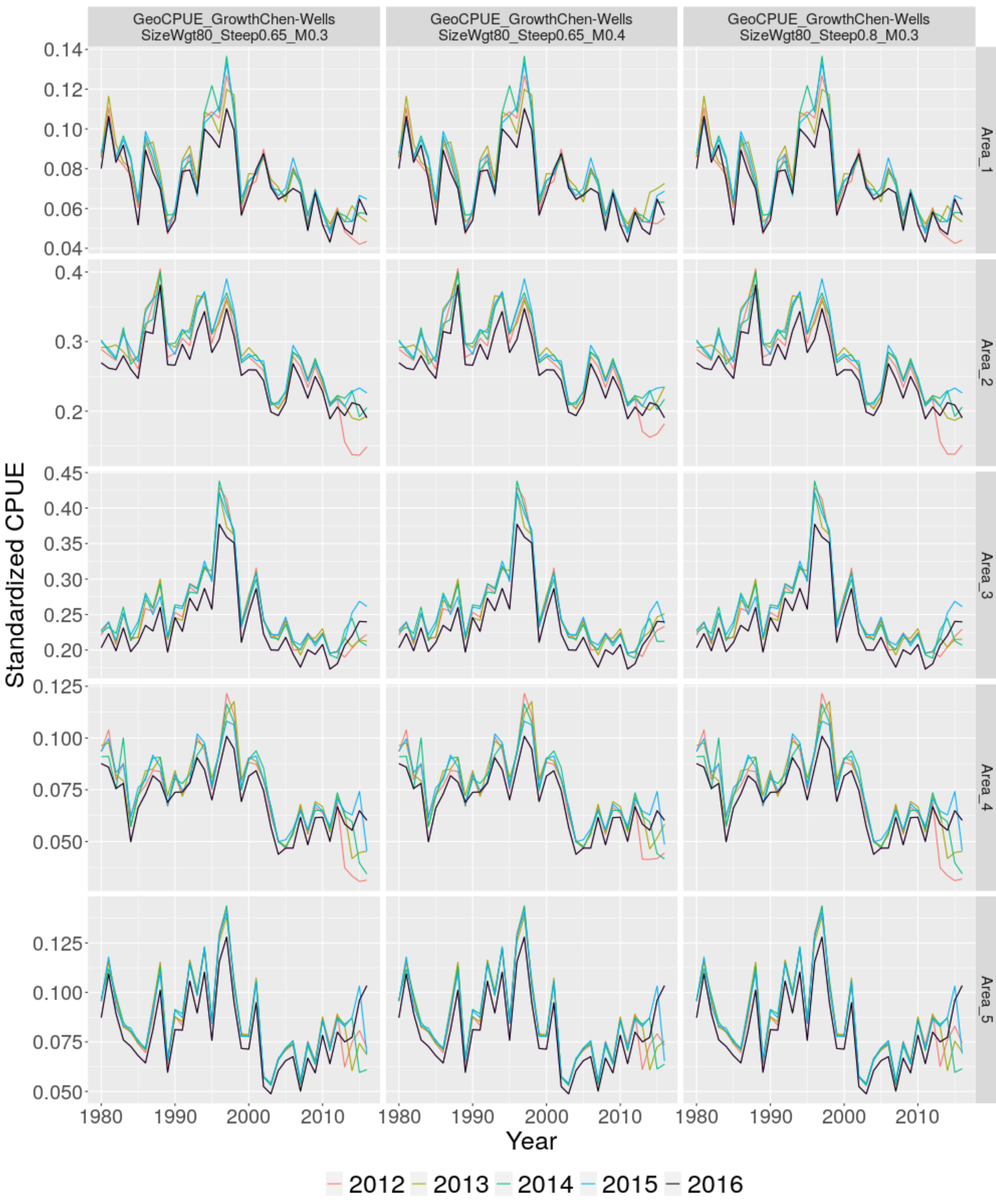


Figure 25: Retrospective forecasting CPUE time-series results for the five index fisheries for each individual model. Annual estimates of the index CPUE (standardized CPUE in each area) determined from the 2018 South Pacific albacore stock assessment (black line) and forecasted CPUE (terminal year 2012 to 2015, color lines).

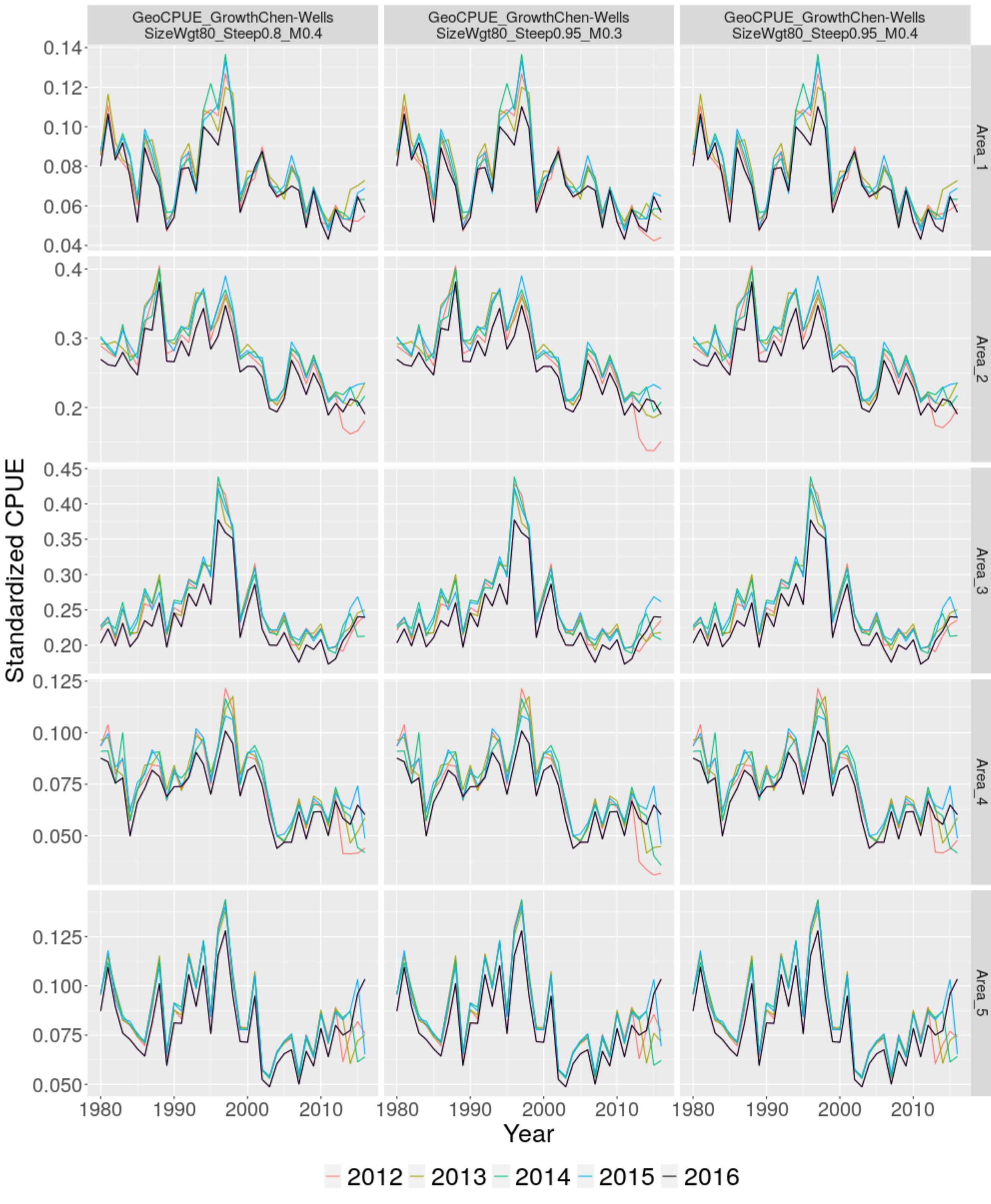


Figure 26: Retrospective forecasting CPUE time-series results for the five index fisheries for each individual model. Annual estimates of the index CPUE (standardized CPUE in each area) determined from the 2018 South Pacific albacore stock assessment (black line) and forecasted CPUE (terminal year 2012 to 2015, color lines).

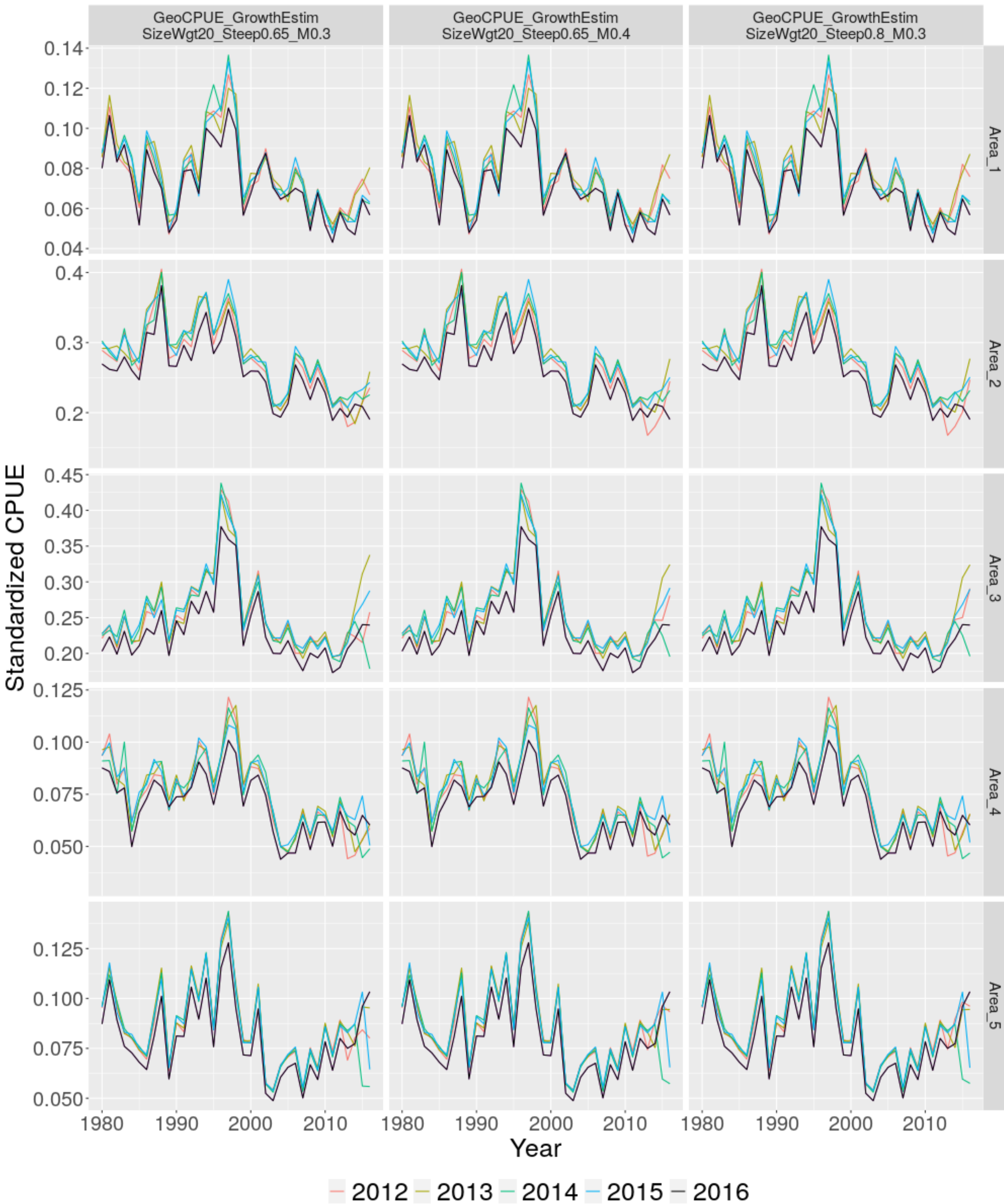


Figure 27: Retrospective forecasting CPUE time-series results for the five index fisheries for each individual model. Annual estimates of the index CPUE (standardized CPUE in each area) determined from the 2018 South Pacific albacore stock assessment (black line) and forecasted CPUE (terminal year 2012 to 2015, color lines).

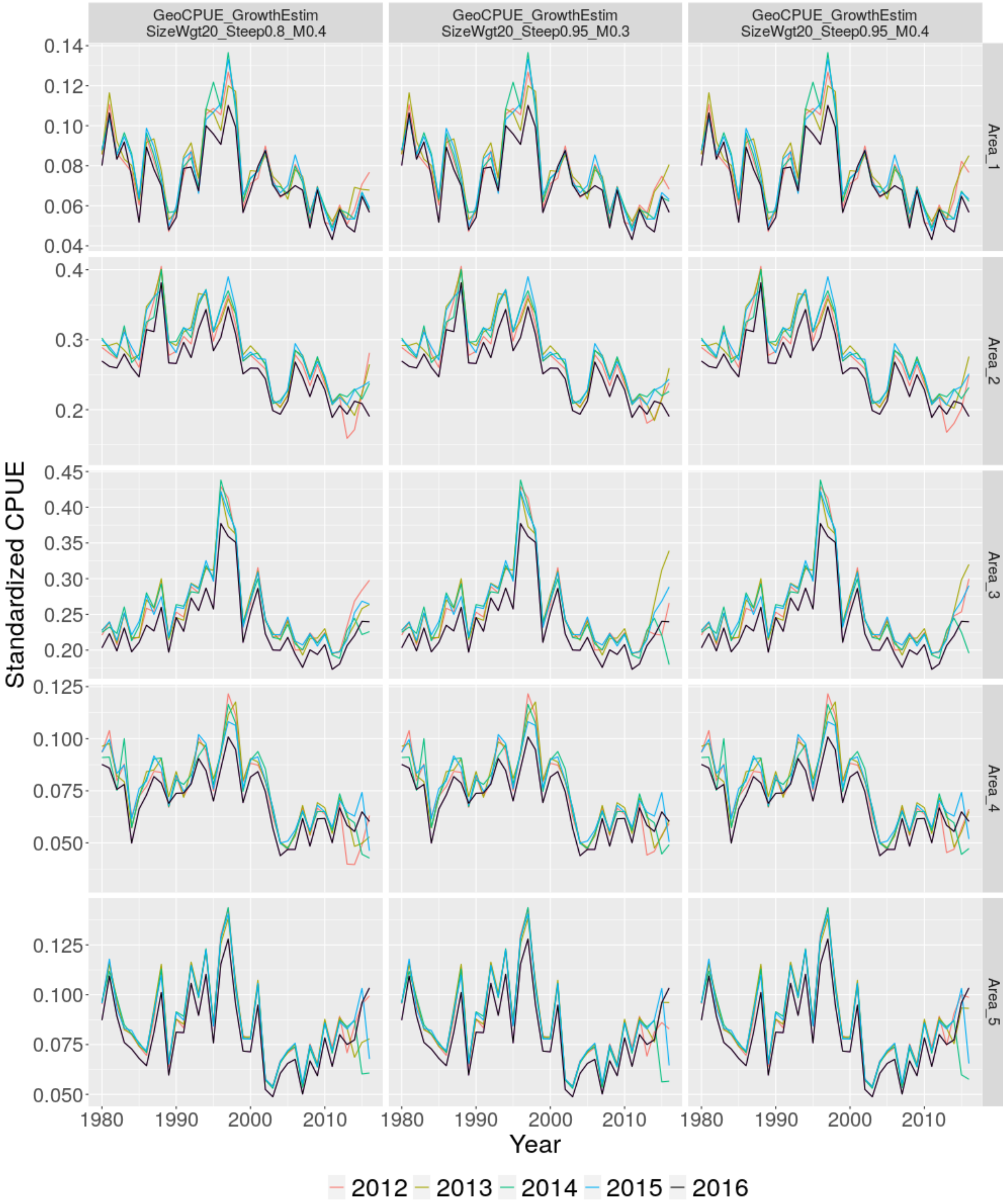


Figure 28: Retrospective forecasting CPUE time-series results for the five index fisheries for each individual model. Annual estimates of the index CPUE (standardized CPUE in each area) determined from the 2018 South Pacific albacore stock assessment (black line) and forecasted CPUE (terminal year 2012 to 2015, color lines).

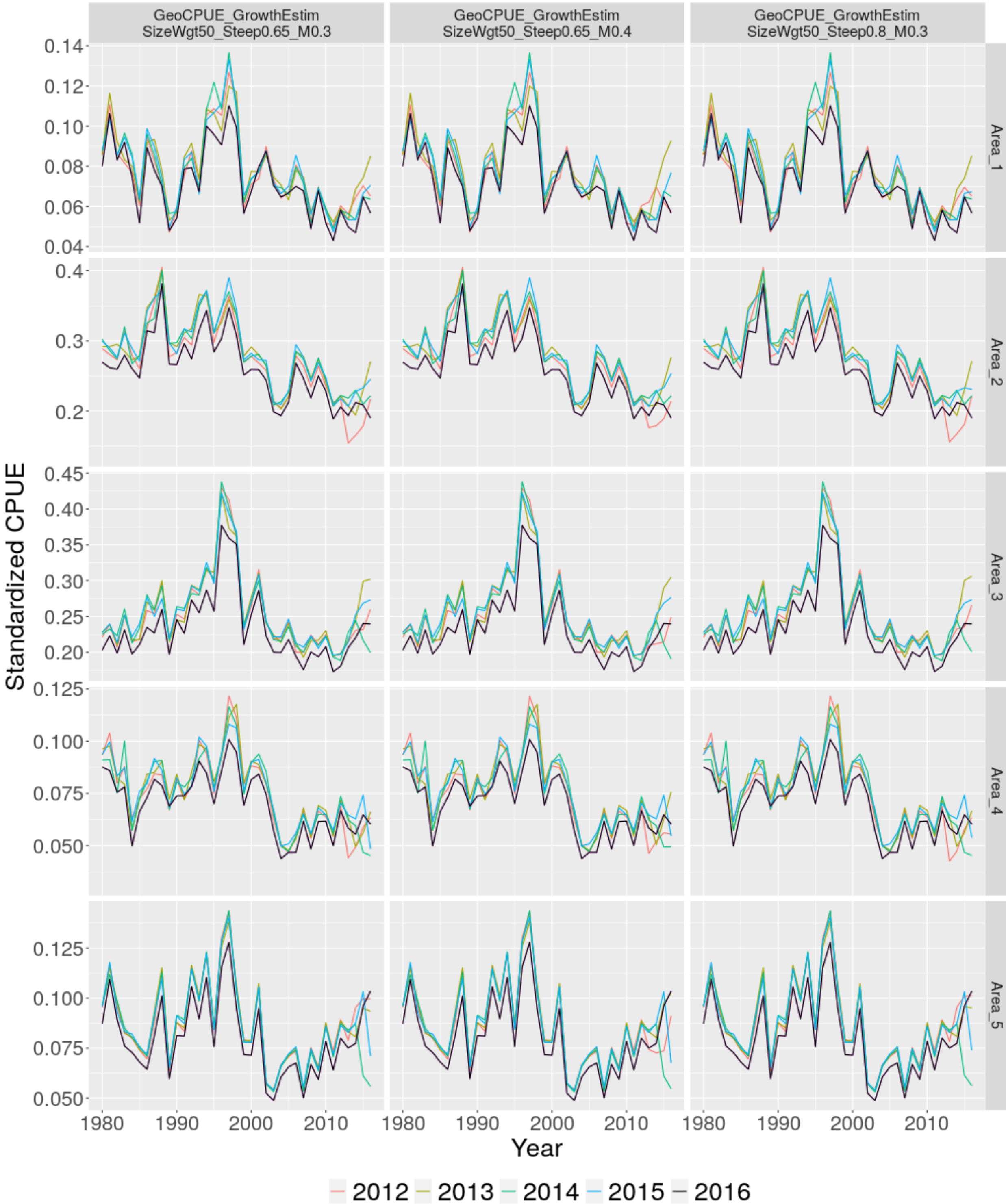


Figure 29: Retrospective forecasting CPUE time-series results for the five index fisheries for each individual model. Annual estimates of the index CPUE (standardized CPUE in each area) determined from the 2018 South Pacific albacore stock assessment (black line) and forecasted CPUE (terminal year 2012 to 2015, color lines).

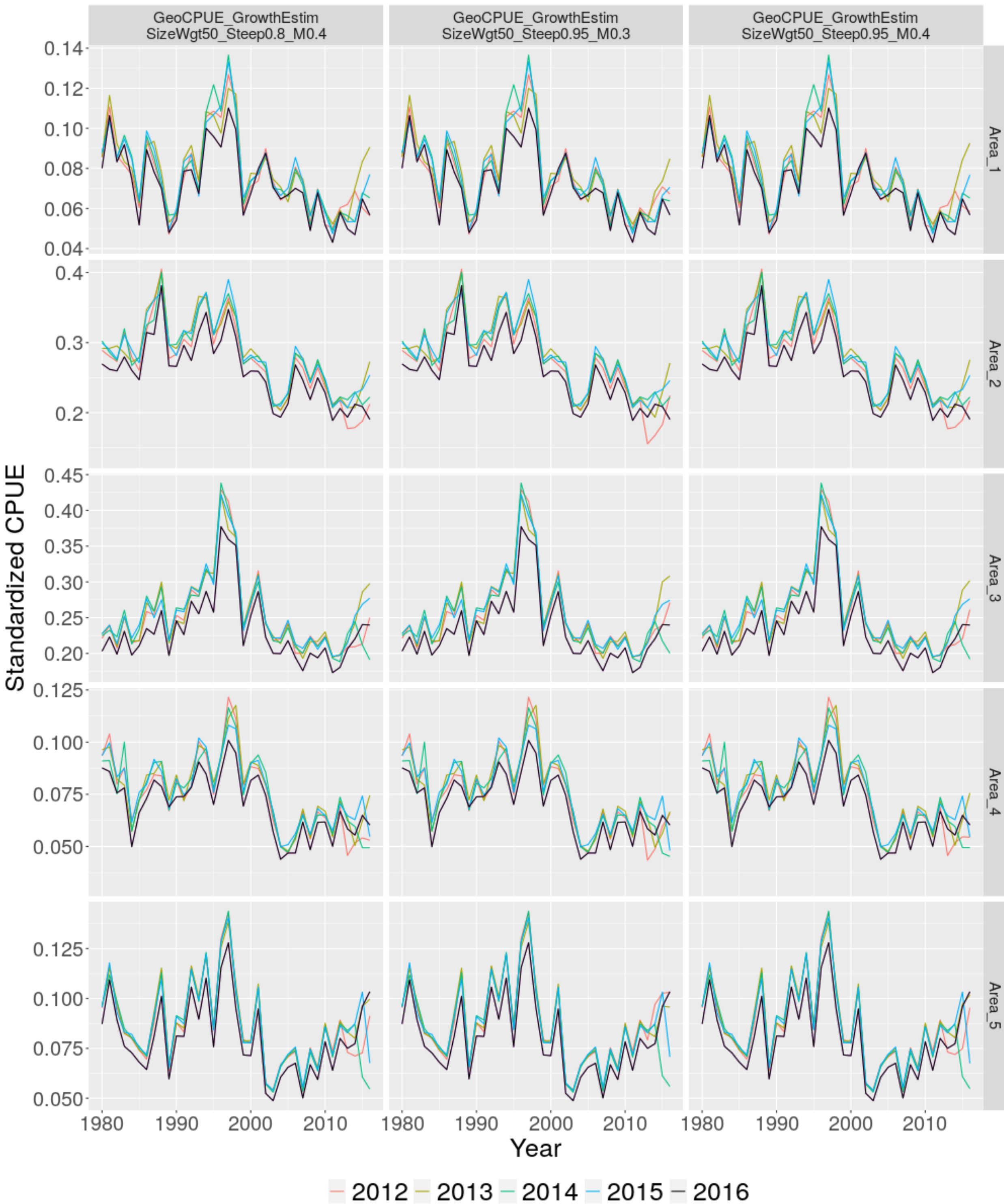


Figure 30: Retrospective forecasting CPUE time-series results for the five index fisheries for each individual model. Annual estimates of the index CPUE (standardized CPUE in each area) determined from the 2018 South Pacific albacore stock assessment (black line) and forecasted CPUE (terminal year 2012 to 2015, color lines).

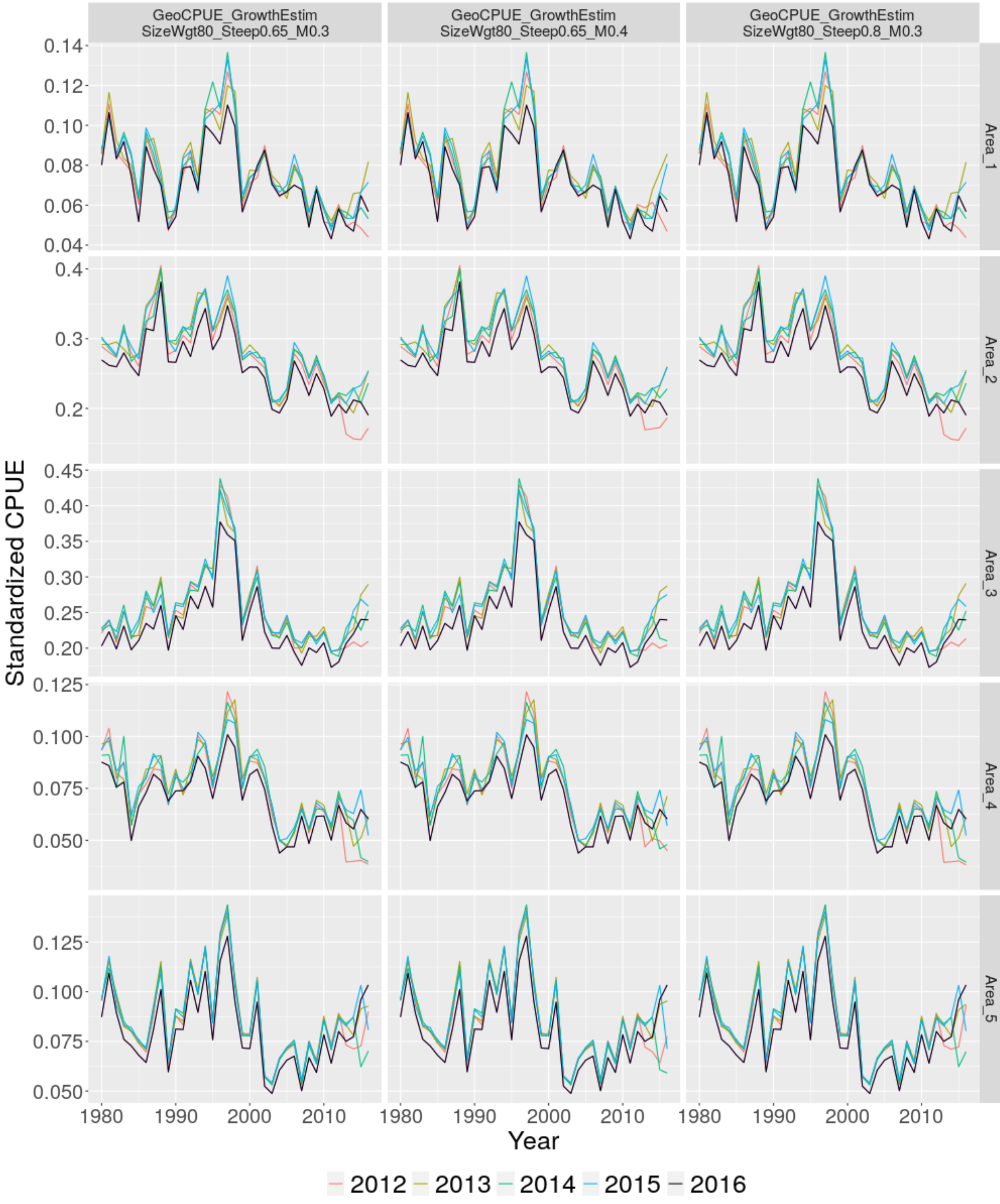


Figure 31: Retrospective forecasting CPUE time-series results for the five index fisheries for each individual model. Annual estimates of the index CPUE (standardized CPUE in each area) determined from the 2018 South Pacific albacore stock assessment (black line) and forecasted CPUE (terminal year 2012 to 2015, color lines).

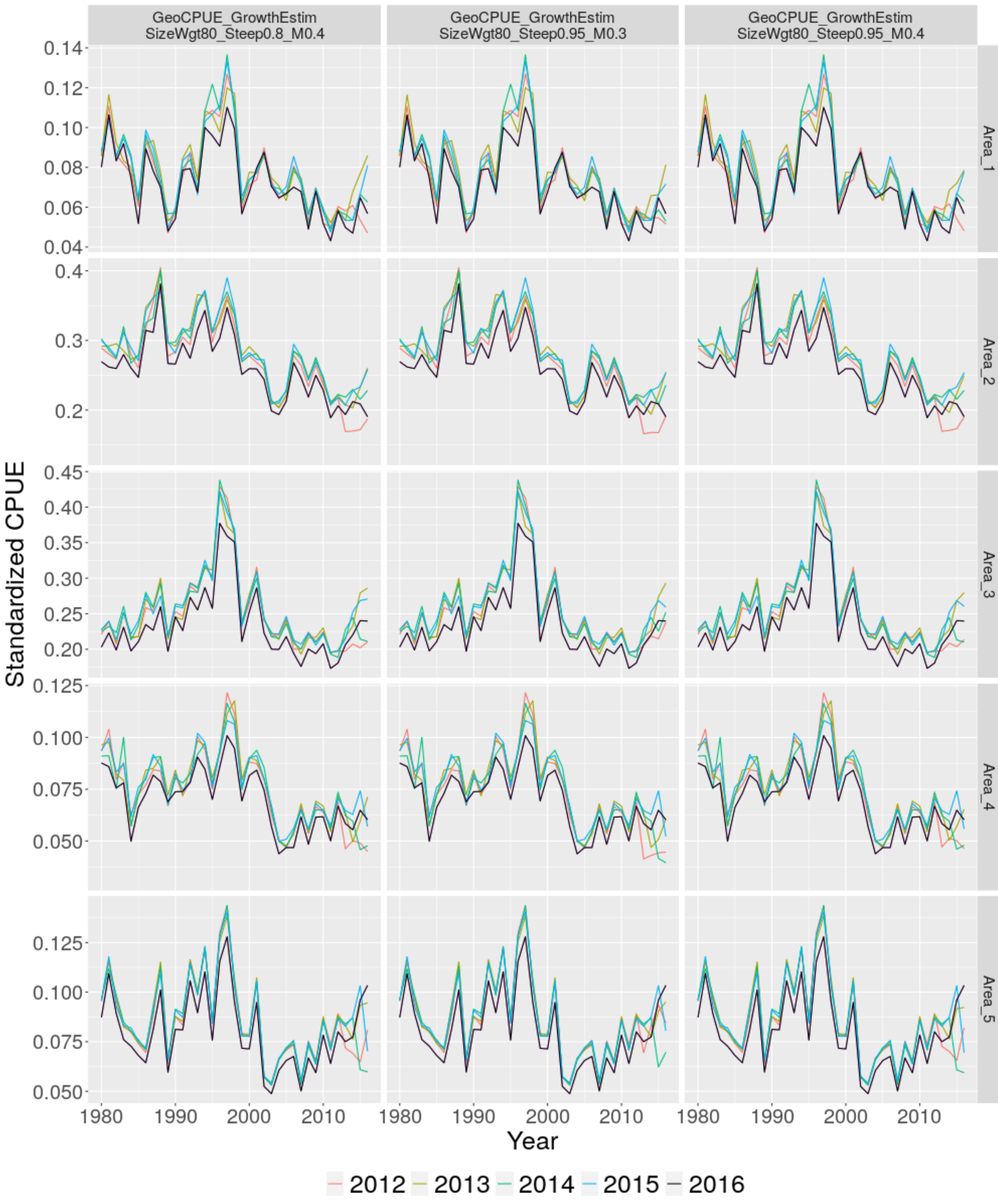


Figure 32: Retrospective forecasting CPUE time-series results for the five index fisheries for each individual model. Annual estimates of the index CPUE (standardized CPUE in each area) determined from the 2018 South Pacific albacore stock assessment (black line) and forecasted CPUE (terminal year 2012 to 2015, color lines).

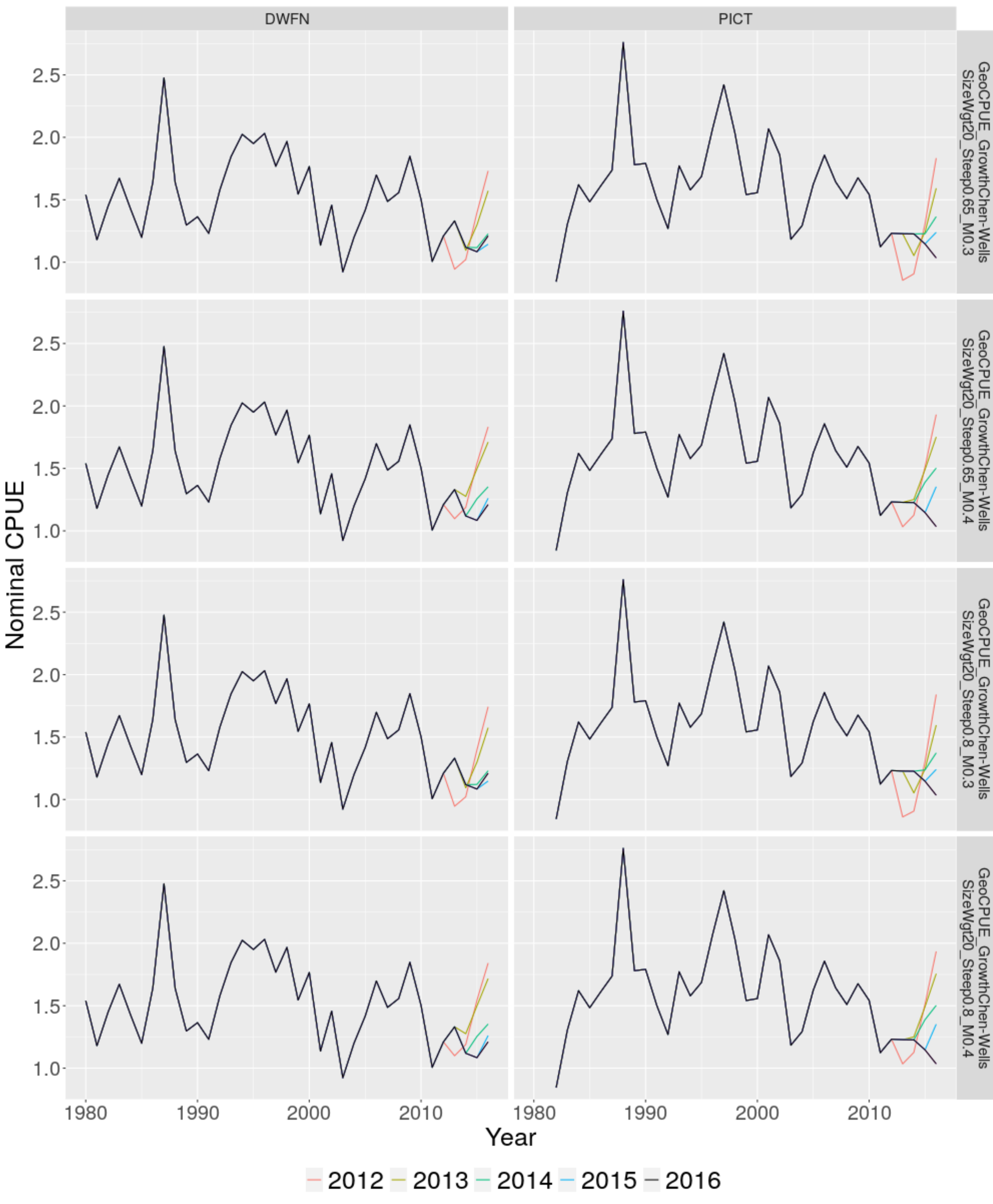


Figure 33: Retrospective forecasting CPUE time-series results for the DWFN and PICT fisheries in area 2 for each individual model. Annual estimates of the DWFN and PICT CPUE (nominal CPUE) determined from the 2018 South Pacific albacore stock assessment (black line) and forecasted CPUE (terminal year 2012 to 2015, color lines).

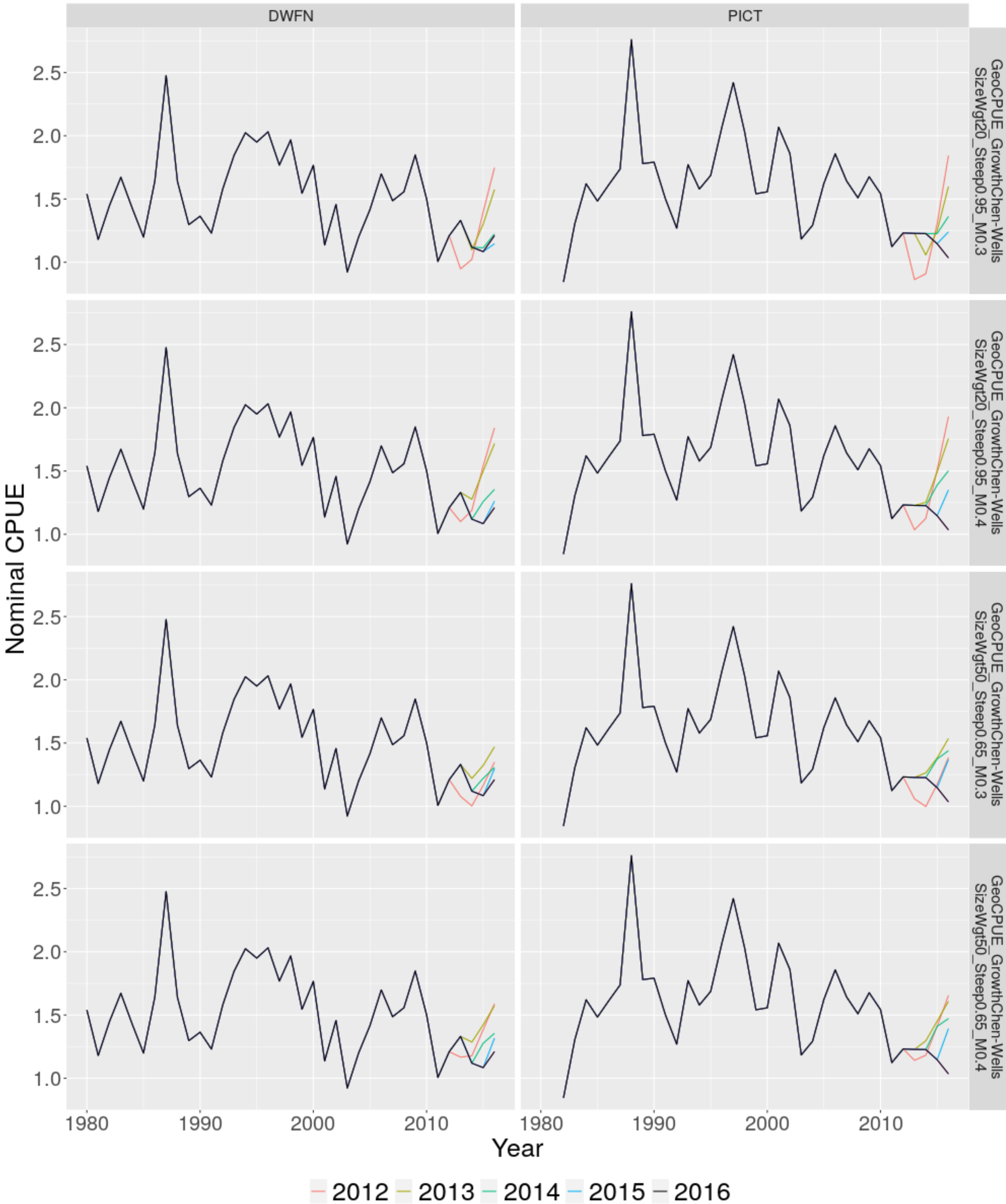


Figure 34: Retrospective forecasting CPUE time-series results for the DWFN and PICT fisheries in area 2 for each individual model. Annual estimates of the DWFN and PICT CPUE (nominal CPUE) determined from the 2018 South Pacific albacore stock assessment (black line) and forecasted CPUE (terminal year 2012 to 2015, color lines).

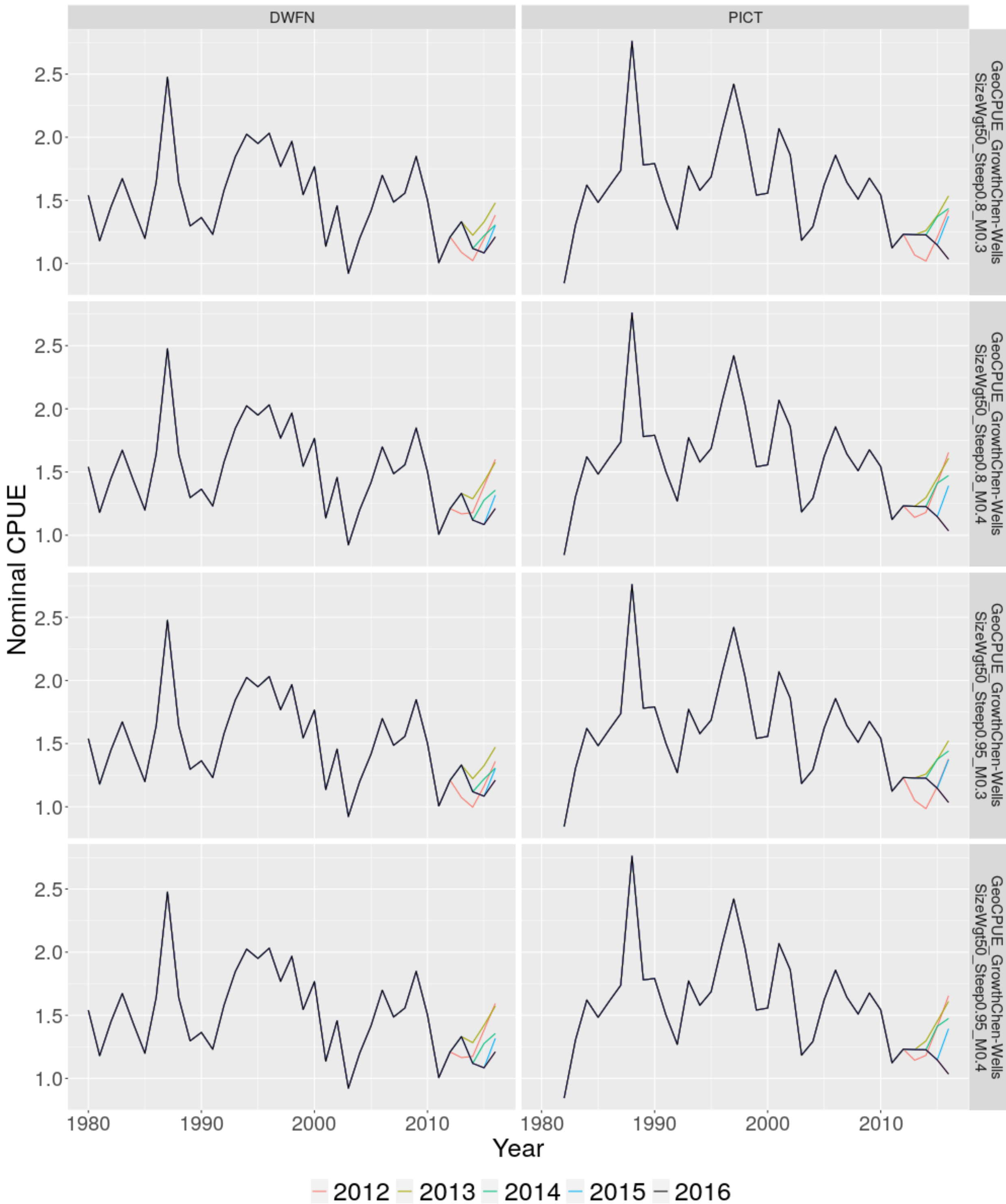


Figure 35: Retrospective forecasting CPUE time-series results for the DWFN and PICT fisheries in area 2 for each individual model. Annual estimates of the DWFN and PICT CPUE (nominal CPUE) determined from the 2018 South Pacific albacore stock assessment (black line) and forecasted CPUE (terminal year 2012 to 2015, color lines).

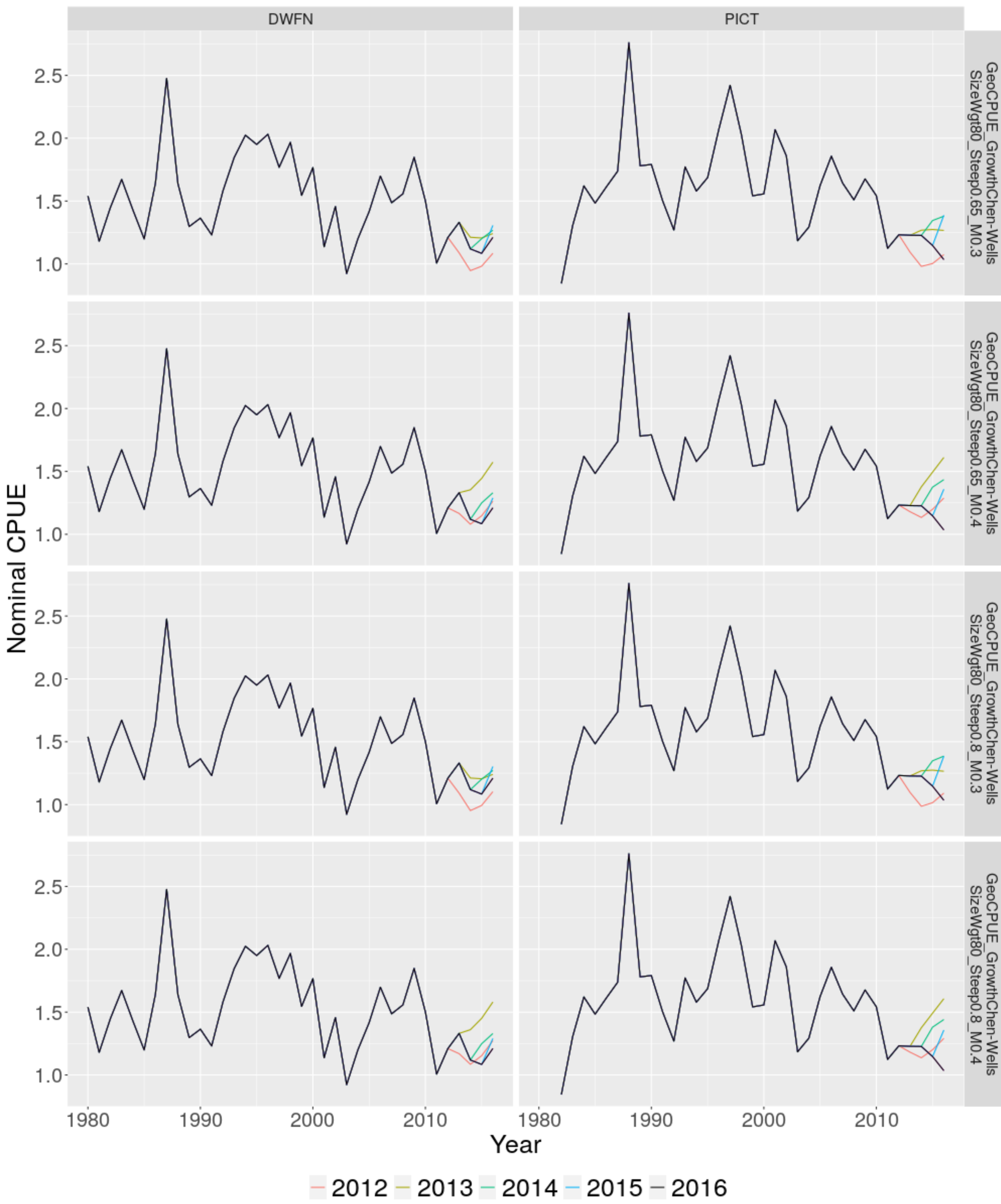


Figure 36: Retrospective forecasting CPUE time-series results for the DWFN and PICT fisheries in area 2 for each individual model. Annual estimates of the DWFN and PICT CPUE (nominal CPUE) determined from the 2018 South Pacific albacore stock assessment (black line) and forecasted CPUE (terminal year 2012 to 2015, color lines).

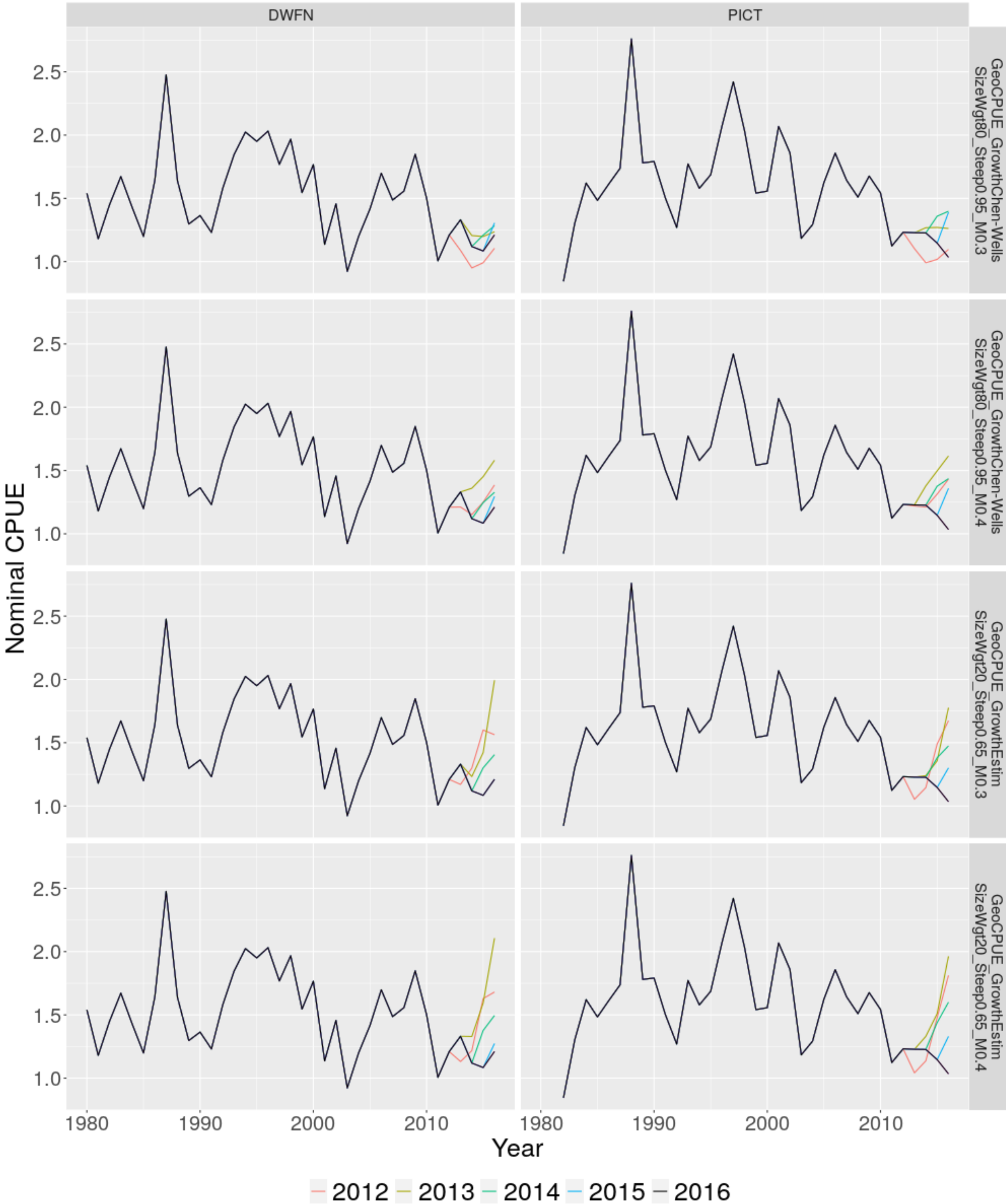


Figure 37: Retrospective forecasting CPUE time-series results for the DWFN and PICT fisheries in area 2 for each individual model. Annual estimates of the DWFN and PICT CPUE (nominal CPUE) determined from the 2018 South Pacific albacore stock assessment (black line) and forecasted CPUE (terminal year 2012 to 2015, color lines).

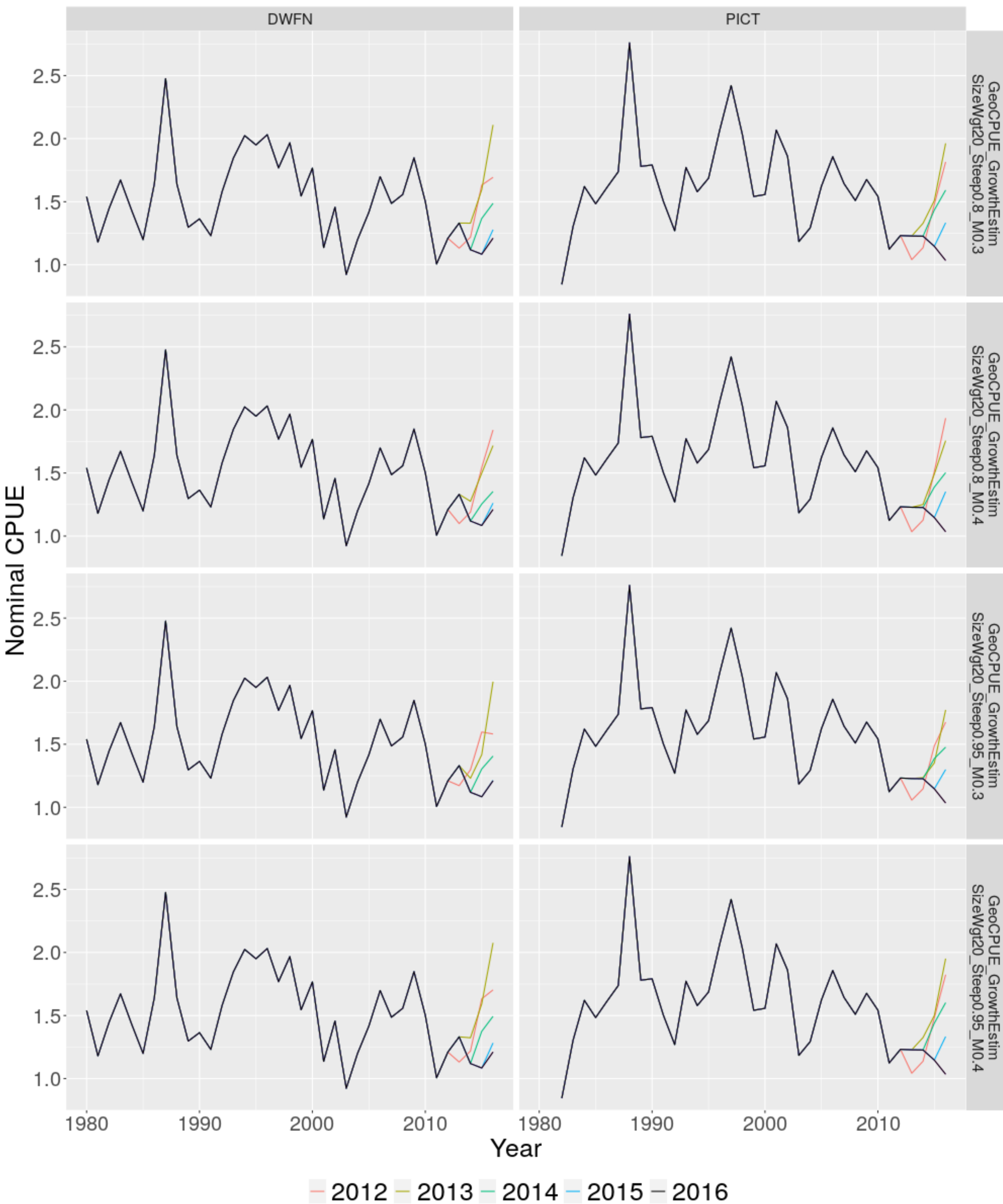


Figure 38: Retrospective forecasting CPUE time-series results for the DWFN and PICT fisheries in area 2 for each individual model. Annual estimates of the DWFN and PICT CPUE (nominal CPUE) determined from the 2018 South Pacific albacore stock assessment (black line) and forecasted CPUE (terminal year 2012 to 2015, color lines).

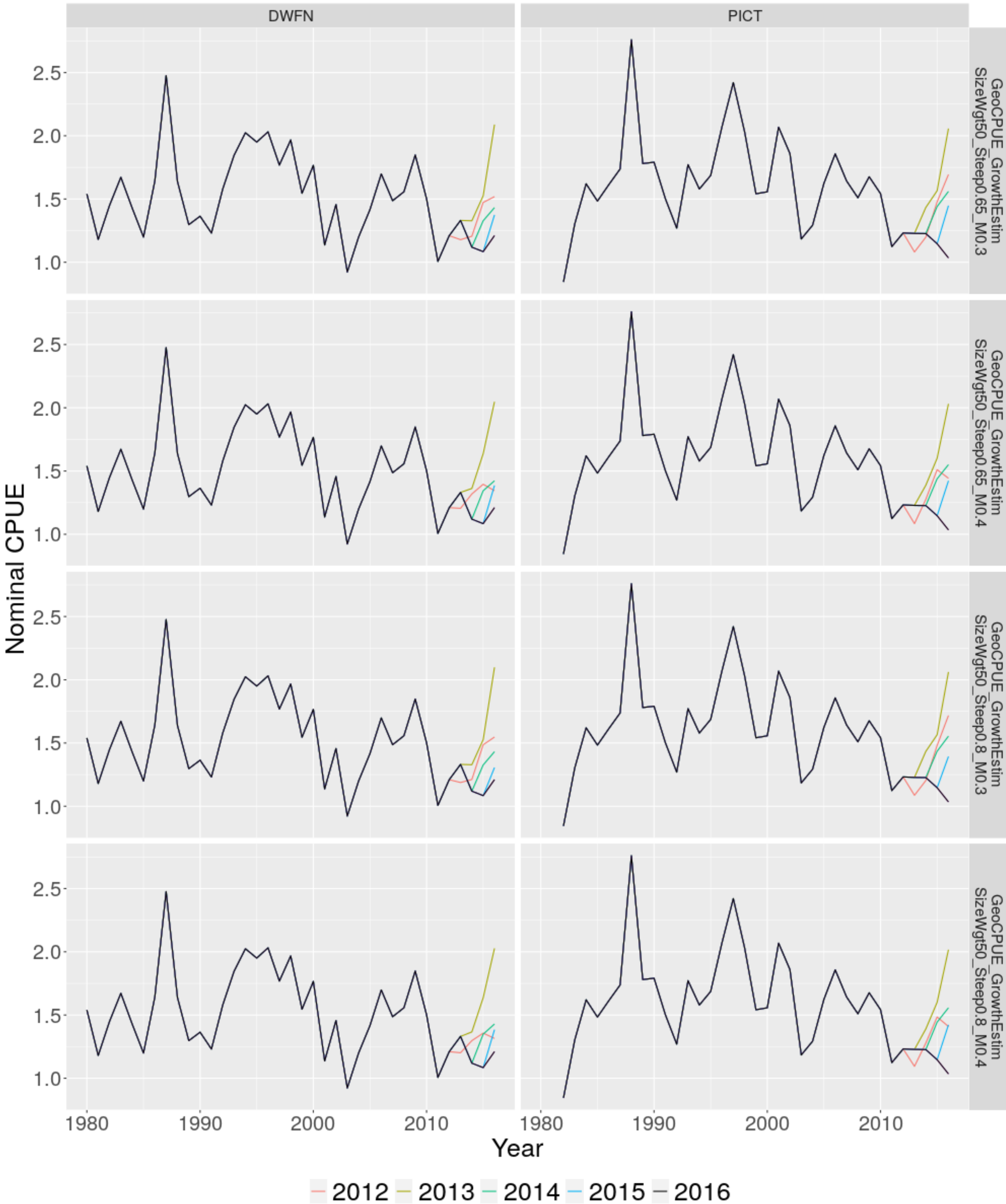


Figure 39: Retrospective forecasting CPUE time-series results for the DWFN and PICT fisheries in area 2 for each individual model. Annual estimates of the DWFN and PICT CPUE (nominal CPUE) determined from the 2018 South Pacific albacore stock assessment (black line) and forecasted CPUE (terminal year 2012 to 2015, color lines).

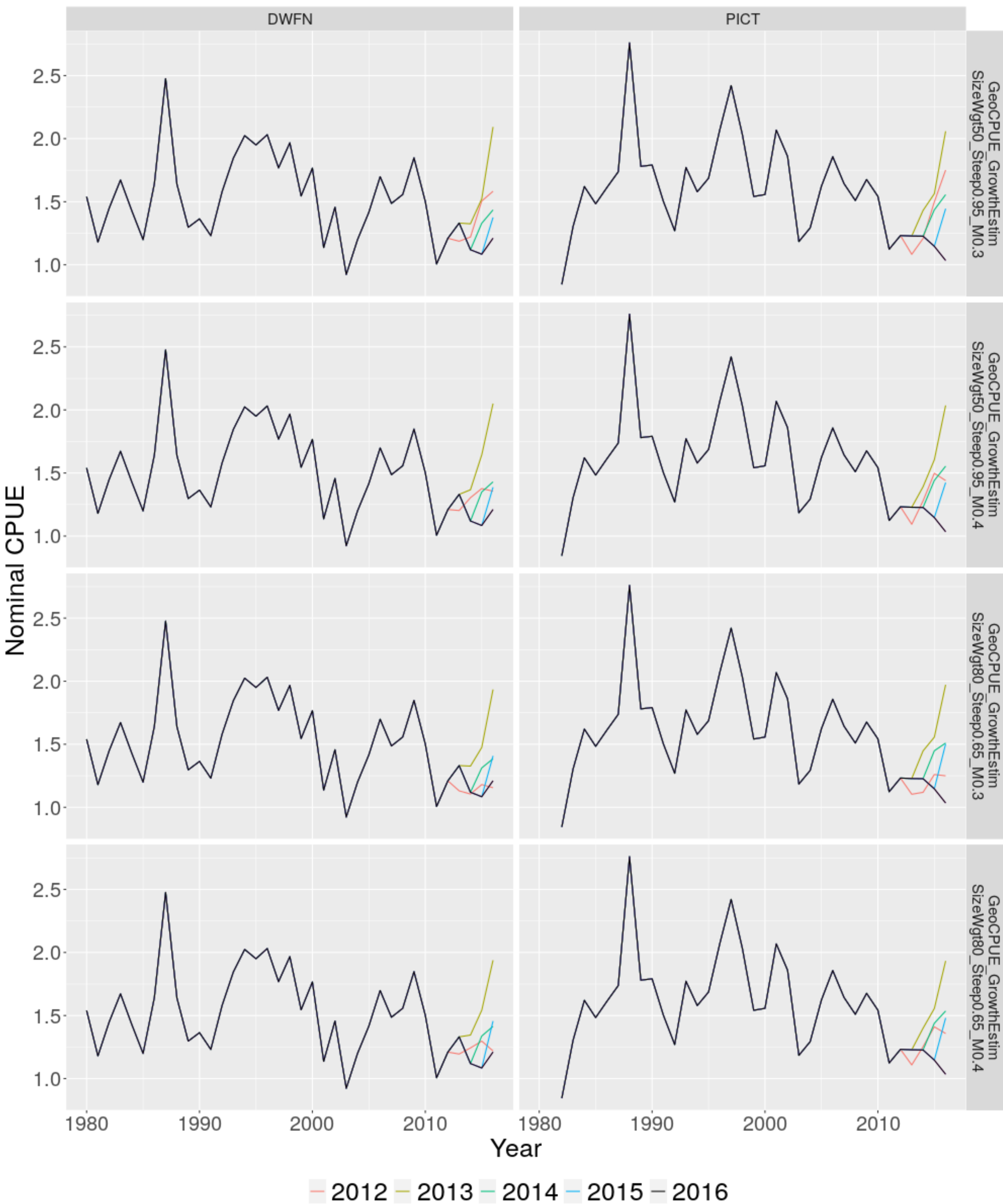


Figure 40: Retrospective forecasting CPUE time-series results for the DWFN and PICT fisheries in area 2 for each individual model. Annual estimates of the DWFN and PICT CPUE (nominal CPUE) determined from the 2018 South Pacific albacore stock assessment (black line) and forecasted CPUE (terminal year 2012 to 2015, color lines).

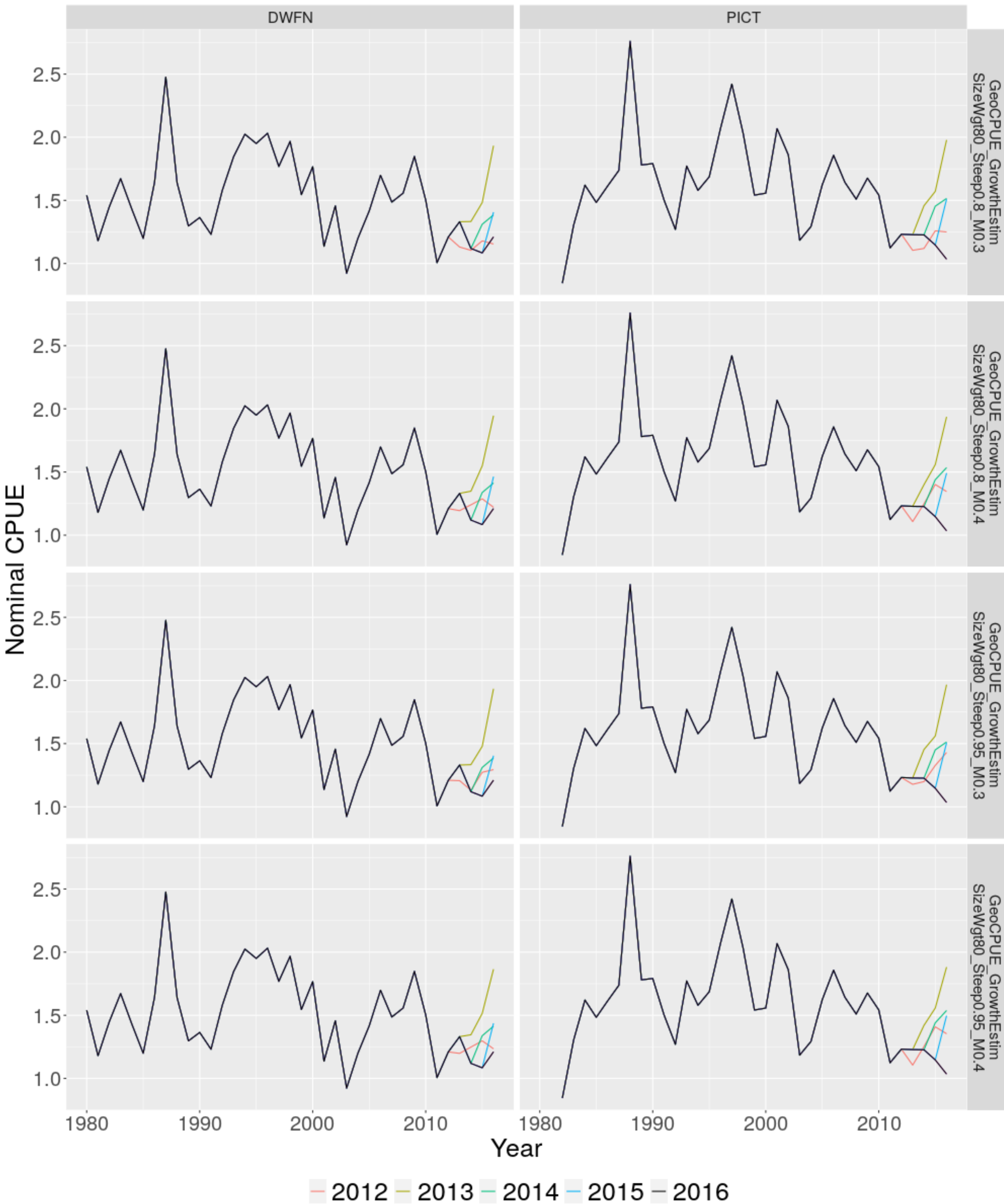


Figure 41: Retrospective forecasting CPUE time-series results for the DWFN and PICT fisheries in area 2 for each individual model. Annual estimates of the DWFN and PICT CPUE (nominal CPUE) determined from the 2018 South Pacific albacore stock assessment (black line) and forecasted CPUE (terminal year 2012 to 2015, color lines).