Reconditioning of the CCSBT Operating Model in 2017

R. M. Hillary, A. L. Preece, and C. R. Davies, CSIRO, Hobart 7004, Australia

N. Takahashi, National Research Institutes of Far Seas Fisheries, Yokohama, Kanagawa, Japan

O. Sakai and T. Itoh, National Research Institutes of Far Seas Fisheries, Shimizu, Shizuoka, Japan

Abstract

The CCSBT Operating Model is scheduled to be reconditioned in 2017, with both new and updated data sources, as part of the scheduled assessment of stock status and for the next round of Management Procedure testing scheduled to begin in 2018. Updated data sources include the Japanese long-line CPUE index (up to and including 2016), the aerial survey index (up to and including 2017), catch by fishery and the age and length composition data for the various fleets. New data sources include close-kin parent-offspring pairs (covering the cohorts 2002 to 2012, not just 2002-2007 as before) and half-sibling pairs (covering the cohorts 2003 to 2011). The Operating Model was run for the reference set and agreed set of sensitivity tests from the 8th OMMP meeting. For the reference set of models, the current level of total reproductive output (relative to the unfished state) is 0.13 (0.11-0.18 80% PI), with recent annual recruitment estimated to be well above the average predicted by the stock-recruit relationship. Projections for the reference set using the Bali Procedure indicated that the CCSBT interim management target of recovery to 20% of the unfished stock by 2035 is achieved with a probability of 91% (using the total reproductive output measure, or 88% if using total biomass aged 10 or older). The data are generally explained well by the reference set, with no obvious consistent trends in the fits to the abundance indices, catch composition, tagging, or the close-kin mark-recapture data. The results across the sensitivity tests are all very consistent, with only the test relating to constant selectivity from age 20 for the Indonesian fishery being clearly more optimistic than all the other tests. In terms of issues relating to data generation for future Management Procedure testing, the likelihood functions in their current format seem more than adequate for this purpose, with the exception of the gene tagging data, which is currently not available, so will not be evaluated until next year.

Introduction

Given the agreed Management Procedure (MP) implementation schedule, the CCSBT Operating Model (OM) is due to be reconditioned in 2017. The previous reconditioning was undertaken in 2014 (Anon., 2014). Given the cessation of the aerial survey after 2017, a new MP is to be developed within the CCSBT and testing will begin in 2018. For this conditioning, there are both new and updated data sources. Structural changes to the Operating Model, including modifications to the adult population dynamics and additional likelihood functions required for new data sources, were undertaken, reviewed and accepted at the 8th OMMP meeting in June 2017 (Anon., 2017a), with the final addition of the half-sibling pair (HSP) data series agreed at a special HSP webinar in late July (Anon., 2017b).

New and updated data sources

New data sources include close-kin mark-recapture (CKMR) parent-offspring pair (POP) (covering the cohorts 2002 to 2012, not just 2002-2007 as before) and HSP data (covering the cohorts 2003 to 2011). The generation of and quality control analyses on these data are detailed in paper CCSBT-ESC/1708/12 (Bravington *et al.*, 2017).

Updated data sources include:

- Japanese long-line core vessels CPUE index up to and including 2016
- Aerial survey index up to and including 2017
- Age composition for the surface and Indonesian fisheries up to and including 2016
- Length composition for the four main long-line fisheries up to and including 2016
- Catch by fishery up to 2016

Structural changes to OM conditioning and projection code

The major change to the population dynamics in the OM conditioning (and projection) code was related to how relative reproductive output-at-age (the *per capita* contribution of each age class to the reproductive population) is defined. Prior to the initial inclusion of the CKMR POP data, this was defined to be the biomass of all fish aged 10+; for the inclusion of the original CKMR POP data it was a time-invariant ogive (between 0 and 1) informed by the CKMR data and available reproductive information on relative fecundity-at-length. Given we now have more CKMR data (both POPs and now HSPs) we expanded this model to include both a control parameter in the OM grid (controlling the degree to which increasing length relates to reproductive success), and the changing distribution of length-at-age over time within the OM. The mathematical details of this change are provided in CCSBT-OMMP/1706/4 (Hillary *et al.*, 2017). With respect to the projections, the final year value of the relative reproductive output-at-age (for a given grid cell) is used to define future values in the projections.

No changes were required to the likelihood function for the POP data. A new likelihood function was required to include the HSP data though. The probability of two animals being a HSP depends on a complicated array of information, including: relative adult abundance-at-age, the total reproductive output (TRO) over time and the adult total mortality rate-at-age. There is additional complexity required to deal with both potential biases in the absolute abundance information in the HSPs (relative to the POPs) and to account for the false negative rate we have estimated to ensure that no false-positive non-HSPs were included in the final HSP data set (Hillary *et al.*, 2017a). Papers specifying the proposed modifications were presented to the OMMP meeting in June 2017 in CCSBT-OMMP/1706/4, and were accepted by the OMMP group for inclusion in the OM. The grid configuration for the reference set of OMs, as agreed at the OMMP meeting is given in Table 1.

Table 1: Grid configuration for the agreed reference set of OMs for the 2017 stock assessment

Parameter	Value	CumulN	Prior	Sampling
H	0.6, 0.7, 0.8	3	Uniform	Prior
M_0	0.35, 0.4, 0.45, 0.5	12	Uniform	ObjFn
M_{10}	0.05, 0.085, 0.12	36	Uniform	ObjFn
Omega (ω)	1	36	Uniform	Prior
CPUE series	w0.5, w0.8	72	Uniform	Prior
CPUE age range	4-18, 8-12	144	0.67, 0.33	Prior
Psi (ψ)	1.5, 1.75, 2	432	0.25, 0.5, 0.25	Prior

An additional change that was required to the OM conditioning and projection code was ensuring that the right scale for the catchability parameter for the LL1 CPUE (q) is maintained between the conditioning and the projection code. The scale of the *estimated* catchability parameter ($\ln(q)$) was changed in 2014 from 'real-space' to one where predicted CPUE is rescaled by the historical mean, so that q is close to 1 instead of on a scale around 1e-6. An issue in the projection code is that scale reverts to real-space (i.e. q * exploitable LL1 abundance) and so the estimated q values will result in CPUE levels of 1e+6 in the future and lose their connection to the historical data. A simple additional nuisance variable is added to the OM in the conditioning phase, and then transferred via the .prj files to the grid files and the projection code to ensure that historical observed CPUE and future simulated CPUE are on the same scale.

Reference set and sensitivity tests

All 432 grid combinations were run, with confirmed estimation on all grid combinations, and then 2,000 samples were taken to generate the reference set of OMs. The issue with some combinations of steepness and the natural mortality parameter crashing, which was identified at OMMP8, was solved by both using an alternative starting estimate of M_4 (a directly estimated parameter) and more iterations in the initial phases of the estimation algorithm.

The list of sensitivity tests agreed at the OMMP8 meeting is reproduced in Table 2 below with edits to clarify updated specifications. There is a total of 20 sensitivity tests relating to issues, such as, over-catch and unaccounted mortality, CPUE interpretation and alternative indices, structural issues such as tag mixing, alternative data sources (like the Piston line survey), excluding data sources (like the POP and/or HSP CKMR data) and alternative weighting scenarios for specific grid parameters.

Table 2. Sensitivity tests for 2017 assessment and stock status advice (reproduced from Table 6, Anon. 2017).

Run name	Conditioning	Projections
UAM1	Added unaccounted catch mortality (UAM) in conditioning: 1000 t of small fish + 1000 t of large fish, ramping up from 0t of each size class in 1990 to 1000t in 2013, and 1000t in each year 2014-2016, in addition to 20% increase in the surface fishery.	Additional catch remains at the same proportion as in 2016.
SFOC40	40% overcatch by Australian surface fishery: ramps up from 1% in 1992 to 40% by 1999 and onwards to 2016. Adjust the age composition as was done for the 20% method.	Continued 40% overcatch in projections
SFO00	No historical additional catch in surface fishery	No future additional catch in surface fishery
LL1 Case 2 of MR	LL1 overcatch based on Case 2 of the 2006 Market Report	
IS20	Indonesian selectivity flat from age 20+	
High_aerialCV	In conditioning set process CV to 0.4	
Aerial2016	Remove the 2016 aerial survey data point	
Upq2008	CPUE q increased by 25% (permanent in 2008 due to individual quota system that went into effect in 2006)	
Omega75	Power function for biomass-CPUE relationship with power ω = 0.75 (retain)	
S00CPUE	Overcatch had no impact on CPUE	
S50CPUE	50% of LL1 overcatch associated with reported effort	
Updownq	Increase in catchability (0.5) in 2009 then returns to normal in 2012 (when the pertinent quota was restored to pre-2009 level)	
GamCPUE	Use the “GAM CPUE” series provided from Australia under the 2017 CCSBT data exchange. This is the monitoring CPUE series 3.	
Base CPUE w/o area 7	As a sensitivity to note a possible concentration effect on CPUE	
Incomplete tag mixing	Sensitivity to incomplete mixing of tagged fish released in the WA and GAB. Increases fishing mortality of tagged fish in the surface fishery by 50% relative to the whole population for fishing season 1 (surface fishery).	
Piston line with high AS cv	Includes the piston-line troll survey index (updated to 2017) included as alternative sensitive to recruitment index (2017 data exchange)	
NoPOP&HSP	Exclude both close-kin data sets (POPs and HSPs)	
NoHSP	Exclude HSP close-kin data	
Psi	Grid sampling using objection function weighting psi	
Noh0.8	Change steepness (*h*) preference weighting to 0.5, 0.5, 0.0 to examine impact of excluding *h*=0.8 on projections.	

One additional sensitivity test that arose from the discussion during the special web meeting in July (Anon.2017b) was running the OM with the HSP scaling parameter (q_{hsp}) fixed at 1, instead of being freely estimated.

Status and fitting diagnostics for the reference set of OMs

Figure 1 shows the estimated relative TRO for the spawning stock and the recruitment for the reference set. For relative TRO (TRO in 2017 relative to TRO initial unfished), the median estimate is 0.13 (0.11-0.17 80% PI). The lowest point is around 2009, with a clear increasing trend from 2012 onwards. Estimates of recruitment from 2009/2010 cohorts have been above the mean level predicted by the stock-recruit relationship – particularly the 2013 recruitment.

Figure 2 summarises the historical estimates of maximum sustainable yield (MSY), the ratio of F to F_{MSY}, and surplus production. MSY has varied as population selectivity (a product of fisheries selectivity over time and relative allocation among fleets) and mean length-at-age has altered over time. The current estimate is between 32,000t-34,000t. The ratio of F to F_{MSY} has been steadily decreasing from ~1.5 in the mid-2000s to a current median of 0.5 (0.38-0.7 80% PI). Surplus production has been highly variable over time, as the stock abundance declined, recruitment varied, and fisheries characteristics changed. The most recent values of surplus production are just above 40,000t, well in excess of current catches, hence the current lower values of the F to F_{MSY} ratio. The current TRO (at 2017) to MSY ratio is also around 0.5 (0.37-0.7 80% PI), with the MSY to unfished TRO ratio estimated to be around 0.27 (0.22-0.32 80% PI) – i.e. the TRO at which MSY is produced is somewhere between a depletion level of 0.22 and 0.32. Figure 3 shows the likelihood profiles for steepness, M_0 and M_{10}, respectively. Figure 4 shows the level plot for the reference grid.

To summarise the fits to the abundance indices, tagging data, CKMR data and the catch composition we initially focussed on the best fitting grid cell (2312321 for information). We did, however, undertake more detailed predictive analyses across the whole grid for the data that are/were simulated for potential use in candidate MPs (CPUE, aerial survey, and the CKMR data). Figure 5 summarises the fit to the Japanese long-line CPUE and aerial survey, respectively. The CPUE are fitted well, with all the observed points sitting within (or just about on the edge of) the predicted 95% CI and with no consistent trends in the fits. The aerial survey data are fitted fairly well, with a few of the points just outside the 95% CI (assume a process error of 0.22) and with no consistent trend in the fits – the only obvious discrepancy is the inability of the OM to fit to the very large 2016 survey data point.

Figure 6 summarises the fit to the tagging data at the release year and recapture age disaggregation level. These fits are aggregated over the individual tagger and release age, but give the clearest indication of the consistency of information on cohort abundance and mortality informed by multiple release and recapture events (over both time and age). The fits are good, with the data mostly being fitted closely (especially the largest recapture events in numbers) and with no clear or consistent trends in the fits themselves. The tagging over-dispersion factor was re-estimated given the updated and new data sets. This was done as follows: for each recapture event (at the full disaggregation level) we calculate the standardised residual; we then calculate the variance in the standardised residuals; this value yields the multiplier by which we would alter the current over-dispersion factor ($\varphi = 1.82$). The estimated value of the dispersion multiplier was 0.998, which is so close to 1 as to suggest we are fine to keep with the current value of 1.82. A more detailed analysis of the trends in the standardised residuals at the full disaggregation level found only one apparently clear trend: that for tagger 6 the residuals were consistently less variable (and, hence, with lower implied over-dispersion factor) than for all the other taggers. There were no apparent trends across release cohort, age or recapture age.

The CKMR POP data are, in the form they are used in the OM, the number of juvenile-adult comparisons (and POP matches) at the level of juvenile cohort/adult capture year/adult capture age (the POP probability is the same for these covariates). There are 1,728 such unique groupings with expected non-zero comparisons and only 77 POPs, so these data are *very* sparse. To summarise the fits to these data, we aggregate them (both the observations and the predictions) to more useful levels. For the POP data, we aggregate to the juvenile cohort (across adult capture year and age) and the adult capture age (across adult capture year and juvenile cohort) levels. The cohort level gives us an indication of whether we are getting the overall adult abundance level right over time, and the adult capture age level covers whether we are getting the age distribution of the adults in the POPs (and, by implication, the relative reproductive output-at-age) about right given the data. Figure 7 summarises the fits to the CKMR POP data at these two aggregation levels. For both, the observed data sit within the approximate 95% CI and with no obvious or consistent trend in the fits.

The CKMR HSP data, in the form they are currently used in the OM, are aggregated at the level of the number of juvenile-juvenile comparisons (and HSP matches) between animals of a given cohort/birth year. We show the fits to these data at the OM level and where we estimate the total number of HSPs found between a reference cohort (the earliest one) and the subsequent cohorts it is compared against. Figure 8 summarises the fits at these two aggregation levels and, as with the CKMR POP data, the data lie (almost exclusively) within the approximate 95%ile and with no obvious or consistent trends in the fits.

Figure 9 summarises the fits to the age composition data for the surface and Indonesian fisheries. As in previous years, the fits to these data are generally very good.

Figure 10 summarises the fits to the length composition data for the other 4 long-line fisheries. As with previous years, the fits to the LL1 fleet are good, with some misfit in the LL2 data, and also with the earliest data from the LL3 and LL4 fleets. As in previous reconditioning exercised, there is very little variation in the fits to both the age and length composition across the grid cells.

For previous reconditioning of the OMs, we have undertaken more detailed predictive analyses of the data that are being considered for inclusion in candidate MPs. In the previous MP, this has meant the long-line CPUE and the aerial survey (Anon., 2011), but we also extended this analysis to the CKMR POP data to explore whether these data display additional process error. Given the agreed set of data series to be used in the next suite of candidate MPs (CPUE, gene tagging, CKMR POP and HSP data) we advanced the original CKMR predictive analyses to look at the data at the various levels of aggregation of interest (juvenile cohort and adult capture age) and also for the HSP data as well. The principle is fairly straightforward:

1) Simulate the data (at the required aggregation level for the CKMR data) from the likelihood model used in the OM
2) Calculate two residuals: the first is the simulated data minus the expected value; the second is the actual data minus the expected value
3) Calculate an appropriate "discrepancy" measure for each of these residuals, Δ^{sim} and Δ^{obs}, and we use the median absolute deviation (which is non-parametric)
4) Do this for each of the 2,000 grid samples and calculate $\mathbb{P}(\Delta^{\text{obs}} > \Delta^{\text{sim}})$

If the simulated data are very similar (not just in terms of expected prediction but also in terms of the variance properties) to the observed data this p-value will be close to 0.5 (i.e. just as likely to be more or less variable than the data). If the data are consistently more variable than the predictions (i.e. over-dispersed/possess process error) then this value will be greater than 0.5; *vice versa* if the data are less variable the predictions. Based on Gelman et al. (1995) it is only when values are outside the range 0.05-0.95 in terms of p-values that there is a strong indication of something not right with the likelihood model. The relative shape of the "cloud" of discrepancy values can also be instructive as to whether the likelihood might be misspecified – for example one assumes a normal distribution but the data appear to show a more fat-tailed distribution, even if their p-values are close to 0.5.

For the long-line CPUE and aerial survey indices, the predictive distributions of the data (Figure 11) across the grid look very similar those predicted for the best fitting grid cell in Figure 5. This suggests that there is good consistency of fit across the grid samples, relative to the best fitting grid cell. As for the p-values, the long-line CPUE is 0.07 and the aerial survey is 0.72 suggesting that we are moderately under-weighting the CPUE and moderately over-weighting the survey. This is not surprising, given the empirical CV in the CPUE residuals is around 0.16 (not the fixed minimum value of 0.2), and the process error in the survey CV in the survey is around 0.36 not 0.22 as currently assumed. This is similar to previous such analyses and has been accepted practice to account for potential additional uncertainty in the CPUE, given the assumptions required to account for the over-catch term, and to give a little more weight to the aerial survey, given it is both fishery independent and the earliest data set to inform on recruitment trends currently in the OM.

Figure 12 summarises the predictive discrepancy statistics and p-values for the CKMR POP and HSP data at the main levels of aggregation: POPs at the juvenile-cohort and adult capture age level; HSPs at the initial cohort level. For the POPs at the cohort level the p-value is 0.91 suggesting the data are more variable than the predictions, though Figure 8 shows the data are all within the 95% CI for the best fitting grid, so this effect is not likely to be strong. For the POPs at the adult capture age level, the p-value is 0.59 suggesting the OM is explaining the data well at this level. For the HSP data at the initial cohort level the p-value is 0.28, suggesting that the data are in fact slightly *less* variable than the predictions. In all three cases, the spread in the discrepancy for the predictions is wider than for the data. A possible reason for this is the algorithm used to combine the individual binomial probabilities together at the relevant aggregation level (Butler and Stephens, 2016) is exact, but assumes that the binomial probabilities themselves are independent. In reality, given the correlation between TRO and adult mortality over time and age this will not be the case. This correlation would likely decrease the variability in the discrepancy, but not accounting for it could result in the kind of wider spread we see in the plots.

For both the abundance indices (CPUE and the aerial survey) and the CKMR POP and HSP data the predictive analyses all look fine – there is nothing obviously troubling about any of the discrepancy distributions. This supports using the currently defined likelihoods for these data for simulation purposes in the next round of MP design and testing. For the gene tagging data, we will perform similar analyses when these data are available in 2018.

Projections for the reference set of OMs

Based on the OM conditioning result for the reference set, future projections were conducted using the Bali procedure (MP3 for the projection code name) to provide an indication as to whether the overall estimate of stock productivity has changed since the previous assessment.

Figure 13 shows historical and projected trajectories of the reference set for recruitment, biomass of age 10+ fish, and TRO. For comparison, those of the previous assessment were also plotted. All trajectories of recruitment and spawners for the 2017 projections have increasing trends with respect to the medians. Compared to the results of the 2014 assessment, the overall increasing trends for the 2017 projection is shifted upward for both recruitment and spawners. For the 2017 projection, the CCSBT interim management target of recovery to 20% of the unfished stock by 2035 is achieved with a probability of 91% (using TRO, or 88% using B10+) (Table 3). This target is achieved by 2025 with a probability of 81%. Projected future TAC trajectories for the reference set are plotted in Figure 14. With respect to the median values, the future TAC continues to increase toward 2035. The average TAC over 2018-2035 is predicted as 22,570t (18,767-25,147t 80% PI) (Table 3).

Summary of sensitivity tests

The results for the reference set of OMs and agreed sensitivity tests are summarise in Table 3 using the following statistics: (I) Relative Total Reproductive Output (Rel.TRO) in 2017, (2) Relative biomass of age 10+ fish (Rel. B10+) in 2017, (3) ratio of current TRO (2017) to TRO at MSY (TRO-to-TROmsy), (4) ratio of TRO at MSY to the unfished level (TROmsy/TRO0), (5) ratio of current F to F at MSY (F-to-FMSY), (6) Maximum Sustainable Yield) (MSY), (7) Relative TRO in 2035, (Rel. TRO (2035)), (8) the probability that biomass of age 10+ fish is greater than 20% of the unfished state in 2035 (the original tuning objective; P(B10+ >0.2B0)@2035), and (9) mean TAC under the Bali Procedure between 2018 and 2035.

Table 3: Summary table for Reference Set and the sensitivity tests from OMMP8 and webinar. Medians are listed first, with the 80%PI included in the bracket as appropriate. Definitions of sensitivity tests are in Table 2 and summary statistics in text above. [1]The piston line could only be run to completion (i.e. convergence of all grid combinations) with the higher aerial survey CV.

Run	Rel. TRO (2017)	Rel. B10+ (2017)	TRO-to-TROmsy (2017)	TROmsy /TRO0	F-to-FMSY (2017)	Median MSY (t) (2017)	Rel. TRO (2035)	P(B10+ > 0.2B0) @ 2035	Mean TAC (2018-2035)
Reference	0.13 (0.11-0.17)	0.11 (0.09-0.13)	0.49 (0.38-0.69)	0.27 (0.22-0.32)	0.5 (0.38-0.66)	33,036	0.3 (0.21-0.46)	0.88	22,570
UAM1	0.13 (0.1-0.17)	0.11 (0.09-0.13)	0.49 (0/37-0.67)	0.27 (0.22-0.32)	0.57 (0.43-0.74)	33,471	0.28 (0.18-0.43)	0.80	22,025
SFOC40	0.14 (0.11-0.18)	0.11 (0.09-0.14)	0.52 (0.38-0.71)	0.27 (0.22-0.32)	0.53 (0.4-0.7)	35,120	0.31 (0.21-0.48)	0.89	22,707
SFOC00	0.12 (0.1-0.16)	0.1 (0.09-0.12)	0.46 (0.35-64)	0.27 (0.22-0.32)	0.48 (0.35-0.63)	30,865	0.29 (0.20-0.45)	0.87	22,319
LL1 Case 2	0.13 (0.11-0.16)	0.11 (0.09-0.13)	0.48 (0.37-0.66)	0.27 (0.22-0.32)	0.5 (0.38-0.63)	33,526	0.31 (0.21-0.47)	0.90	22,627
IS20	0.18 (0.15-0.22)	0.14 (0.12-0.17)	0.64 (0.46-0.97)	0.28 (0.23-0.33)	0.41 (0.3-0.57)	34,304	0.38 (0.26-0.59)	0.96	23,224
High Aerial CV	0.12 (0.1-0.16)	0.11 (0.09-0.14)	0.47 (0.35-0.67)	0.27 (0.22-0.32)	0.58 (0.43-0.78)	32,799	0.26 (0.16-0.41)	0.72	21,745
No AS 2016	0.13 (0.1-0.16)	0.11 (0.09-0.14)	0.47 (0.36-0.66)	0.27 (0.22-0.32)	0.59 (0.44-0.78)	33,140	0.26 (0.17-0.40)	0.74	21,455
Upq2008	0.11 (0.1-0.15)	0.09 (0.08-0.12)	0.42 (0.35-0.65)	0.27 (0.22-0.32)	0.56 (0.42-0.75)	32,552	0.26 (0.17-0.42)	0.73	22,635
Omega 75	0.12 (0.1-0.16)	0.1 (0.08-0.13)	0.46 (0.35-0.65)	0.27 (0.22-0.32)	0.49 (0.36-0.65)	33,799	0.31 (0.21-0.48)	0.88	21,847
S00CPUE	0.15 (0.12-0.19)	0.12 (0.1-0.15)	0.55 (0.41-0.76)	0.27 (0.22-0.32)	0.46 (0.35-0.6)	34,126	0.33 (0.23-0.52)	0.94	22,665
S50CPUE	0.12 (0.1-0.15)	0.1 (0.08-0.12)	0.45 (0.41-0.76)	0.27 (0.22-0.32)	0.54 (0.4-0.71)	32,458	0.28 (0.19-0.44)	0.82	22,444
Updownq	0.13 (0.11-0.17)	0.11 (0.09-0.13)	0.49 (0.38-0.69)	0.27 (0.22-0.32)	0.5 (0.38-0.66)	33,036	0.3 (0.21-0.47)	0.88	22,569
GAM CPUE	0.14 (0.12-0.18)	0.12 (0.1-0.14)	0.53 (0.43-0.76)	0.27 (0.22-0.32)	0.51 (0.36-0.62)	32,774	0.31 (0.22-0.47)	0.91	23,168
CPUE w/o A7	0.12 (0.1-0.15)	0.1 (0.08-0.12)	0.45 (0.35-0.62)	0.27 (0.22-0.32)	0.54 (0.4-0.71)	32,734	0.29 (0.19-0.44)	0.83	22,246
Tag mixing	0.13 (0.11-0.17)	0.11 (0.09-0.14)	0.49 (0.38-0.68)	0.27 (0.22-0.32)	0.48 (0.36-0.64)	33,165	0.31 (0.22-0.53)	0.90	22,540
Piston Line[1]	0.14 (0.11-0.2)	0.13 (0.1-0.18)	0.54 (0.4-0.81)	0.27 (0.22-0.32)	0.59 (0.44-0.8)	33,086	0.35 (0.22-0.53)	0.93	23,499
No HSPs	0.13 (0.11-0.17)	0.11 (0.09-0.13)	0.49 (0.38-0.68)	0.27 (0.22-0.32)	0.5 (0.38-0.66)	33,039	0.30 (0.21-0.47)	0.88	22,565
No POPs/HSPs	0.12 (0.1-0.15)	0.1 (0.08-0.11)	0.47 (0.34–0.61)	0.28 (0.22-0.33)	0.52 (0.4-0.67)	34.168	0.29 (0.19-0.45)	0.79	23,148
Psi (ObjFn)	0.13 (0.11-0.17)	0.11 (0.09-0.13)	0.49 (0.38-0.69)	0.27 (0.22-0.32)	0.5 (0.38-0.65)	33,064	0.30 (0.21-0.47)	0.88	22,601
No h = 0.8	0.13 (0.1-0.16)	0.11 (0.09-0.13)	0.44 (0.36-0.58)	0.31 (0.27-0.32)	0.57 (0.44-0.67)	32,512	0.28 (0.20-0.43)	0.83	22,220
q(HSP) = 1	0.15 (0.12-0.18)	0.12 (0.1-0.14)	0.54 (0.4-0.75)	0.27 (0.22-0.32)	0.48 (0.36-0.65)	33,396	0.31 (0.21-0.5)	0.92	24,585

In terms of relative TRO (depletion) the results are very consistent. The median estimates are between 0.11 and 0.15 with only the IS20 test clearly higher (median level of 0.18). Lower quantiles never dip below 0.10 and most range up to around 0.17 to 0.18. The relative biomass of age 10+ fish shows similar stability, albeit with median estimates and overall probability intervals being on average around 0.02 to 0.03 lower than relative TRO. Interestingly, the case where both sources of the CKMR data are not included (no POPs or HSPs) the answers are only slightly less optimistic than for the reference set. The current TRO to the TRO at MSY ratio is also consistent across almost all the trials (median levels between 0.45 and 0.55) – again, only for the IS20 test is it clearly different and higher (median of 0.64). The relative TRO level at which MSY is produced (TRO_{msy}/TRO_0) is *very* consistent – the median and range for almost all trials is 0.27 (0.22-0.32) apart from the IS20 and no CKMR data trial where it was 0.28 (0.23-0.33), and for the no steepness of 0.8 run where it was 0.31 (0.27-0.33). Current median F to F_{msy} ratios mostly range between 0.45 and 0.55 with only the IS20 trial having a clearly lower (and more optimistic) value of 0.41. MSY is very consistent across the trials, with a median range between around 30,000t to 34,000t and low variation across grid samples. In terms of the projection results across the sensitivity tests, median relative TRO levels by 2035 are projected to be between 0.26 and 0.34 (with the lowest of the lower 10%ile being 0.19). In terms of the original tuning objective - the probability of the relative level of the biomass of age 10+ fish being greater than 20% by 2035 – values are generally in the 0.8 to 0.9 range apart from the two trials relating to discarding the 2016 aerial survey, or with a higher value of the aerial survey CV, and the Upq2008 trial where they are 0.74, 0.72, and 0.73, respectively. Expected levels of the TAC across the years 2018 to 2035 are consistent, varying between around 22,000t to 26,000t.

Discussion

The CCSBT OM was reconditioned in 2017 to include new and updated data sources. The reference set and associated sensitivity tests were agreed at the OMMP meeting in June (Anon., 2017a) and were successfully run, including projections using the Bali Procedure. In terms of the reference set, current levels of median relative TRO are estimated to be 0.13 (0.11-0.17 80% PI); recent recruitment is estimated to be well above the expected level, especially 2013; the ratios of the TRO and fishing mortality to their MSY counterparts are 0.49 and 0.5, respectively; and recent surplus production is just above 40,000t (a historical high). In terms of projections, using the Bali Procedure, by 2035 the median (and 80% PI) for relative TRO is 0.3 (0.21-0.46); the probability that the biomass of age 10+ fish is above 20% of the unfished state (the 2011 tuning objective) is 0.88; the year in which the probability that the relative TRO is above 20% of the unfished state with a probability of 0.7 is 2023; and the mean TAC between 2018 and 2035 is 22,570t.

The data are generally explained well, with the only notable instances of misfit some years for the early length frequency data (as in previous reconditioning work) and for the 2016 aerial survey point (which the OM under-estimates). The CKMR data, both POP and HSP, are explained well at all relevant aggregation levels, which suggests we are getting adult abundance and the relative reproductive output of each of the adult age classes about right, given these data. Detailed predictive analyses for the data currently in the OM and likely to be used in the next round of candidate MPs (long-line CPUE, CKMR data) suggested that the current likelihood structures will certainly be adequate to simulate them in projections.

Across the sensitivity tests, medians (and ranges) of relative TRO and the biomass of age 10+ fish are consistent – 0.12 to 0.15 for the former, 0.1-0.13 for the latter. The only different looking trials are the “IS20” and “Fix qhsp = 1” trials – both results are, in general, more optimistic in their depletion statistics. As with the depletion statistics, the MSY ratios (both TRO and fishing mortality) are broadly consistent with the reference set. Estimates of the ratio of TRO at MSY to the unfished level are *very* consistent, with the only real difference being for the “no h = 0.8” trial, where this ratio is around 0.31, not 0.27. Estimates of MSY range between 31,000t and 35,000t. With respect to projections, the results are also consistent with the reference set, with only the “no 2016 AS”, “high aerial CV” and “Upq2008” trials resulting in slightly lower levels of biomass rebuilding, and with none failing the tuning objective (whether TRO or age 10+ biomass based). Future levels of TAC are likely to be between 22,000t to 26,000t.

An interesting outcome of the sensitivity trials has been that now, as opposed to when the CKMR POP data were first included, there is only a slightly less optimistic outlook for the case where we remove the CKMR data altogether. This might seem odd initially, but two things to remember are: (i) we are not really performing a test without the data *entirely*, given how many structural changes we have made to the OM (M_{10} range, relative reproductive output model etc.) *because* of the CKMR data, (ii) we now have more of the non-CKMR data (CPUE, surveys etc.) than we had in 2012. The first is arguably the most influential change, but the second is instructive also as it suggests that - with the appropriate structural changes in the OM – with more recent optimistic data there is a consistency across the various data sources now that was not apparent in 2012.

The reconditioning of the OM suggests that recent signals are positive, there is a clear upward trend in the adult population, recent recruitment is above the expected level, and current levels of fishing mortality suggest future rebuilding will be somewhat faster than initially envisaged in 2011. There is a marked consistency across the suite of sensitivity trials which, while positive, may have implications for considering robustness tests for MP testing. In relation to the Bali Procedure’s performance across the sensitivities, in all cases the 2011 rebuilding objective was met (and exceeded, sometimes significantly) and so would the same objective if referenced in terms of relative total reproductive output.

Acknowledgements

Authors Hillary, Preece and Davies were funded by both the Australian Fisheries Management Authority and CSIRO Oceans and Atmosphere. Authors Takahashi, Sakai, and Itoh were funded by both the Fisheries Agency of Japan (FAJ) and Japan Fisheries Research and Education Agency (FRA).

References

Anonymous (2011) Report of the 16th meeting of the Extended Scientific Committee. Bali, Indonesia. CCSBT.

Anonymous (2014) Report of the 19th meeting of the Extended Scientific Committee. Auckland, New Zealand. CCSBT.

Anonymous (2017a) Report of the 8th Operating Model and Management Procedure group. Seattle, USA. CCSBT.

Anonymous (2017b) Report of the July 2017 HSP web meeting.

Bravington, M.V. *et al* (2017) SBT close-kin mark-recapture with parent-offspring and half-sibling pairs: update on genotyping, kin-finding and model development. CCSBT-ESC/1708/12.

Butler, K, and Stephens, M. A. (2016) The distribution of a sum of independent binomial random variables. *Methodol. Comput. Appl. Probab.* DOI 10.1007/s11009-016-9533-4

Gelman, A., Carlin, B. J., Stern, H. S., and Rubin, H. S. (1995) Bayesian data analysis. Chapman and Hall, London, UK.

Hillary, R. M., Preece, A. L., and Davies, C. R. (2017) Updates required for new data sources and reconditioning of the CCSBT OM. CCSBT-OMMP/1706/4.

Figures

Figure 1: Relative level of total reproductive output (left) and recruitment (right) for the reference set of OMs and covering the years 1931-2017.

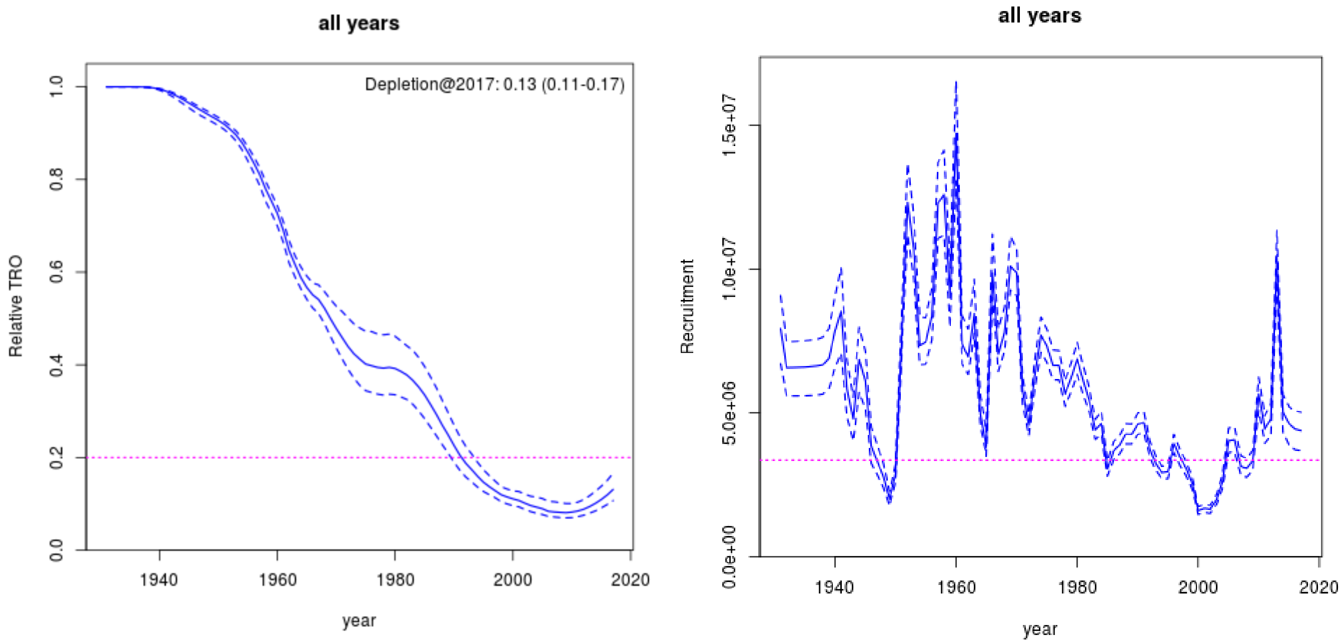

Figure 2: Summary of MSY (top), the ratio of F to F_{MSY} (middle), and surplus production (bottom) for the reference set of OMs. The surplus production is estimated by adding catch in year t and total biomass difference in year t from year t-1 together.

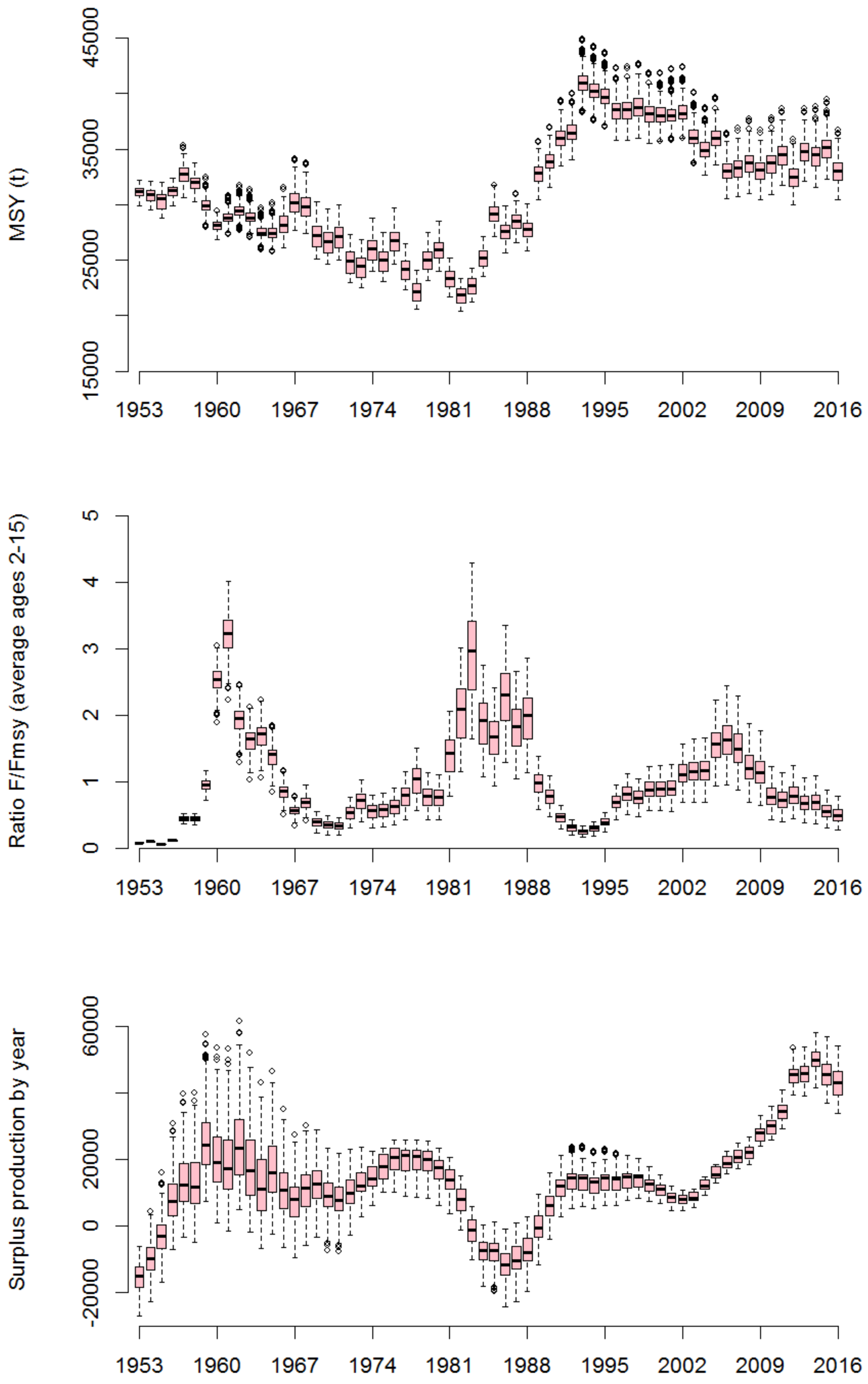

Figure 3: Likelihood profiles for steepness (top left), M_0 (top right) and M_{10} (bottom).

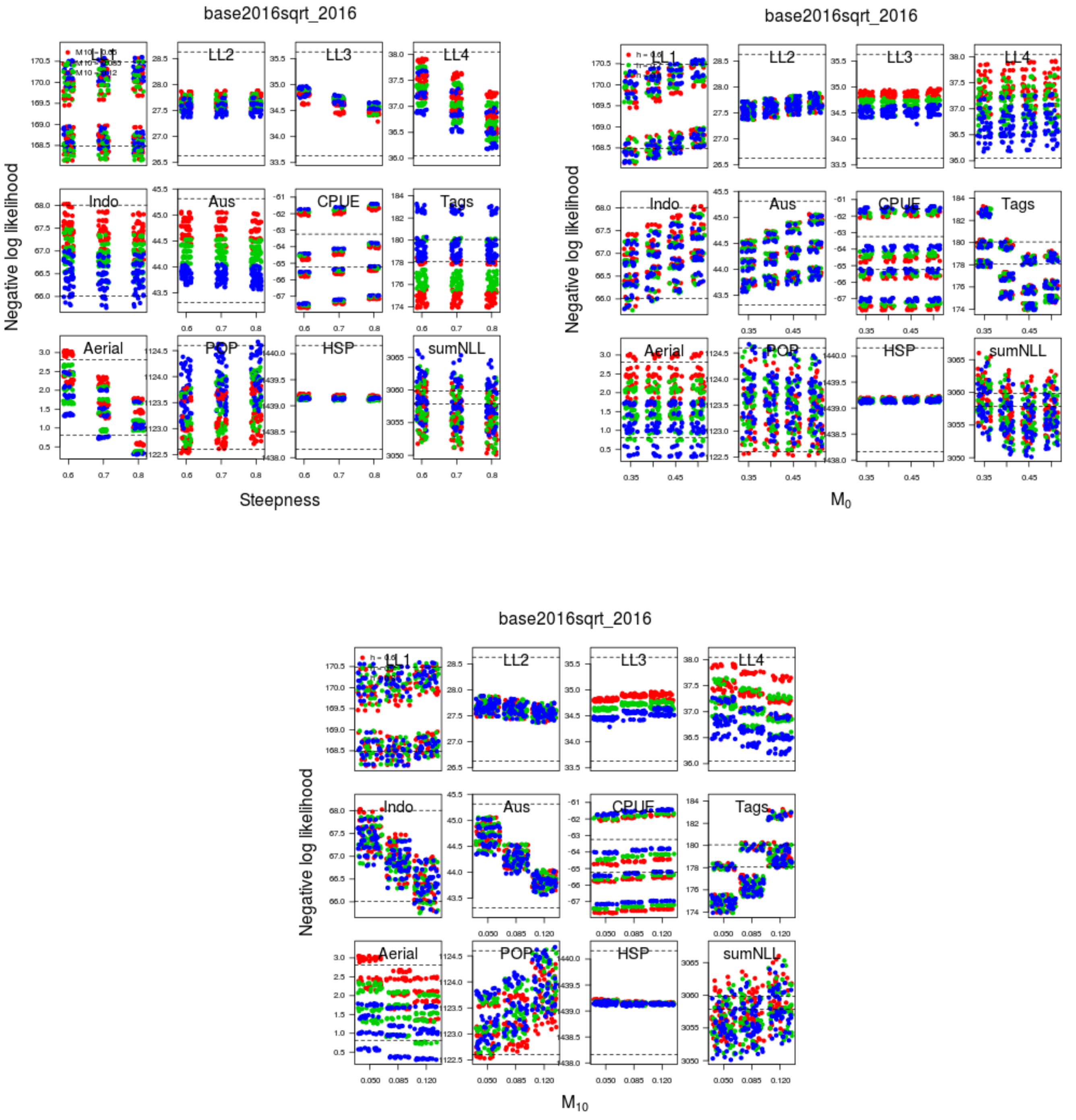

Figure 4: level plot for the grid parameters in the reference set of OMs.

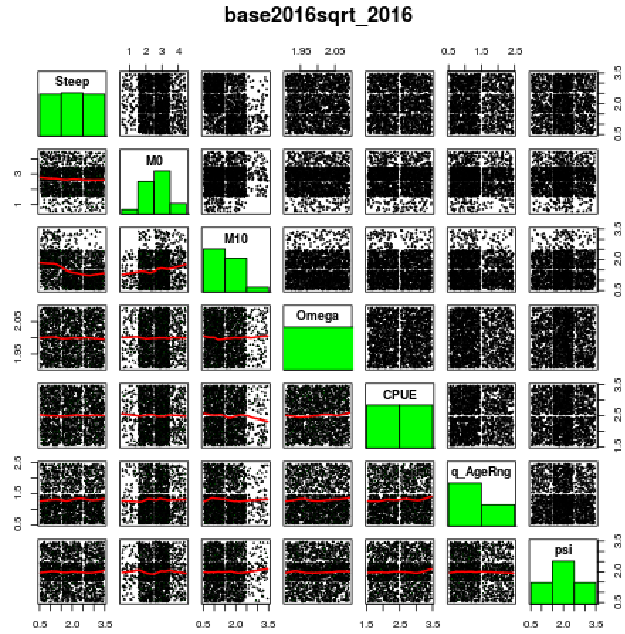

Figure 5: Observed (magenta circles) and precdicted mean and approximate 95%iles (blue solid and dotted lines) for the Japanese long-line CPUE (left) and the aerial survey (right).

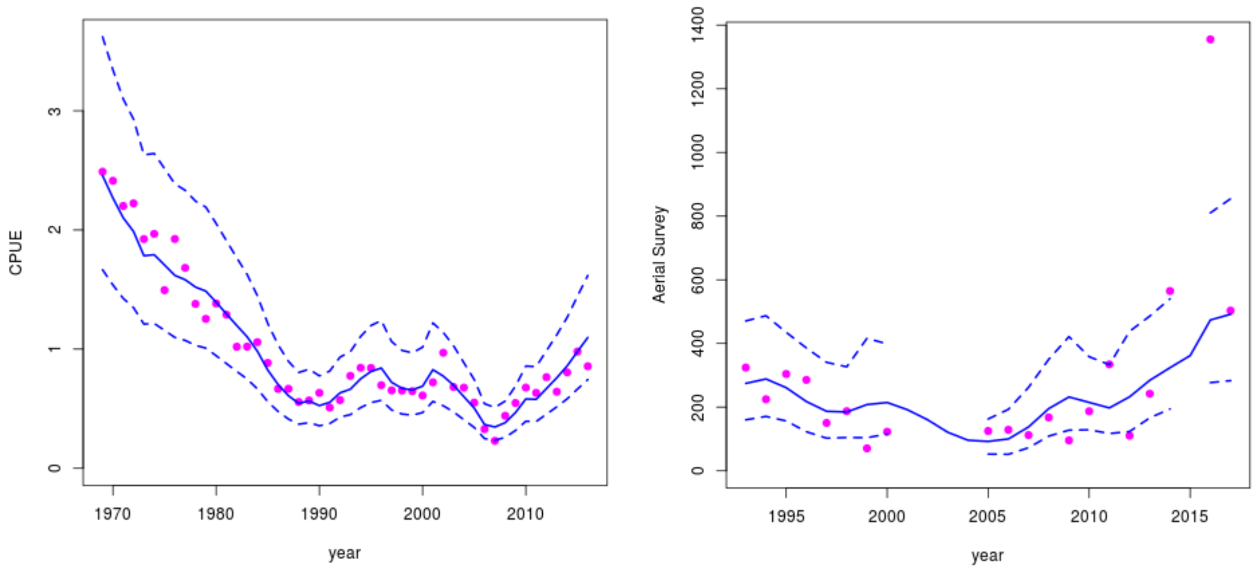

Figure 6: Fits to the tagging data, aggregated across taggers and at the release year and recapture age level.

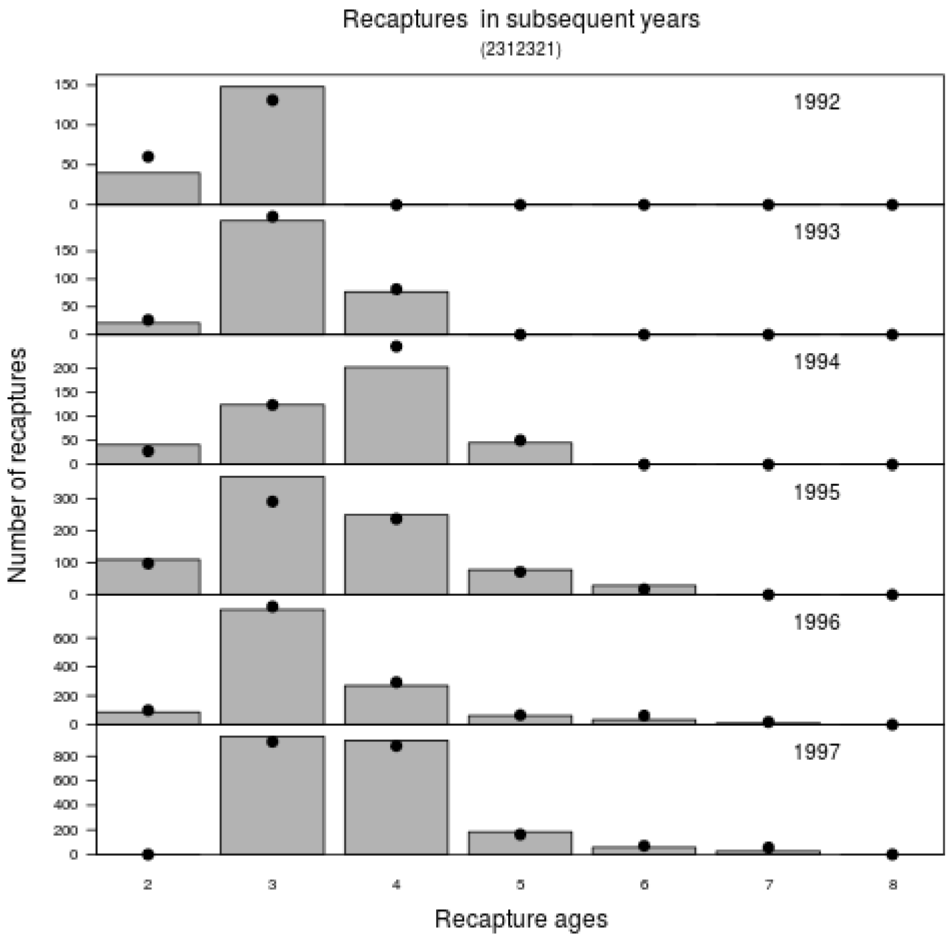

Figure 7: Fits to the CKMR POP data at the cohort (left) and adult capture age (right) aggregation level.

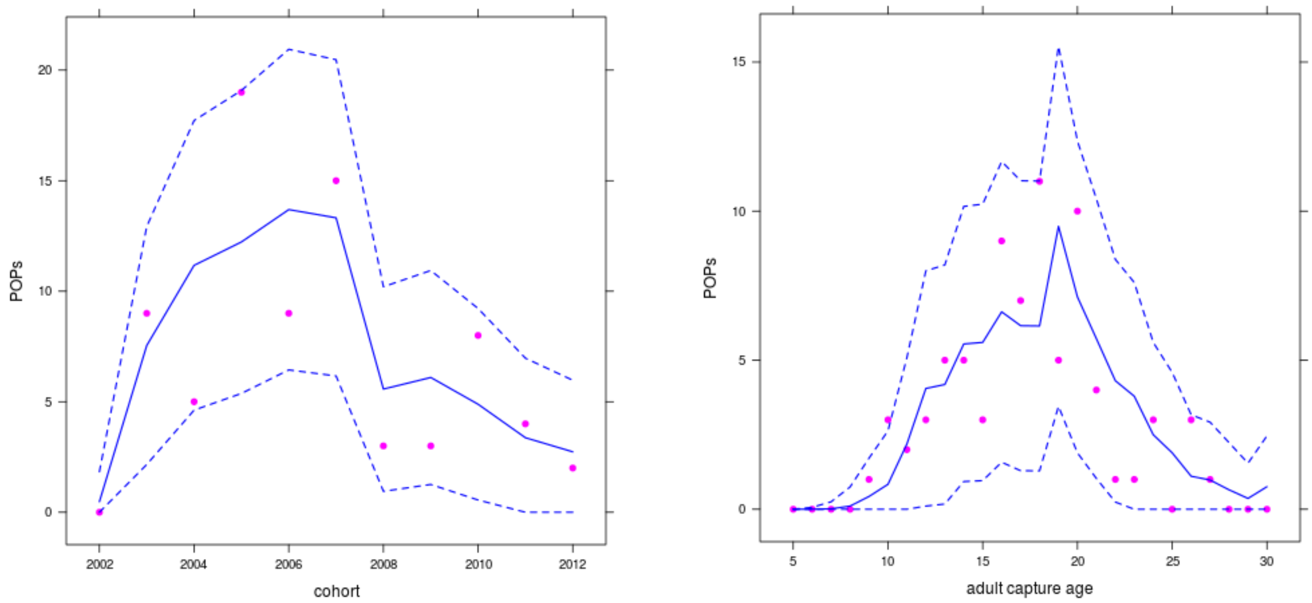

Figure 8: Fits to the CKMR HSP data at the full disaggregation (left) and initial cohort (right) aggregation level (the initial cohort is the oldest animal in the juvenile comparison group).

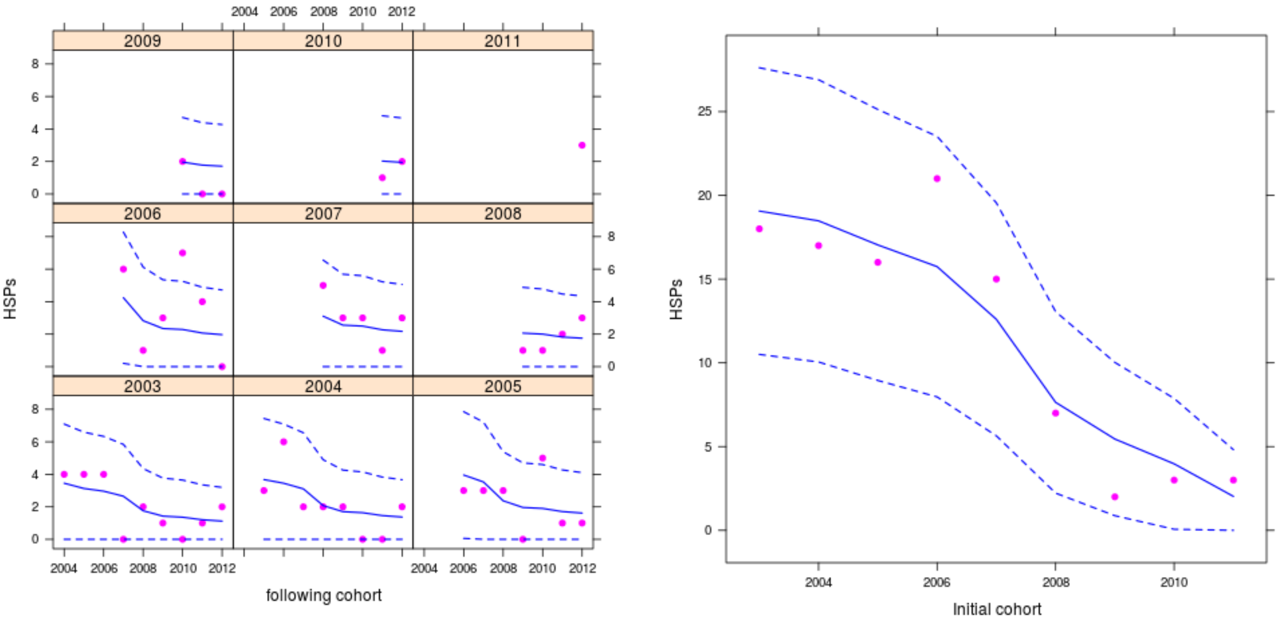

Figure 9: Fits to the age frequency data for the surface (left) and Indonesian (right) fisheries.

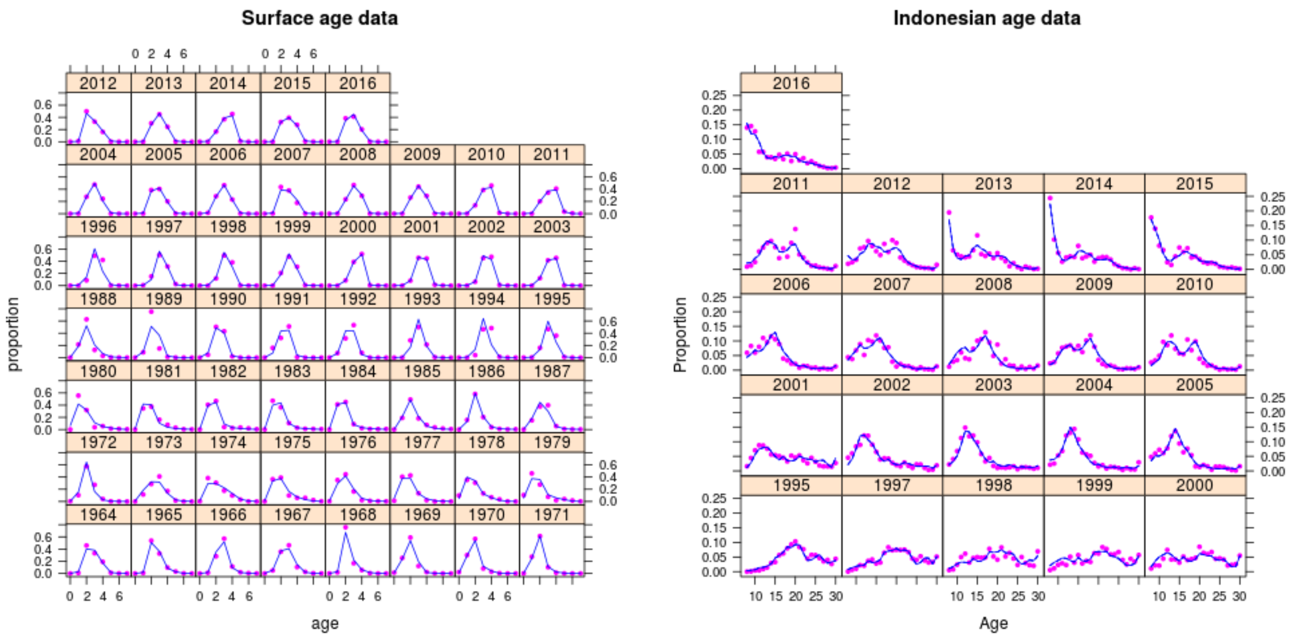

Figure 10: Fits to the length frequency data for four other long-line fisheries (LL1, LL2, LL3 and LL4).

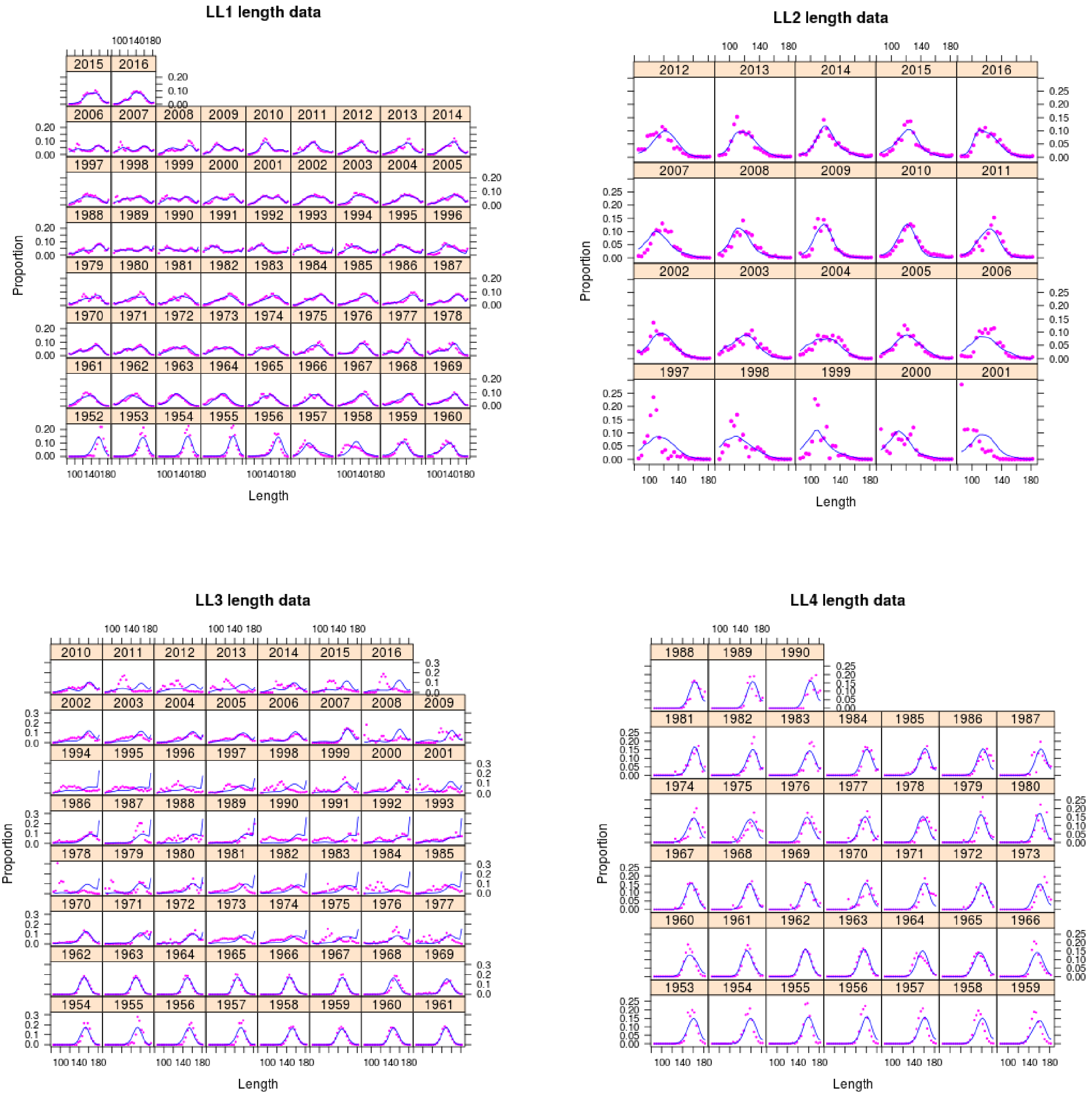

Figure 11: Predictive distribution (top) and the associated observed and predicted discrepancy statistics (bottom) and p-values for the long-line CPUE (left) and the aerial survey (right)

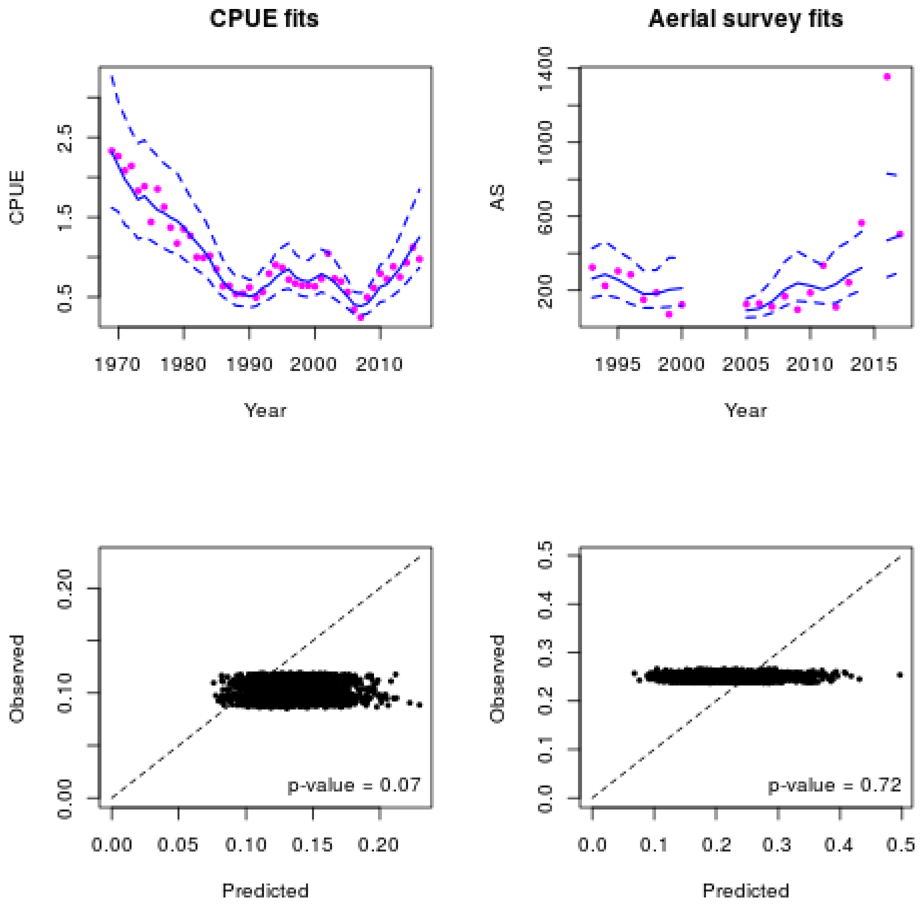

Figure 12: Predictive discrepancy statistics (and p-values) for the CKMR POP data at the cohort aggregation level (POPc, top left), the adult capture age aggregation level (POPa, top right), and the CKMR HSP data at the initial cohort reference level (HSPc1. bottom left). The x-axis is the predicted discrepancy and the y-axis the observed.

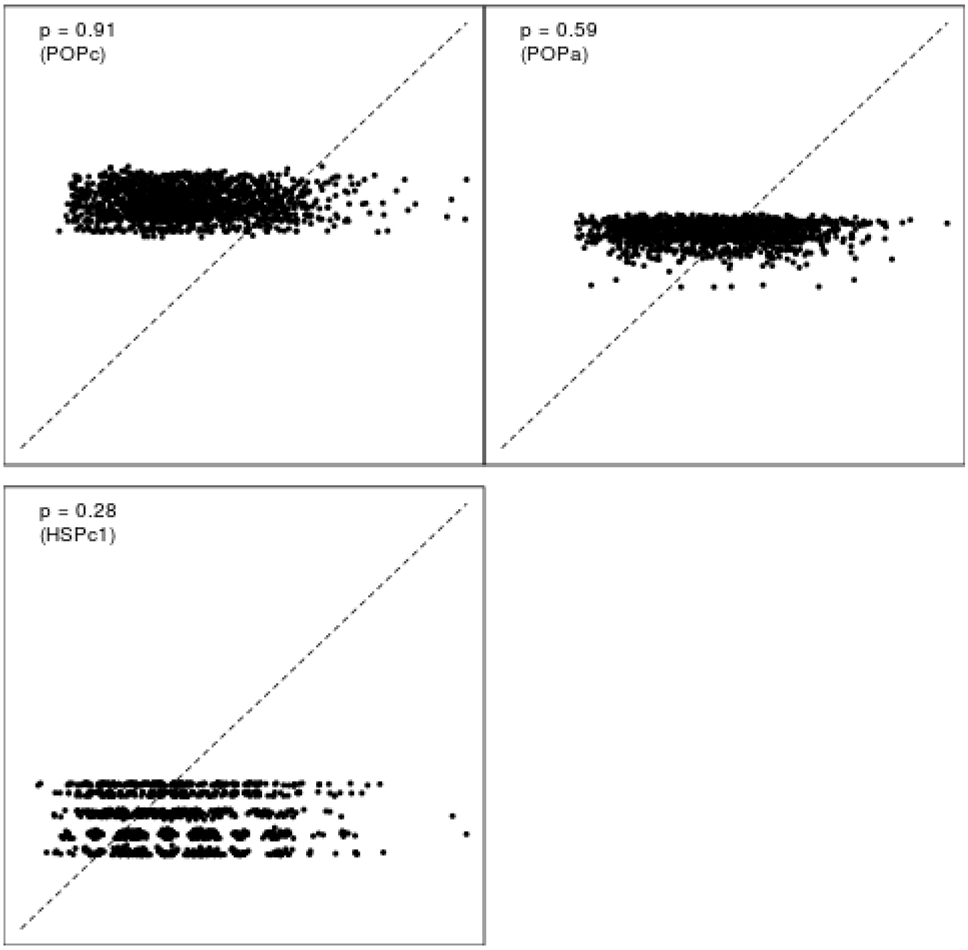

Figure 13: Historical and projected trajectories of the reference set for a) recruitment, b) biomass of age 10+ fish, and c) total reproductive output (TRO). The red line with the pink region represents the median and 90% probability intervals of the 2017 reference set (current assessment). The blue line with the light blue region represents those for the 2014 reference set (previous assessment). The dotted lines indicate the boundaries of the conditioning and projections.

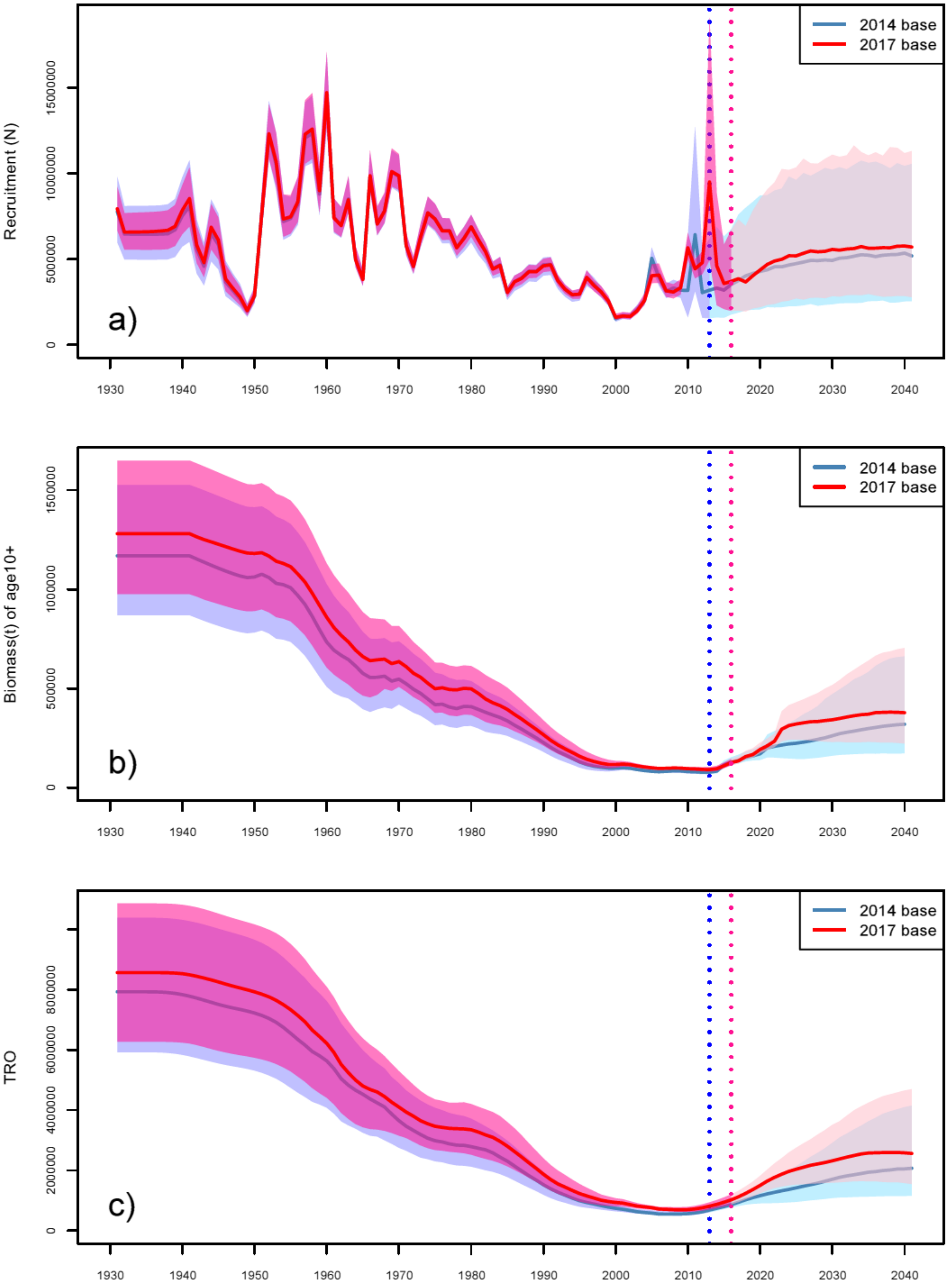

Figure 14: Projected future TAC trajectories for the reference set projections. The bold green line with the greenish yellow region represents the median and 90% probability intervals. The thin greenish lines represent worm plots for each simulation trial.