Information paper for the IOTC Working Party on Methods 2019: IOTC–2019–WPM10-INF02

Prepared for the IOTC MSE Task Force (March 2019, Ispra, Italy)

Indian Ocean Bigeye Tuna

MSE Update March 2019

Dale Kolody (email: dale.kolody@csiro.au)

Paavo Jumppanen

CSIRO, Australia

Summary

This working paper describes developments on the IOTC bigeye reference set and robustness test operating models since the 2018 WPTT and WPM, with some summary MP evaluation results. Due to time constraints, and a pervasive configuration error in the conditioning files (that was identified late in the process), the comparison of fractional grids and repeated convergence issues is based on flawed models, but the generic inferences are expected to remain valid.

Key points include:

- It was not possible to produce an alternative BET growth curve that sensibly merged the western Indian Ocean tag growth increment data with the eastern Indian Ocean otolith data using the existing statistical approaches (i.e. the data are too incompatible). An alternative ad hoc growth curve was produced by combining the two growth curves with a high weighing on the Western curve for younger ages and a high weighting on the Eastern curve for older ages. When combined with the higher CL sample size assumption and low M, the ad hoc growth curve was associated with implausible population dynamics (poor fit to early CPUE combined with dubiously high initial depletion). This suggests that growth uncertainty may well be important, but we omitted this scenario from further investigation because it is not a defensible scenario.
- Additional attention was given to the issue of numerical stability and model convergence in 2019. Instead of simply rejecting models that failed to meet the gradient-based convergence criterion, the minimization was automatically repeated from jittered initial parameter values, until convergence (maximum absolute gradient < 0.01) was achieved (or at least 10 minimization failures occurred). All configurations were able to meet this criteria for BET (though this was not the case for YFT).
- The automated minimization was also used to replicate (3 times) convergence to examine minimization sensitivity to the in initial conditions. Within a model configuration, the standard deviation of the final objective function was ~20 likelihood units (with several values of 100-1000+). However, the CV of stock status characteristics (MSY and B/BMSY) within a configuration were an order of magnitude smaller than the stock status variability among models (based on the lowest likelihood attained within each configuration). There was no evidence that a better likelihood was associated with a lower gradient among those models that reached the gradient threshold. Only the lowest objective function model was retained for the OMs.
- A 288 model ensemble was initially intended as the reference case OM for this meeting (but subsequently found to contain an error in the application of the regional scaling factors and CPUE weightings). Comparing the full factorial grid (288 model) with fractional grids, we note that:
    - The fractional 144 model grid (which would allow all main effects and two way interactions to be estimable) appears essentially identical in terms of stock status estimates.
    - The fractional 72 model grid (with only main effects estimable) resembles the full grid in terms of stock status estimates except that the distribution is somewhat polymodal.
    - The MP evaluation graphics are virtually indistinguishable among the three grids. i.e. It appears that the MP selection process would have likely reached the same conclusion regardless of how many models were in the OM grid.

- Conditioning and MP results are presented for OMgridB19.5 - a 7 factor fractional factorial grid of 144 models. We propose retaining the grid for the TCMP 2019 reference case OM, subject to feedback from the MSE Task Force with respect to:
    - factors and levels to include
    - fractional factorial design (main effects + 2-way interactions proposed)
    - model plausibility filtering (notably with respect to the SS3 catch penalty)
- The reference case MP evaluation performance appears very similar to the previous iteration, and the tuning objectives set by the 2018 TCMP appear to cover a reasonable range of sensible behaviour.
- A number of robustness scenarios are presented, which degrade the performance of the MPs in a qualitatively predictable manner. It is not clear that these results are all plausible or helpful for the purposes of MP selection. We propose not to present any of them to the 2019 TCMP.

Summary points are presented for discussion and/or endorsement from the IOTC MSE task force.

The issue of evaluating plausibility of models within a large grid remains unresolved. We speculate that the current diagnostics and ad hoc inspection process should identify gross outlier behaviour in the system features that are likely to be relevant for MP evaluation, however, we expect that undesirable characteristics might be evident in some models if they were explored in detail.

Introduction

This paper represents a minimalist progress update on key technical elements of the IOTC yellowfin MSE project to obtain feedback in preparation for the 2019 IOTC TCMP, WPM and WPTT. The intended audience is already familiar with the scope of the work and technical jargon. Other interested parties should consult the more accessible project reports found in https://github.com/pjumppanen/niMSE-IO-BET-YFT/.

The operating model is derived from the most recent bigeye stock assessment (Langley 2016).

OM ensembles examined in this iteration

Table 1 lists the OM configurations and rationale. Grid factor details are provided in Table 2, with elaboration of new options provided in the text below (other options are explained in earlier documents).

Table 1. Operating Model definitions. The OMs are listed in the order discussed in the text, reflecting the sequence of development.

OM Ensemble	Rationale (factor abbreviations are defined in Table 2)

OMgridB19.0	72 models with 9 factors in a fractional factorial design. Intent was to screen features for relative importance. Factors h70, h80, h90 (SR steepness) M10, M08, M06 (M) t0001, t10 (tag-weight) q0, q1 (catchability trend) iH, iC (LL CPUE standardization method) iR1, iR2 (regional scaling factors applied correctly) gr1, gr2 (growth curve) ess10, CLRW (CL assumed sample sizes) SL, SD (longline selectivity function) (i2 - LL CPUE CV 0.2 only)
OMgridB19.1	(similar to OMgridB19.3, contained a configuration error and is not reported except in the context of convergence reliability)
OMgridB19.2	Subset of **OMgridB19.4 -** 72 models with 7 factors in a fractional factorial design that quantifies main effects (2 way interactions are confounded). Contains errors in CPUE data processing, but is retained for discussion of convergence and fractional design. h70, h80, h90 M10, M08, M06 t0001, t10 q0, q1

	iH, iC i1, i3 ess10, CLRW (gr1)
OMgridB19.3	subset of **OMgridB19.4 -** Same 7 factors as OMgrid19.2, but with 144 model fractional factorial design that enables estimation of main effects and 2 way interactions.
OMgridB19.4	Same 7 factors as OMgrid19.2, but with 288 models in the full factorial design (all interactions are calculated). Contains errors in CPUE data processing, but is retained for discussion of convergence and fractional design.
OMgridB19.5	Proposed reference case OM - 144 models with 9 factors in a fractional factorial design with all 2-way interactions. Factors: h70, h80, h90 (SR steepness) M10, M08, M06 (M) t0001, t10 (tag-weight) q0, q1 (catchability trend) iH, iC (LL CPUE standardization method) iR1, iR2 (regional scaling factors) ess10, CLRW (CL assumed sample sizes) (i2 - longline CPUE CV = 0.2 only) (gr1 - original growth curve only) (SL - logistic longline selectivity only)

OMrobB19.5.over	A robustness scenario with consistent 10% overcatch for all fleets (catch is accurately reported) (conditioning is unchanged from OMrefB19.5)
OMrobB19.5.iuu10	A robustness scenario with consistent 10% *unreported* overcatch for all fleets (catch is accurately reported) (conditioning is unchanged from OMrefB19.5)
OMrobB19.5.qTrend3	A robustness scenario with a longline CPUE catchability trend of 3% per year in projections (conditioning is unchanged from OMrefB19.5)
OMrobB19.4.ICV3	A robustness scenario with a longline CPUE CV of 0.3 (aggregate annualized) auto-correlation = 0.5.
OMrobB19.4.recShock (not presented this iteration)	A robustness scenario with 8 consecutive quarters of poor recruitment (55% of expected values, similar to estimates for YFT in the early 2000s). (conditioning is unchanged from OMrefB19.4)
OMrobB19.4.impErrCV10 (not presented this iteration)	A robustness scenario in which each fishery has a 40% catch implementation error CV (independent by year and fishery). This corresponds to an annual aggregate CV >10%. (conditioning is unchanged from OMrefB19.4)
OMrobB19.4.under (not presented this iteration)	A robustness scenario in which TACs are ignored for 10 years (fishing mortality constant at current levels) before the TAC is taken without error (conditioning is unchanged from OMrefB19.4)

OMrobB19.4.ICVxxx (not presented this iteration)	An exploratory scenario with a very small longline CPUE CV (aggregate annual =) to explore what might be achievable with a good abundance index
OMrobB19.4.recCVxxx (not presented this iteration)	An exploratory scenario with a very small recruitment CV (aggregate annual =) to illustrate the effect of this assumption

Table 2. Model specification abbreviations. Bold indicates the BET assessment assumption(s). Some abbreviations may relate to additional explorations that were either not completed, reported in earlier iterations, or pertain to YFT.

Abbreviation	Definition
	Stock-recruit function (*h* = steepness)
h70	**Beverton-Holt, *h* = 0.7**
h80	**Beverton-Holt, *h* = 0.8**
h90	**Beverton-Holt, *h* = 0.9**
Rh70	Ricker, *h* = 0.7
Rh80	Ricker, *h* = 0.8
Rh90	Ricker, *h* = 0.9
	CPUE area-weighting factors
iR1	preferred estimate from Hoyle (2018)
iR2	alternate from Hoyle (2018) in which Region 2 represents the lowest proportion of the total vulnerable biomass
	mean Age-length relationship (growth curve)
gr1	**original from assessment**
gr2	speculative alternate curve produced for this report
	Recruitment deviation penalty
sr4	σ_R = 0.4
sr6	**σ_R = 0.6**
sr8	σ_R = 0.8
	Future recruit failure
r55	3 years of poor recruitment (2019-2022); mean dev = -0.55, consistent with 2015 YFT assessment

	Natural mortality scaling factor relative to SA baseline level
M10	**1.0**
M08	0.8
M06	0.6
	Tag recapture data weighting (tag composition and negative binomial)
t00	$\lambda = 0$
t0001	$\lambda = 0.0001$
t001	$\lambda = 0.01$
t01	**$\lambda = 0.1$**
t10	**$\lambda = 1.0$**
t15	$\lambda = 1.5$
	Assumed longline CPUE catchability trend (compounded)
q0	**0% per annum**
q1	1% per annum
q3	3% per annum
q5	5% per annum
	Tropical CPUE standardization method
iH	**Hooks Between Floats**
iC	Cluster analysis
	CPUE observation error
i1	annual $\sigma_{CPUE} = 0.2$
i2	**annual $\sigma_{CPUE} = 0.2$**
i3	annual $\sigma_{CPUE} = 0.3$
	Tag mixing period

x3	3 quarters
x4	**4 quarters**
x8	8 quarters
	Longline selectivity (in conditioning)
SL	**Stationary, logistic, shared among areas**
SD	Stationary, double normal, shared among areas
S4	LL selectivity independent among areas
NS	Temporal variability estimated in 10 year blocks
ST	Logistic selectivity trend estimated over time
Sdev	15 years of recent selectivity deviations estimated
Sspl	Cubic spline function (to admit possibility of dome-shape)
	Size composition input Effective Sample Sizes (ESS)
ESS2	ESS = 2, all fisheries
ESS5	ESS = 5, all fisheries
ESS10	**ESS = 10, all fisheries**
CLRW	ESS = One iteration of re-weighting; the output ESS from a reference case assessment specification (capped at 100)
CL75	ESS = One iteration of re-weighting; the output ESS from a reference case assessment specification raised to the power of 0.75 (capped at 100)

New assumptions in the March 2019 iteration

Alternate bigeye growth curve

The WPTT 2018 requested an additional growth curve in the BET OM uncertainty grid, partly in recognition of the stock status sensitivity identified in the WCPFC BET assessment, when that growth curve was updated (McKechnie et al 2017). The assessment adopted the growth curve from Eveson et al. (2012), and it was recognized that the growth curve of Farley (2006) offers an alternative with substantial contrast, because it included annual otolith counts from ages that are considerably greater than those examined from the RTTP tagging programme, and resulted in a considerably higher estimate of L_∞.

Paige Eveson (CSIRO, pers. comm.) attempted to estimate a new IOTC BET growth curve that integrates the Farley (2016) otolith data (slightly revised using new methods, Jess Farley, CSIRO, pers. comm.) with the tagging data used for Eveson et al. (2012). However, the data sources are very inconsistent (e.g. Figure 1), and the results were implausible (e.g. Figure 2). The young length-at-age estimates from Eveson (2012) cannot be reconciled with the much larger fish of similar ages estimated by Farley (2016). The two stanza, *VBlogK*, growth curve attempts to explain the poor fit to the otolith data as a high degree of measurement error and/or assigns a substantial number of small fish to a negative age. Removing the younger otolith data (< 6 years) from Farley (2006) did not resolve the problem.

There are a number of mechanisms contributing to the bigeye growth uncertainty at this time:

- Farley (2006) otolith samples were obtained from longline catches in the Eastern Indian Ocean, and it would not be surprising if growth rates differ from the western region, where the tagging data come from. Daily ageing indicates that bigeye growth in the western Indian Ocean is considerably slower than in other oceans.
- The bulk of the juvenile catches and tag recoveries come from the purse seine fisheries in the Western Indian Ocean. Size-based selectivity probably explains some of the observed length-at-age inconsistency, especially for juvenile fish, e.g. if slower growing fish tend to remain around FADs longer, they will be over-represented in the FAD fishery and under-represented in the longline fishery. Note that only the two oldest fish (aged with daily otolith counts) from the western tagging study, appear to be consistent with Farley (2006). It would be interesting to know which fishery these came from.
- Temporal variability has also been noted in tuna growth, such that the different sample periods could also contribute to apparent inconsistencies.
  Methods for bigeye daily and annual ageing are currently in the process of being reviewed and refined. It is considered possible that the annual age could be biased slightly since the algorithm used to convert zone counts to a fractional age was developed for the Western and Central Pacific Ocean rather than the Indian Ocean (which could affect the early part of the growth curve, but not L_∞). As noted in Farley et al. (2006), the estimate of t_0 is higher than previous studies due to the absence of annual age data for fish < 75 cm resulted in a higher estimated length-at-age for young fish (≤ 2 years).

At this time, there is not an alternative bigeye growth curve that is plausibly consistent with all of the data. In the interest of following up on the WPTT request, we simply merged the Eveson (2012) and Farley (2006) growth curves, with an ad hoc differential age-dependent weighting (i.e. Eveson 2012 for younger ages, transitioning to Farley 2006 for older ages), as shown in Figure 3.

As discussed below, the results suggest that the alternate growth curve has an appreciable effect on stock status estimates, however, we advise against using it in the reference case OM at this time, because the ad hoc nature of the curve is not very defensible.

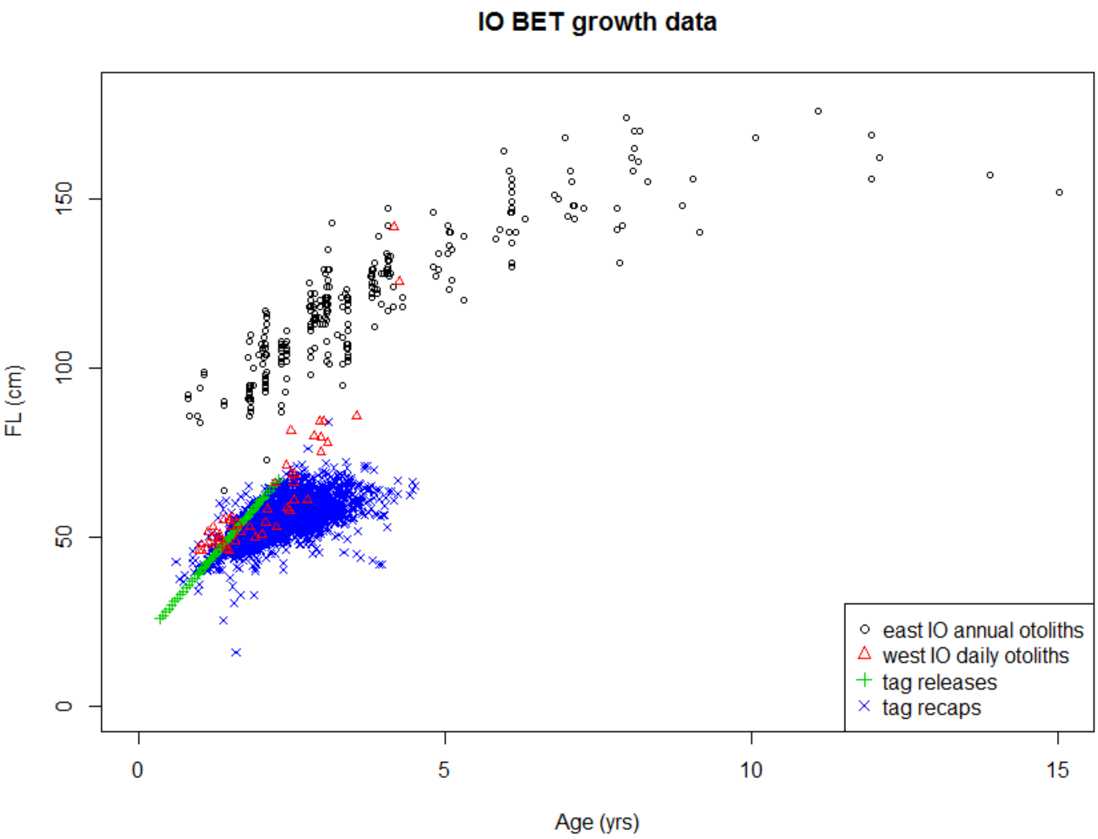


Figure 1. Comparison of Indian Ocean bigeye age/growth data from western and eastern locations (Paige Eveson and Jessica Farley, pers. comm.). Tag release ages were estimated by fitting a linear model through the west IO daily otolith data, then plugging release length into the model to estimate release age. The recapture age is the estimated release age plus the time at liberty.

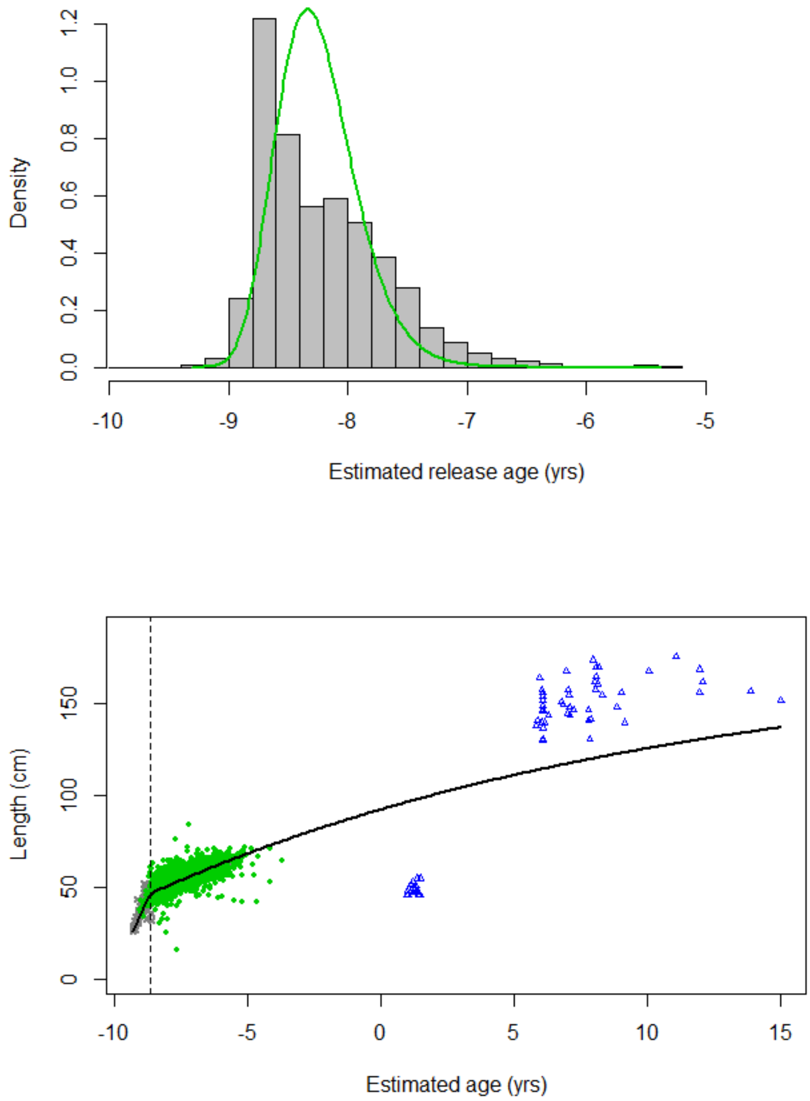


Figure 2. Typical result of attempting to integrate the western and eastern bigeye age and growth data in a single model of the structure used in the past (Paige Eveson, pers. comm.). Eastern otoliths of age <6 years are omitted.

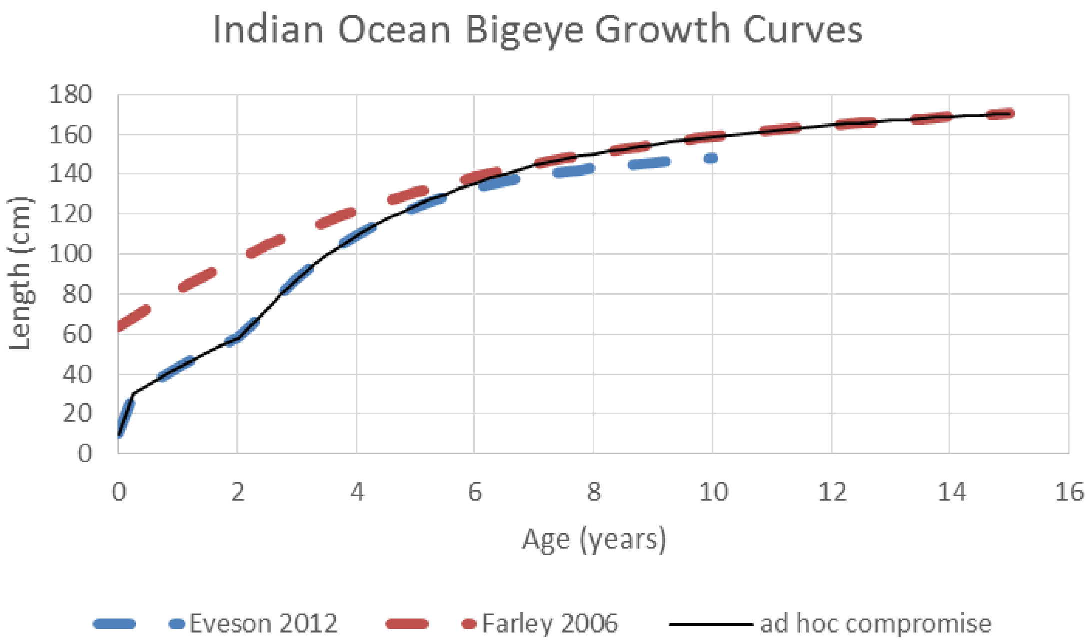


Figure 3. The alternative growth curve (ad hoc compromise) explored in the OM, combining features of the existing western and eastern Indian Ocean growth curves.

CPUE series assumptions

Hoyle (2018) calculated dozens of new candidate CPUE regional scaling factors for bigeye (and yellowfin), all of which corrected an earlier problem (ignoring the change in 5x5 degree surface area due to changing latitude). The trends associated with the different CPUE targeting assumptions (HBF vs cluster analysis) are shown in Figure 4. The effect of the 1% per year catchability trend option is shown in Figure 5. Figure 6 shows the effect of the 3 regional scaling calculation methods (that were considered as viable alternatives in the 2018 YFT assessment):

- iR1 = pr_7994_m8
- iR2 = pr_7594_m8
- iR3 = pr_8000_m8

For consistency with bigeye, options iR1 and iR2 were adopted. Arguably, iR3 probably should have been the alternative option, since it has greater contrast to iR1.

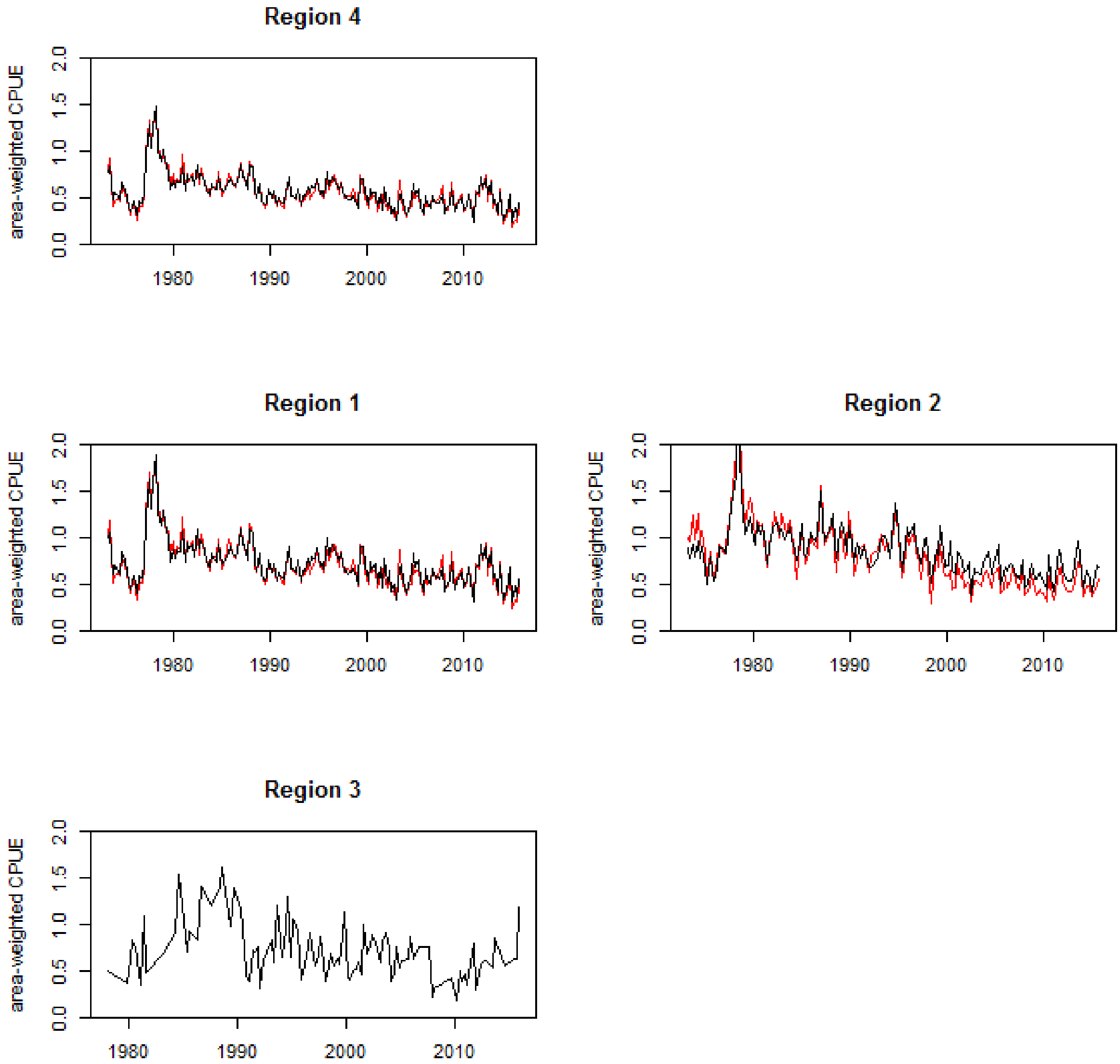


Figure 4. CPUE trends with clustering assumptions (black) and HBF (red) for the targeting assumption.

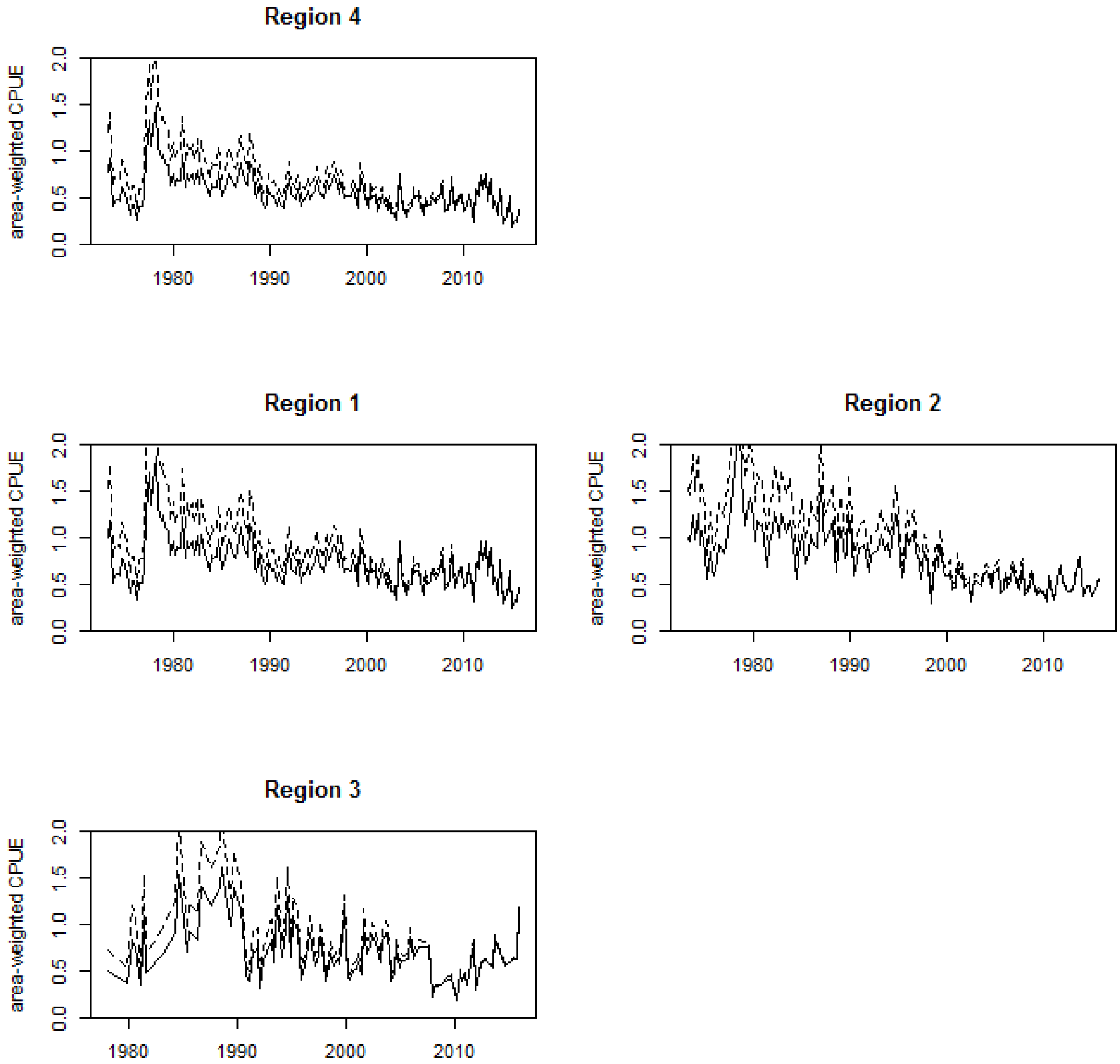


Figure 5. The default assessment CPUE series (solid line) and the 1% per year catchability trend series (broken line).

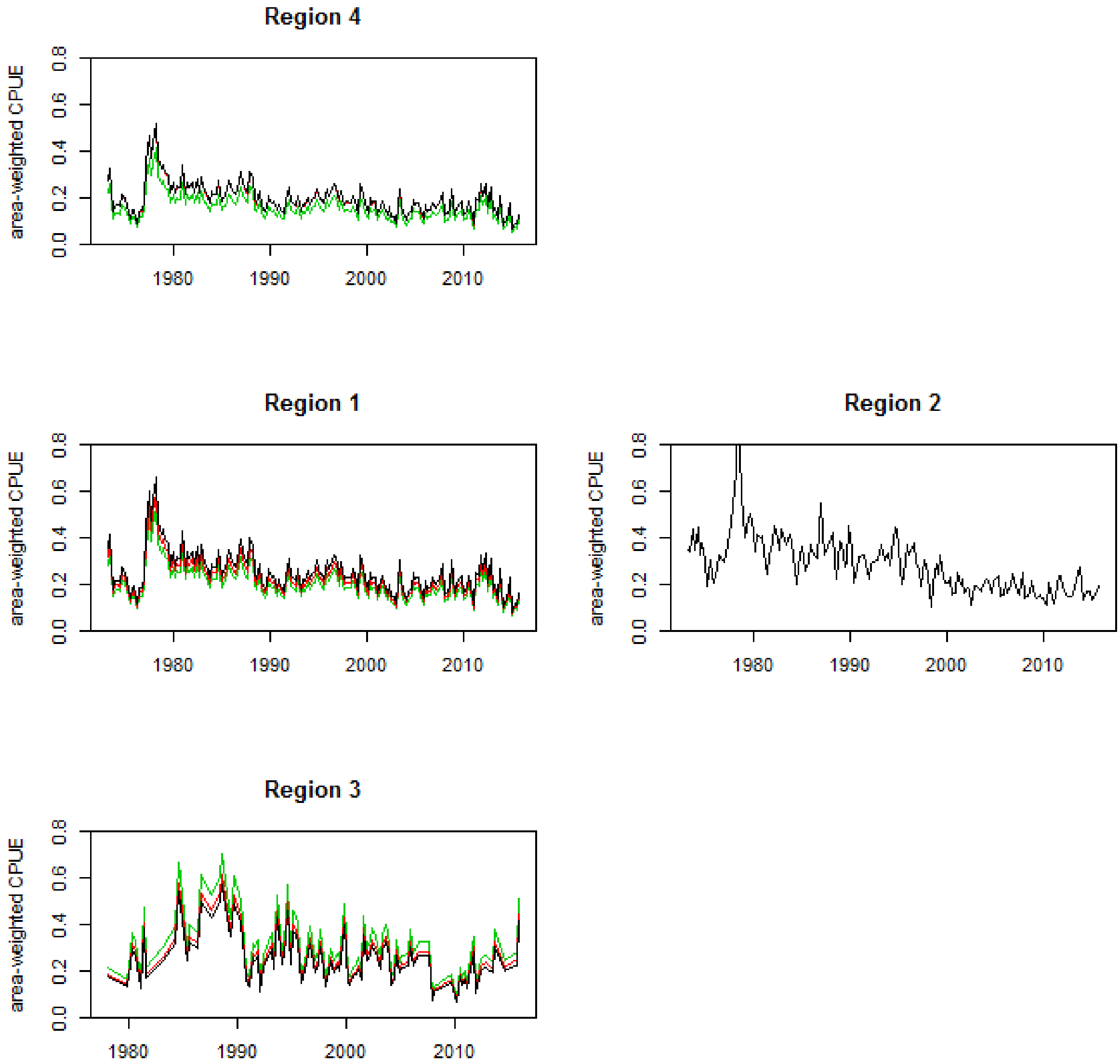


Figure 6. Comparison of the assessment CPUE series with the assessment area-scaling factors (black) and two alternates weighting factors (red, green).

Dome-shaped longline selectivity

A double normal longline selectivity function was included as an alternative dimension in OMgrid19.0, in the interest of being consistent with YFT. However, this resulted in additional parameter bounds problems that there was not time to resolve. Since this was not a specific development request for this iteration, it is not discussed further.

Fractional Factorial Experimental Design

The concept of fractional factorial design was developed in recognition that it is not practical to run experiments with every possible combination of interactions among a large number of factors, and even if it were possible, appreciable 3 way (and higher) interactions tend to be very rare. The principles of fractional factorial design have been used for OM development dating back to at least Schweder (1997). The approach appears to be most common and accessibly described for situations focused on interactions among 2 level factors (three level factors are generally included to identify whether non-linear responses are important). We used the R package "planor" to propose fractional designs for mixtures of 2 and 3 level factors.

Three grids were defined to explore the implications of fractional factorial design and propose the new reference case OM (the two fractional grids are a subset of the full grid and did not require re-running). Unfortunately, specification errors were identified too late to repeat the exercise. But we consider the results to be informative with respect to the requirements for OM design.

- OMgridB19.2 - 72 model fractional design with main effects estimable
- OMgridB19.3 - 144 model fractional design with main effects and all 2-way interactions estimable
- OMgridB19.4 - 288 model full factorial design

The cost of the fractional design is the confounding of the interaction terms. We are not certain that the experimental design principles that enable parameter estimation are necessarily synonymous with the needs of MP testing, but the results below suggest that it works well in this case. We do not know if this is a reasonable generalization for OMs - it could be specific to the BET situation, or possibly an artefact of the random way in which the confounding was assigned.

Figure 7 compares aggregate stock status characteristics among the 3 grids. The 144 and 288 model grids are very similar, while the 72 model grid is generally similar but somewhat polymodal. Figure 8 shows that MP evaluations (tuned for the 288 model grid) are essentially identical for all three.

Figure 9 - Figure 13 compare several diagnostics partitioned by OM factor levels for the 3 grids. The properties are broadly similar, however we might reach different conclusions about some factors. Notably the full factorial grid (B19.4), suggests that the marginal effect of the CPUE weighting factor (i1 vs i3) is probably not important, while the fractional grids (B19.2 and 19.3) suggest that there likely is a minor effect. However, this is purely an artefact of the confounding in the fractional design, because i1 and i3 were actually identical (erroneously).

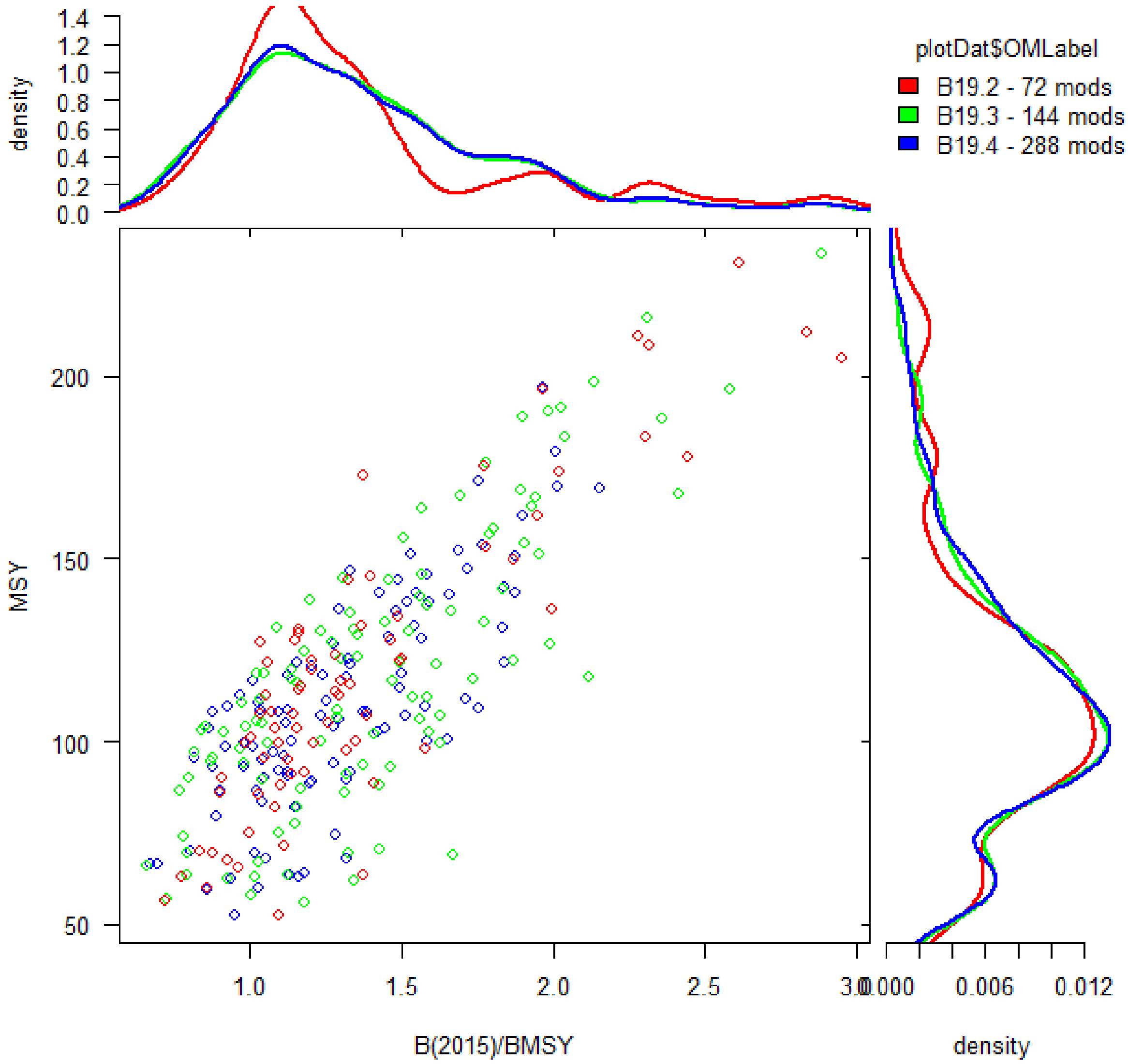
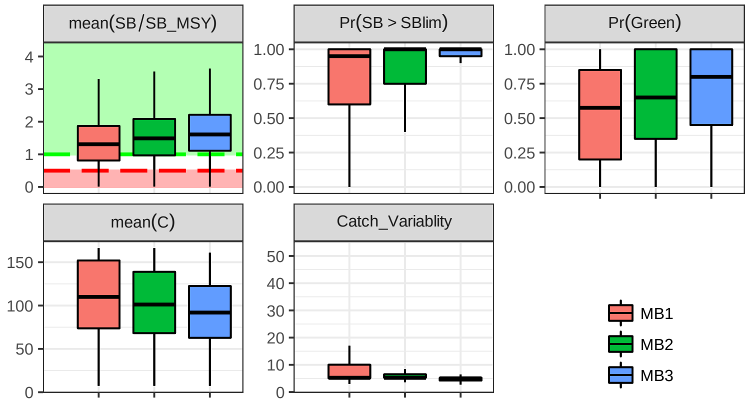
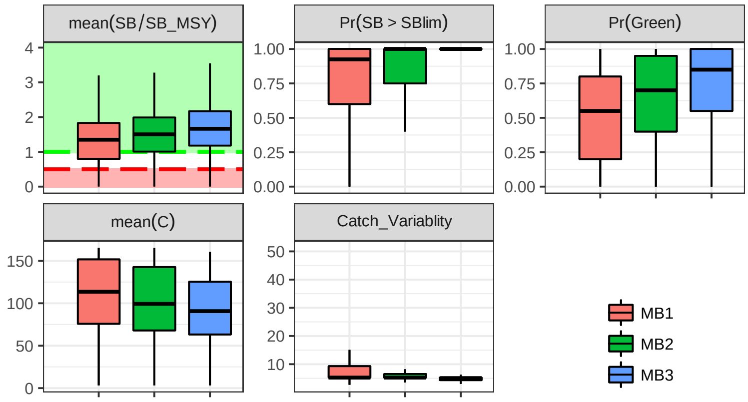


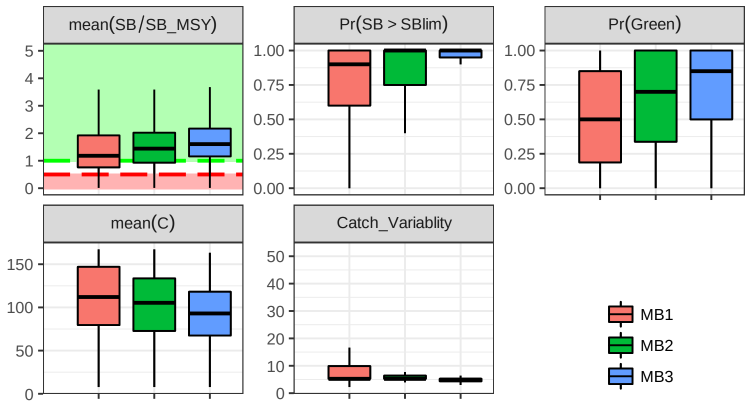
Figure 7. Scatterplot and marginal distributions for stock status summary statistics from grids B19.2, B19.3 and B19.4. Note that B19.3 includes points labelled B19.3 and B19.2 (red + green), and gridB19.4 includes all points.



72 model grid (main effects only)

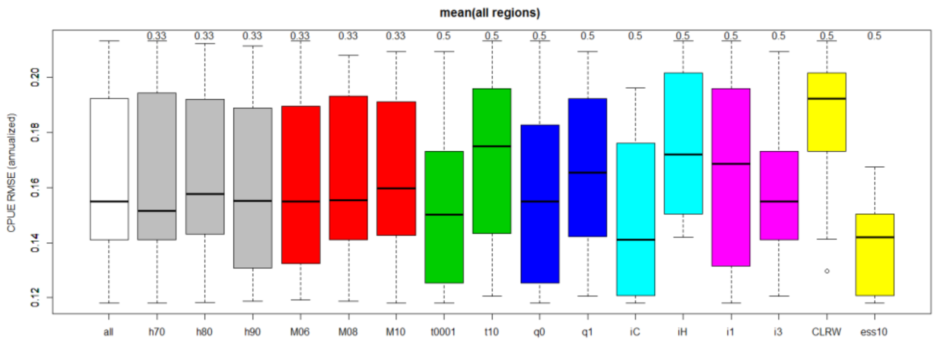


144 model grid (main effects and 2-way interactions)

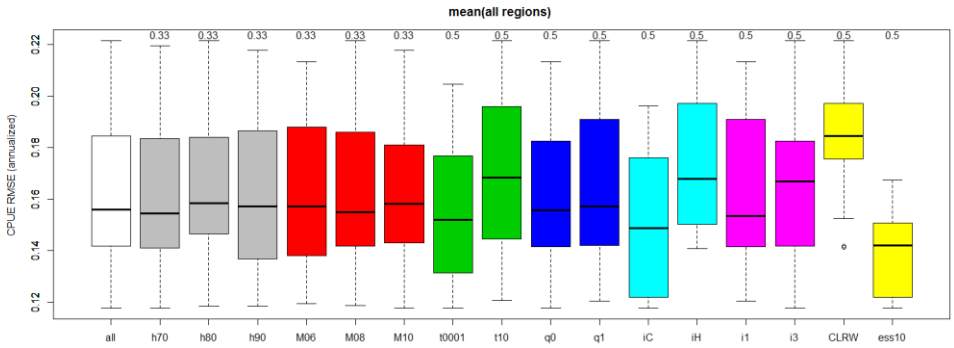


288 model grid (all interactions)

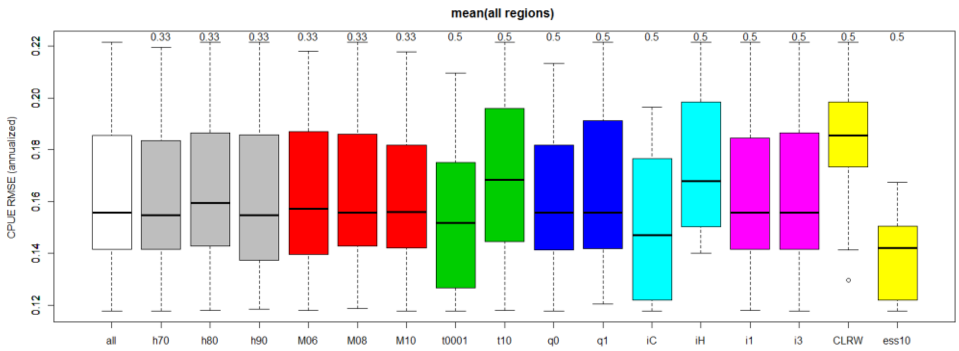
Figure 8. BET MP evaluation results for the full and fractions grids OMgridY19.2, OMgridY19.3, OMgridY19.4. (Note these grids contain model specification errors, but remain informative about fractional factorial design)



B19.2

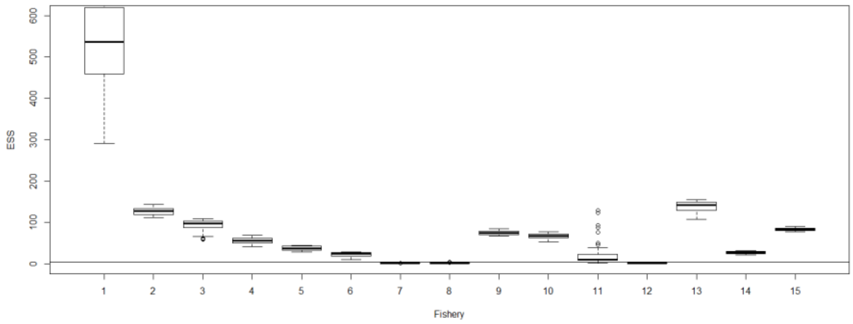


B19.3

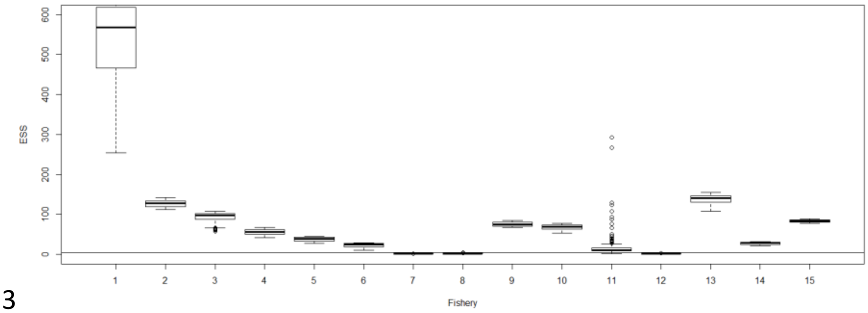


B19.4

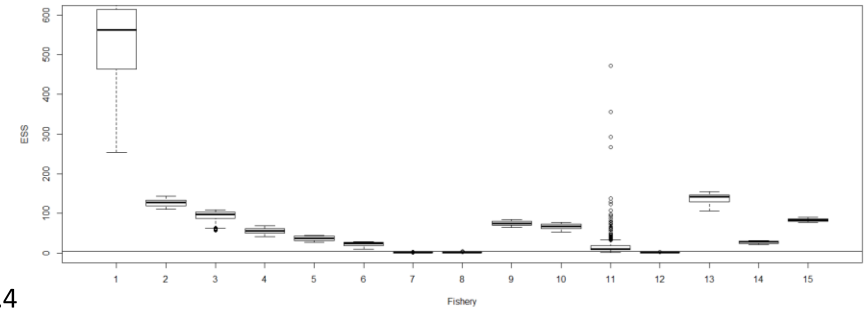
Figure 9. Comparison of CPUE fit among BET OM ensembles. (Note these grids contain specification errors, but are informative about fractional factorial design)



B19.2

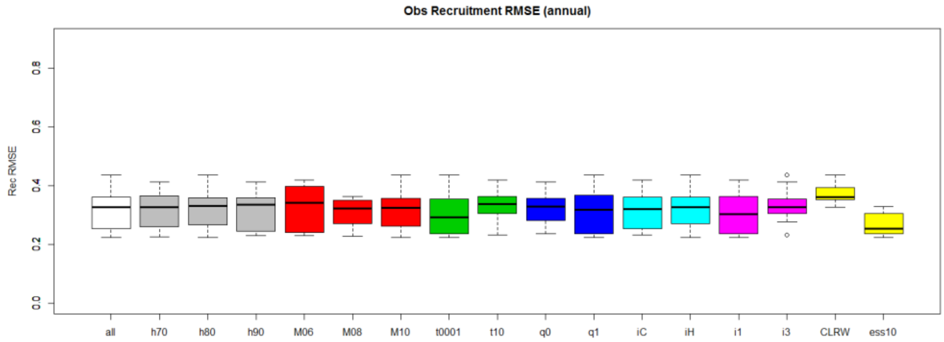


B19.3

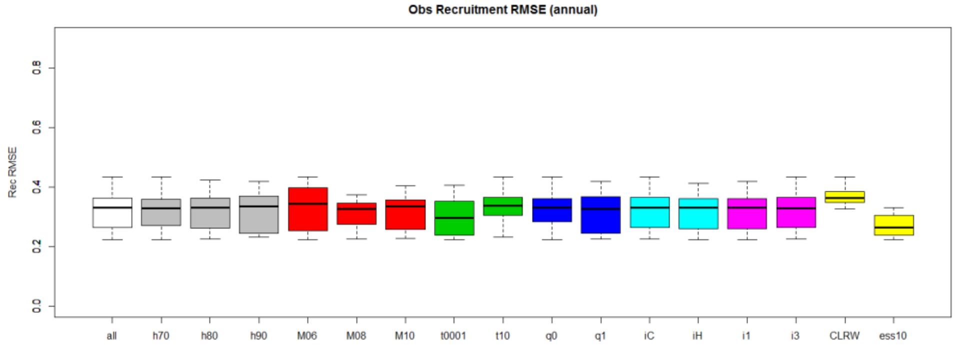


B19.4

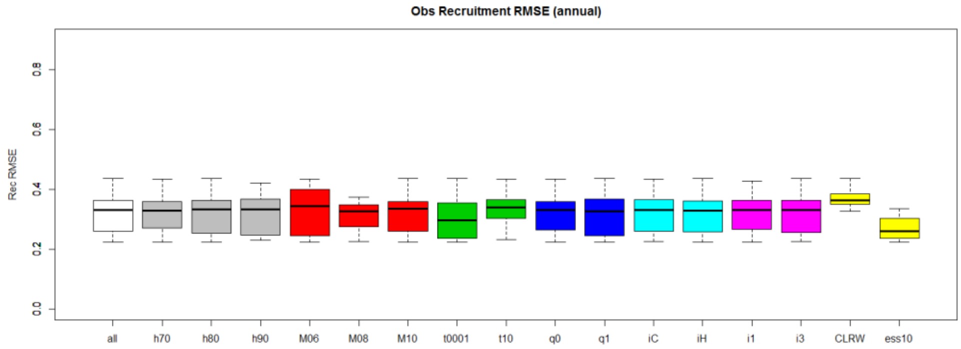
Figure 10. Comparison of CL fit among BET OM ensembles. (Note these grids contain specification errors, but are informative about fractional factorial design)



B19.2

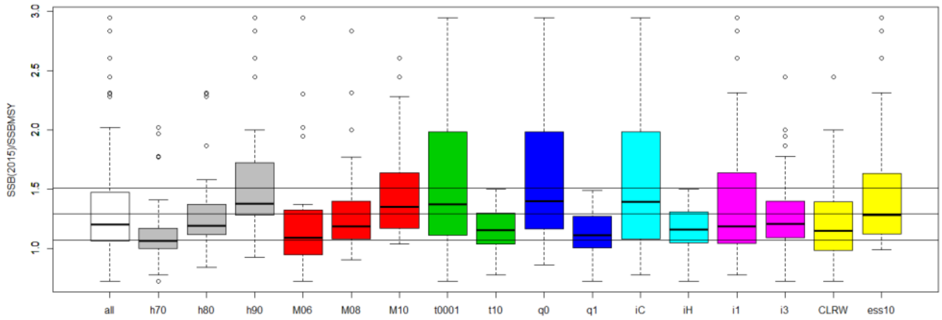


B19.3

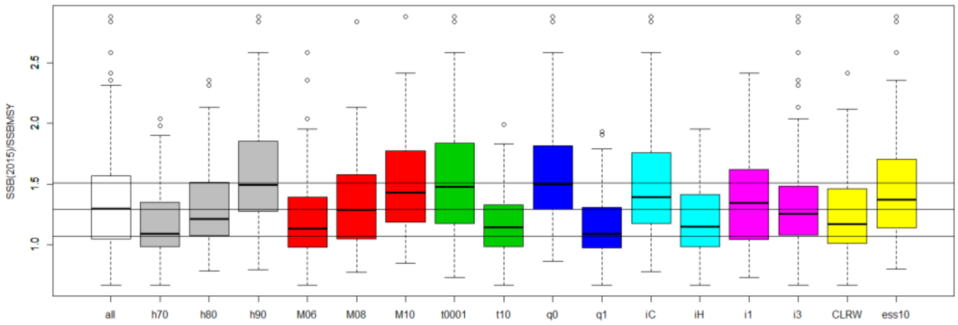


B19.4

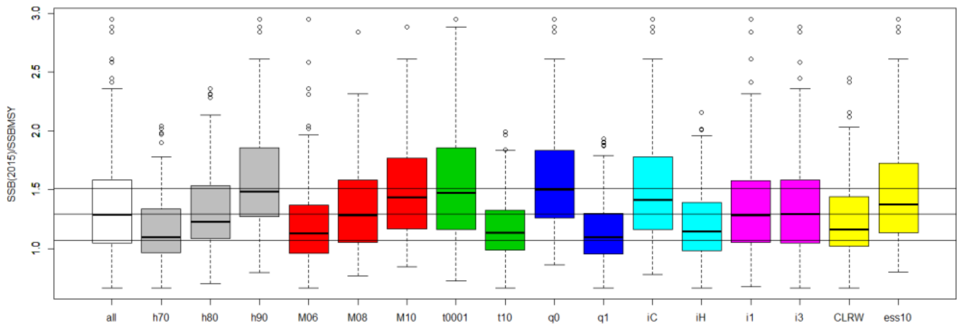
Figure 11. Comparison of Recruitment Deviation magnitude among BET OM ensembles. (Note these grids contain specification errors, but are informative about fractional factorial design)



B19.2

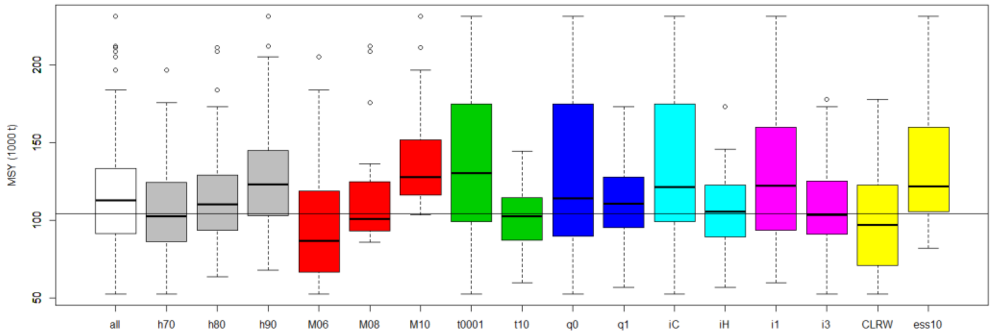


B19.3

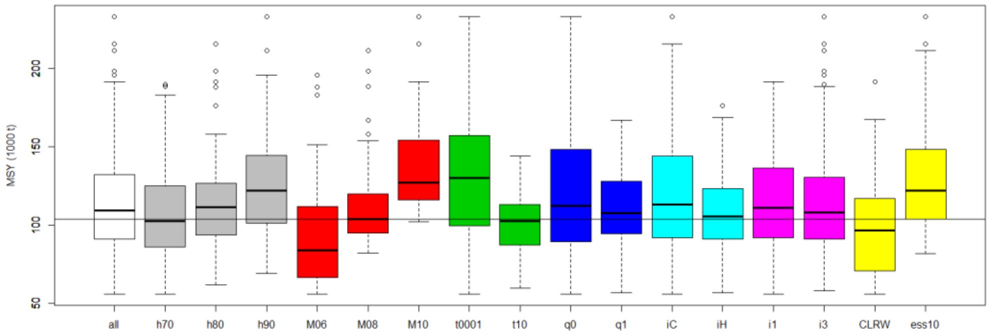


B19.4

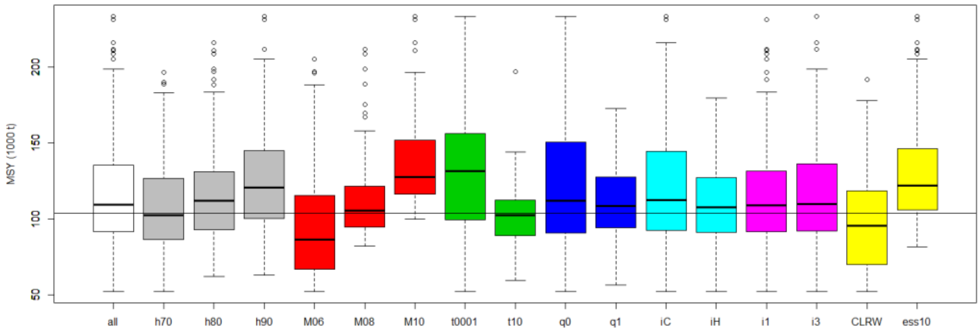
Figure 12. Comparison of estimated current depletion among BET OM ensembles. (Note these grids contain specification errors, but are informative about fractional factorial design)



B19.2



B19.3



B19.4

Figure 13. Comparison of estimated MSY among BET OM ensembles. (Note these grids contain specification errors, but are informative about fractional factorial design)

Parameter estimation sensitivity to initial values

It is recognized that parameter estimation is often sensitive to initial conditions in SS3 assessments (e.g. due to very flat and/or polymodal likelihood surfaces), and the bigeye model is no exception. It would be very difficult to ever conclude that one of these highly parameterized models has truly reached the global minimum, so we have usually assumed that a model reaching the satisfactory convergence criterion (absolute value of the maximum gradient of the objective function with respect to the parameters < 0.01) should provide informative results, even if it is not perfect. This may be a risky approach if one is relying on a single (or small number) of models, but it is hoped that the problem should not be serious if the MSE is designed to be robust to many models. There is an additional concern that a large number of model failures in a particular parameter space could substantially distort the overall distribution of the OM ensemble. To examine these assumptions in more detail, the minimization was repeated from jittered initial parameter values, until 3 separate runs converged for each model configuration. A limit of ~10 fitting failures was imposed (but overridden in some cases). From this exercise we note:

- The summary results presented below are based on a flawed grid specification, but were qualitatively consistent when repeated with other grids (not shown). There was considerable variability in the number of attempts required to achieve convergence (and note that for YFT this was sometimes not achievable with >20 attempts). Around 1000 model fittings were required to achieve 3 successful convergences for all elements of the 288 model bigeye grid.
- There is non-trivial variability in the stock status results due to minimization variability, but this variability is much less than the variability observed among model configurations:
    - The objective function deviation from the minimum (within a model configuration) had a mean of 24 likelihood units (maximum 1681). Note that model A is ~20X more likely than model B if the (negative log) likelihood of model A is ~3 units lower than model B.
    - The mean (across all models) of the CV of the MSY variability (within model configurations) is small (3.8%) relative to the CV among all model configurations (with the best objective function value) 31%.
    - Similarly, the mean (across all models) of the CV of the B/BMSY variability (within model configurations) is small (~3.0%) relative to the CV among all models (with the best objective function value) ~19%
- The OMs are based on the converged models with the lowest objective function value. In so far as MSY and the depletion ratio can be interpreted as indicators of the effect of the numerical convergence problems, it is not obvious that the distribution of these "best" models differs appreciably from the results that would have been achieved from a random sample of converged models (Figure 14). Among all the converged models, there is no obvious indication that lower (maximum absolute value of the) gradients are associated with better objective function values (Figure 15). Thus there does not appear to be any reason to adopt a more restrictive gradient threshold for retention (presumably it could be somewhat relaxed).
- The models with persistent convergence problems were often (possibly always) associated with very high exploitation rates (for at least one age-quarter-region strata), resulting in a runtime catch penalty that SS uses to steer the function minimizer away from dubious results (F > 2.9 for bigeye and yellowfin). We speculate that this penalty introduces a steep gradient, that can conflict with other steep gradients from the likelihood terms. Thus the model may struggle to find the flat bottom in the steep-sided valley between these opposing terms. Even if the true minimum is found, a non-trivial catch penalty probably suggests that

something is probably not realistic. This might indicate that the minimizer has failed to find a plausible parameter space, or it might indicate a fundamental problem in model assumptions. e.g. if M is too high (or there is CPUE hyper-depletion), the model may estimate that there are simply not enough fish in some age-quarter-region strata to support the observed catches. This seems to be consistent with the retrospective pattern observed for YFT in 2018.

- SS3 also reports a "crash penalty" which was always 0, and is presumably only relevant for Pope's approximation to the Baranov F.
- On the basis of these results, and given the time constraints, we opted to obtain successful convergence twice per model specification in subsequent grids.

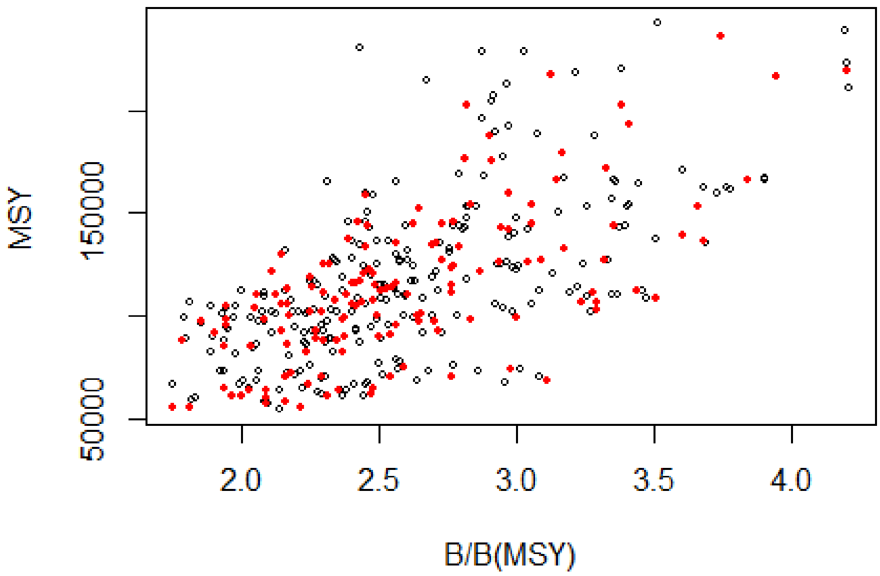


Figure 14. Comparison of stock status characteristics from all (432) of the converged models from the bigeye OM ensemble OMgridB19.1. Red points indicate the (144) best objective function value for each configuration.

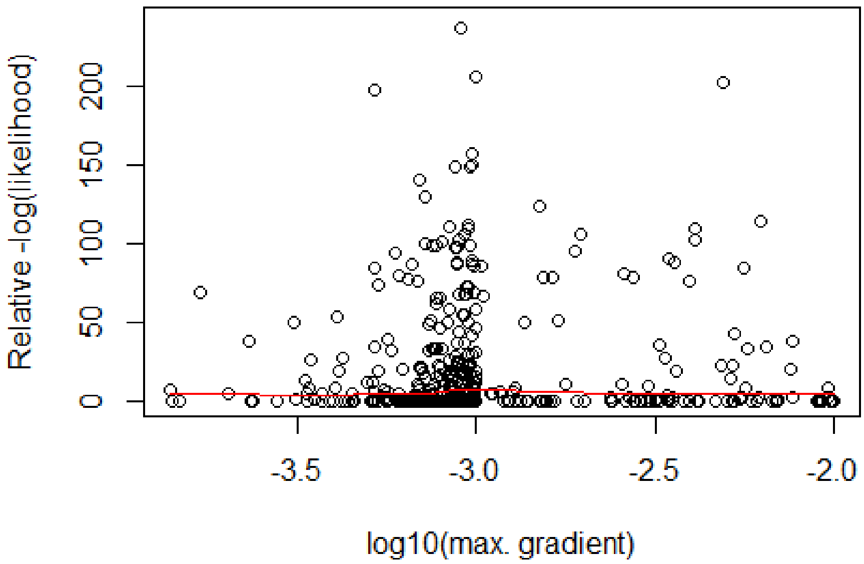


Figure 15. OMgridB19.1 comparison of the relative objective function values (i.e. L - minimum(L) within the 3 converged models for each of the 144 configurations) and the maximum gradient (for models meeting the minimum gradient convergence criteria). One point is off-scale (1681).

Parameters on Bounds

As in previous iterations of OM development, the configuration files for the bigeye OMs had several bounds and prior distributions relaxed relative to the original assessment, to reduce unintended consequences of these somewhat arbitrary constraints. This relaxation presumably has consequences for the minimization speed and sensitivity to initial conditions in some cases. In OMgridB19.1, there were a total of 170 lower parameter bound warnings, related to selectivity, movement and initial fishery depletion. These are probably not important, because 0 is a plausible solution, and a parameter will tend to approach 0 (or a negative bound in log-space) if this is the preferred solution (e.g. the difference in movement rate between 0.001 and 0.0001 should not have a material impact on model dynamics). There were 4 upper parameter bound warnings in OMgridB19.1, related to movement. We are also not concerned about these bounds, as SS3 appears to impose an upper bound of 1 using internal re-scaling. A movement rate of 100% seems unlikely, but it does not represent a logical problem and it would be hard to justify a particular alternative value given the current evidence.

However, simply expanding bounds and priors is not always a satisfactory solution. Figure 16 shows the dubious selectivity and poor fit to the size composition data for the pole and line fishery when bounds are overly relaxed. This model seems to have reached some internal SS3 limit and resulted in selectivity that is not rescaled to a maximum of 1.0. In this case, we opted for a moderately informative prior, but retained the relaxed bounds. There was no systematic consideration of what this prior should be, but given that it is a relatively small fishery, we assume that the influence on the overall assessment should be small.

Initial fishery depletion represents another potential problem. Several models from OMgridB19.1 estimate that the population is substantially depleted at the beginning of the model time period. This results in a poor fit to the early CPUE data (e.g. Figure 17), and is not very believable. This was only observed as an interaction among the low M, relatively high weighting on the size composition data, and the alternate growth curve. For the reasons discussed above, we opted to drop the alternate growth curve from the reference case OM at this time.

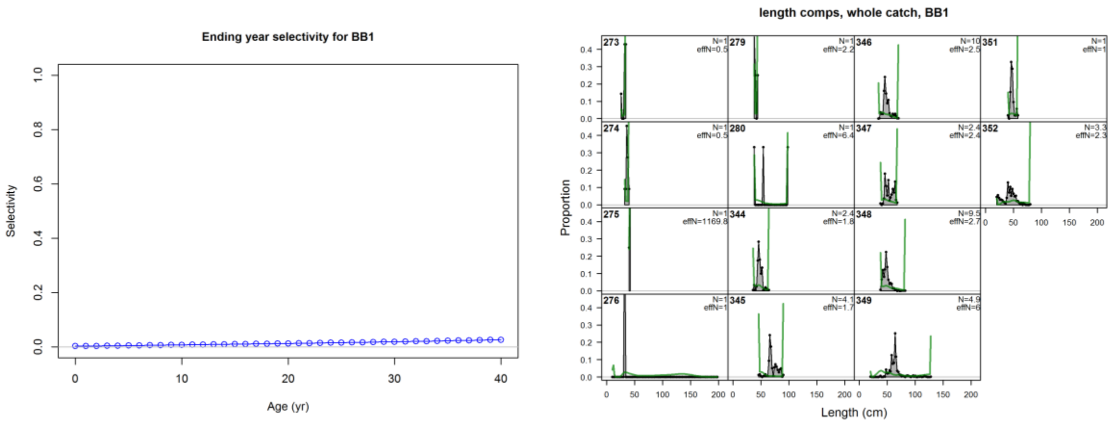


Figure 16. Example of implausible pole and line fishery selectivity hitting an extreme bound, and the corresponding poor fit to the (very limited) size composition data.

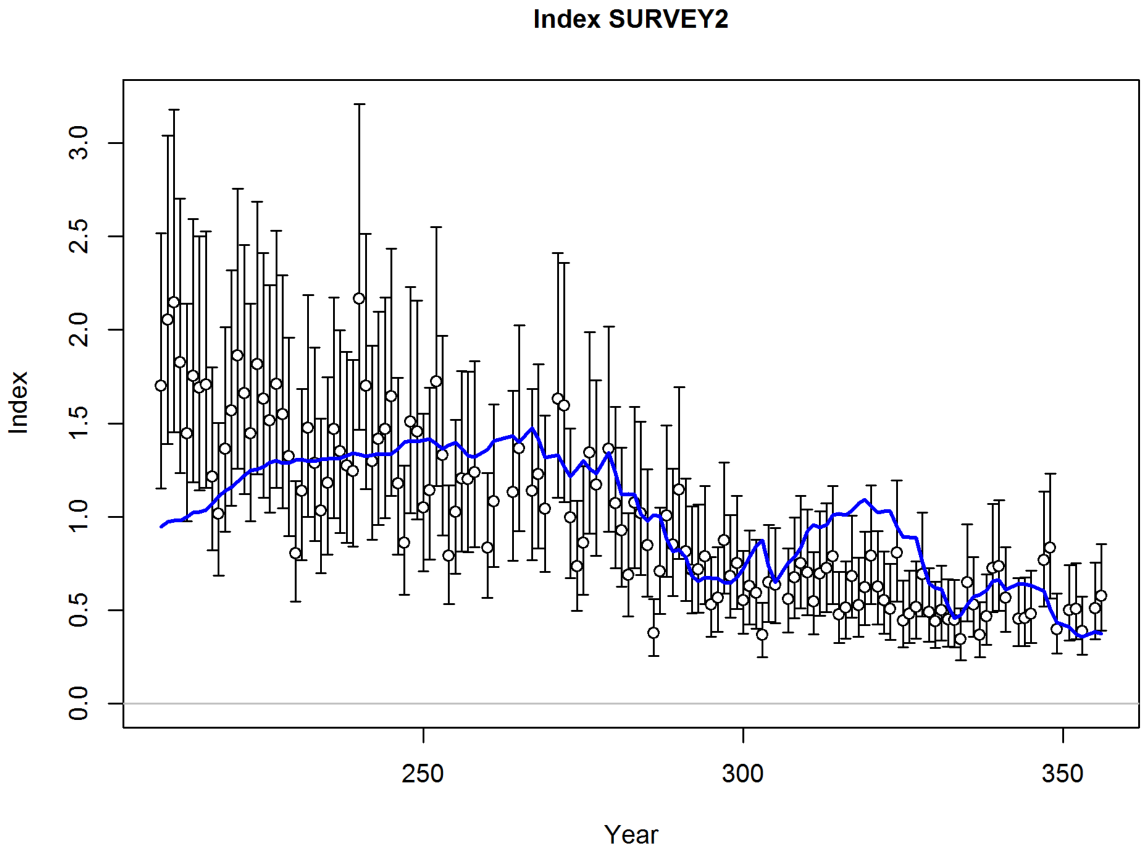


Figure 17. Example fit to the CPUE data from a model that hit the initial fishery depletion bound.

Bigeye Reference set OM conditioning results

OMgridB19.0 - 72 model ensemble, fractional factorial design with 9 factors, only the main effects are not aliased. The intent was an initial screening design to quantify important main factor effects, and drop some factors if appropriate in a subsequent grid with 2 way interactions. On the basis of this grid and other arguments, the growth and longline selectivity options were removed, but recognize that they might merit future consideration in a robustness context.

OMgridB19.5 - proposed reference case OM, 144 model ensemble, fractional factorial design with all main effects and 2-way interactions estimable. R package "planor" describes the aliasing in OMgridB19.5:

```
********** Prime  2  design **********
--- Solution  1  for prime  2  ---
UNALIASED TREATMENT EFFECTS
t.val ; q.val ; i.val ; iRWt.val ; ess.val ; t.val:q.val ; t.val:i.val ; t.val:iRWt.val ; t.val:ess.val ; q.val:i.val ; q.val:iRWt.val ;
q.val:ess.val ; i.val:iRWt.val ; i.val:ess.val ; iRWt.val:ess.val
ALIASED TREATMENT EFFECTS
nil
TREATMENT AND BLOCK EFFECTS CONFOUNDED WITH BLOCK EFFECTS
nil
UNALIASED BLOCK EFFECTS
nil
--- Synthesis on the aliased treatment effects for prime  2  ---
     unaliased trt.aliased blc.aliased
[1,]        15           0           0
********** Prime  3  design **********
--- Solution  1  for prime  3  ---
UNALIASED TREATMENT EFFECTS
h.val ; M.val ; h.val:M.val ; h.val:M.val^2
ALIASED TREATMENT EFFECTS
nil
TREATMENT AND BLOCK EFFECTS CONFOUNDED WITH BLOCK EFFECTS
nil
UNALIASED BLOCK EFFECTS
nil
```

Figure 18 - Figure 19 show the multidimensional scatterplot summaries for these two grids, and Figure 20 - Figure 25 compare summary diagnostics for these two grids, partitioned by model assumption.

Figure 26 shows the distribution of catch likelihoods for OMgrid19.5. It indicates a bimodal distribution, which might be useful as a plausibility indicator. A comparison of the full grid, with two subsets based on catch likelihood filtering, indicates that the most pessimistic configurations are disproportionately affected by the catch penalties (Figure 27).

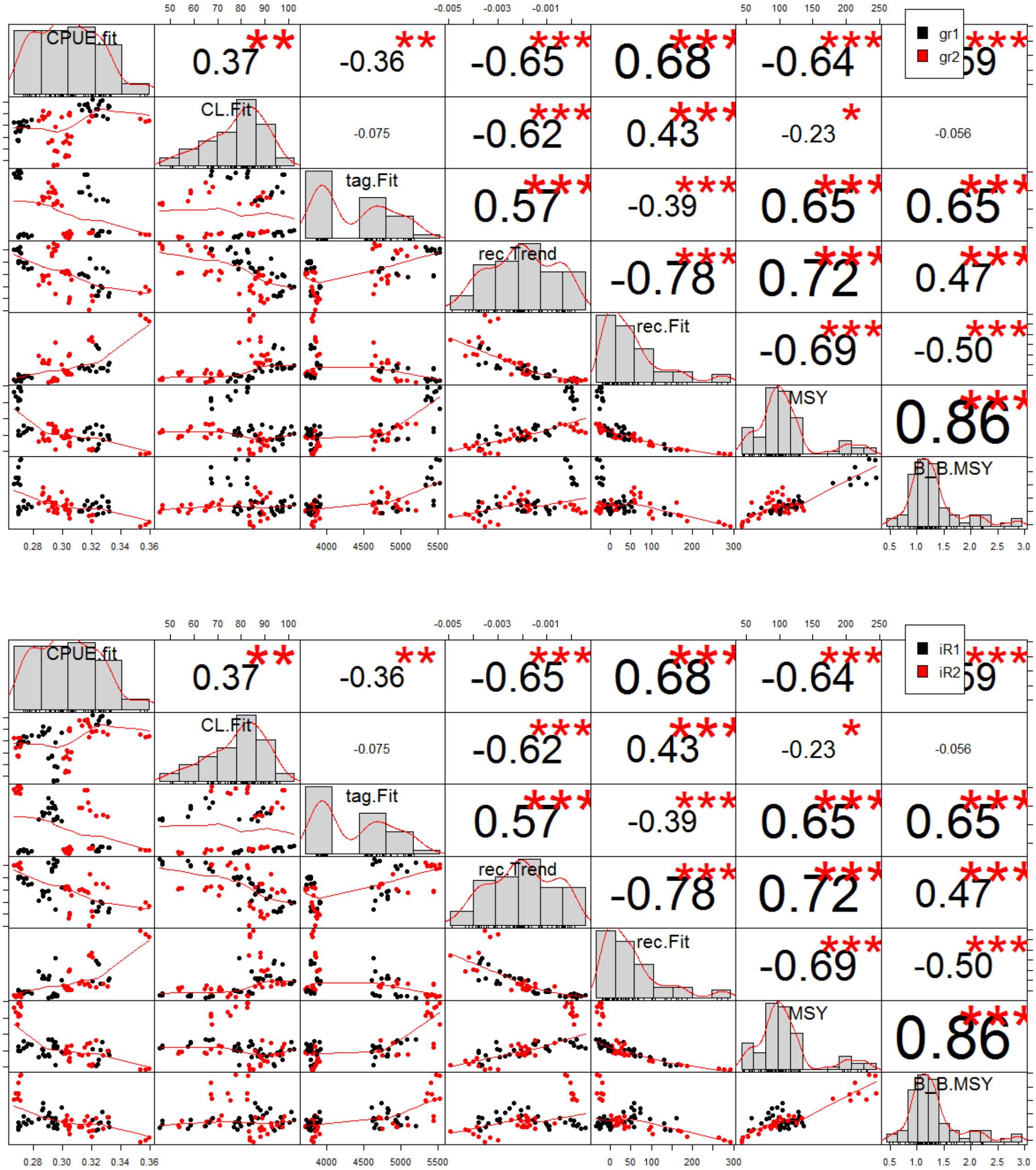


Figure 18. Multiway comparison of OMgridB19.0 model characteristics, partitioned by the dimensions that are proposed for simplification - growth (top) and CPUE area-weightings (bottom).

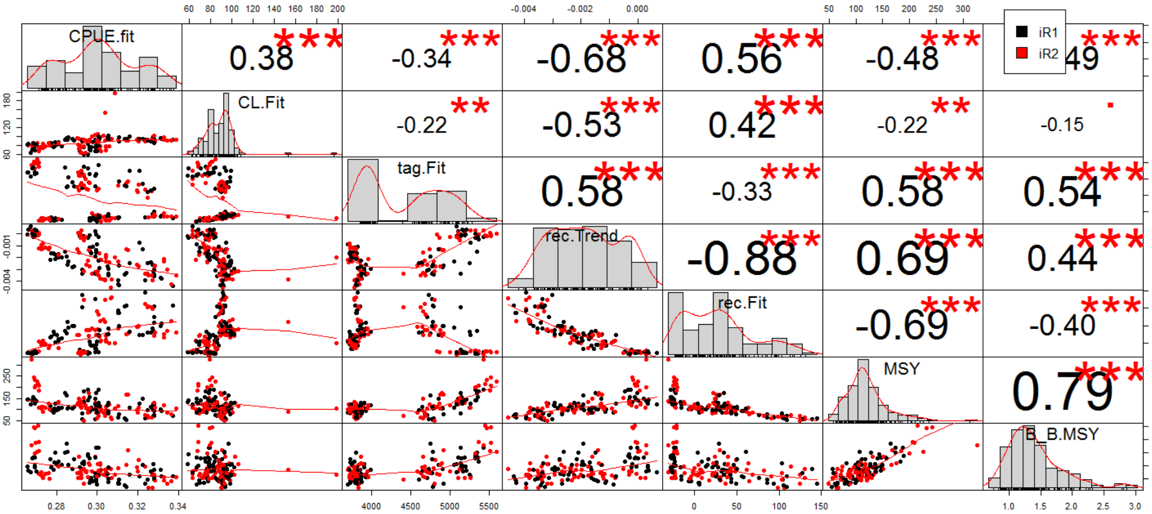
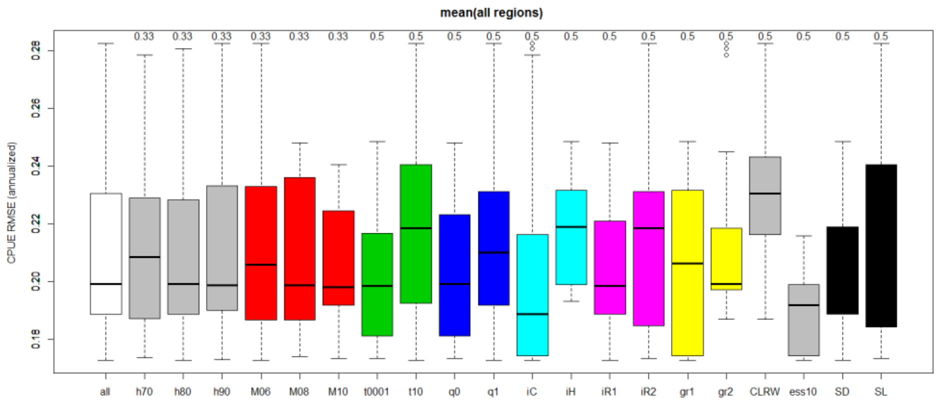
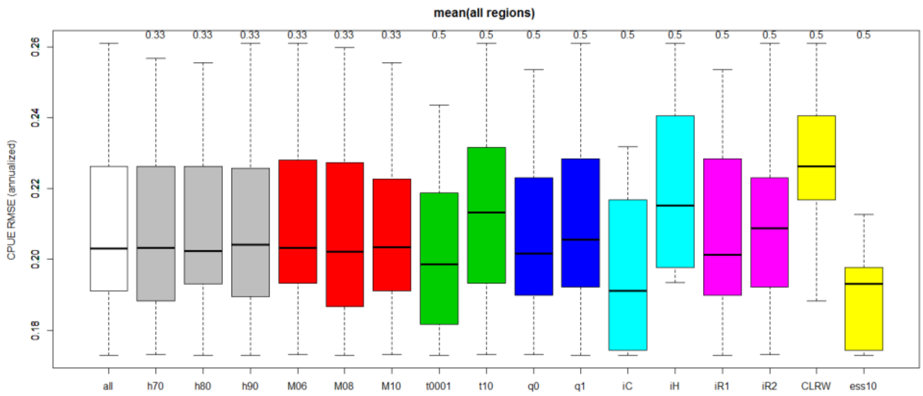


Figure 19. Multiway comparison of OMgridB19.5 model characteristics, partitioned CPUE area-weightings.

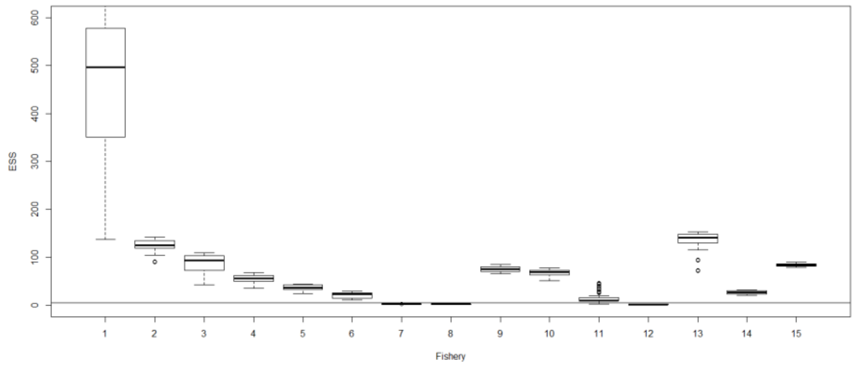


B19.0

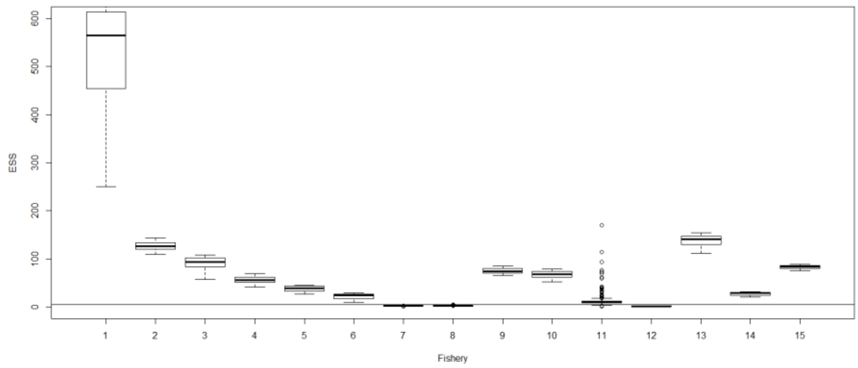


B19.5

Figure 20. Comparison of CPUE fit among BET OM ensembles.

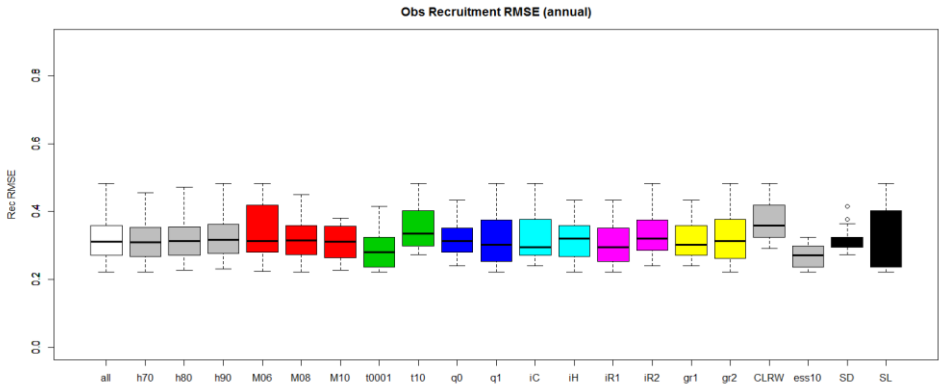


B19.0

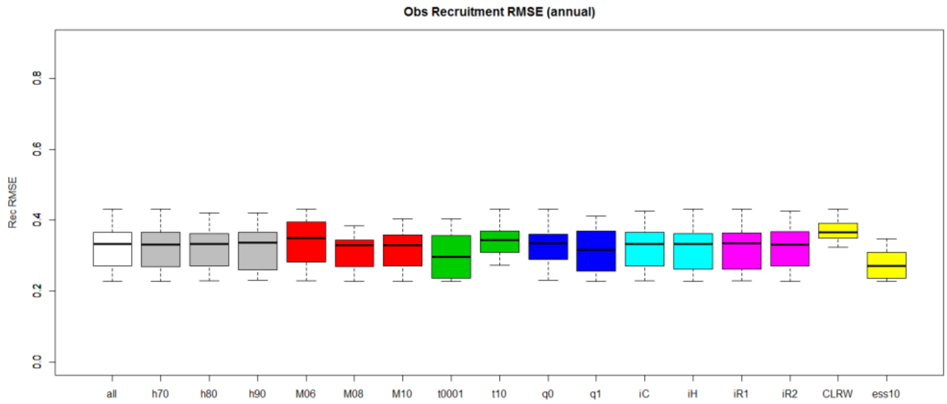


B19.5

Figure 21. Comparison of CL fit among BET OM ensembles.

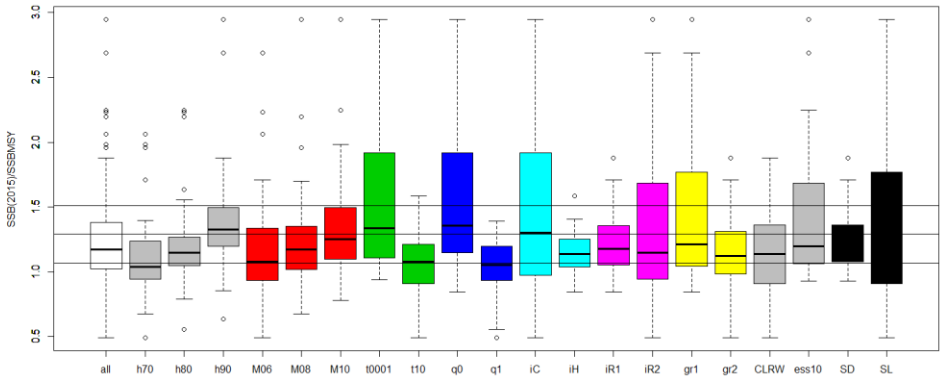


B19.0

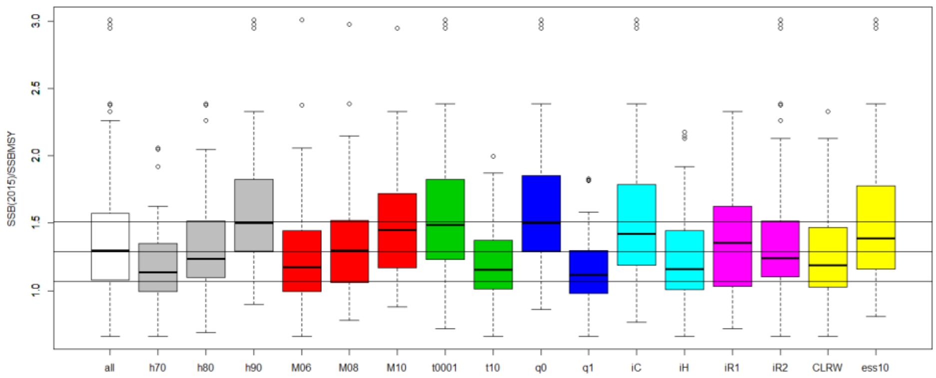


B19.5

Figure 22. Comparison of Recruitment Deviation magnitude among BET OM ensembles.

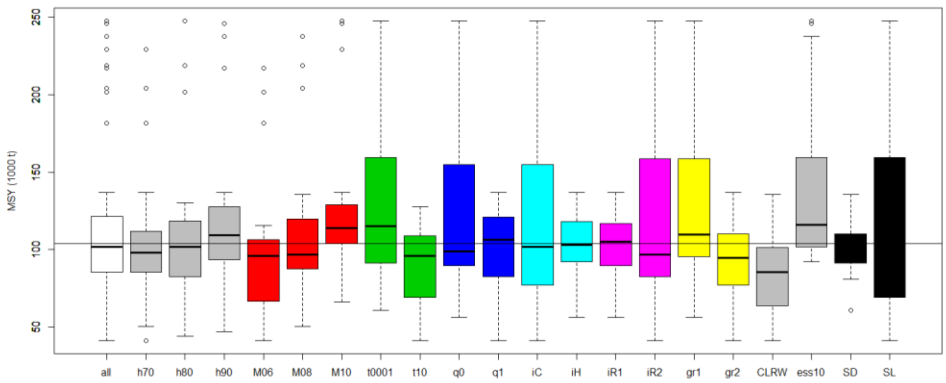


B19.0

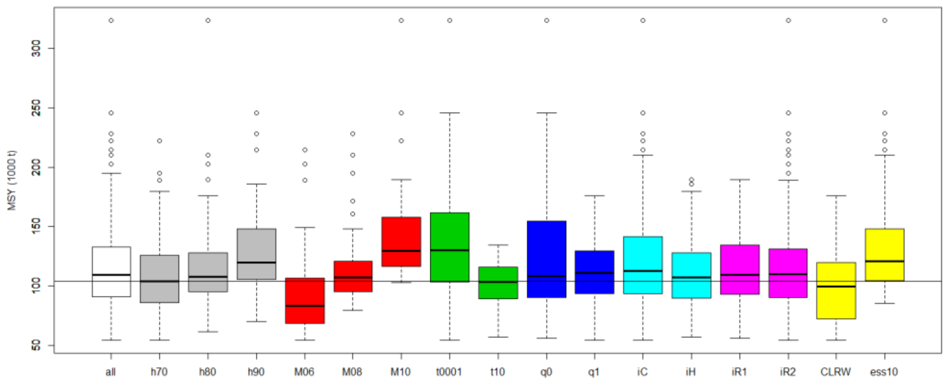


B19.5

Figure 23. Comparison of estimated current depletion among BET OM ensembles.

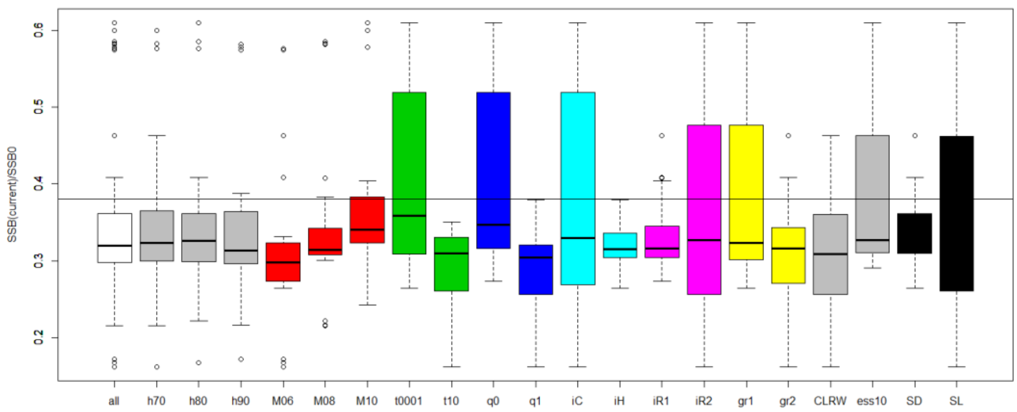


B19.0

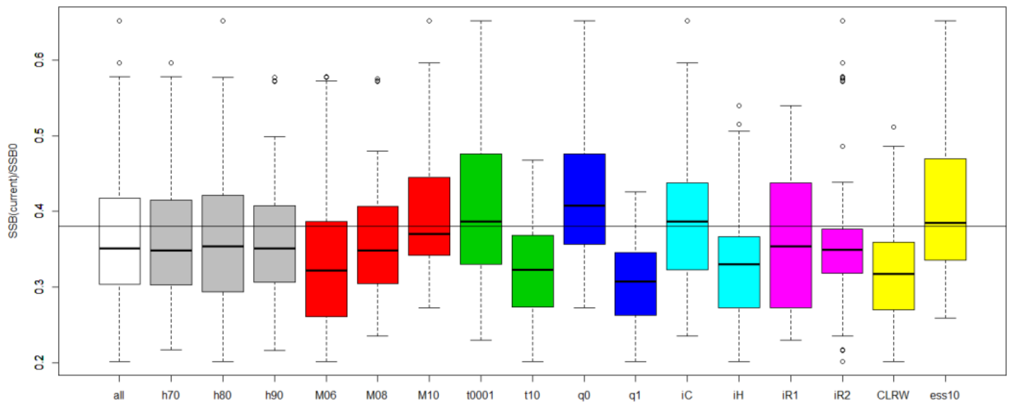


B19.5

Figure 24. Comparison of estimated MSY among BET OM ensembles.



B19.0



B19.5

Figure 25. Distribution of depletion estimates for OMgridB19.0 and OMgridB19.5

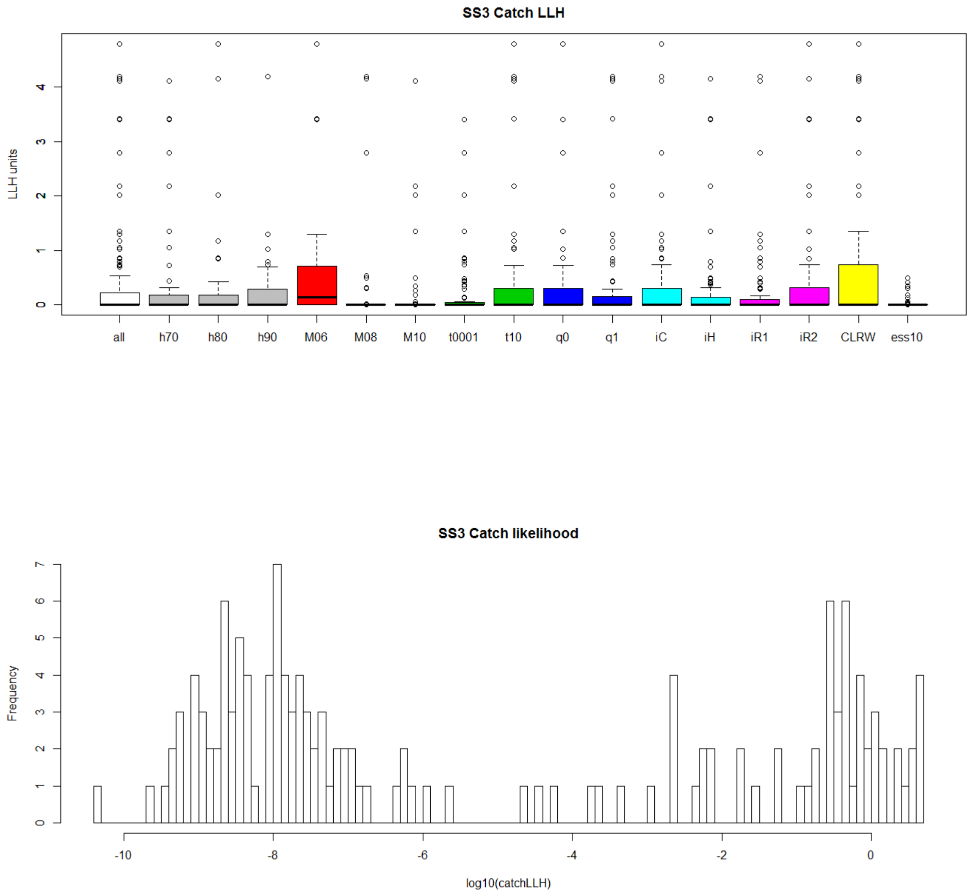


Figure 26. Distribution of catch likelihood terms for OMgridB19.5.

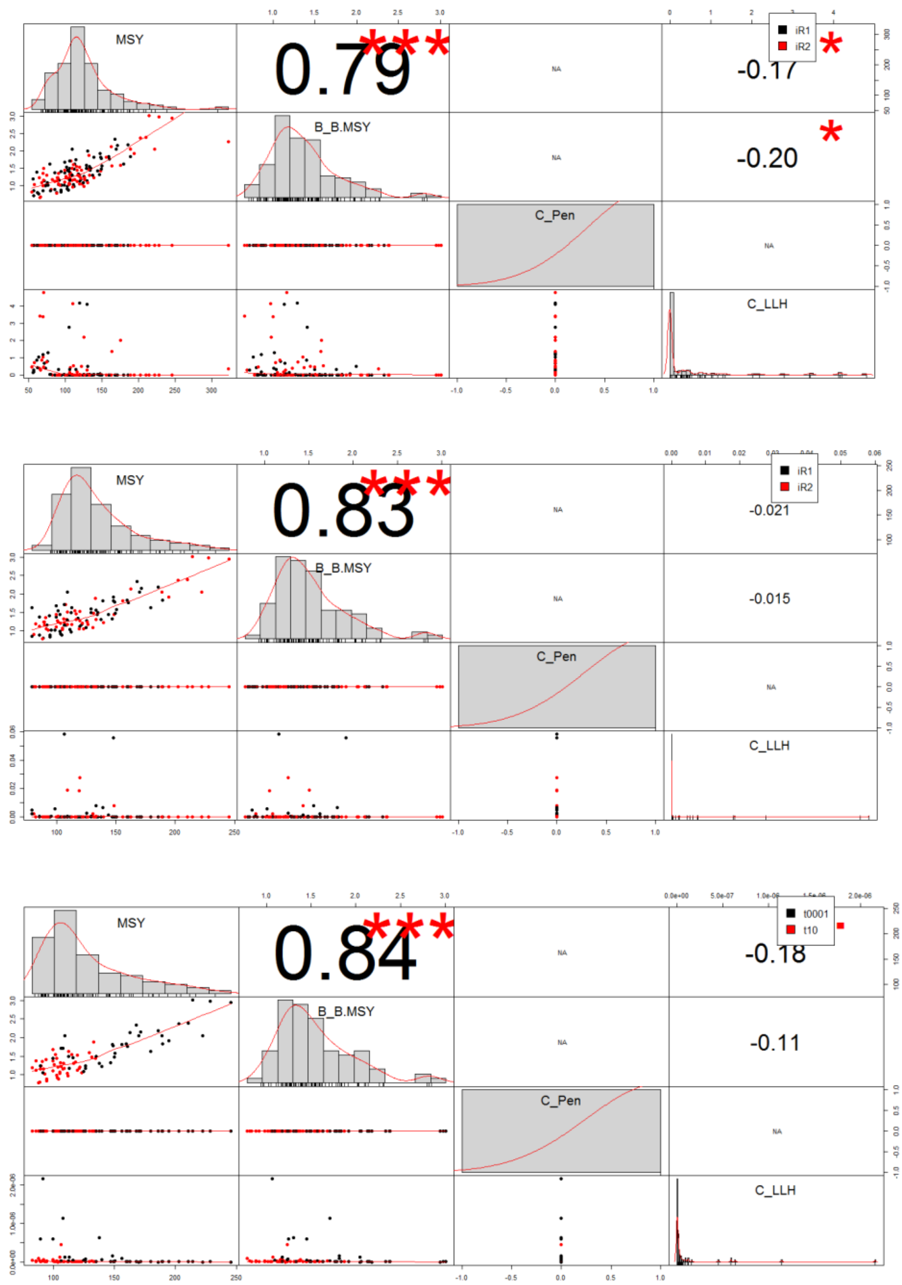


Figure 27. Multiway comparison of OMgridB19.5 model characteristics, comparing stock status and indices of numerical problems with catch calculations. Top panel is the full 144 model grid, middle panel is the subset of 104 models in which the catch penalty < 0.1, bottom panel is 84 models with catch penalty < 0.00001.

Revised Projection Assumptions

The reference case OM projection assumptions were updated as recommended by the WPTT and WPM in 2018 (attachment 1), including:

- MP-based management was set to start in 2021, and the bridging catches for the intervening years were updated from the WPTT 2018 figures (2017 catch).
- The annual aggregate CV = 0.2 and auto-correlation = 0.5 was used (it was identified as a mistake in the previous iteration, but was actually correct). At this time, the OM only outputs the aggregate annual CPUE for the MP to use. The annualized mean post-fit CPUE RMSE suggests that 0.2 is near the upper end of the variability observed among the OM ensemble. This observation error could be made model specific, but we would be reluctant to expect a level of precision higher than 0.2.

Figure 28 illustrates an example stochastic CPUE time series with the specified level of error (and the level of variability that would be observed in 4 independent quarterly series that would result in the annual aggregate characteristics). Further work may be warranted to determine if the MP needs to simulate independent series, e.g. to test the implications of missing observations in the future.

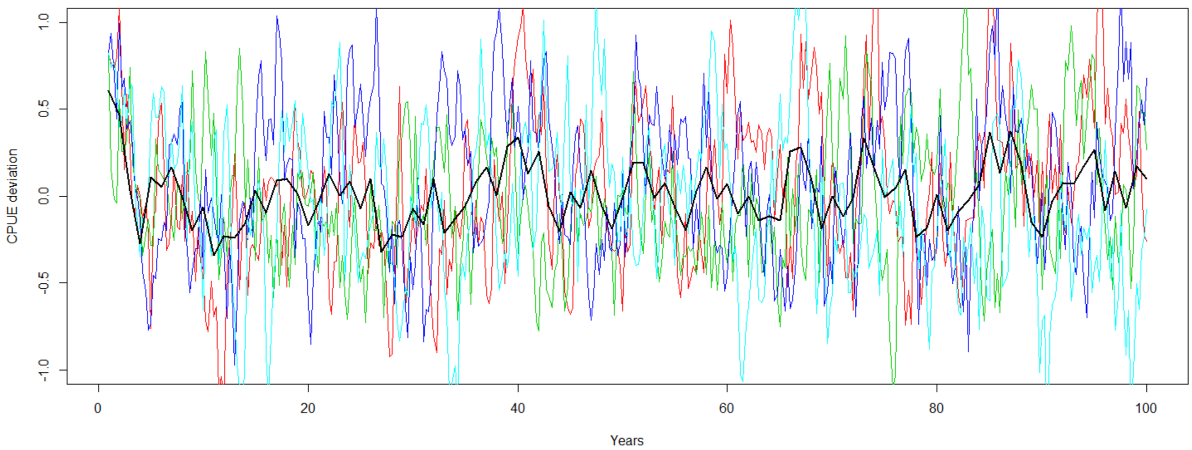


Figure 28. Simulated CPUE error time series yielding an annual aggregate (mean) CPUE series with a CV of 0.2 and auto-correlation of 0.5 (black line). Coloured lines represent 4 independent quarterly CPUE series errors that yield this aggregate characteristic.

Bigeye Reference case OMgrid19.5 MP evaluation results

The standard TCMP MP outputs are presented for 1 model-based and 1 CPUE-based MP for each of the 2018 tuning objectives in Figure 29 - Figure 41 (and Table 3 - Table 4), from which we note:

- MP performance is qualitatively similar to previous iterations. There is a high degree of performance variability for any given MP, but the median performance suggests that the stock should remain in a comfortable area, with catches remaining around recent levels.
- The most aggressive tuning objective will ramp up catches and tend to push SB down in the long term, while the more conservative tuning objectives will keep catch and SB very stable on average.
- We do not identify any obvious need to provide more performance contrast through additional tuning objectives for the 2019 TCMP. However, there is scope for improving the example MPs. Notably the data-based MPs exhibit some undesirable features in the long term (tendency to keep increasing catches followed by a population crash).

Figure 36 - Figure 46 show MP evaluation summary statistics partitioned by conditioning assumption. Qualitatively, it appears that the relative importance of the factors examined is similar in terms of the MP performance and OM conditioning results (Figure 20 - Figure 25).

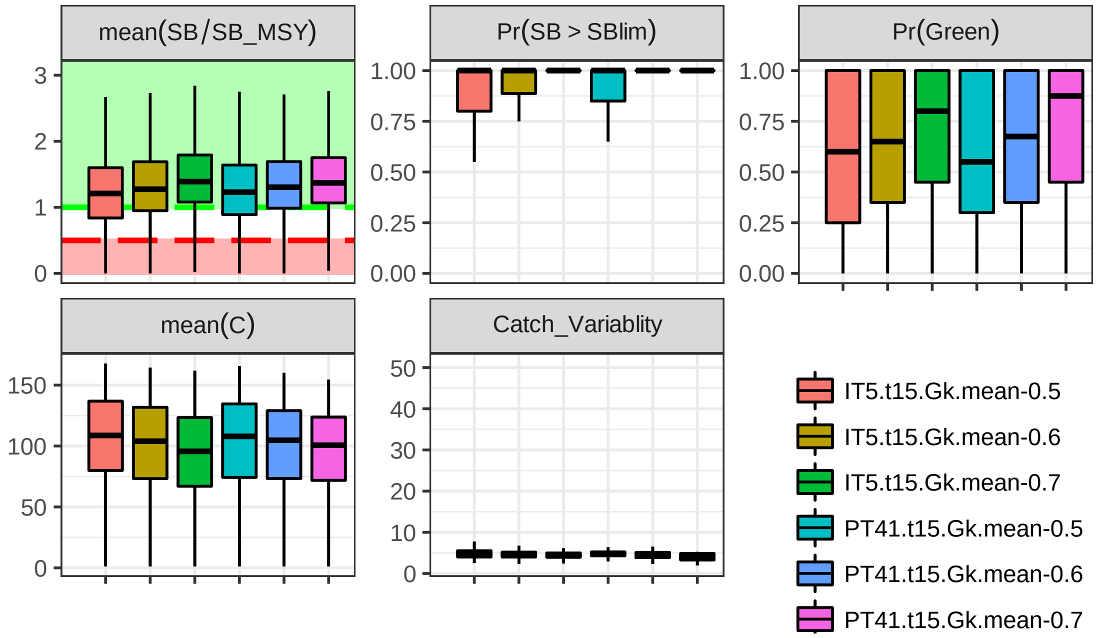


Figure 29. Bigeye reference case (B19.5). Boxplots comparing candidate MPs with respect to key performance measures averaged over the period 2019 - 2038. Horizontal line is the median, boxes represent 25th - 75th percentiles, thin lines represent 10th - 90th percentiles. Red and green horizontal lines represent the interim limit and target reference points for the mean SB/SBMSY performance measure. The horizontal dashed black line is 2016 catch.

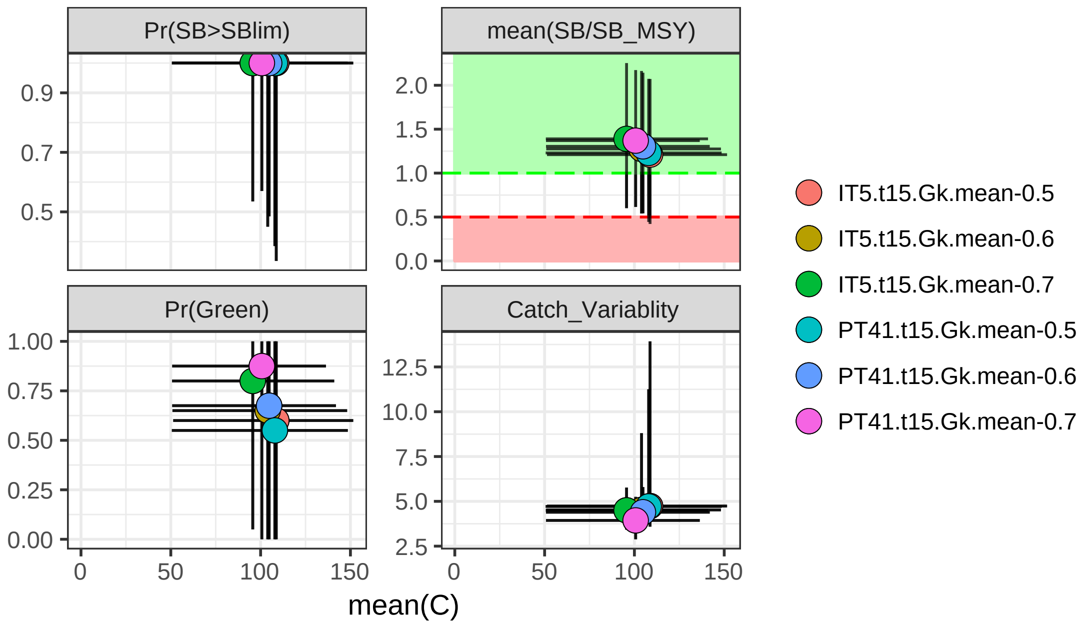


Figure 30. Bigeye reference case (B19.5). Trade-off plots comparing candidate MPs with respect to catch on the X-axis, and 4 other key performance measures on the Y-axis, each averaged over the period 2019 - 2038. Circle is the median, lines represent 10th-90th percentiles. Red and green horizontal lines represent the interim limit and target reference points for the mean SB/SBMSY performance measure. The dashed vertical black line is 2016 catch.

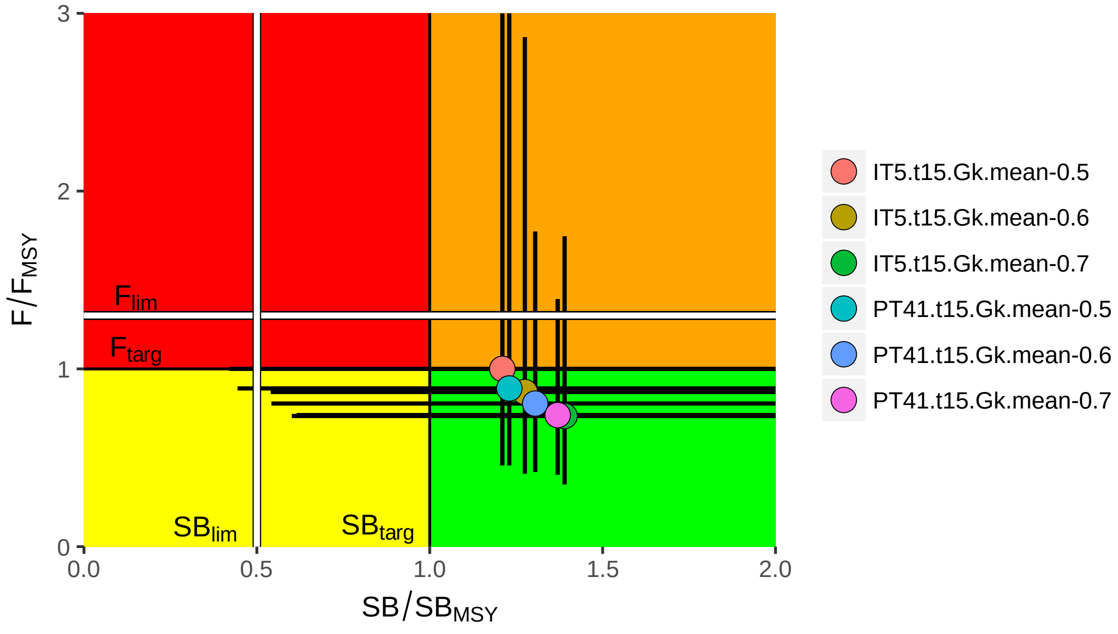


Figure 31. Bigeye reference case (B19.5). Kobe plot comparing candidate MPs on the basis of the expected 20 year average (2019-2038) performance. Circle is the median, lines represent 10th-90th percentiles.

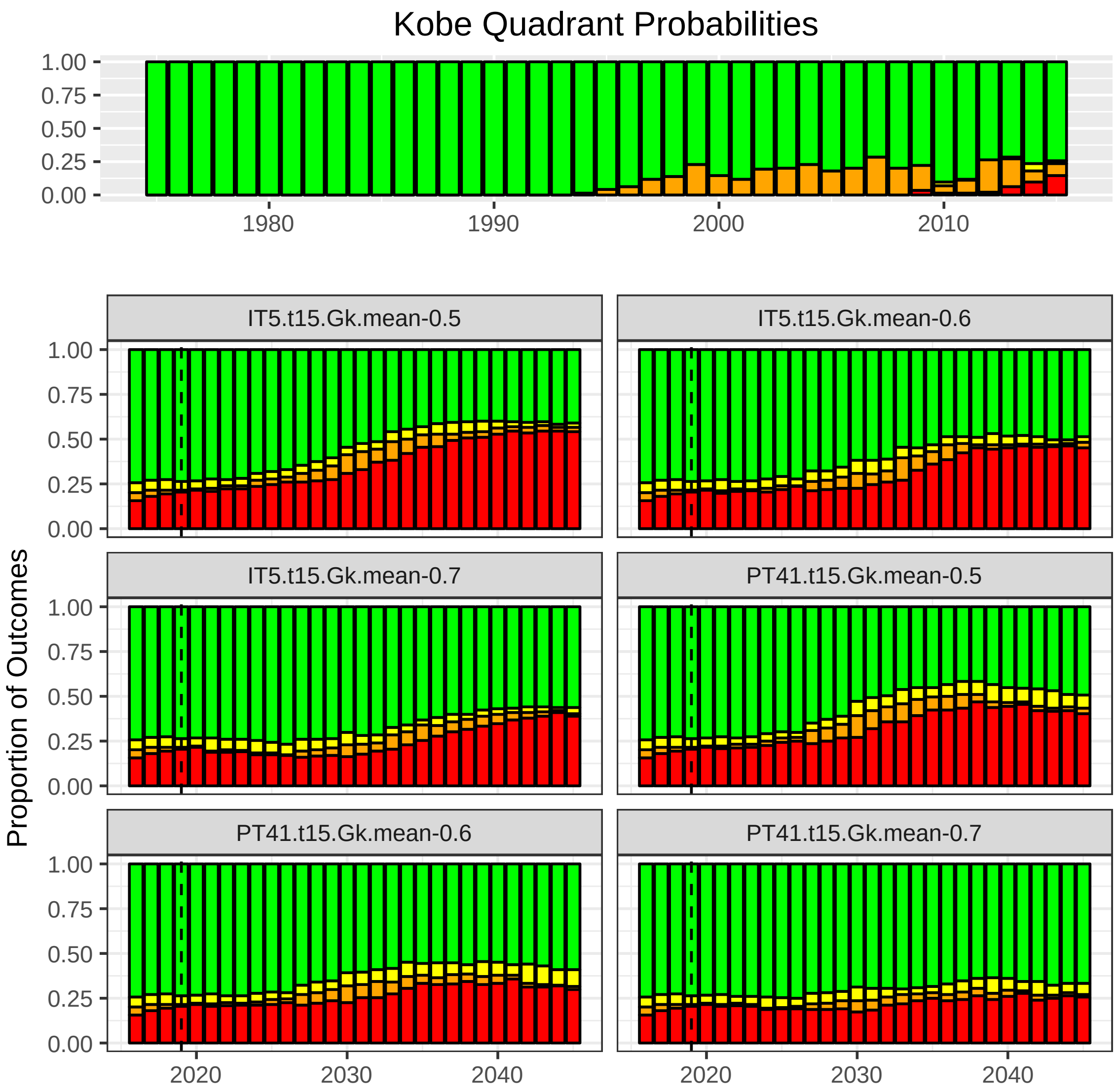


Figure 32. Bigeye reference case (B19.5). Proportion of simulations in each of the Kobe quadrants over time for each of the candidate MPs. Historical estimates are included in the top panel. The lower panels are projections, with the first MP application indicated by the broken vertical line (2019).

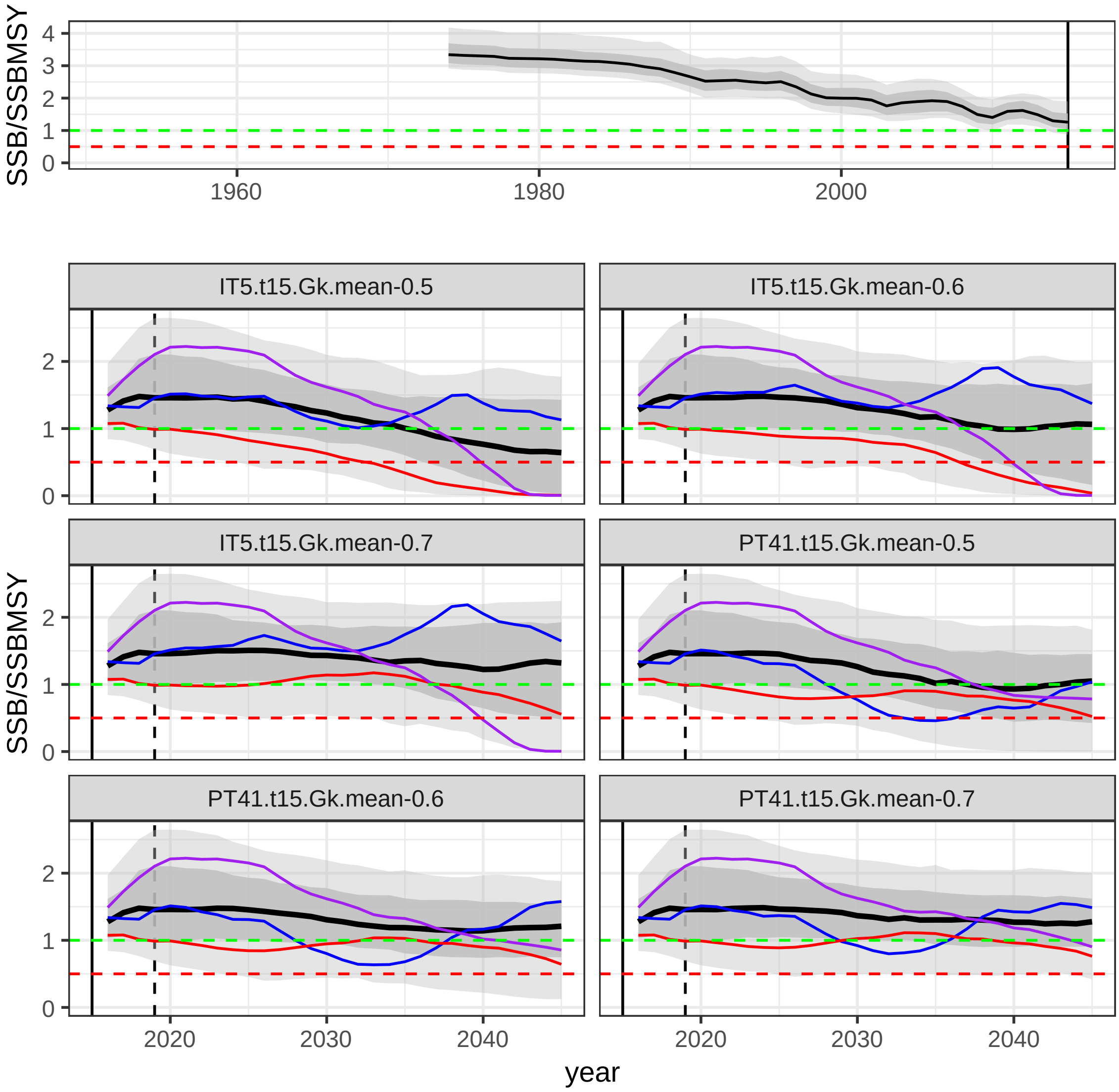


Figure 33. Bigeye reference case (B19.5). Time series of spawning stock size for the candidate MPs. The top panel represents the historical estimates from the reference case operating model, and lower plots represent the projection period. The solid vertical line represents the last year used in the historical conditioning. The broken vertical line represents the first year that the MP is applied. The median is represented by the bold black line, the dark shaded ribbon represents the 25th-75th percentiles, the light shaded ribbon represents the 10th-90th percentiles. Thick broken lines represent the interim target (green) and limit (red) reference points. The 3 thin coloured lines represent examples of individual realizations (the same OM scenarios across MPs and performance measures), to illustrate that individual variability greatly exceeds the median.

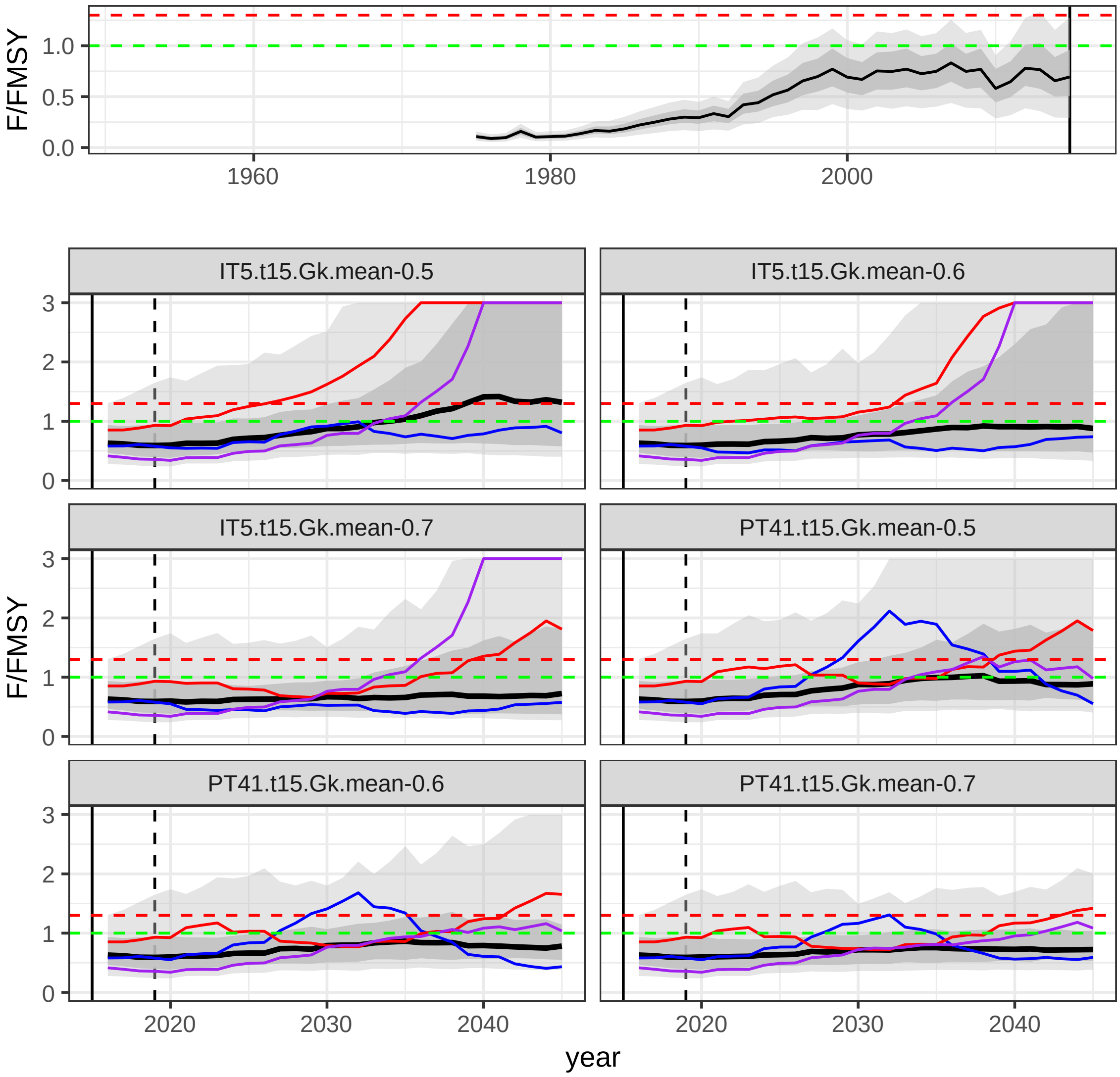


Figure 34. Bigeye reference case (B19.5). Time series of fishing intensity (Upper bound truncated at F = 3) for the candidate MPs. The top panel represents the historical estimates from the reference case operating model, and lower plots represent the projection period. The solid vertical line represents the last year used in the historical conditioning. The broken vertical line represents the first year that the MP is applied. The median is represented by the bold black line, the dark shaded ribbon represents the 25th-75th percentiles, the light shaded ribbon represents the 10th-90th percentiles. Thick broken lines represent the interim target (green) and limit (red) reference points. The 3 thin coloured lines represent examples of individual realizations (the same OM scenarios across MPs and performance measures), to illustrate that individual variability greatly exceeds the median.

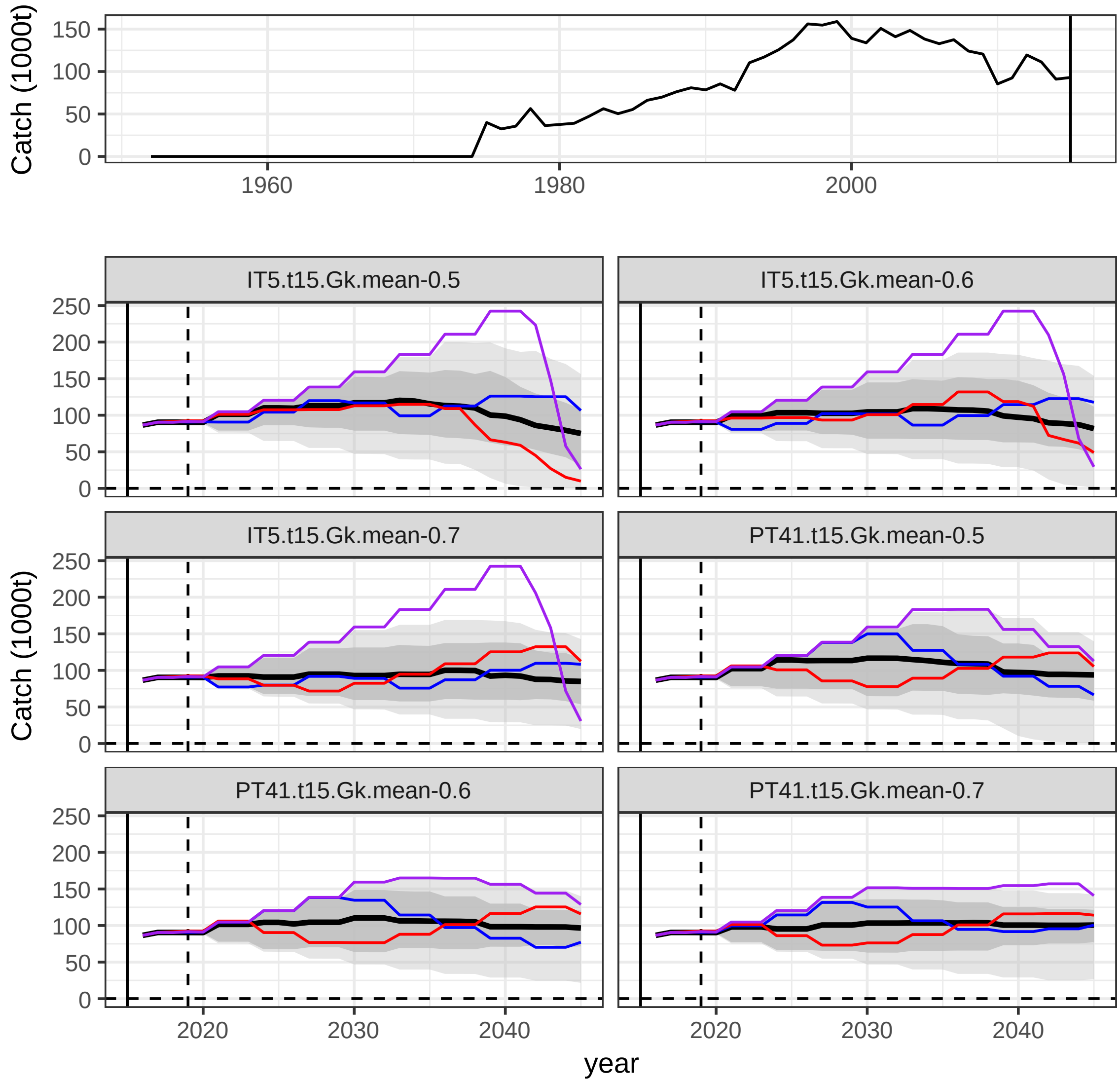


Figure 35. Bigeye reference case (B19.5). Time series of catch for the candidate MPs. The top panel represents the historical estimates from the reference case operating model, and lower plots represent the projection period. The solid vertical line represents the last year used in the historical conditioning. The broken vertical line represents the first year that the MP is applied. The median is represented by the bold black line, the dark shaded ribbon represents the 25th-75th percentiles, the light shaded ribbon represents the 10th-90th percentiles. The broken black horizontal line represents recent (2016) catch. The 3 thin coloured lines represent examples of individual realizations (the same OM scenarios across MPs and performance measures), to illustrate that individual variability greatly exceeds the median.

Table 3. Bigeye Reference case (B19.5). Performance of candidate MPs with respect to key performance measures (averaged over the period 2019-2038) for grid B19.0. Shading indicates the relative performance (darker = better).

	Performance Measure				
Management Procedure	SB/SB_{MSY}	Prob(Green)	Prob(SB>limit)	Mean Catch	Catch Variability
PT41.t15.Gk.mean-0.5	1.23 (0.89-1.64)	0.57	0.85	107.9 (74.3-134.6)	4.72
PT41.t15.Gk.mean-0.6	1.31 (0.99-1.69)	0.64	0.89	104.8 (73.4-129.0)	4.39
PT41.t15.Gk.mean-0.7	1.37 (1.07-1.75)	0.71	0.90	100.7 (71.9-123.7)	3.94
IT5.t15.Gk.mean-0.5	1.21 (0.84-1.60)	0.56	0.82	108.7 (80.0-136.8)	4.74
IT5.t15.Gk.mean-0.6	1.27 (0.95-1.69)	0.62	0.86	104.0 (73.3-131.8)	4.52
IT5.t15.Gk.mean-0.7	1.39 (1.08-1.79)	0.70	0.89	95.6 (67.0-123.3)	4.50

Table 4. Bigeye reference case (B19.5). Candidate MP performance for standard IOTC performance measures for the 20 year period 2019-2038

Status : maximise stock status		**20 year average**					
		PT41.t15.Gk. mean-0.5	**PT41.t15.Gk. mean-0.6**	**PT41.t15.Gk. mean-0.7**	**IT5.t15.Gk. mean-0.5**	**IT5.t15.Gk. mean-0.6**	**IT5.t15.Gk. mean-0.7**
Mean spawner biomass relative to pristine	SB/SB_0	0.34	0.36	0.38	0.33	0.35	0.39
Minimum spawner biomass relative to pristine	SB/SB_0	0.21	0.25	0.27	0.16	0.20	0.26
Mean spawner biomass relative to SBMSY	SB/SB_{MSY}	1.23	1.31	1.37	1.21	1.27	1.39
Mean fishing mortality relative to FMSY	F/F_{tar}	0.89	0.80	0.74	1.00	0.87	0.73
Mean fishing mortality relative to target	F/F_{MSY}	0.89	0.80	0.74	1.00	0.87	0.73
Probability of being in Kobe green quadrant	SB,F	0.57	0.64	0.71	0.56	0.62	0.70
Probability of being in Kobe red quadrant	SB,F	0.32	0.26	0.21	0.35	0.29	0.22
Safety : maximise the probability of remaining above low stock status (i.e. minimise risk)							
Probability of spawner biomass being above 20% of SB0	SB	0.77	0.82	0.85	0.74	0.79	0.83
Probability of spawner biomass being above BLim	SB	0.85	0.89	0.90	0.82	0.86	0.89
Yield : maximise catches across regions and gears							
Mean catch (1000 t)	C	107.93	104.75	100.69	108.70	103.97	95.63
Mean relative CPUE (aggregate)	C	0.93	0.90	0.86	0.94	0.88	0.80
Mean catch relative to MSY	C/MSY	0.80	0.84	0.89	0.77	0.83	0.90
Stability: maximise stability in catches to reduce commercial uncertainty							
Mean absolute proportional change in catch	C	4.72	4.39	3.94	4.74	4.52	4.50
% Catch coefficient of variation	C	0.18	0.16	0.15	0.19	0.17	0.16
Probability of shutdown	C	0.05	0.04	0.04	0.04	0.04	0.04

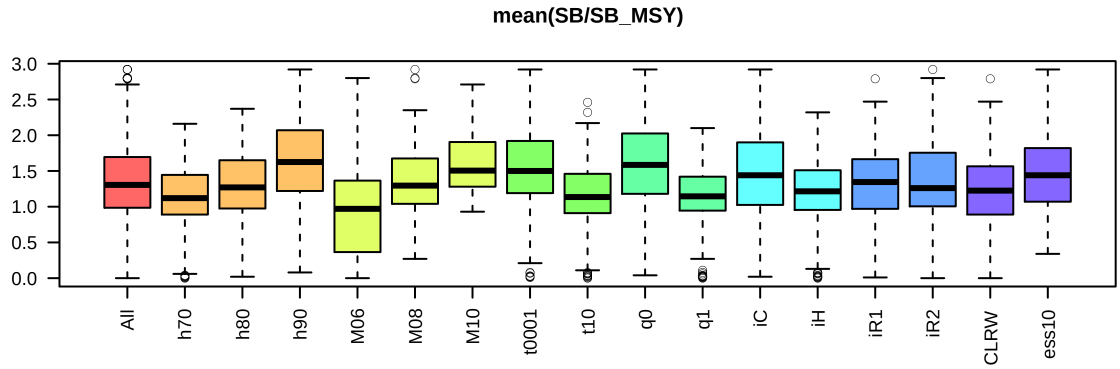


Figure 36. Bigeye reference case (B19.5). Mean SSB / SSBMSY estimates, partitioned by assumptions (all models are encompassed within an individual colour set).

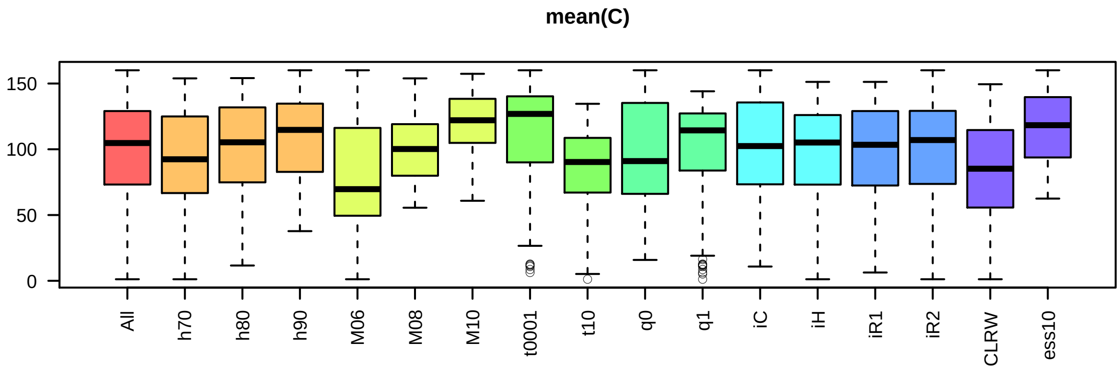


Figure 37. Bigeye reference case (B19.5). Mean Catch estimates, partitioned by assumptions (all models are encompassed within an individual colour set).

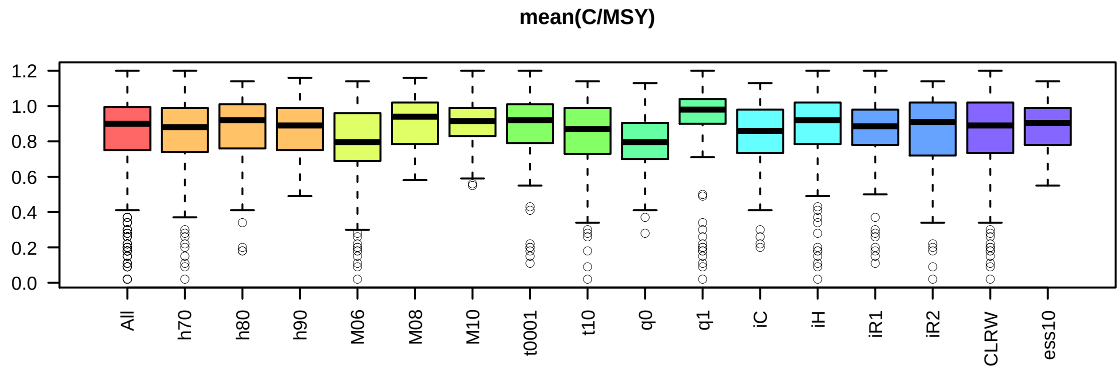


Figure 38. Bigeye reference case (B19.54). Mean Catch/MSY estimates, partitioned by assumptions (all models are encompassed within an individual colour set).

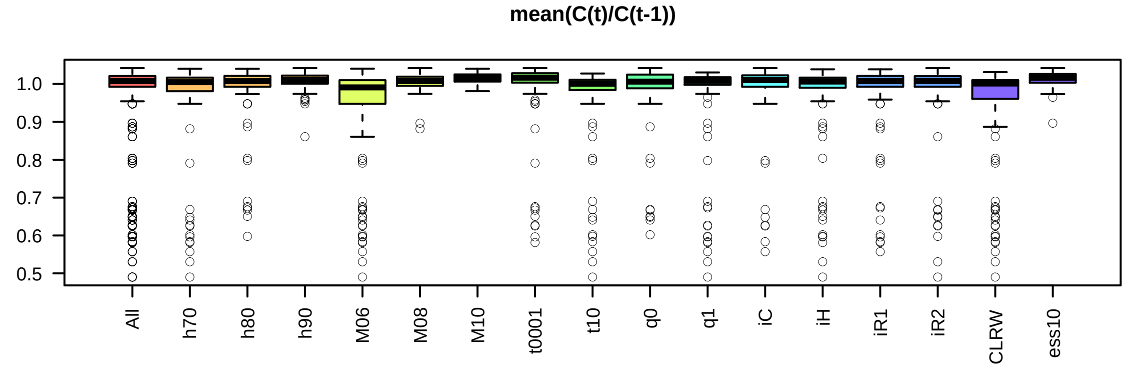


Figure 39. Bigeye reference case (B19.5). C(t)/C(t-1) estimates, partitioned by assumptions (all models are encompassed within an individual colour set).

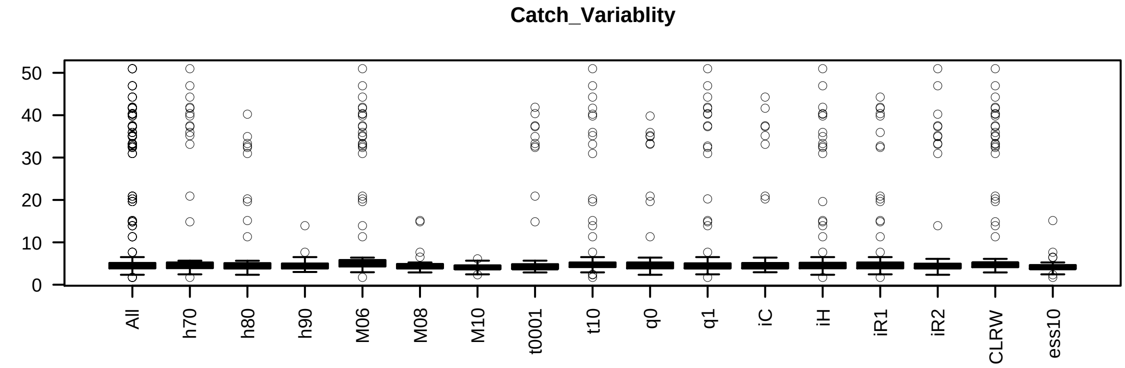


Figure 40. Bigeye reference case (B19.5). Catch Variability estimates, partitioned by assumptions (all models are encompassed within an individual colour set).

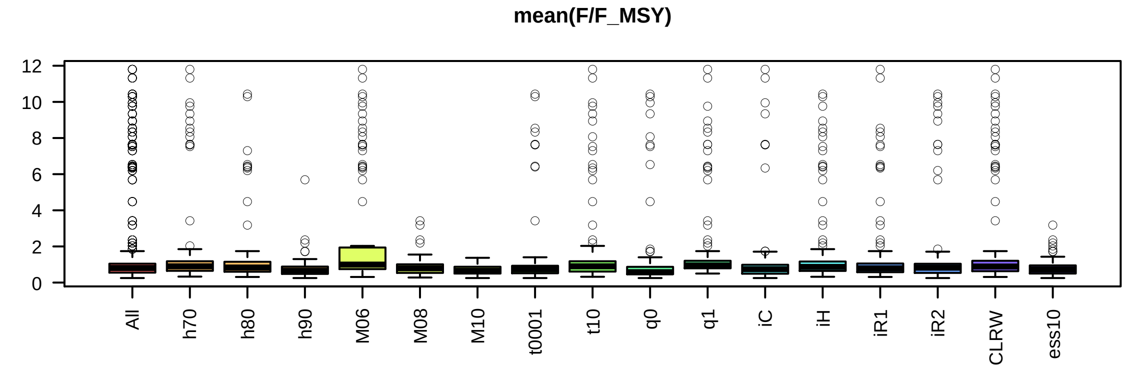


Figure 41. Bigeye reference case (B19.5). F/FMSY estimates, partitioned by assumptions (all models are encompassed within an individual colour set).

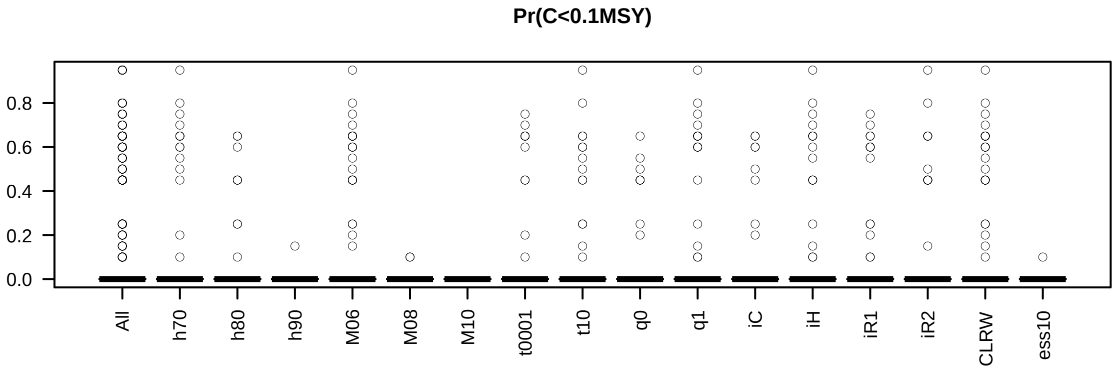


Figure 42. Bigeye reference case (B19.5). Probability Catch < 0.1MSY estimates, partitioned by assumptions (all models are encompassed within an individual colour set).

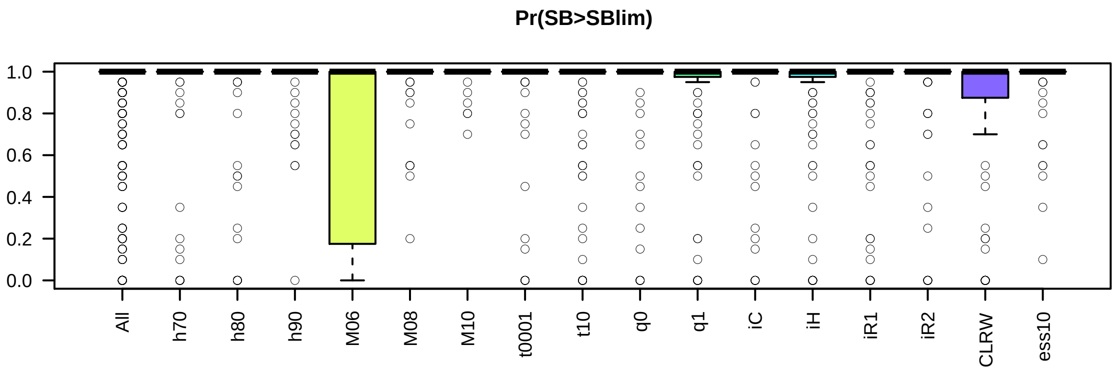


Figure 43. Bigeye reference case (B19.5). Probability SB > SBlim estimates, partitioned by assumptions (all models are encompassed within an individual colour set).

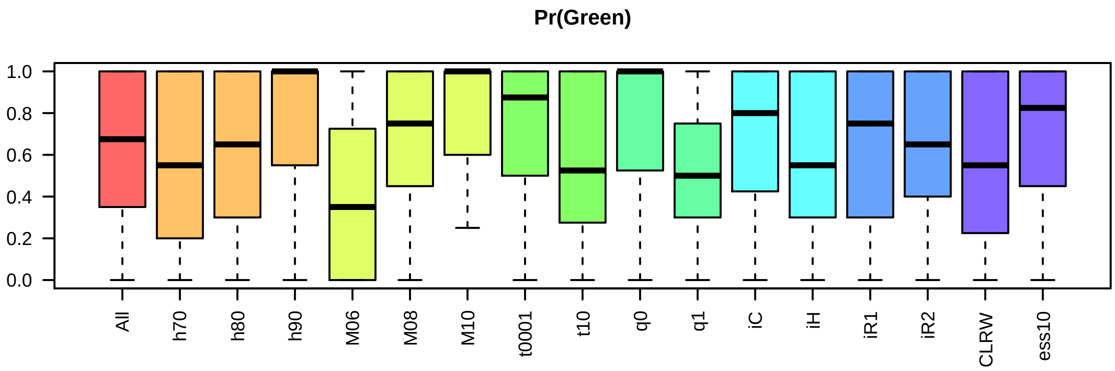


Figure 44. Bigeye reference case (B19.5). Probability in Green Kobe estimates, partitioned by assumptions (all models are encompassed within an individual colour set).

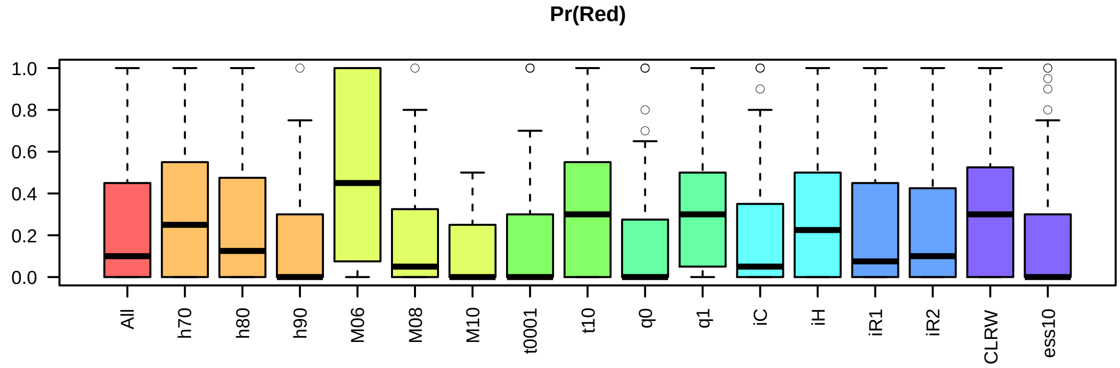


Figure 45. Bigeye reference case (B19.5). Probability in Red Kobe estimates, partitioned by assumptions (all models are encompassed within an individual colour set).

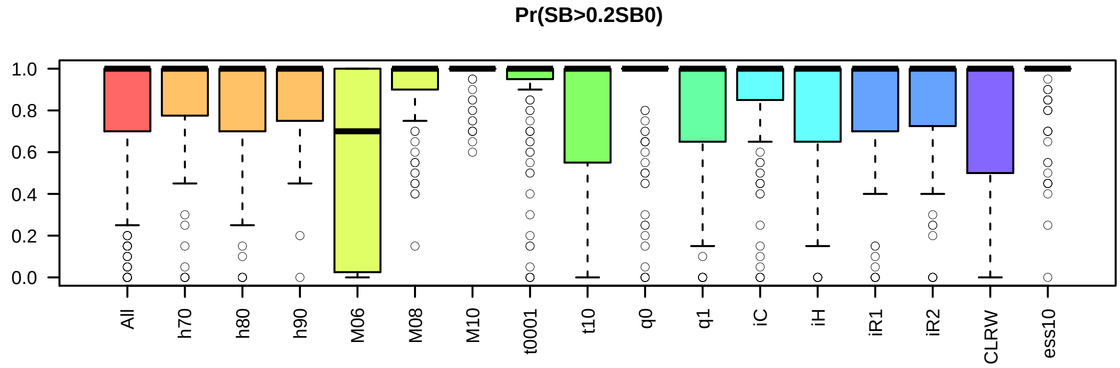


Figure 46. Bigeye reference case (B19.5). Probability SSB > 0.2 SSB0 estimates, partitioned by assumptions (all models are encompassed within an individual colour set).

Bigeye MP Robustness tests

Aside from the new uncertainty dimensions discussed in the reference case above, seven additional robustness scenarios were proposed in the WPTT/WPM 2018. The first four were addressed by changing the OM projection assumptions (Figure 47 - Figure 68), and provided results that are qualitatively consistent with expectations. Notably a 3% per year catchability trend in CPUE is eventually going to cause a stock collapse. But this is not a very insightful test. If this is considered a plausible scenario, the Commission needs to seriously consider collecting new and more informative assessment data. The last three scenarios were not addressed for reasons discussed below.

1. Annual aggregated CPUE CV = 0.3 (auto-correlation = 0.5)
2. 10% reported over-catch (projections only; reference case conditioning)
3. 10% unreported over-catch (projections only; reference case conditioning)
4. 3% LL catchability trend (projections only; reference case conditioning)

5. Spatial Structure- possibly additional area around eastern INDONESIA, another in the Bay of Bengal Region and the area around Oman (other area stratification as is).

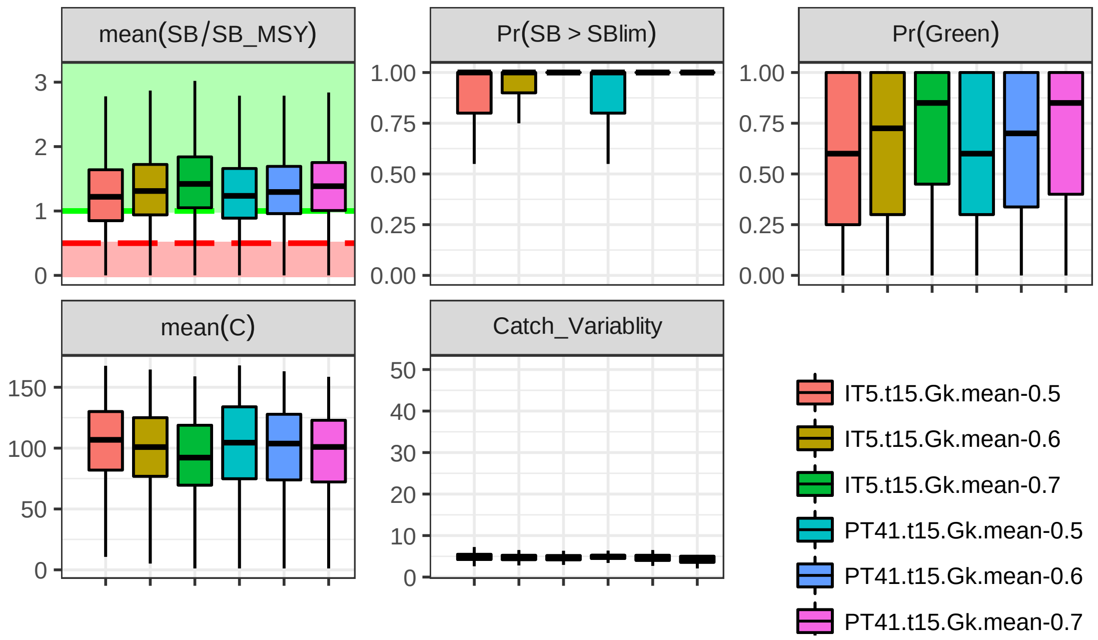
    - This is a major restructuring of model assumptions that would require considerable data processing. The request would require a more explicit definition, and an explanation of the rationale seems warranted. Given that the assessment already appears to be over-parameterized, with unstable parameter estimation, adding more complexity with areas that probably cannot be reliably linked with tags or longline CPUE series does not seem like a fruitful avenue for exploration.

6. Non stationary M, linf and K in the projections.

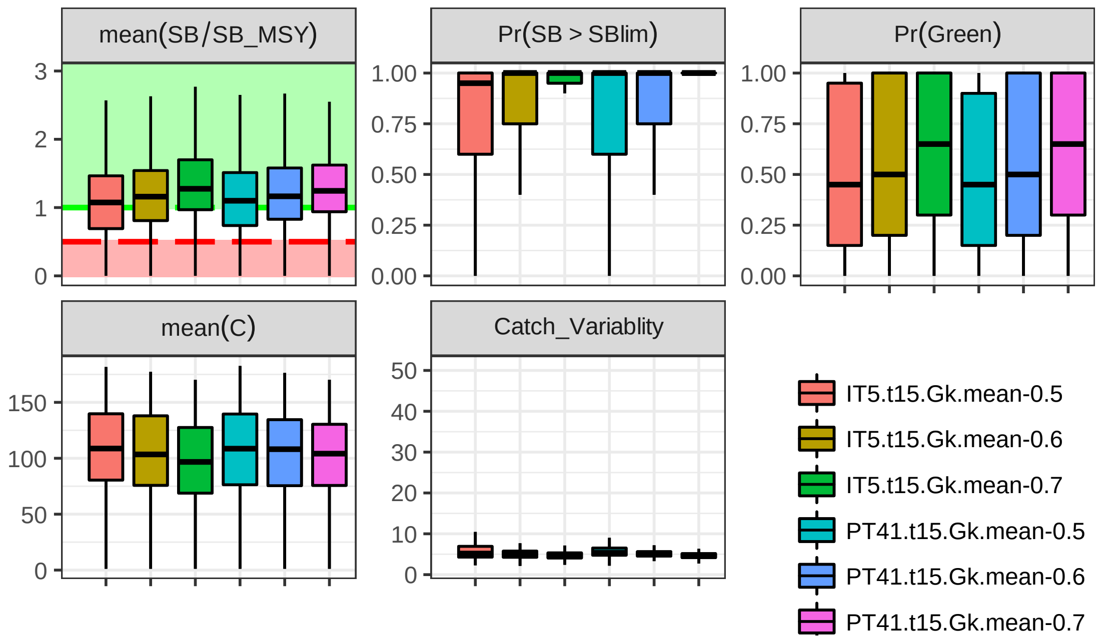
    - While untested to date, the OM includes the functionality to add some or all of these features. However, before venturing down this route, we would like the proponents to provide specific requests (e.g. magnitude of change, time series structure) ideally with some empirical justification.

7. Stock Structure (based on ongoing IO stock structure project).

    - While untested to date, the OM was originally designed to include this functionality. However, this adds considerable complexity in terms of conditioning assumptions and the provision of summary statistics etc. This was deferred as a longer term goal, ideally informed by the ongoing Indian Ocean stock structure project.

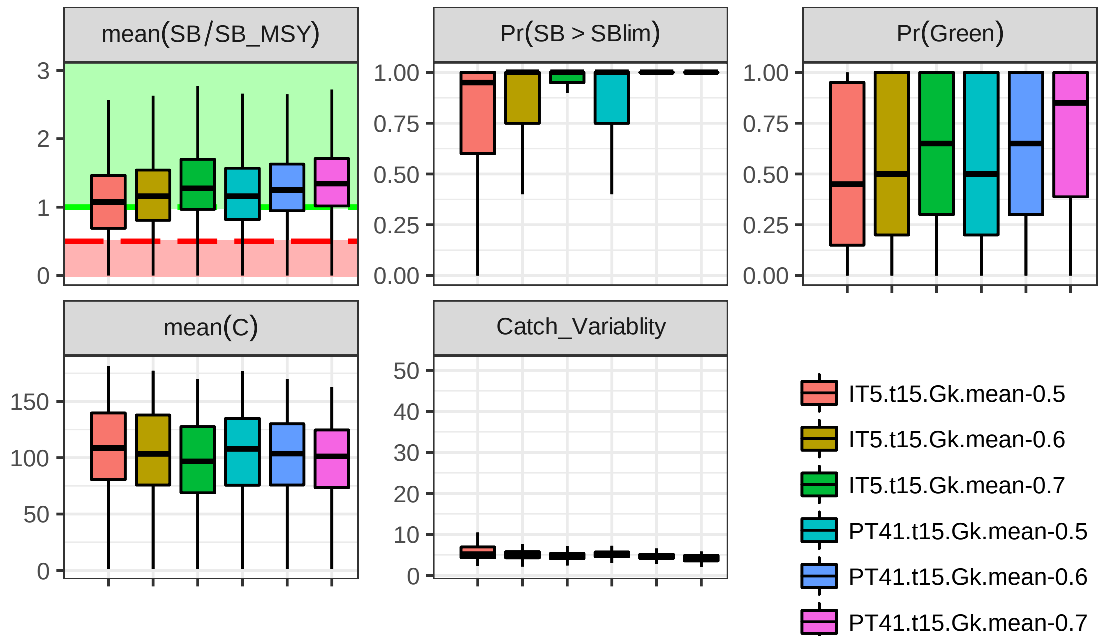


a)

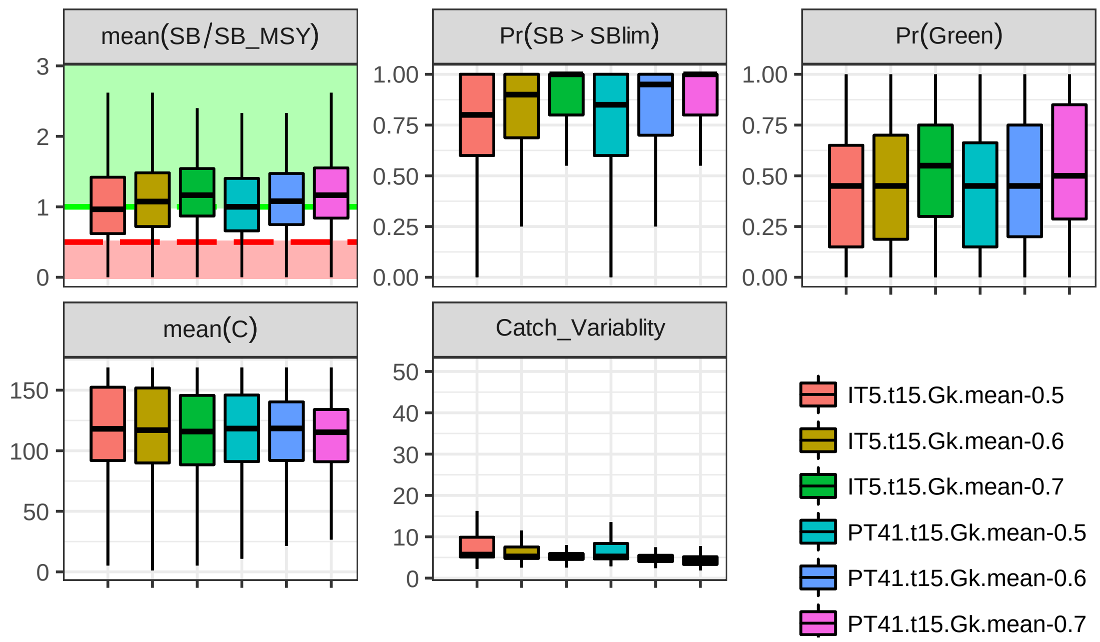


b)

Figure 47. Bigeye robustness cases (B19.5) : a) annual aggregate CPUE CV=0.3 and auto-correlation=0.5; b) 10% reported over-catch. Boxplots comparing candidate MPs with respect to key performance measures averaged over the period 2019 - 2038. Horizontal line is the median, boxes represent 25th - 75th percentiles, thin lines represent 10th - 90th percentiles. Red and green horizontal lines represent the interim limit and target reference points for the mean SB/SBMSY performance measure. The horizontal dashed black line is 2016 catch.

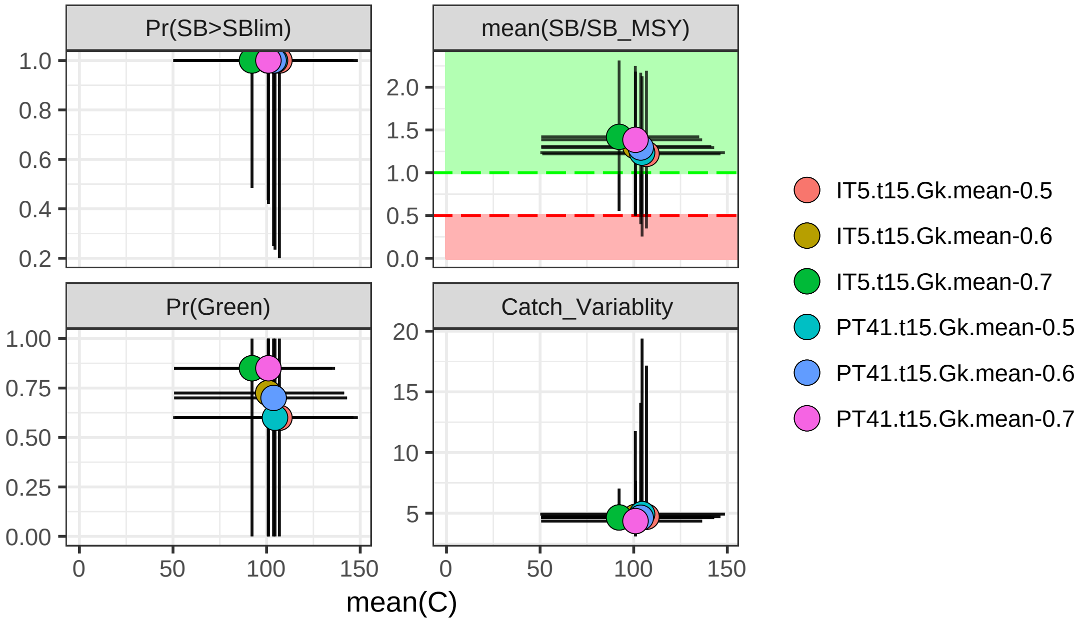


c)

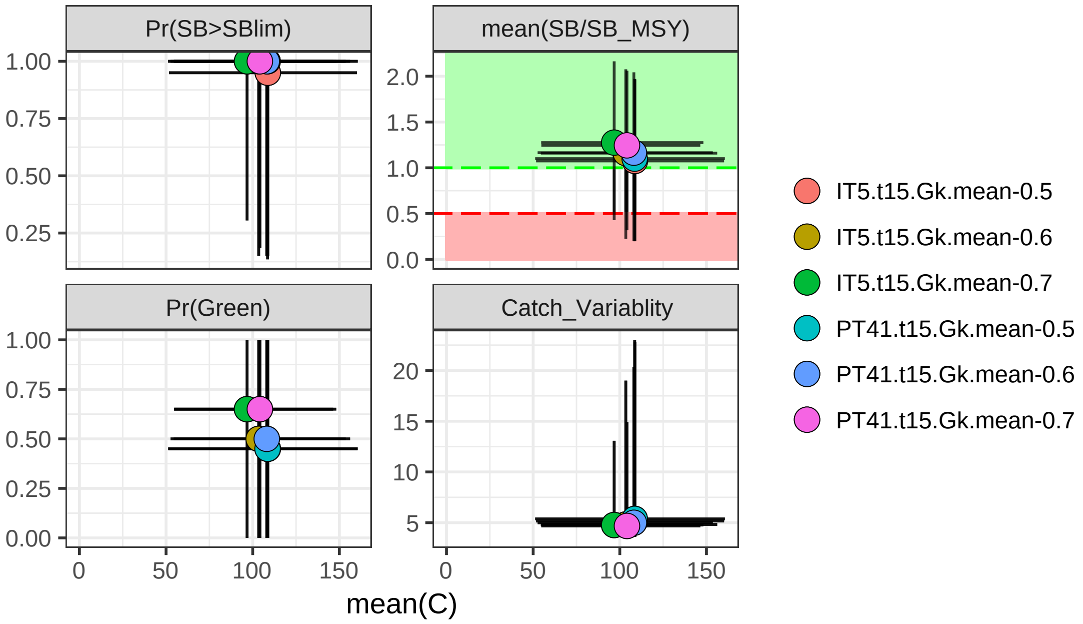


d)

Figure 48. Bigeye robustness cases (B19.5) : c) 10% unreported over-catch; d) 3% p.a. longline catchability trend. Boxplots comparing candidate MPs with respect to key performance measures averaged over the period 2019 - 2038. Horizontal line is the median, boxes represent 25th - 75th percentiles, thin lines represent 10th - 90th percentiles. Red and green horizontal lines represent the interim limit and target reference points for the mean SB/SBMSY performance measure. The horizontal dashed black line is 2016 catch.

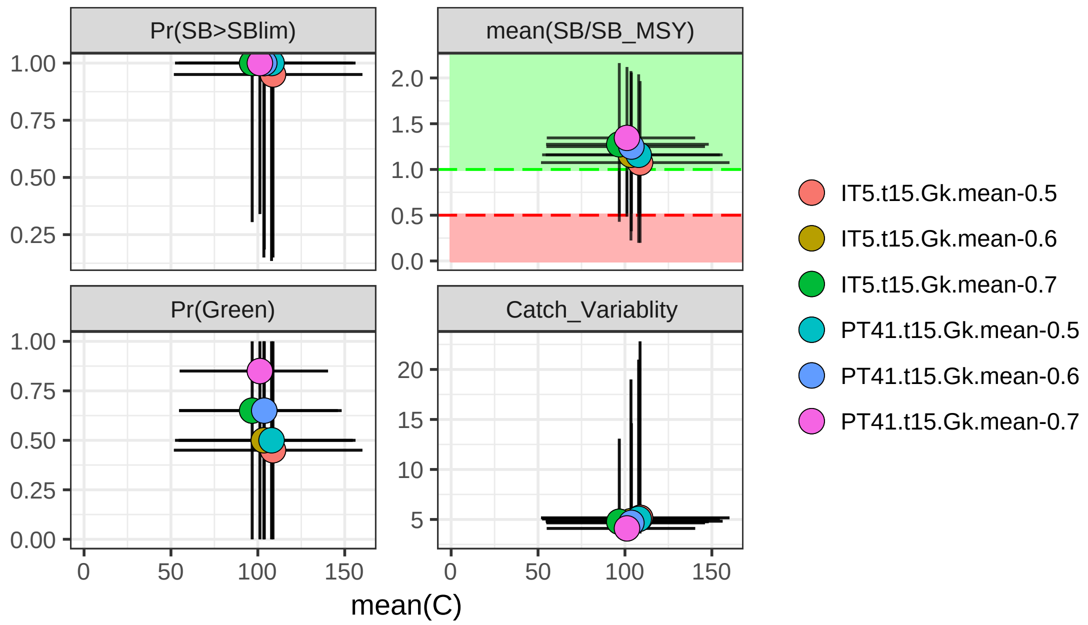


a)

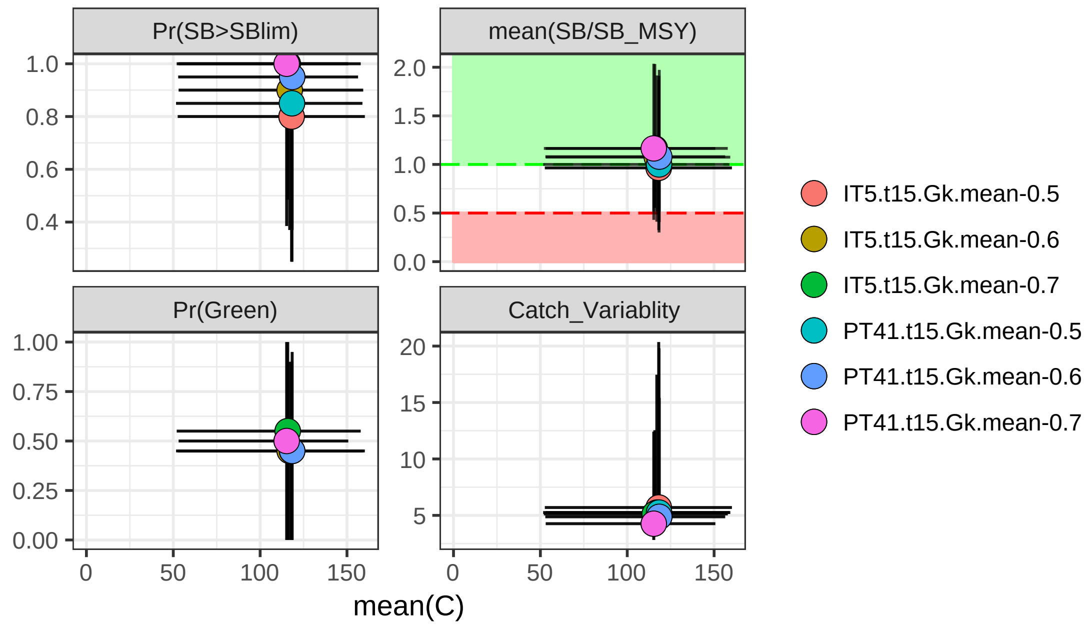


b)

Figure 49. Bigeye robustness cases (B19.5) : a) annual aggregate CPUE CV=0.3 and auto-correlation=0.5; b) 10% reported over-catch. Trade-off plots comparing candidate MPs with respect to catch on the X-axis, and 4 other key performance measures on the Y-axis, each averaged over the period 2019 - 2038. Circle is the median, lines represent 10th-90th percentiles. Red and green horizontal lines represent the interim limit and target reference points for the mean SB/SBMSY performance measure. The dashed vertical black line is 2016 catch.

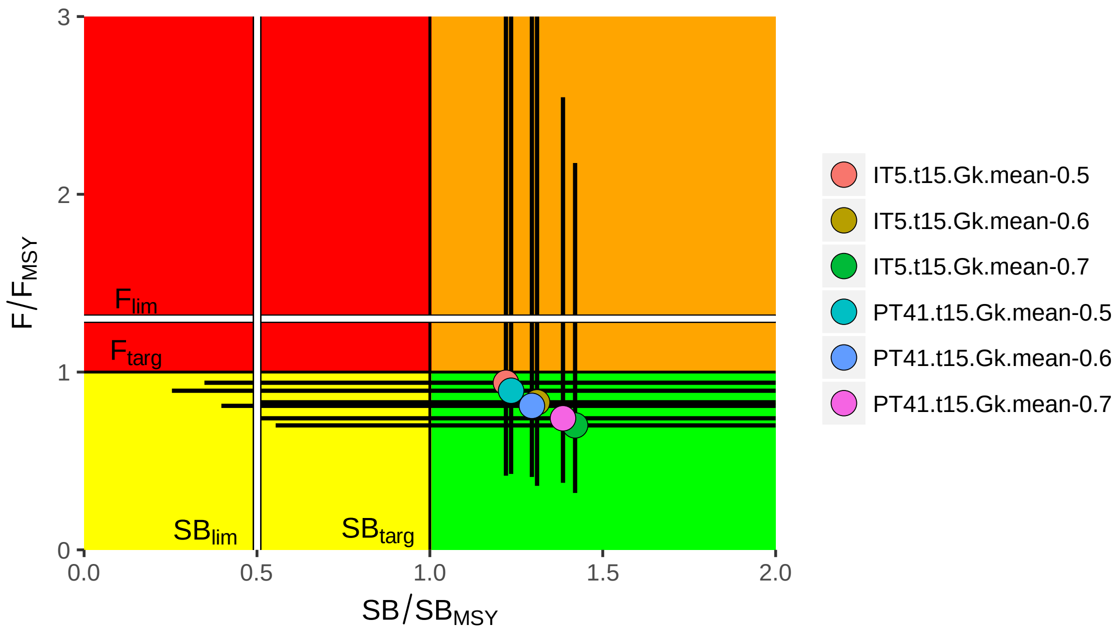


c)

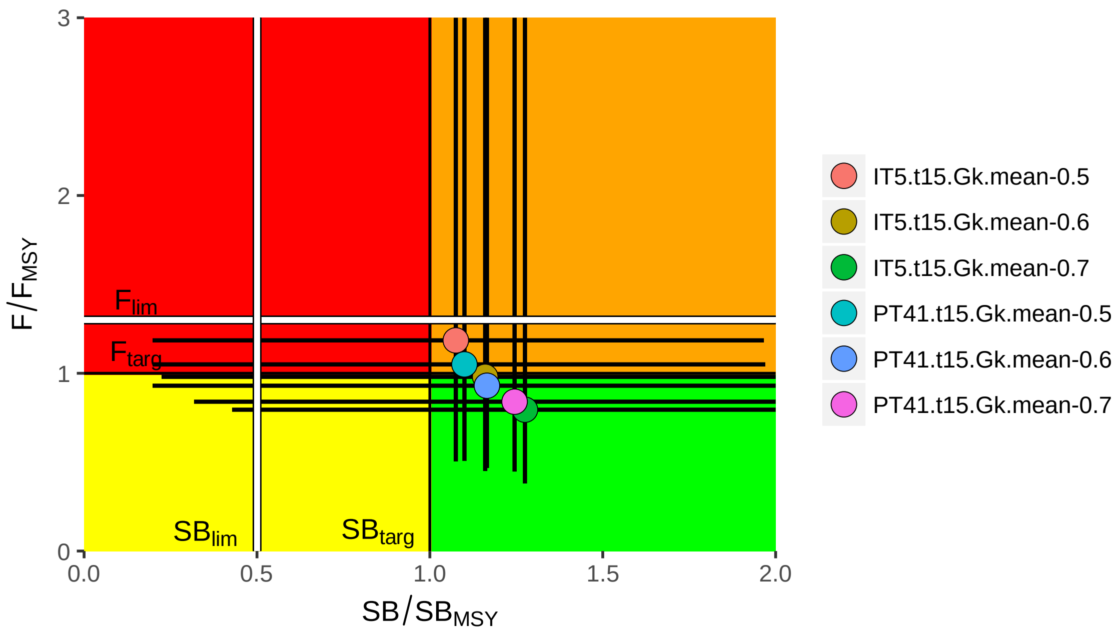


d)

Figure 50. Bigeye robustness cases (B19.5) : c) 10% unreported over-catch; d) 3% p.a. longline catchability trend. Trade-off plots comparing candidate MPs with respect to catch on the X-axis, and 4 other key performance measures on the Y-axis, each averaged over the period 2019 - 2038. Circle is the median, lines represent 10th-90th percentiles. Red and green horizontal lines represent the interim limit and target reference points for the mean SB/SBMSY performance measure. The dashed vertical black line is 2016 catch.

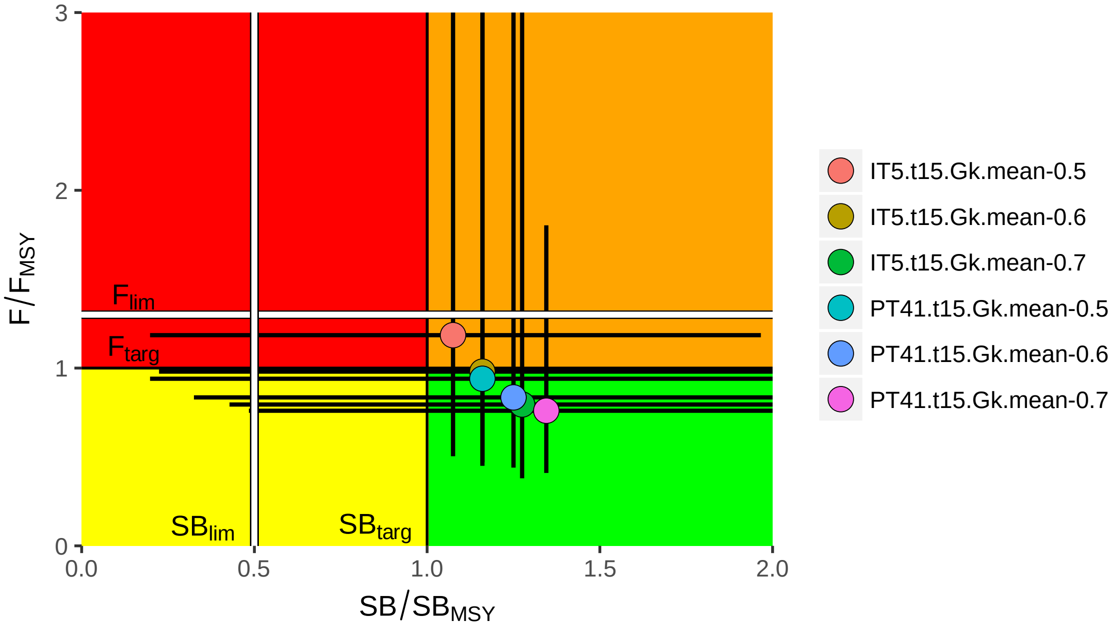


a)

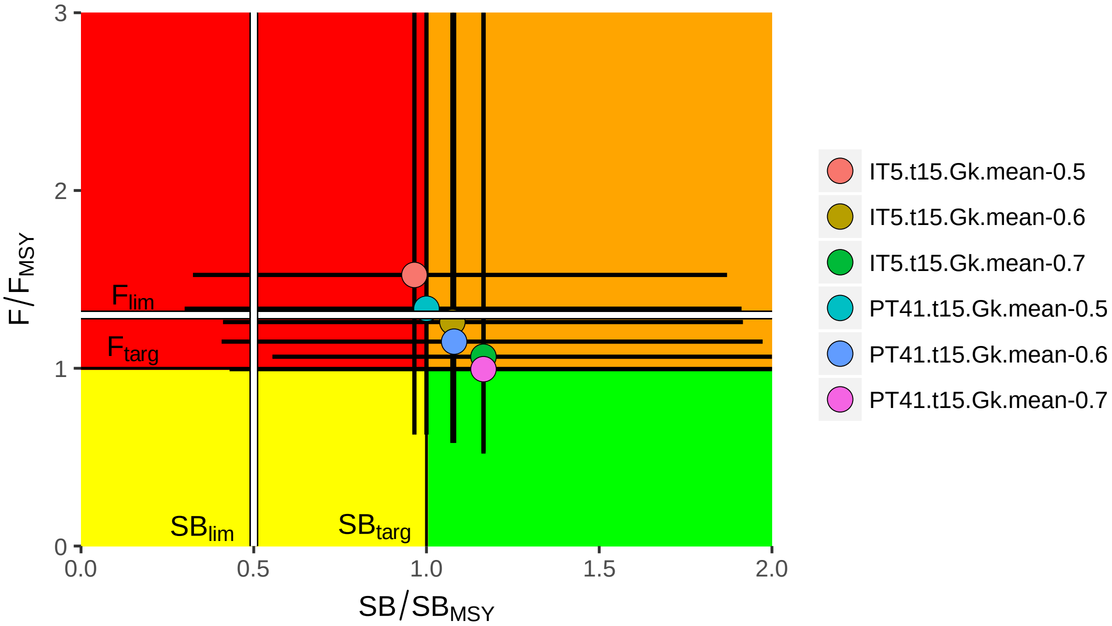


b)

Figure 51. Bigeye robustness cases (B19.5) : a) annual aggregate CPUE CV=0.3 and auto-correlation=0.5; b) 10% reported over-catch. Kobe plot comparing candidate MPs on the basis of the expected 20 year average (2019-2038) performance. Circle is the median, lines represent 10th-90th percentiles.



c)



d)

Figure 52. Bigeye robustness cases (B19.5) : c) 10% unreported over-catch; d) 3% p.a. longline catchability trend. Kobe plot comparing candidate MPs on the basis of the expected 20 year average (2019-2038) performance. Circle is the median, lines represent 10th-90th percentiles.

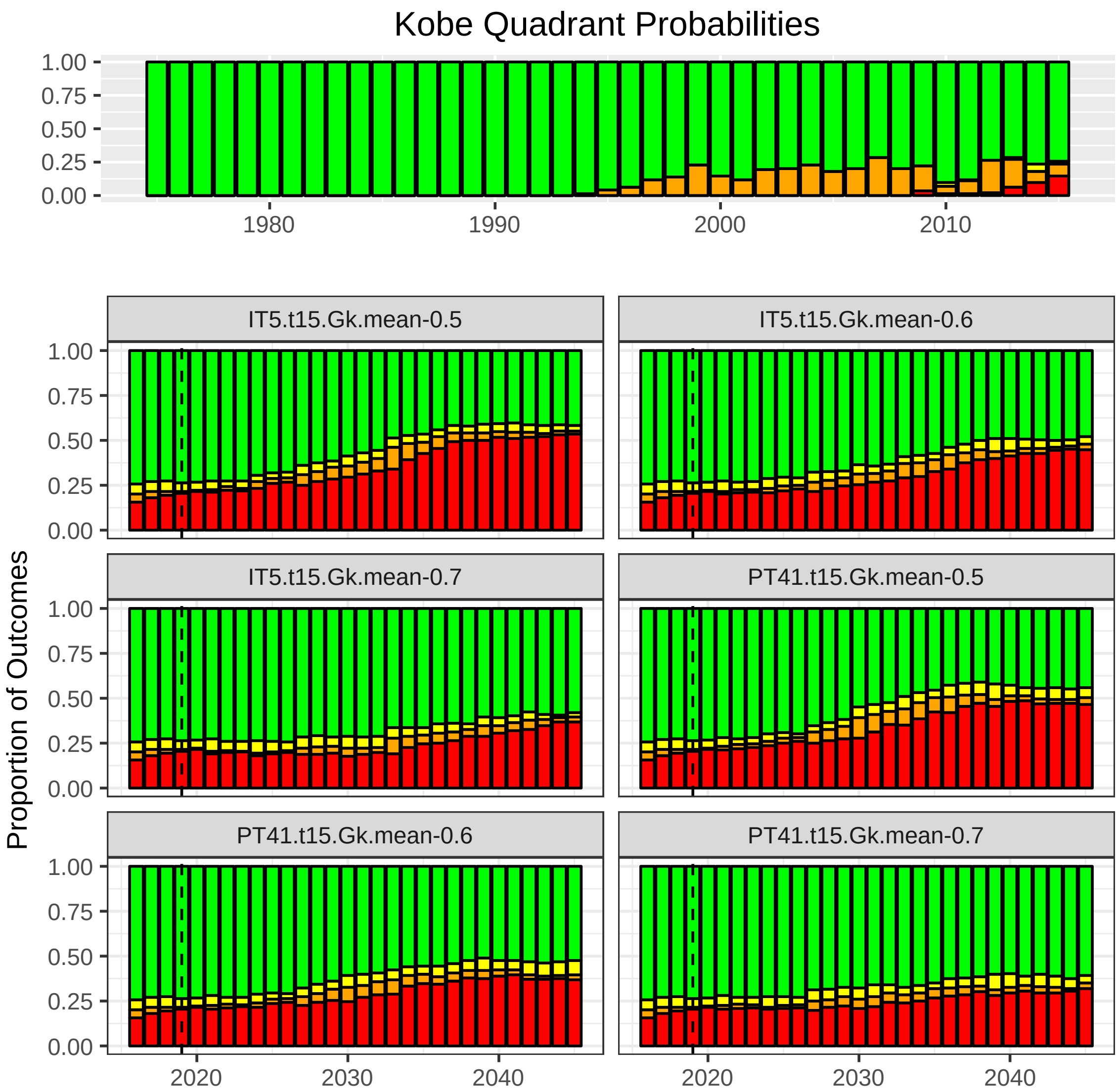


Figure 53. Bigeye robustness cases (B19.5) a) annual aggregate CPUE CV=0.3 and auto-correlation=0.5. Proportion of simulations in each of the Kobe quadrants over time for each of the candidate MPs. Historical estimates are included in the top panel. The lower panels are projections, with the first MP application indicated by the broken vertical line (2019).

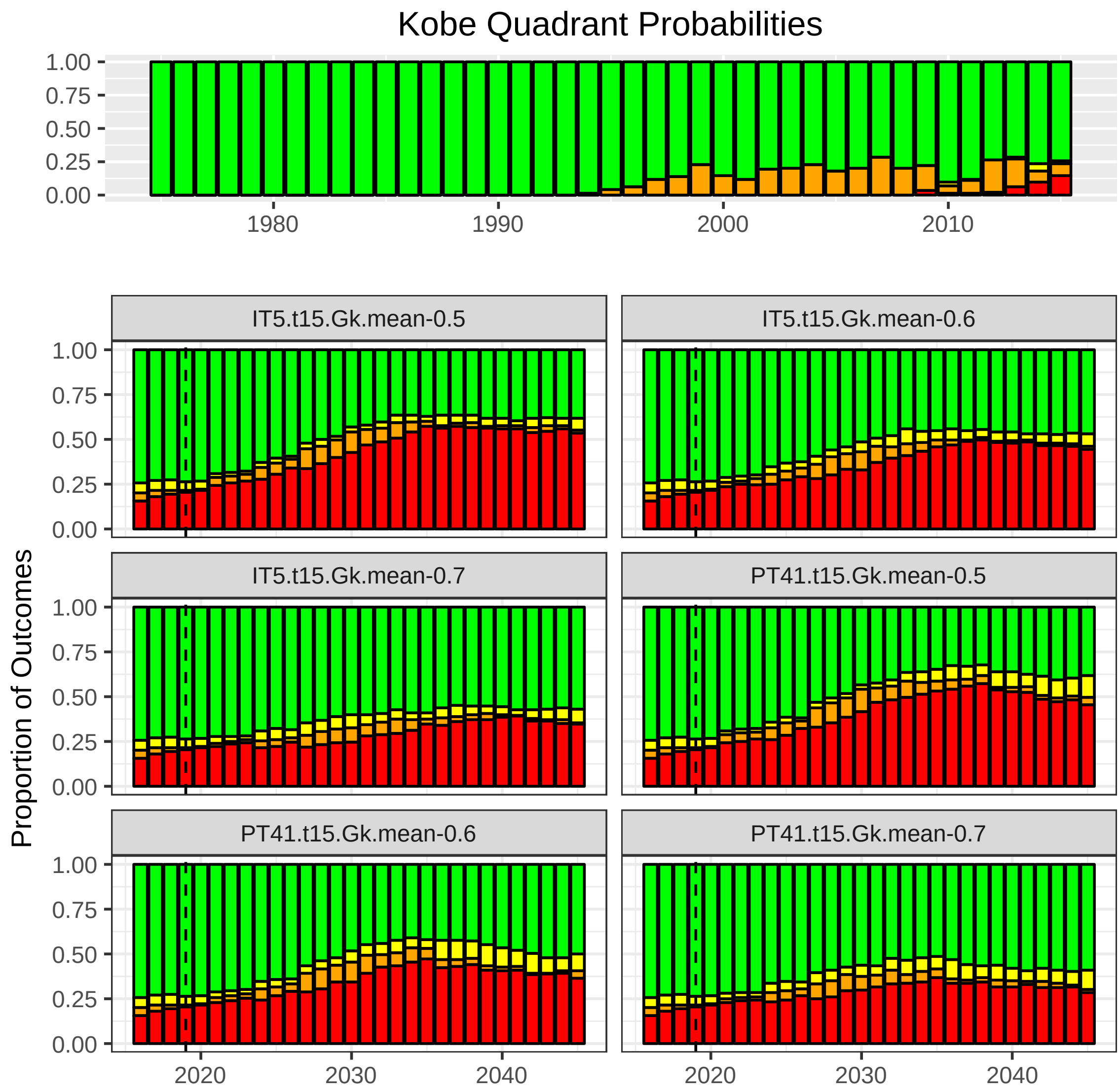


Figure 54. Bigeye robustness cases (B19.5) b) 10% reported over-catch. Proportion of simulations in each of the Kobe quadrants over time for each of the candidate MPs. Historical estimates are included in the top panel. The lower panels are projections, with the first MP application indicated by the broken vertical line (2019).

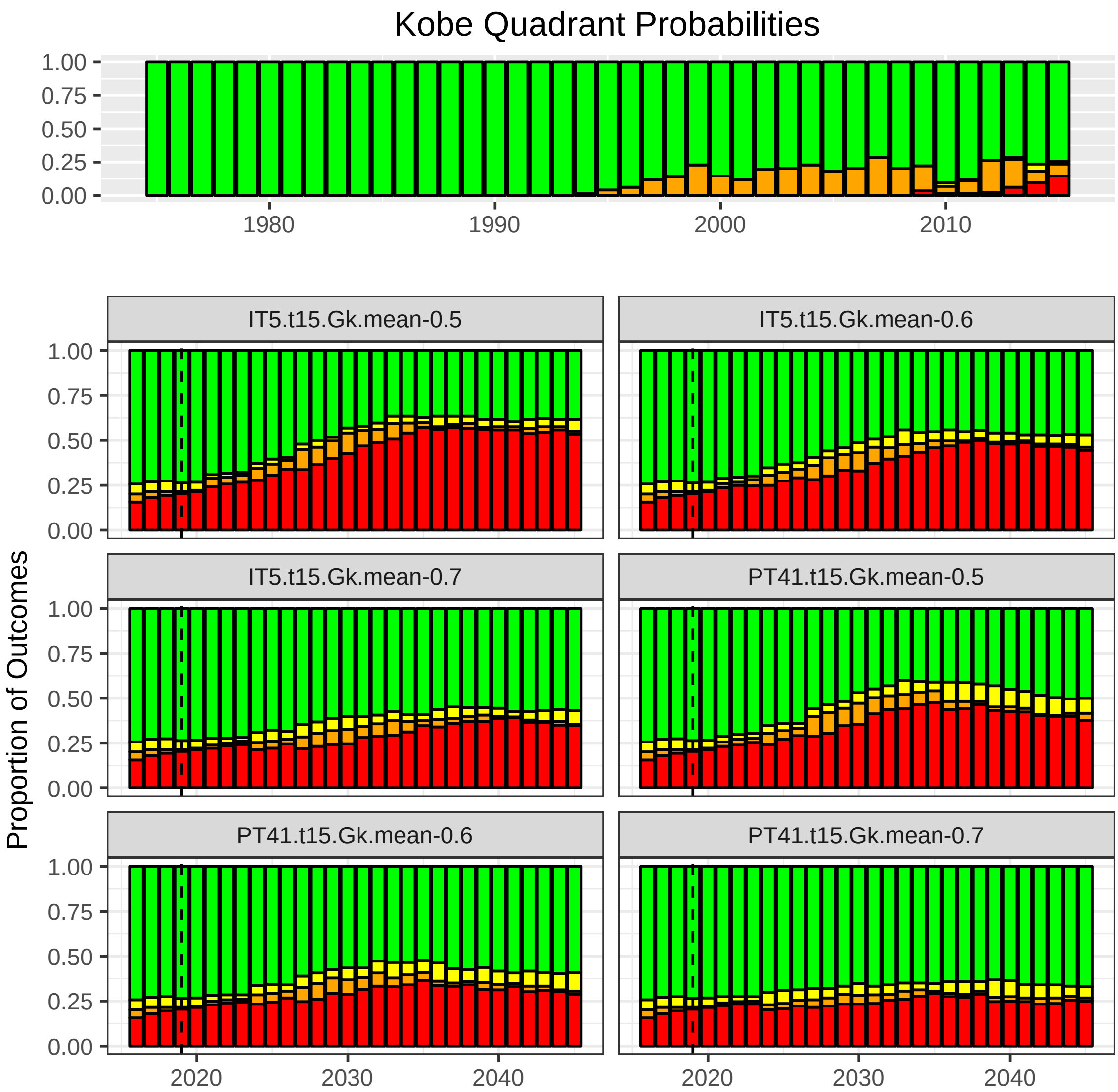


Figure 55. Bigeye robustness cases (B19.5) c) 10% unreported over-catch. Proportion of simulations in each of the Kobe quadrants over time for each of the candidate MPs. Historical estimates are included in the top panel. The lower panels are projections, with the first MP application indicated by the broken vertical line (2019).

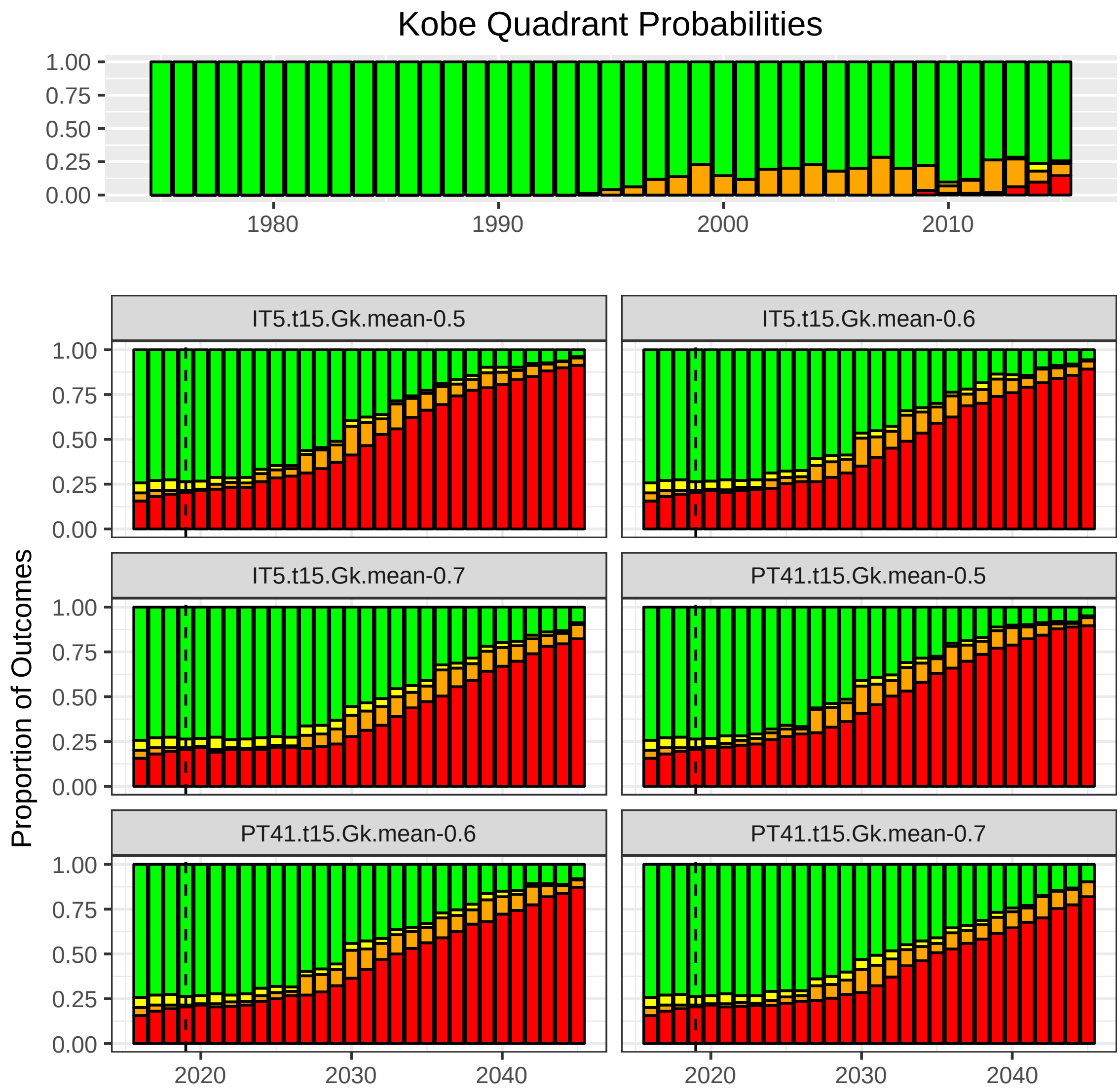


Figure 56. Bigeye robustness cases (B19.5) d) 3% p.a. longline catchability trend. Proportion of simulations in each of the Kobe quadrants over time for each of the candidate MPs. Historical estimates are included in the top panel. The lower panels are projections, with the first MP application indicated by the broken vertical line (2019).

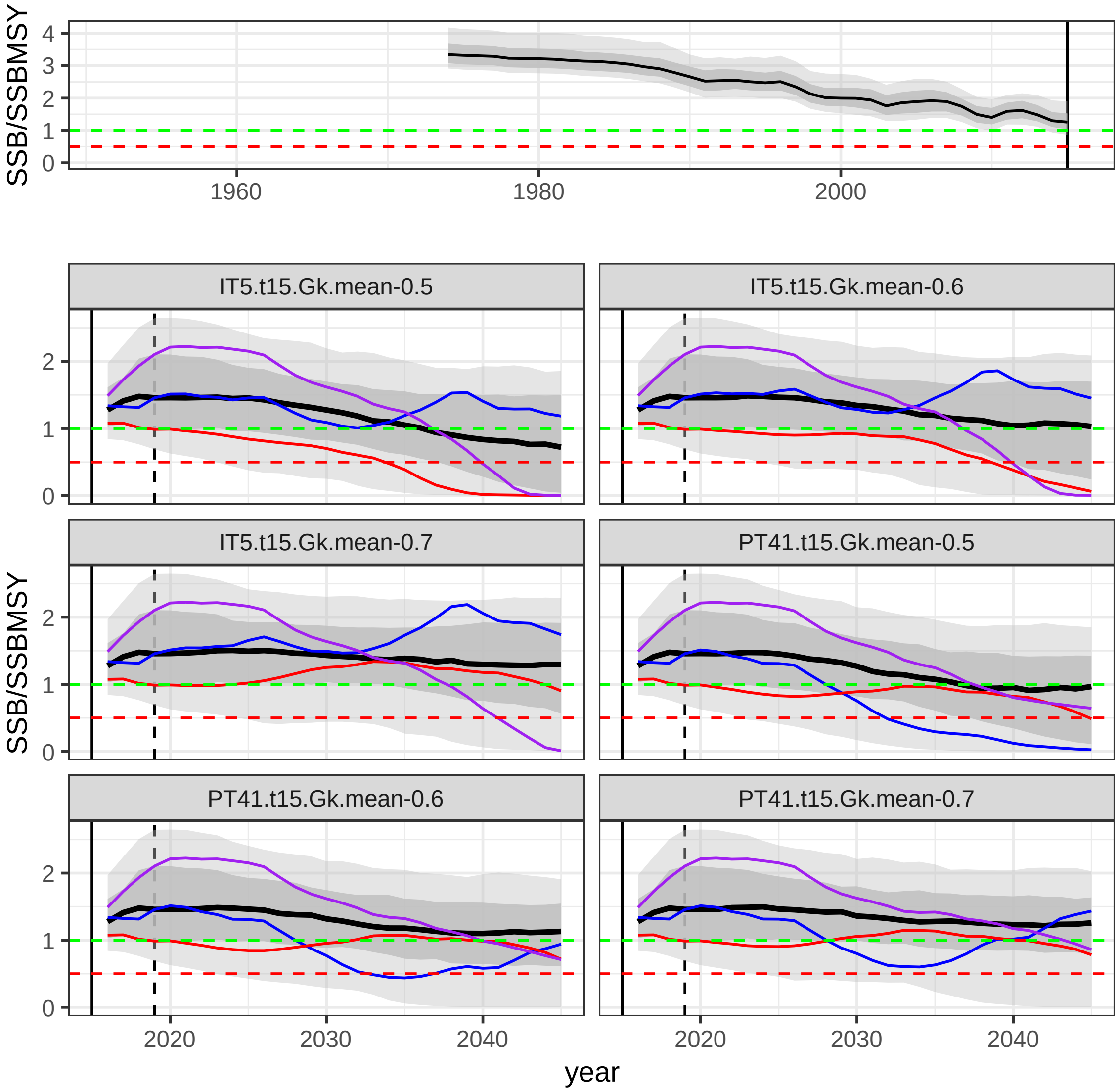


Figure 57. Bigeye robustness cases (B19.5) a) annual aggregate CPUE CV=0.3 and auto-correlation=0.5. Time series of spawning stock size for the candidate MPs. The top panel represents the historical estimates from the reference case operating model, and lower plots represent the projection period. The solid vertical line represents the last year used in the historical conditioning. The broken vertical line represents the first year that the MP is applied. The median is represented by the bold black line, the dark shaded ribbon represents the 25th-75th percentiles, the light shaded ribbon represents the 10th-90th percentiles. Thick broken lines represent the interim target (green) and limit (red) reference points. The 3 thin coloured lines represent examples of individual realizations (the same OM scenarios across MPs and performance measures), to illustrate that individual variability greatly exceeds the median.

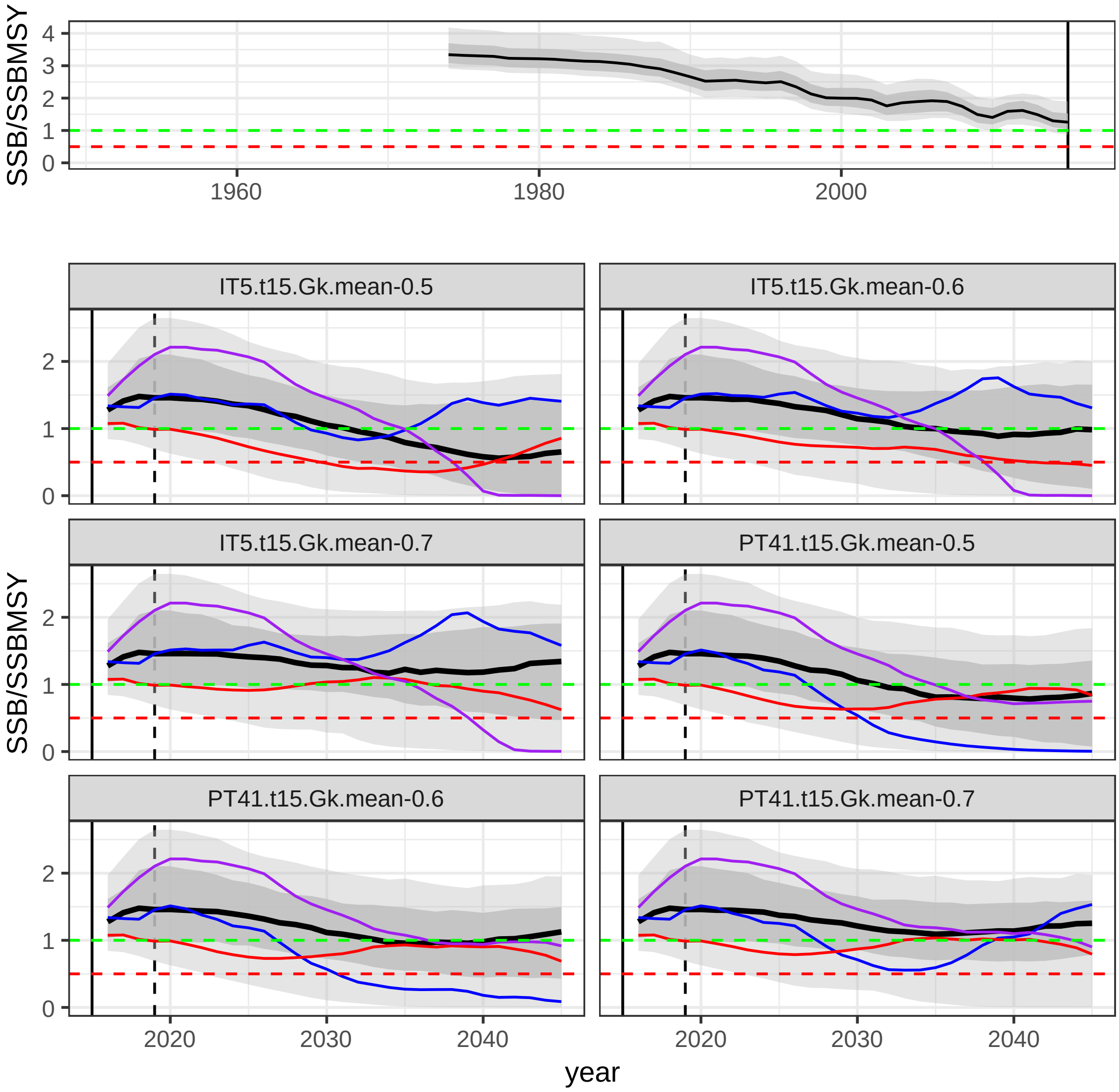


Figure 58. Bigeye robustness cases (B19.5) b) 10% reported over-catch. Time series of spawning stock size for the candidate MPs. The top panel represents the historical estimates from the reference case operating model, and lower plots represent the projection period. The solid vertical line represents the last year used in the historical conditioning. The broken vertical line represents the first year that the MP is applied. The median is represented by the bold black line, the dark shaded ribbon represents the 25th-75th percentiles, the light shaded ribbon represents the 10th-90th percentiles. Thick broken lines represent the interim target (green) and limit (red) reference points. The 3 thin coloured lines represent examples of individual realizations (the same OM scenarios across MPs and performance measures), to illustrate that individual variability greatly exceeds the median.

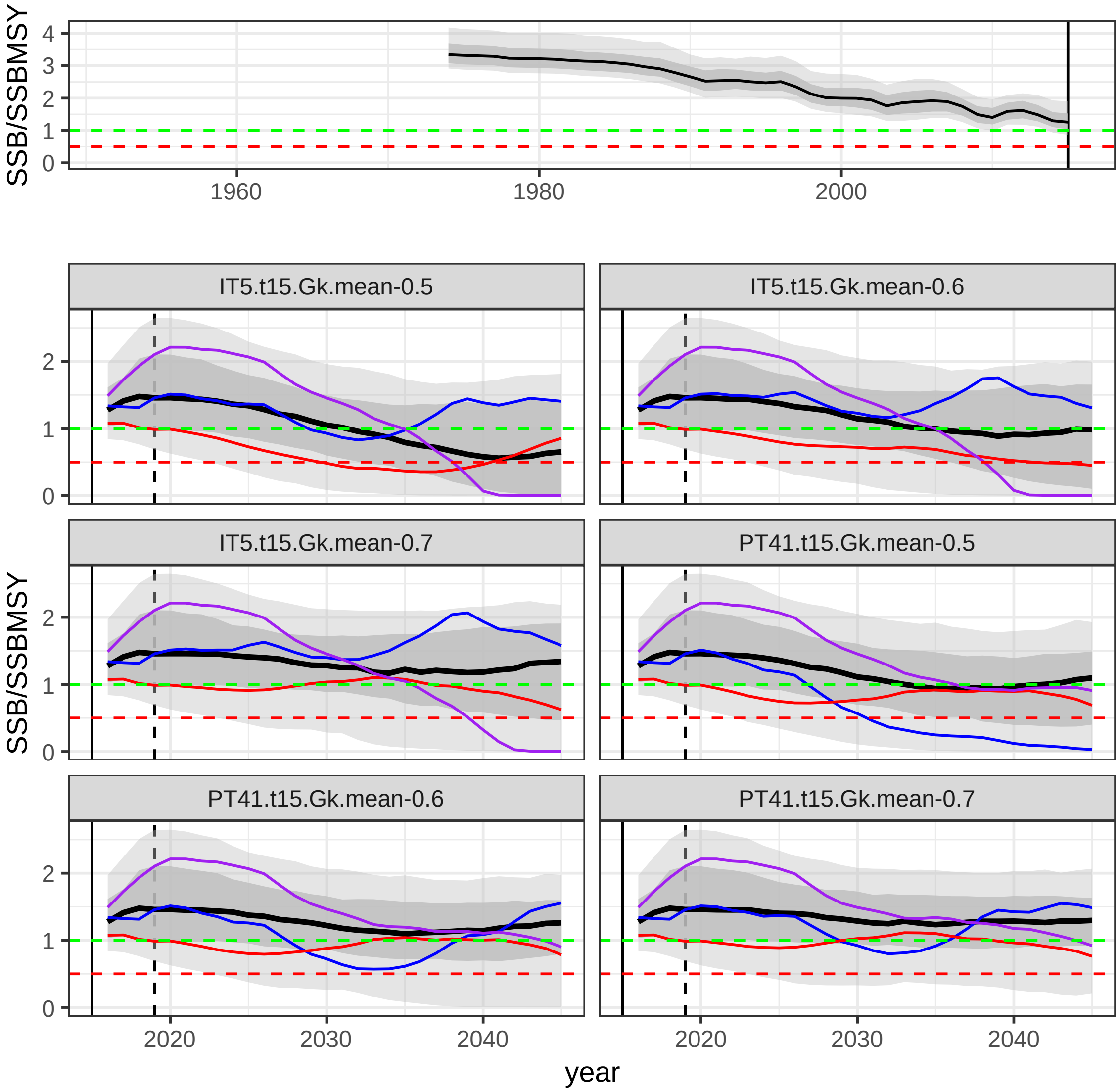


Figure 59. Bigeye robustness cases (B19.5) c) 10% unreported over-catch. Time series of spawning stock size for the candidate MPs. The top panel represents the historical estimates from the reference case operating model, and lower plots represent the projection period. The solid vertical line represents the last year used in the historical conditioning. The broken vertical line represents the first year that the MP is applied. The median is represented by the bold black line, the dark shaded ribbon represents the 25th-75th percentiles, the light shaded ribbon represents the 10th-90th percentiles. Thick broken lines represent the interim target (green) and limit (red) reference points. The 3 thin coloured lines represent examples of individual realizations (the same OM scenarios across MPs and performance measures), to illustrate that individual variability greatly exceeds the median.

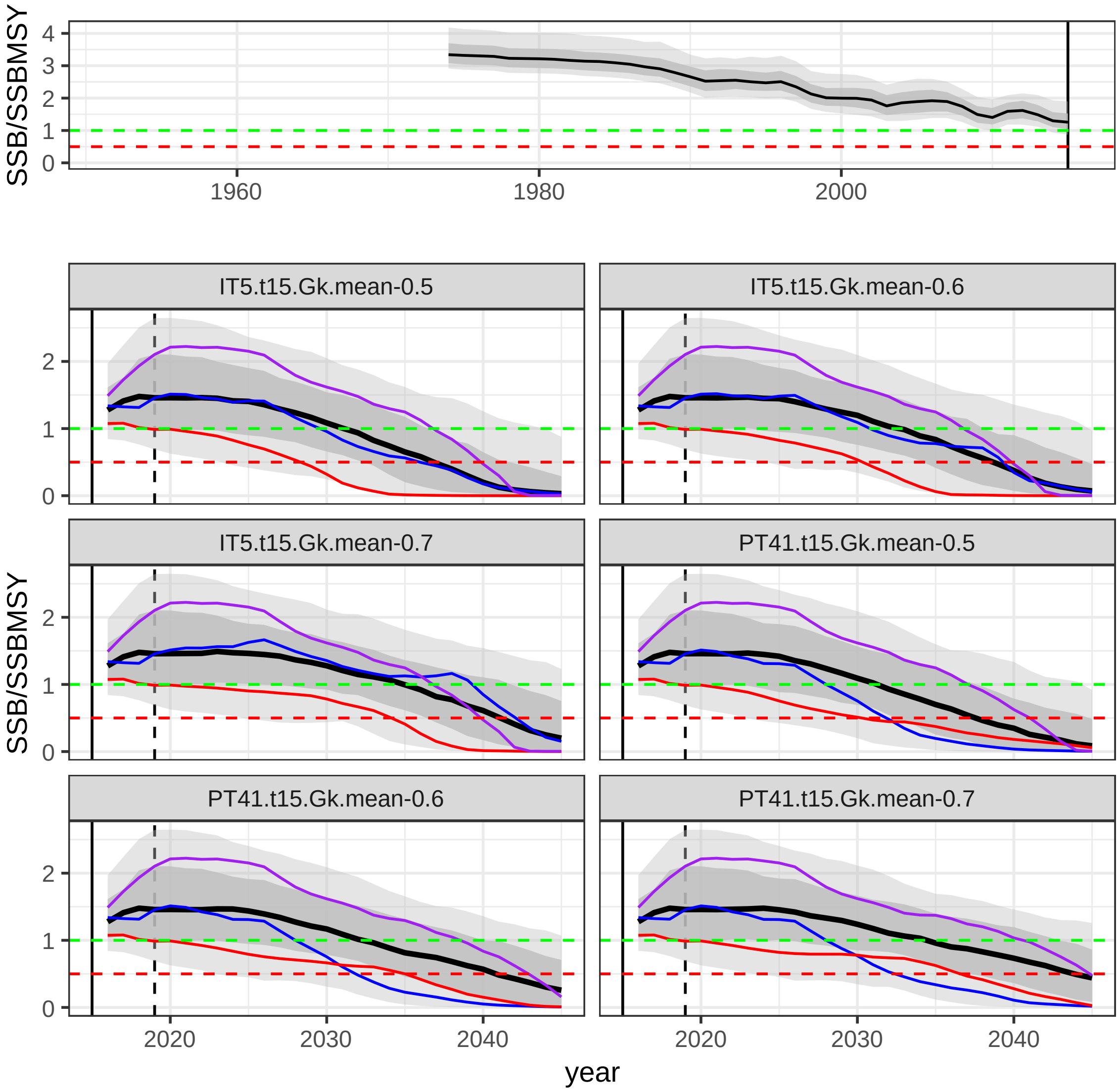


Figure 60. Bigeye robustness cases (B19.5) d) 3% p.a. longline catchability trend. Time series of spawning stock size for the candidate MPs. The top panel represents the historical estimates from the reference case operating model, and lower plots represent the projection period. The solid vertical line represents the last year used in the historical conditioning. The broken vertical line represents the first year that the MP is applied. The median is represented by the bold black line, the dark shaded ribbon represents the 25th-75th percentiles, the light shaded ribbon represents the 10th-90th percentiles. Thick broken lines represent the interim target (green) and limit (red) reference points. The 3 thin coloured lines represent examples of individual realizations (the same OM scenarios across MPs and performance measures), to illustrate that individual variability greatly exceeds the median.

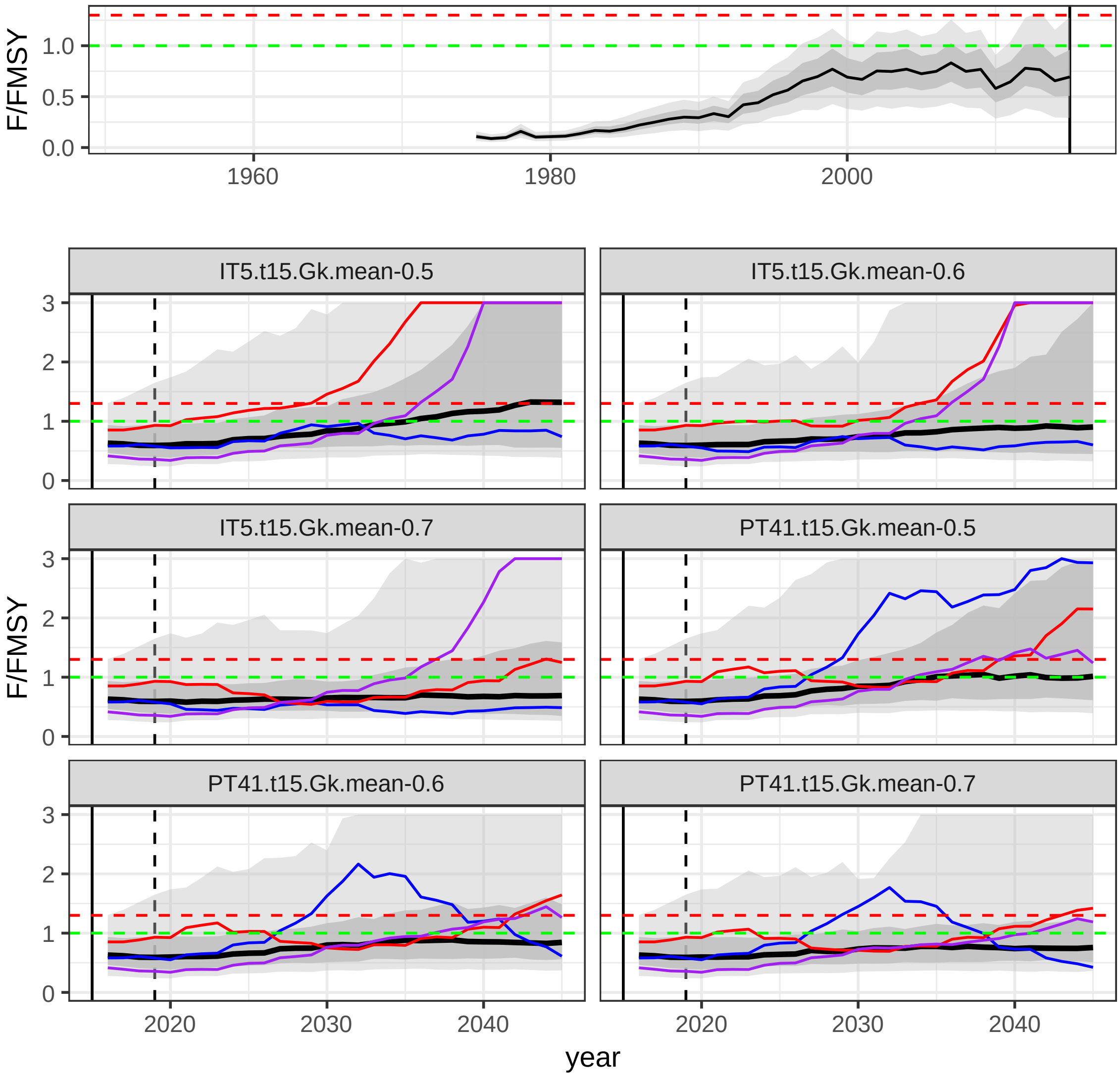


Figure 61. Bigeye robustness cases (B19.5) a) annual aggregate CPUE CV=0.3 and auto-correlation=0.5. Time series of fishing intensity (Upper bound truncated at F = 3) for the candidate MPs. The top panel represents the historical estimates from the reference case operating model, and lower plots represent the projection period. The solid vertical line represents the last year used in the historical conditioning. The broken vertical line represents the first year that the MP is applied. The median is represented by the bold black line, the dark shaded ribbon represents the 25th-75th percentiles, the light shaded ribbon represents the 10th-90th percentiles. Thick broken lines represent the interim target (green) and limit (red) reference points. The 3 thin coloured lines represent examples of individual realizations (the same OM scenarios across MPs and performance measures), to illustrate that individual variability greatly exceeds the median.

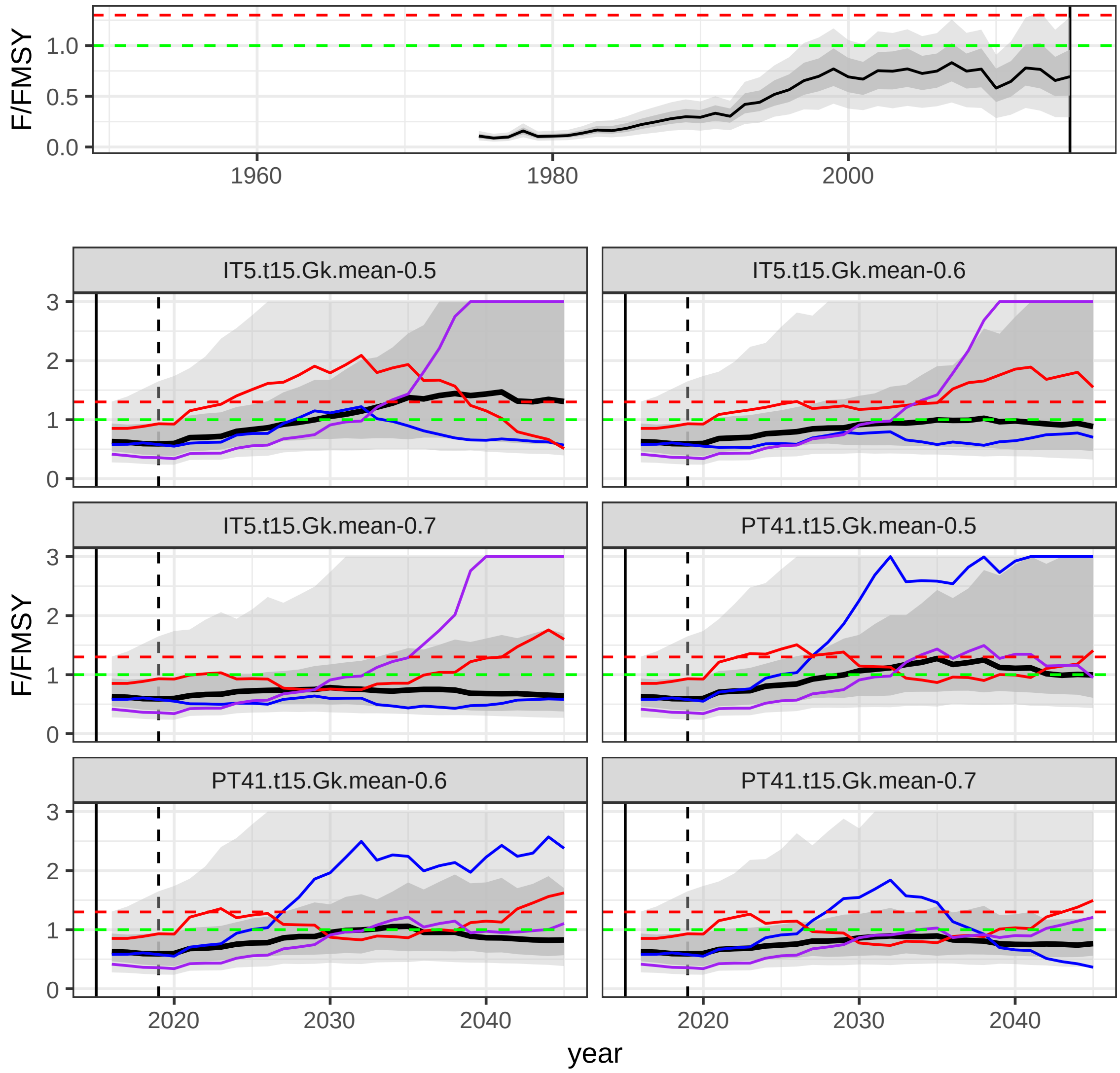


Figure 62. Bigeye robustness cases (B19.5) b) 10% reported over-catch. Time series of fishing intensity (Upper bound truncated at F = 3) for the candidate MPs. The top panel represents the historical estimates from the reference case operating model, and lower plots represent the projection period. The solid vertical line represents the last year used in the historical conditioning. The broken vertical line represents the first year that the MP is applied. The median is represented by the bold black line, the dark shaded ribbon represents the 25th-75th percentiles, the light shaded ribbon represents the 10th-90th percentiles. Thick broken lines represent the interim target (green) and limit (red) reference points. The 3 thin coloured lines represent examples of individual realizations (the same OM scenarios across MPs and performance measures), to illustrate that individual variability greatly exceeds the median.

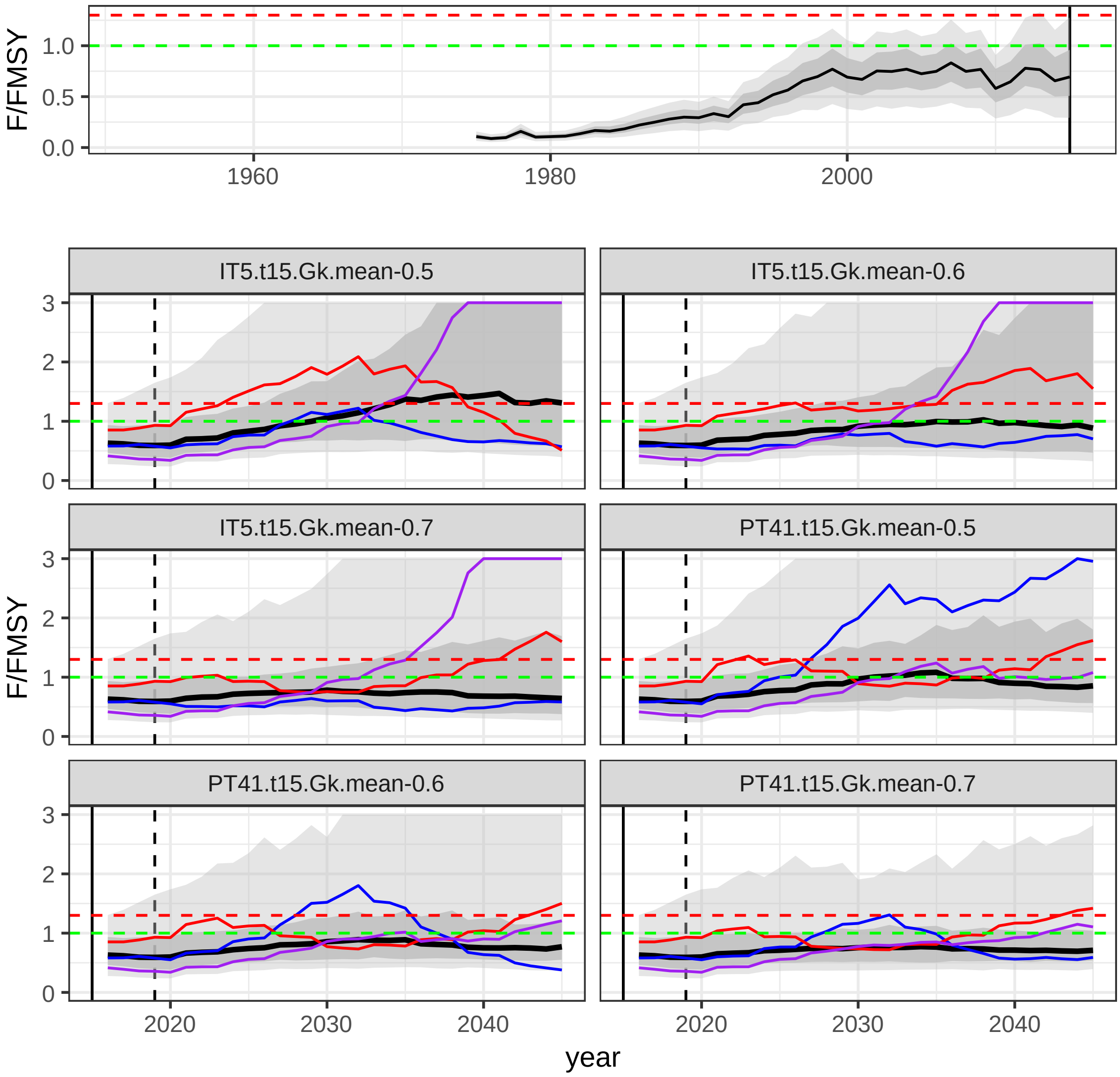


Figure 63. Bigeye robustness cases (B19.5) c) 10% unreported over-catch. Time series of fishing intensity (Upper bound truncated at F = 3) for the candidate MPs. The top panel represents the historical estimates from the reference case operating model, and lower plots represent the projection period. The solid vertical line represents the last year used in the historical conditioning. The broken vertical line represents the first year that the MP is applied. The median is represented by the bold black line, the dark shaded ribbon represents the 25th-75th percentiles, the light shaded ribbon represents the 10th-90th percentiles. Thick broken lines represent the interim target (green) and limit (red) reference points. The 3 thin coloured lines represent examples of individual realizations (the same OM scenarios across MPs and performance measures), to illustrate that individual variability greatly exceeds the median.

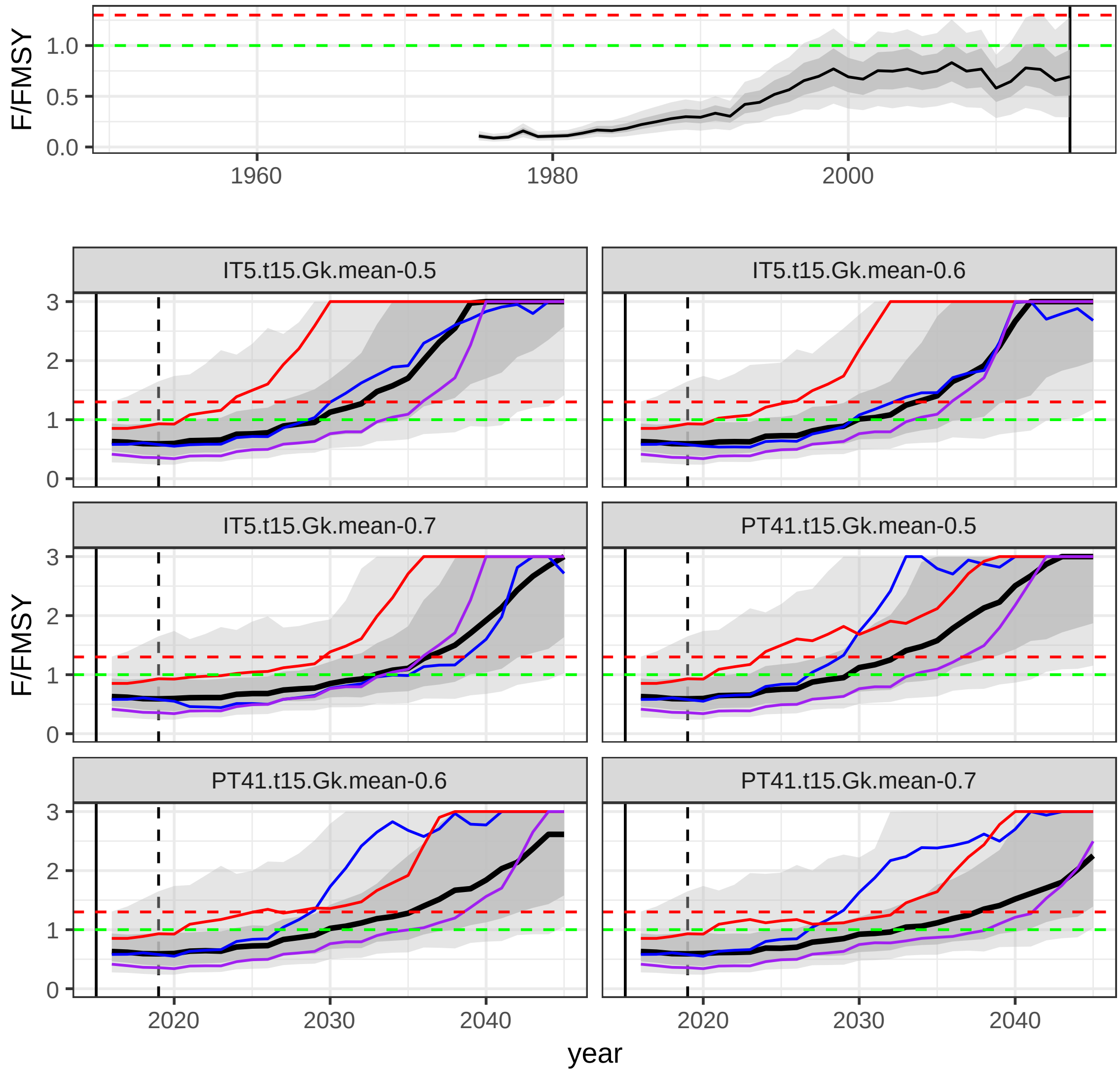


Figure 64. Bigeye robustness cases (B19.5) d) 3% p.a. longline catchability trend. Time series of fishing intensity (Upper bound truncated at F = 3) for the candidate MPs. The top panel represents the historical estimates from the reference case operating model, and lower plots represent the projection period. The solid vertical line represents the last year used in the historical conditioning. The broken vertical line represents the first year that the MP is applied. The median is represented by the bold black line, the dark shaded ribbon represents the 25th-75th percentiles, the light shaded ribbon represents the 10th-90th percentiles. Thick broken lines represent the interim target (green) and limit (red) reference points. The 3 thin coloured lines represent examples of individual realizations (the same OM scenarios across MPs and performance measures), to illustrate that individual variability greatly exceeds the median.

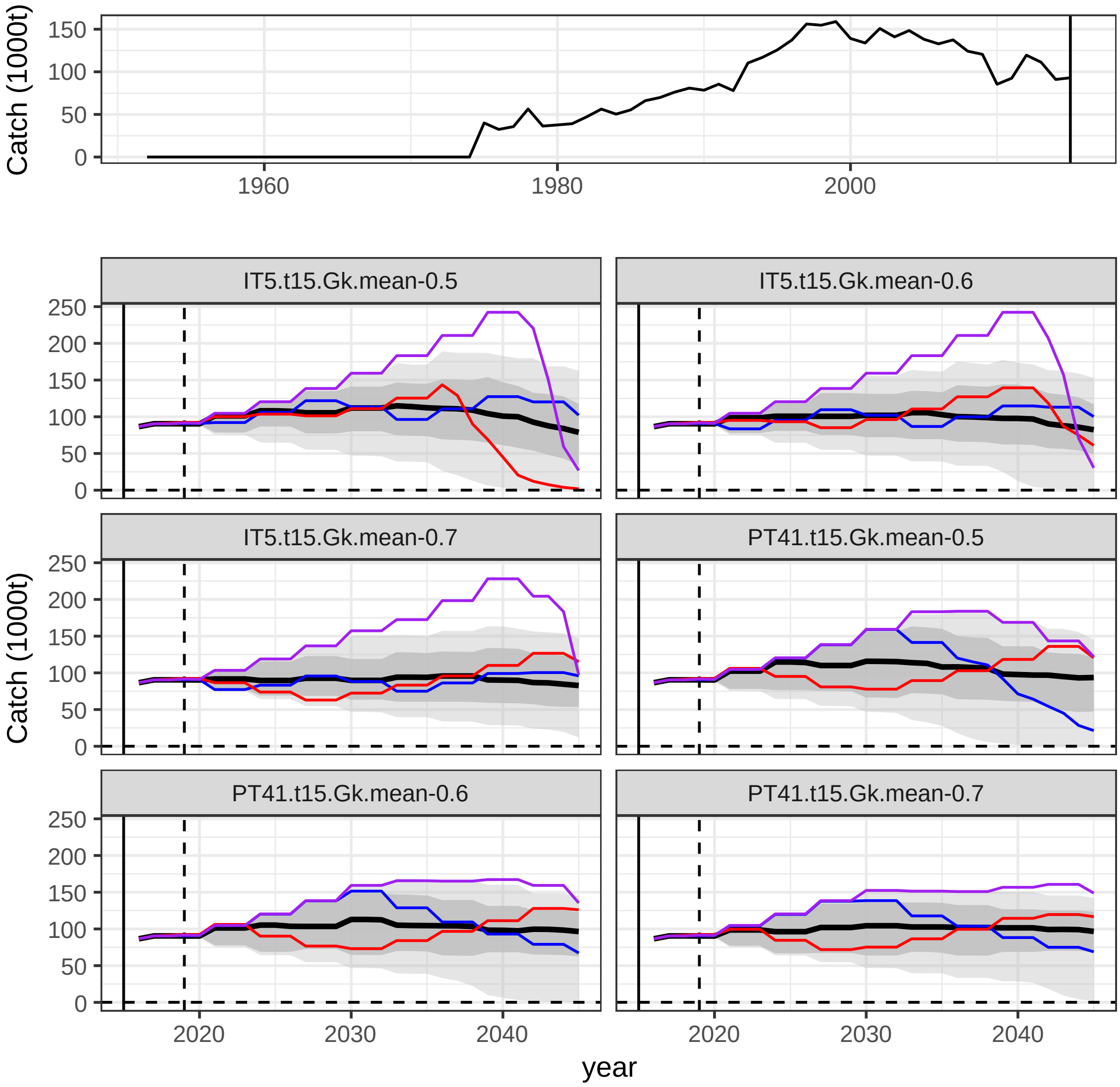


Figure 65. Bigeye robustness cases (B19.5) a) annual aggregate CPUE CV=0.3 and auto-correlation=0.5. Time series of catch for the candidate MPs. The top panel represents the historical estimates from the reference case operating model, and lower plots represent the projection period. The solid vertical line represents the last year used in the historical conditioning. The broken vertical line represents the first year that the MP is applied. The median is represented by the bold black line, the dark shaded ribbon represents the 25th-75th percentiles, the light shaded ribbon represents the 10th-90th percentiles. The broken black horizontal line represents recent (2016) catch. The 3 thin coloured lines represent examples of individual realizations (the same OM scenarios across MPs and performance measures), to illustrate that individual variability greatly exceeds the median.

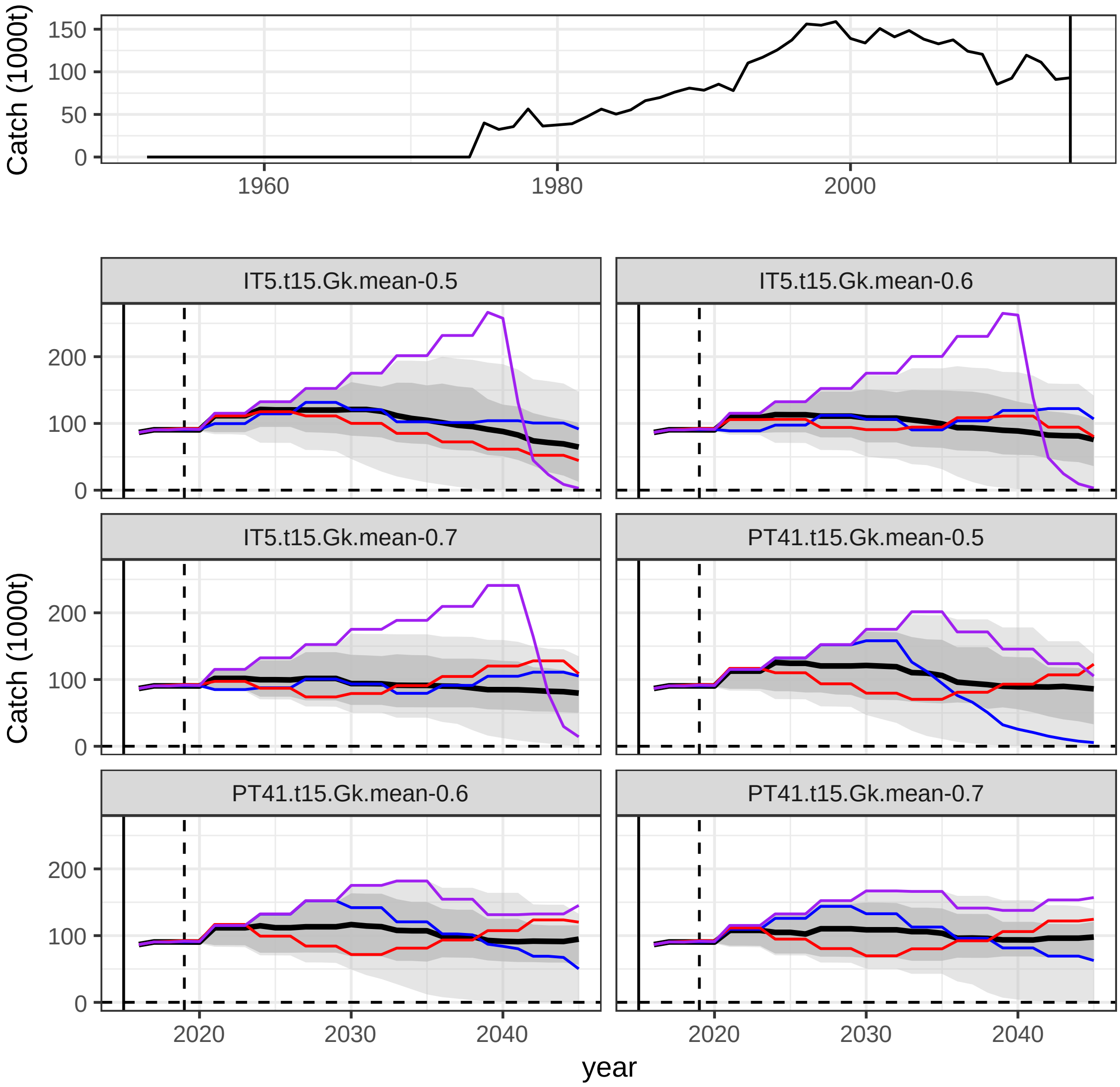


Figure 66. Bigeye robustness cases (B19.5) b) 10% reported over-catch. Time series of catch for the candidate MPs. The top panel represents the historical estimates from the reference case operating model, and lower plots represent the projection period. The solid vertical line represents the last year used in the historical conditioning. The broken vertical line represents the first year that the MP is applied. The median is represented by the bold black line, the dark shaded ribbon represents the 25th-75th percentiles, the light shaded ribbon represents the 10th-90th percentiles. The broken black horizontal line represents recent (2016) catch. The 3 thin coloured lines represent examples of individual realizations (the same OM scenarios across MPs and performance measures), to illustrate that individual variability greatly exceeds the median.

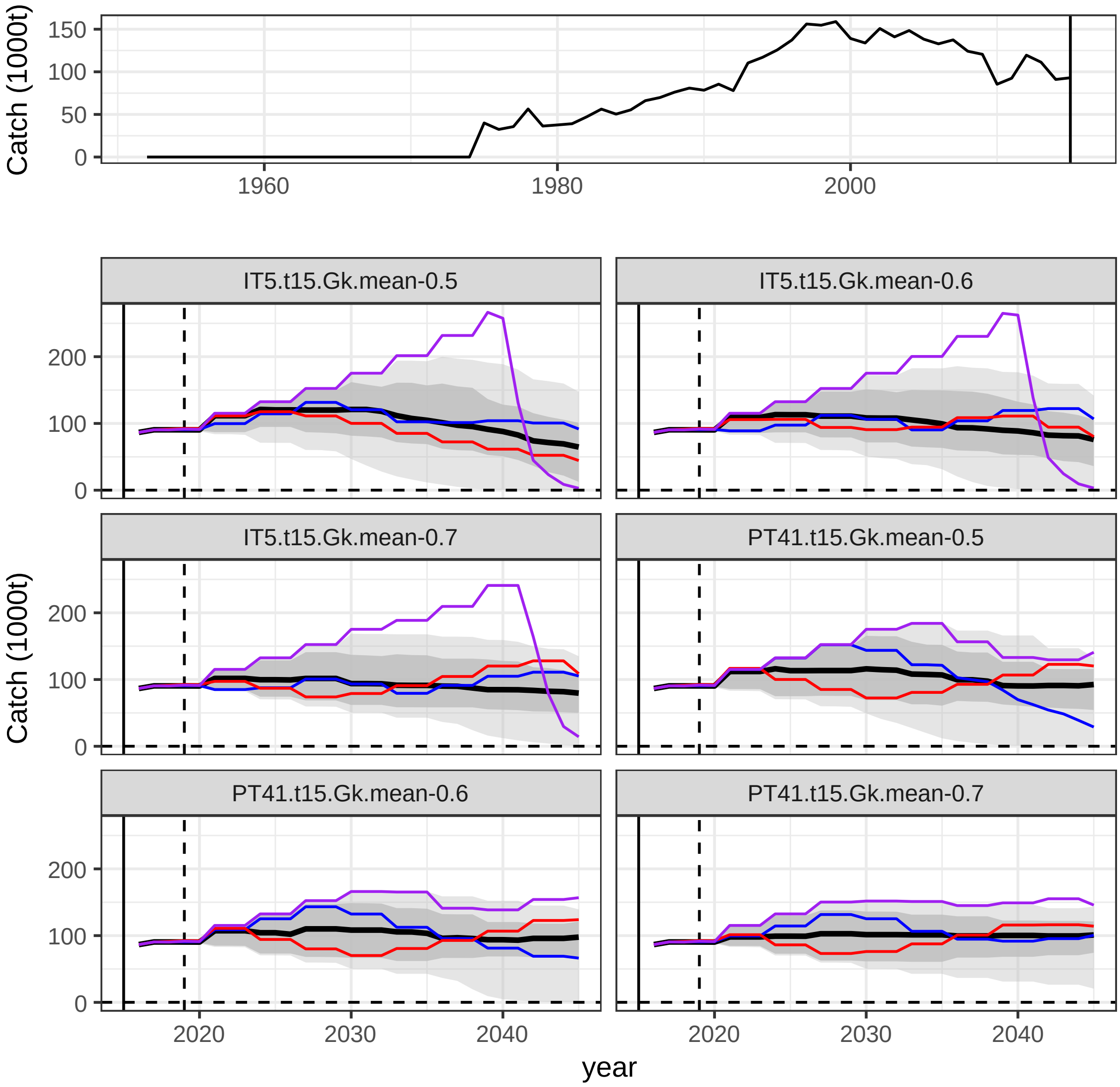


Figure 67. Bigeye robustness cases (B19.5) c) 10% unreported over-catch. Time series of catch for the candidate MPs. The top panel represents the historical estimates from the reference case operating model, and lower plots represent the projection period. The solid vertical line represents the last year used in the historical conditioning. The broken vertical line represents the first year that the MP is applied. The median is represented by the bold black line, the dark shaded ribbon represents the 25th-75th percentiles, the light shaded ribbon represents the 10th-90th percentiles. The broken black horizontal line represents recent (2016) catch. The 3 thin coloured lines represent examples of individual realizations (the same OM scenarios across MPs and performance measures), to illustrate that individual variability greatly exceeds the median.

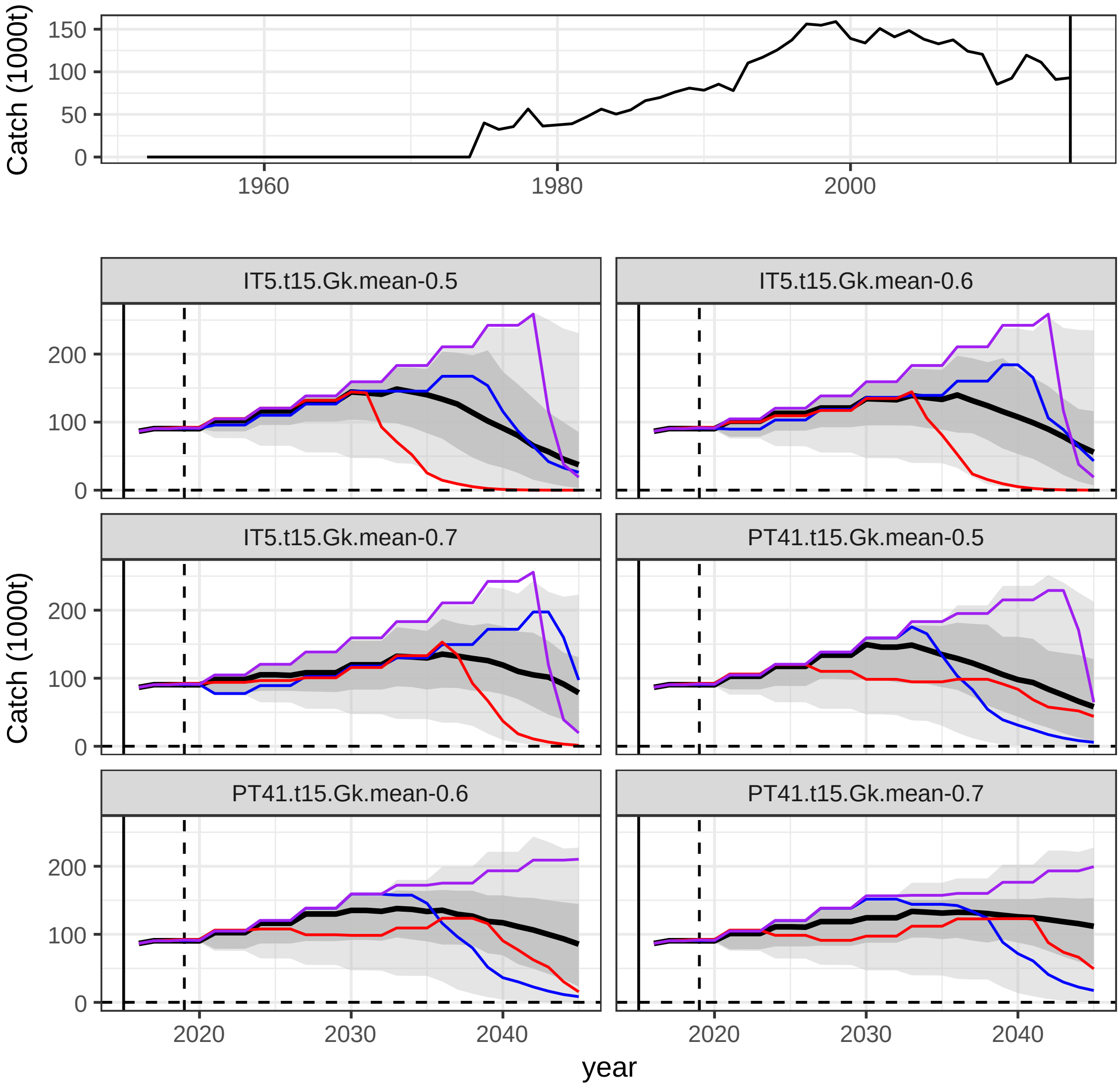


Figure 68. Bigeye robustness cases (B19.5) d) 3% p.a. longline catchability trend. Time series of catch for the candidate MPs. The top panel represents the historical estimates from the reference case operating model, and lower plots represent the projection period. The solid vertical line represents the last year used in the historical conditioning. The broken vertical line represents the first year that the MP is applied. The median is represented by the bold black line, the dark shaded ribbon represents the 25th-75th percentiles, the light shaded ribbon represents the 10th-90th percentiles. The broken black horizontal line represents recent (2016) catch. The 3 thin coloured lines represent examples of individual realizations (the same OM scenarios across MPs and performance measures), to illustrate that individual variability greatly exceeds the median.

Discussion Points for the IOTC MSE Task Force:

1. Based on these results, we recommend that the reference set BET OM ensemble for the 2019 TCMP should consist of the following grid options and levels:

- h70, h80, h90
- M10, M08, M06
- t0001, t10
- q0, q1
- iH, iC
- ess10, CLRW
- iR1, iR2 (this seems to be the least important uncertainty dimension)
- gr1 (original growth curve only)

2. We recommend dropping the alterative growth curve at this time, because it is an ad hoc hybrid that is not statistically defensible. It can lead to implausible stock dynamics estimates. But bigeye growth is probably worth revisiting at a future date, in relation to inter-lab ageing method comparisons that are currently underway. The current options suggest that there may be non-trivial spatial/temporal variability in growth that cannot be properly represented in the current conditioning or projection software.

3. We recommend using a fractional factorial design that results in a manageable number of conditioned models in the ensemble of around 50 - 150, includes all interactions between the 3 level assumptions h and M, allows all main effects of all 2 level options to be estimable, and includes all 2-way interactions. We would suggest dropping the alternative regional scaling assumption if the dimensionality needs to be reduced. Dropping the requirement for 2 way interactions is probably preferable to dropping assumption options that people believe to be important.

4. We recommend retaining the repeated convergence procedure to minimize the probability of accepting outliers due to extreme numerical convergence problems.

5. The best method for evaluating model plausibility remains a topic for discussion. We seek feedback on:

- automating diagnostics for large ensembles
- use of the catch penalty as a plausibility criterion
- using informative priors to avoid bounds problems

6. Preparation for the 2019 TCMP:

a. Agree OM reference set definition and contingencies.
b. Do the TCMP 2018 tuning levels cover a sufficiently desirable part of the management trade-off space, or should we define additional levels?
c. Do not present any robustness scenario results to the 2019 TCMP.
d. Tune a range of MPs for behaviour that is more stable, and/or more responsive to new data.

7. Preparation for the WPTT/WPM 2019:

e. Evaluate MP results for new tuning levels from 2019 TCMP. Attempt to refine MP performance if TCMP provides additional management objective insights, or evidence from other studies suggests that additional information can be extracted using different approaches.
f. Review issues of simulating CPUE observation error in relation to spatial and temporal variability and potentially missing observations.

8. There is no funding identified for bigeye and yellowfin MSE scientific and technical support beyond Dec 2019.

Acknowledgements:

This work was jointly funded by the Global Environment Facility - Common Oceans - Areas Beyond National Jurisdiction Tuna Project with the technical cooperation of the Food and Agriculture Organization of the United Nations), and Australia's Commonwealth Scientific and Industrial Research Organization (CSIRO - Oceans & Atmosphere). The work does not necessarily reflect the views of the funding organizations.

References:

Eveson, P., Million, J., Sardenne, F., Le Croizier, G. 2012. Updated growth estimates for skipjack, yellowfin and bigeye tuna in the Indian Ocean using the most recent tag-recapture and otolith data. IOTC-2012-WPTT14-23.

Farley, JH, Clear, NP, Leroy, B, Davis, TLO, Farley, McPherson, G, 2016. Age, growth and preliminary estimates of maturity of bigeye tuna, *Thunnus obesus*, in the Australian region. Marine and Freshwater Research 57: 713–724.

Hoyle, SD, 2018. Indian Ocean tropical tuna regional scaling factors that allow for seasonality and cell areas. IOTC-2018-WPM09-13

Langley, A, 2016. Stock assessment of bigeye tuna in the Indian Ocean for 2016 — model development and evaluation. IOTC-2016-WPTT18-20

McKechnie, S, Pilling, G, Hampton, J. 2017. Stock assessment of bigeye tuna in the western and central Pacific Ocean. WCPFC-SC13-2017/SA-WP-05 Rev1.

Schaefer, K. M., and Fuller, D.W. 2006. Estimates of age and growth of bigeye tuna (*Thunnus obesus*) in the eastern Pacific Ocean, based on otolith increments and tagging data. Bulletin of the Inter-American Tropical Tuna Commission, 23: 35–76.

Schweder, T, Hagen, GS and Hatlebakk E, 1998. On the effect on cod and herring fisheries of retuning the Revised Management Procedure for minke whaling in the greater Barents Sea. Fisheries Research 37, 77– 95.

Appendix 1. Extracts from the 2018 Methods and Tropical Tuna Working Party reports relevant to bigeye and yellowfin MSE

Methods Working Group (2018) draft report

1. BIGEYE TUNA AND YELLOWFIN TUNA MSE: UPDATE

1.1 Review of Operating Models based on WPM and SC feedback, including possible robustness tests

1. The WPM **NOTED** that two presentations were made in this section, summarizing three related working papers, IOTC–2018–WPM09–09 (BET OM definition), IOTC–2018–WPM09–10 (YFT OM definition) and IOTC–2018–WPM09–11 (BET and YFT MP Evaluations).
2. The WP **NOTED** that the MSE for both species is being pursued in the strict sense of MP in which the MP consists of simulation-tested combination of data collection, analysis methods and HCR (which makes this work different to the SKJ assessment, when no specification on data and analyses methods was made)
3. The WPM **NOTED** paper IOTC–2018–WPM09–09 which provided an update on the IOTC Bigeye Tuna MSE Operating Model Development. The following abstract was provided by the authors: *"This paper summarizes progress on the development of Operating Models (OMs) for IOTC bigeye (BET) tuna. Additional background detail on recent software developments is provided in the yellowfin (YFT) companion paper (Kolody and Jumppanen 2018f). MP evaluation updates for BET and YFT are described in Kolody and Jumppanen (2018a). This paper builds on the work presented and reviewed at the IOTC informal MSE Working Group in March 2018 (Kolody and Jumppanen 2018d,e), and represents the first time that the formal IOTC WPTT and WPM have the opportunity to review the substantial BET OM developments since the phase 1 work was completed in 2016. (See paper for full abstract):*
4. The WPM **SUGGESTED** the following changes to the reference case OM grid:
   - CPUE variability set to a level that would result in an annual CV of 0.2 (retaining auto-correlation of 0.5)
   - Extend bridging catches, with first TAC in 2021
   - Additional uncertainty dimensions:
     1. alternative growth function (noting the large effect on the recent WCPFC bigeye assessment). WPTT asked to review and specify the most appropriate alternative. This could be a robustness scenario.
     2. alternative regional CPUE scaling factors
     3. alternative historical catch series. Proposals were discussed, but the options were thought to either not represent a large change from the preferred series, or were difficult to justify as plausible.
5. The WPM **RECOMMENDED** exploring partially-confounded experimental design as a computationally tractable method for expanding the number of uncertainty dimensions and the main interactions (at the expense of losing higher order interactions). It should be adopted if if it is not found to have a significant reduction in full grid uncertainty.
6. The WPM **SUGGESTED** the following priorities for robustness scenarios:
   - Annual aggregated CPUE CV = 0.3 (auto-correlation = 0.5)
   - 10% reported over-catch (projections only; reference case conditioning)

- 10% unreported over-catch (projections only; reference case conditioning)
- 3% LL catchability trend (projections only; reference case conditioning)
- Spatial Structure- possibly additional area around eastern INDONESIA, another in the Bay of Bengal Region and the area around Oman (other area stratification as is).
- Non stationary M, linf and K in the projections.
- Stock Structure (based on ongoing IO stock structure project).

7. The WPM **NOTED** that some of these robustness tests should be considered long-term ambitions, which would require more specific definitions and input from the secretariat and external parties, and would likely delay the current development timeline.
8. The WPM **NOTED** that some of the effects tested separately in the Robustness scenarios could eventually happen simultaneously and at least some scenarios should consider these effects in combination (e.g. catch misreporting and recruitment failure in the same simulation). However, it was further noted that an MP cannot be expected to handle every adverse situation and "exceptional circumstances" procedures are applicable in the worst cases
9. The WPM **NOTED** paper IOTC–2018–WPM09–10 which provided an update on the IOTC Yellowfin Tuna Operating Model Development. The following abstract was provided by the authors:

   "This paper summarizes progress on the development of Operating Models (OMs) for IOTC yellowfin (YFT) tuna. MP evaluation updates for yellowfin and bigeye tunas are described in Kolody and Jumppanen (2018a). This paper builds on the work presented and reviewed at the IOTC informal MSE Working Group in March 2018 (Kolody and Jumppanen 2018d,e).

   The latest version of the MSE software is publicly available from github, with a recently updated technical description and user manual (https://github.com/pjumppanen/niMSE-IO-BET-YFT/).The BET and YFT MSE projection software has undergone several changes in the past year, with a substantial rewrite to improve memory usage and parallel processing, which greatly improves MP evaluation speed. Most of these changes to the computational engine are not visible to the end user. (See paper for full abstract):"

10. The WPM **NOTED** the high uncertainty and large number of implausible models in the uniformly weighted grid of the YFT Reference set OMs. It was recognised that the proposed approach of sampling the uniform grid with respect to the central tendency of the assessment was not ideal, but represented a pragmatic path forward.
11. The WPM **DISCUSSED** the alternative option of filtering plausible models in relation to habitat constraints as was used for albacore, and noted the following disadvantages in this case:
    - It is not obvious that a meta-analysis of the productivity of 3 or 4 other YFT populations would provide more valuable insight about productivity than the arguments employed within the IOTC assessment process.
    - The YFT MSY distribution forms a long-tailed continuum, unlike the disjointed polymodal distribution for ALB
    - Unlike ALB, the YFT distribution also had many models that were implausibly unproductive (not only over-productive)
12. The WPM **SUGGESTED** the following changes to the YFT reference set OM grid, and expected that the WPTT would refine these recommendations, particularly with respect to insights from the new YFT assessment:
    - CPUE variability set to a level that would result in an annual CV of 0.2 (retaining auto-correlation of 0.5)

- Extend bridging catches, with first TAC in 2021
- Additional uncertainty dimensions:
    i. alternative growth function (noting the large effect on the recent WCPFC bigeye assessment). WPTT will be asked to review and specify the most appropriate alternative. This could be a robustness scenario.
    ii. alternative regional CPUE scaling factors
    iii. alternative historical catch series. Proposals were discussed, but the options were thought to either not represent a large change from the preferred series, or were difficult to justify as plausible.
    iv. It was noted that a new YFT catch data series will be discussed for the assessment at the WPTT, which is probably appropriate for the OM as well

- Sample the OM grid using the bi-variate sampling approach (sampling with respect to the central tendency of MSE and SB(current)/SB(MSY), but with variance assumptions that are compatible with the distributional characteristics of the BET grid (for consistency)

13. The WPM **RECOMMENDED** exploring partially-confounded experimental design as a computationally tractable method for expanding the number of uncertainty dimensions and the main interactions (at the expense of losing higher order interactions). It should be adopted if if it is not found to have a significant reduction in full grid uncertainty.

14. The WP **SUGGESTED** the following priorities for robustness scenarios:

- Annual aggregated CPUE CV = 0.3 (auto-correlation = 0.5)
- 10% reported over-catch (projections only; reference case conditioning)
- 10% unreported over-catch (projections only; reference case conditioning)
- 3% LL catchability trend (projections only; reference case conditioning)
- dome-shaped longline selectivity (noting potential for interaction with M and growth)

15. The WPM **NOTED** that some of these robustness tests should be considered long-term ambitions, which would require more specific definitions and input from the secretariat and external parties, and would likely delay the current development timeline.

16. The WPM **NOTED** that some of the effects tested separately in the Robustness scenarios could eventually happen simultaneously and at least some scenarios should consider these effects in combination (e.g. catch misreporting and recruitment failure in the same simulation). However, it was further noted that an MP cannot be expected to handle every adverse situation and "exceptional circumstances" procedures are applicable in the worst cases

17. The WPM **NOTED** that alternative MP tuning levels should be adopted to add contrast to the results for the TCMP02.

Working Party on Tropical Tuna (2018) draft report

Bigeye

1. The WPTT **NOTED** paper IOTC–2018–WPM09–09 which provided an update on IOTC bigeye tuna operating model development, October 2018, including the following summary provided by the authors:

   "This paper summarizes progress on the development of Operating Models (OMs) for IOTC bigeye (BET) tuna. Additional background detail on recent software developments is provided in the yellowfin (YFT) companion paper (Kolody and Jumppanen 2018f). MP evaluation updates for BET and YFT are described in Kolody and Jumppanen (2018a). This paper builds on the work presented and reviewed at the IOTC informal MSE Working Group in March 2018 (Kolody and Jumppanen 2018d,e), and represents the first time that the formal IOTC WPTT and WPM have the opportunity to review the substantial BET OM developments since the phase 1 work was completed in 2016."

2. The WPTT reviewed and **ENDORSED** the progress to date on MSE for bigeye tuna while recognizing the discussions held at TCMP and the advice of WPM, but **INDICATED** the need to consider some additional uncertainty dimensions in the bigeye tuna MSE workplan agreed by WPM.

3. In particular, WPTT **ENCOURAGED** that the MSE work consider the importance of an alternative growth curve for bigeye tuna. The WPTT **SUGGESTED** the growth curve estimated by Farley et. al. (2016) is based on a broader size range (up to 160cm+) and may have a more plausible Linf value (~178 cm) than the Eveson (2015) model currently used in the OM. Furthermore, the Farley et. al. (2016) growth curve is derived from samples from the eastern Indian Ocean so may provide additional information on growth from a different region. However, the WPTT acknowledged that the Farley et. al. (2016) growth function may not describe well the length-at-age for fish smaller than 70cm LJFL which is the size range of most of the tagged fish for which the model estimates age.

4. Therefore, the WPTT **SUGGESTED** either anchoring the growth curve to a plausible age at zero length, or preferably combining the data from the Farley et al. (2006) growth curve with the Eveson (2015) and fitting both Von-Bertalanffy Growth Function (VBGF) and multi-stanza growth models to determine the best model fit.

5. The WPTT expressed some concern in combining size at age data from different time periods to estimate a single growth curve due to the potential for temporal shifts in growth, but also **NOTED** that the inclusion of an additional growth curve was to capture a plausible range of uncertainty in growth.

6. The WPTT **NOTED** that there may be a need to revise the number of age classes used in the models when using a different growth curve due to shift in the distribution of size at age

Yellowfin

1.1 Update on Management Strategy Evaluation Progress

7. The WPTT **NOTED** paper IOTC–2018–WPM09–10, which provided an update on the development of the operating model for IOTC yellowfin tuna (October 2018).

8. The WPTT **NOTED** paper IOTC–2018–WPM09–11, which provided an update on IOTC bigeye and yellowfin management procedure evaluation progress (October 2018), including the following abstract provided by the authors:

> *" This document presents MP evaluation results for bigeye and yellowfin tunas, using the new operating models (OMs) proposed in Kolody and Jumppanen (2018a, b) and the new tuning levels requested by TCMP (2018). The results of various robustness scenarios are included, at this point largely to help facilitate the discussion of their role in the MP development and selection process and how they should be presented to the TCMP."*

9. The WPTT reviewed and **ENDORSED** the progress to date on MSE for yellowfin tuna while recognizing the discussions held at TCMP and the advice of WPM, but **INDICATED** the need to alter some of the assumptions used in the operating model grid and consider some additional uncertainty dimensions in the yellowfin tuna MSE workplan agreed by WPM.

10. The WPTT **NOTED** the need to modify the assumed time required to achieve mixing of tagged YFT with the untagged population to 4 quarters (from 3 quarters) based on decisions taken for the 2018 YFT stock assessment. Further, the WPTT **ENCOURAGED** that the MSE work consider the importance of also assuming the time needed for mixing of the tagged and untagged populations of 8 quarters for use in examining robustness of MPs to this assumption.

11. The WPTT **ENCOURAGED** that the MSE work consider the importance of alternative growth for yellowfin tuna based on the growth model estimated by Dortel (2014) for use in examining robustness of yellowfin MPs to alternative growth models.

12. The WPTT further **ENCOURAGED** that the MSE work also consider the importance of adding the Purse Seine Free School CPUE as documented in IOTC–2018–WPTT20–36_Rev1, assuming a 1% per year cumulative increase in catchability (q) for the time period, for use in examining robustness of yellowfin MPs.

13. The WPTT also **NOTED** that the decisions taken for the 2018 YFT assessment regarding short-term and chronic tag loss differed from the YFT Operating Model grid and **RECOMMENDED** that the 2018 YFT assessment assumptions be mimicked in the Operating Model grid.

14. The WPTT **NOTED** that the proposed new uncertainty dimensions would be evaluated with respect to plausibility and impact before deciding whether to assign them to the OM reference set or robustness trials. The informal MSE working group will review these decisions in March 2019.