COMMISSION
FIFTEENTH REGULAR SESSION
Honolulu, Hawaii, USA
10 – 14 December 2018

Performance indicators for comparing management procedures using the MSE modelling framework - revised following discussion at WCPFC-SC14.

WCPFC15-2018-15
25 October 2018

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

1. Oceanic Fisheries Programme, The Pacific Community
2. Te Takina Ltd.

This paper is an updated version of WCPFC-SC14-2018/MI-WP-04 that was presented at WCPFC-SC14. WCPFC-SC14 recommended that this paper be forwarded to WCPFC15.

All changes to the paper have been notified in the text. The main changes are that some of the candidate performance indicators which cannot be calculated currently are retained in the suite of performance indicators and will be developed further instead of being moved to the monitoring strategy. In particular, the performance indicator that relates to maximising SIDS revenues from resource rents should continue to be developed.

Contents

Executive Summary for WCPFC

A key element of the harvest strategy approach is the development and use of a range of performance indicators for evaluating the relative performance of candidate management procedures. The WCPFC13 Summary Report Attachment M (WCPFC, 2017) includes an initial list of performance indicators for Tropical Purse Seine Fisheries for the purpose of the evaluation of HCRs. The original list included 20 proposed indicators, 11 of which were suggested for inclusion by the SWG. The indicators have four categories: Biological, Economic, Social and Ecosystem.

This paper calculates a demonstration set of performance indicators from Attachment M (WCPFC, 2017). The indicators are generated from the demonstration management strategy evaluation (MSE) framework for skipjack that is under development (Scott et al., 2018a). The indicators are calculated over three time periods (short-, medium- and long-term). The length of the time periods are based on the number of management cycles, currently assumed to be three years. It is not possible to calculate all of the indicators in Attachment M (WCPFC, 2017) and it is noted that some indicators are challenging to interpret and therefore may not be appropriate.

The MSE framework considers multiple sources of uncertainty resulting in a distribution of values for each indicator. This distribution can offer additional information on the performance of a management procedure that is not captured by only considering a single summary value, e.g. the median.

Methods for comparing the relative performance of management procedures using the performance indicators are explored. In particular, there is often a trade-off in performance between different management procedures and it is important that this is captured by the values of indicators. This is made more challenging when there are a large number of indicators. It is therefore advantageous to reduce the number of performance indicators to those that capture the main results of interest and discard those that provide duplicate information. The key indicators to be considered will depend on the stock under analysis and the agreed objectives for the fishery. Comparing the relative performance of management procedures requires scaling or transforming the indicator values so that they can be easily compared across management procedures. Approaches include scaling to one of the candidate management procedures, to a 'status quo' scenario or a historical period.

The indicators calculated here represent a provisional set of proposed performance indicators and more indicators can be developed during development of the harvest strategy if necessary. For example, an indicator based on the mean weight of an individual in the stock is presented here.

It is important to note that the values for the indicators presented here are preliminary and are not intended to be used for management purposes. Also, the harvest control rules used in the analysis were chosen to provide contrast in the indicators and are not necessarily candidates for the fishery.

This paper has been updated following discussion at WCPFC-SC14. All changes have been notified in the text. The main changes are that some of the candidate performance indicators which cannot be calculated currently are retained in the suite of performance indicators and will be developed further instead of being moved to the monitoring strategy.

	Objective type	MOW4 Strawman	Performance Indicator (WP14)	Calculated
1	Biological	Maintain SKJ (and YFT and BET) biomass at or above levels that provide fishery sustainability throughout their range	Probability of $SB/SB_{F=0} > 0.2$ as determined from MSE.	Y
2	Economic	Maximise economic yield from the fishery	Predicted effort relative to E_{MEY} (to take account of multi-species considerations, SKJ, BET and YFT may be calculated at the individual fishery level). B_{MEY} and F_{MEY} may also be considered at a single species level.	Y
3	Economic	Maximise economic yield from the fishery	Average expected catch (may also be calculated at the assessment region level)	Y
4	Economic	Maintain acceptable CPUE	Average deviation of predicted SKJ CPUE from reference period levels	Y
5	Economic	Maximise SIDS revenues from resource rents	Proxy: average value of SIDS / non-SIDS catch	N
6	Economic	Catch stability	Average annual variation in catch	Y
7	Economic	Stability and continuity of market supply	Effort variation relative to reference period level (may also be calculated at the assessment region level)	Y
8	Economic	Stability and continuity of market supply	Probability of and deviation from $SB/SB_{F=0} > 0.5$ (SKJ) in the short-, medium- and long-term as determined from MSE (may also be calculated at the assessment region level)	Y
9	Social	Food security in developing states (import replacement)	As a proxy: average proportion of CCMs-catch to total catch for fisheries operating in specific regions	N
10	Social	Avoid adverse impacts on small scale fishers	• MSY of SKJ, BET, YFT • Possible information on other competing fisheries targeting SKJ (may also be calculated at the assessment region level) • Any additional information on other fisheries / species as possible	N
11	Ecosystem	Minimise bycatch	Number of FAD sets; Expected catch of other species	N

Table 1: Summary of proposed performance indicators for the purse seine skipjack tuna fishery (WCPFC, 2017). The *Calculated* column notes whether or not the indicator can be calculated using the current operating models.

1 Introduction

Performance indicators are an important component of a harvest strategy. During the harvest strategy approach, candidate management procedures (MPs) are evaluated using management strategy evaluation (MSE, Figure 1) (Punt et al., 2014; Scott et al., 2016a, 2018a). Performance indicators are used to evaluate how well a candidate MP is expected to perform and enable the selection of a preferred option from a range of candidate procedures. They are interpreted in relation to reference points and management objectives. A reference point often implies that a specific target value is desired or limit should be avoided. Reference points may not be available for all management objectives since very often it is wanted to maximise something relative to some other objective rather than achieve a specific value. In this case performance is measured relative to other management objectives rather than against a defined reference point (Scott et al., 2016b). It is common to compare the performance of candidate MPs in a relative sense, i.e. "MP A outperforms MP B on performance indicator X" (Scott et al., 2016c).

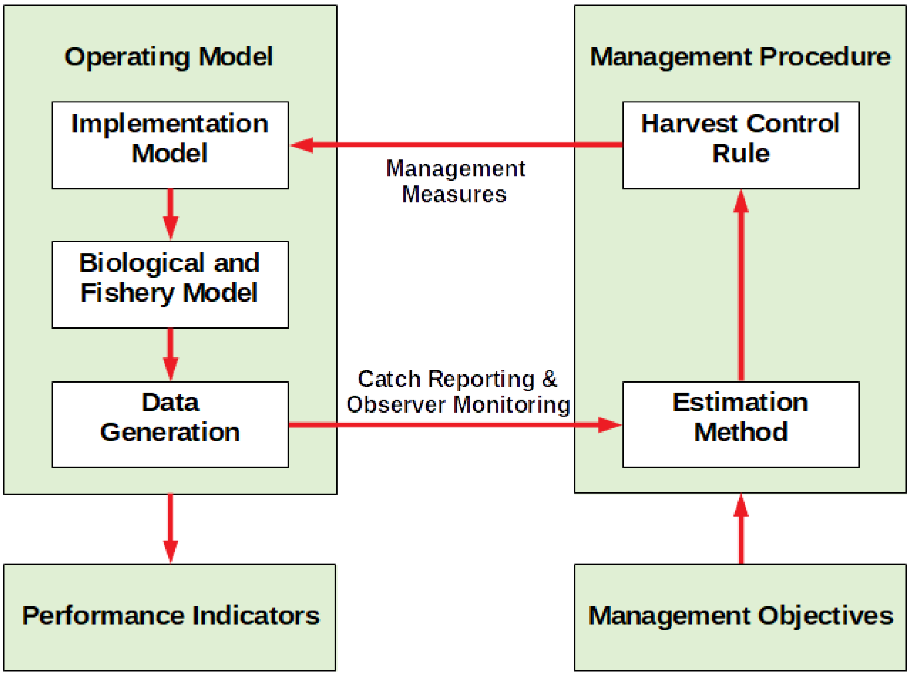

Figure 1: Conceptual diagram of the MSE framework (after Punt et al. (2014)).

The suite of performance indicators that will be used to evaluate the MPs should be agreed by the relevant stakeholders. This is an iterative process and the desired indicators can change as the harvest strategy continues to be developed. An important signal from the indicators is the trade-offs between them i.e. some candidate MPs will score highly on some indicators but less well on others. For example, an MP that results in relatively high catches may have a lower probability of stock sustainability. Another important property of an indicator is that it is easy to communicate and understand.

The report of the second WCPFC Management Objectives Workshop proposed ’strawman’ objectives for WCPO fisheries (WCPFC, 2013). The WCPFC13 Summary Report Attachment M includes an initial suite of performance indicators for Tropical Purse Seine Fisheries for the purpose of the evaluation of HCRs (Scott et al., 2016b; WCPFC, 2017). SPC was requested to continue the work on developing MPs based on the suggested indicators as much as possible. The calculation of the results over the short-, medium- and long-term was also requested. The original list included 20 proposed indicators, 11 of which were suggested for inclusion by the SWG (Table 2). The indicators have four categories: Biological, Economic, Social and Ecosystem.

In this paper the initial proposed suite of indicators are calculated where possible using a demonstration set of results from a 'proof of concept' MSE based on skipjack (see Section 2). Not all of the indicators can be calculated and some are challenging to interpret (see Section 3). Following discussions at the WCPFC-SC14, it is suggested that some of these are not considered further as part of the suite of performance indicators. Instead, they should remain as part of the monitoring strategy. It is also suggested that some of the indicators that cannot be calculated should remain in the suite of performance indicators but require further development.

Methods for analysing and presenting the indicators and the trade-offs between them are explored (see Section 4). The performance indicators, including methods for the calculation, are summarised in Annexes A and E. A description of the demonstration MSE, including summary outputs, is presented in Annexes B and C.

This paper therefore:

- Reviews the initial proposed suite of performance indicators;
- Presents details on how indicators can be calculated using a demonstration set of results for skipjack;
- Identifies proposed indicators which can not be calculated or which provide information that is already captured by other indicators.

2 Calculation of the performance indicators

In this section the calculation of each of the proposed performance indicators in Table 2 is described and discussed. The indicators are calculated using the outputs from a demonstration MSE for skipjack (Section B). For reasons that will be described it is not possible to calculate all of the proposed indicators (Section 3).

Three HCRs are tested (Section B.2). It is important to note that the chosen HCRs do not represent actual candidate HCRs for skipjack. They have been deliberately chosen so that their relative performance differs sufficiently and they provide contrast in the performance indicators.

As requested, indicators are calculated over three time periods: short-, medium- and long-term (WCPFC, 2017). Here these time periods are based on the number of management cycles in each period. In this demonstration the management cycle is 3 years, i.e. the MP is called every 3 years and the resulting future fishing level determined through the harvest control rule (HCR) is fixed for the following 3 years. In this example, projections start in 2016 with the MP first being called in 2017. Short-term is taken to be the first two management cycles (2017-2019, 2020-2022), medium-term is taken to be the next two management cycles (2023-2025, 2026-2028) and long-term is taken to be the final two management cycles (2029-2031, 2032-2034).

The projections performed within the MSE are stochastic and consider multiple sources of uncertainty (see Section B). However, in Table 2 there is no mention of the distribution of the values of the indicator implying that only the central value, e.g. the median, should be considered. This means that there is no consideration of risk and uncertainty which is potentially important. For example, two HCRs may produce the same median catches but the distributions of the catches may be different which may influence which HCR was preferred. To illustrate this point the distributions of the indicator values have been considered here by including the 20th and 80th percentiles in the following tables and plots.

As mentioned above, performance indicators can be interpreted in relation to reference points. When an appropriate reference point is not available for a particular management objective the indicator is used to evaluate the relative performance of the MP rather than the actual performance. This means that when selecting an MP, attention should be paid to its performance relative to the performance of the other candidate MPs as well as the actual projected outcomes. For example, when considering an indicator that is based on expected catches it is important that the relative expected catches of MPs are considered as well as the actual projected expected catches. To allow the comparison of the relative performance it may be necessary to transform the results of some indicators. There are a number of options for the transformation including scaling the indicator values so that they are relative to the performance in a historical time period, relative to a 'status quo' management scenario or relative to one of MPs. The transformation method will need to be discussed by stakeholders. In this section of the report the untransformed values are presented.

The indicators considered below do not represent the final set of performance indicators. Alternative indicators can be tested and adopted throughout the process of developing a harvest strategy. For example, indicators that explicitly consider risk and uncertainty can be developed and tested.

It should be noted that the figures and values presented here do not represent preliminary results. They are provided only to act as a 'proof of concept' and to illustrate how the performance indicators may be calculated.

Recommendation:

- Base the short-, medium- and long-term time periods on the number of management cycles (e.g. for skipjack each time period is 2 management cycles of 3 years);
- Consider the distribution of indicator values as well as the central measure.

2.1 Indicator 1. Maintain SKJ biomass at or above levels that provide fishery sustainability throughout their range

The indicator is the probability of $SB/SB_{F=0} > 0.2$ where 0.2 is the LRP for skipjack. The operating model (OM) reports the biomass (SB) and the unfished biomass ($SB_{F=0}$) by season and region (Scott et al., 2018a). The indicator is based on the mean annual biomass which is calculated by taking the mean biomass over the seasons. Although it is possible to calculate the indicator for each of the five regions in the OM here the total biomass across all regions is considered. The unfished biomass is taken as the mean of the unfished biomass over the years 2005 to 2014 (note that here this is not a moving window).

The probability of $SB/SB_{F=0} > 0.2.$ in a particular year is calculated as the proportion of the simulations in a year in which $SB/SB_{F=0} > 0.2$ (Figure 18):

$$PI1_y = \frac{\sum_{n=1}^{N} SB/SB_{F=0,n,y} > 0.2}{N}$$

where $PI1_y$ is the probability of $SB/SB_{F=0} > 0.2$ in year y and N is the number of simulations.

The indicator in each time period is then calculated as the average probability over the different time periods (short-, medium-, long-term).

$$PI1 = \sum_{y=y_1}^{y_2} Pr_y/Y$$

where y_1 and y_2 are the start and end years of the time period and Y is in the number of years in the time period.

The results show that all HCRs have an indicator value of 1 for each time period meaning that the $SB/SB_{F=0}$ is always greater than 0.2 (Table 3). This means that, in this particular case, this indicator is not useful for deciding between these particular HCRs with this set of OMs. However, this is not guaranteed to always be the case and alternative OMs and HCRs may give different probabilities. The indicator values are also reflected in the box plot of the distribution of $SB/SB_{F=0}$ over the three time periods which also allows performance to be related to the TRP (Figure 2).

There is contrast in the performance of the three HCRs despite this indicator having the same value (Figures 2 and 17). *HCR1* tends to have higher values of $SB/SB_{F=0}$ than the other HCRs and also a wider distribution of those values. *HCR2* tends to have lower values of $SB/SB_{F=0}$ but is still always above 0.2. *HCR3* has the narrowest distribution of values. There is very little difference between the three time periods.

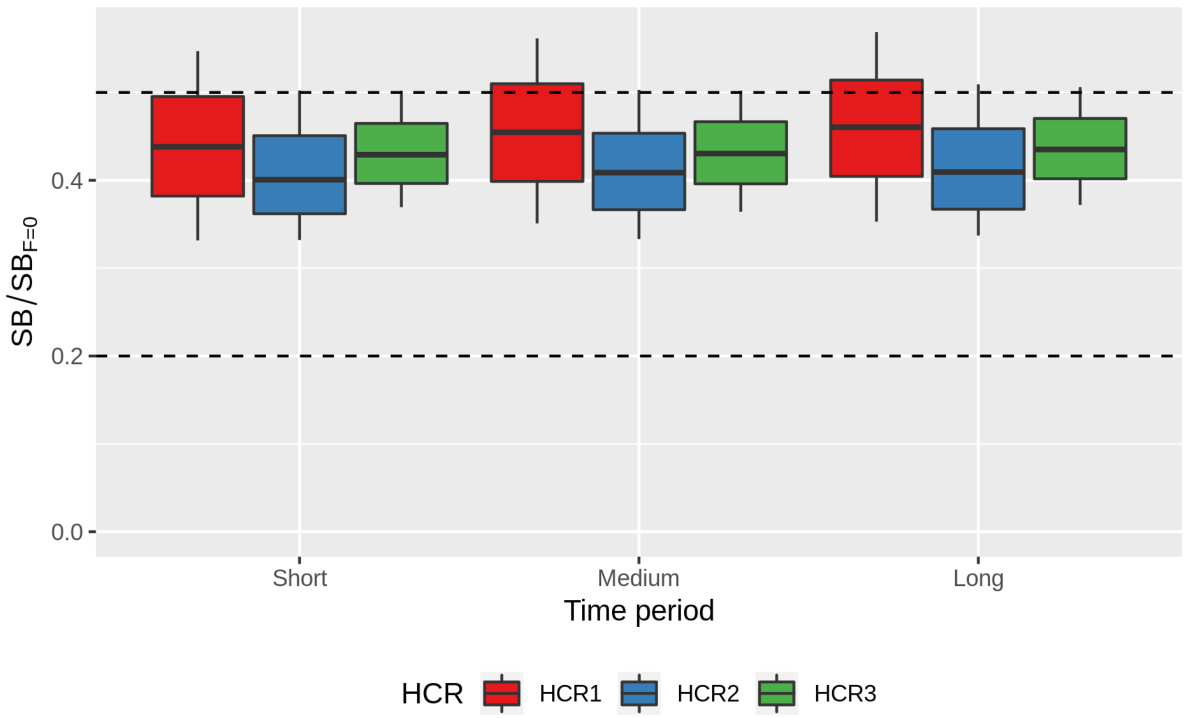

Figure 2: Distribution of $SB/SB_{F=0}$ in the three time periods (short-, medium- and long-). The boxes capture the 20-80th percentiles, the vertical lines capture the 5-95 percentile and the horizontal line is the median. The dashed horizontal lines at 0.2 and 0.5 reflect the LRP that performance indicator 1 is measured against the interim TRP respectively.

2.2 Indicator 2. Maximise economic yield from the fishery (predicted effort relative to E_{MEY})

This indicator is based on the predicted effort, E, relative to E_{MEY} (Figure 19). Additionally, the predicted SB relative to SB_{MEY} and fishing mortality, F, relative to F_{MEY} may also be considered. Calculating this indicator therefore requires values for E_{MEY}, B_{MEY} and F_{MEY}. Here we focus on E and SB. To illustrate the calculation of this indicator E_{MEY} and SB_{MEY} are taken to be the values of E and SB in 2010.

Regarding the effort based indicator, the OM reports effort at the fishery level which means it could be calculated at the fishery, regional and overall total level. However, it is not always possible to combine the fishing effort from the different fisheries in the OM as the historical effort of each fishery may or may not have been standardised (this does not affect the simulations as internally the effort is scaled). The effort based indicator is therefore calculated for the purse seine fisheries operating in regions 2, 3 and 5, excluding the associated purse seine fishery in region 5 which has been standardised. The biomass indicator is calculated for the whole region.

The indicator is concerned with E and B relative to E_{MEY} and B_{MEY}. It is assumed that the closer E and B are to E_{MEY} and B_{MEY} the better the MP is considered to be performing and the further away E and B are from E_{MEY} and B_{MEY} the worse the MP is considered to be performing. To reflect this the indicator is calculated as the absolute difference between E/E_{MEY} and 1. This measure is symmetrical, i.e. being greater than E_{MEY} is considered to be the same as being less than E_{MEY}. This means that the lower the value of the indicator the better the MP is considered to be performing in relation to this indicator. The indicator is calculated for each simulation, n, in each year, y:

$$PI2_{n,y} = |E_{n,y}/E_{MEY} - 1|$$

The average within the three time periods is then taken to give a distribution of N indicator values for each time period (Figure 3):

$$PI2_n = \frac{\sum_{y=y_1}^{y_2} PI2_{n,y}}{Y}$$

The same calculation is performed for the SB based indicator. The median value and 20th and 80th percentiles of the distributions are calculated (Table 3).

Noting the illustrative nature of these analyses, the biomass based indicator, *HCR1* can be thought of as performing relatively worse than the other two HCRs as it has higher values and also a wider distribution. *HCR2* performs the best with the lowest value. This is also the case when considering the effort based indicator, the only difference being that the indicator for *HCR1* has the narrowest distribution. This means that in relation to this indicator, *HCR2* performs the best and *HCR1* performs the worst. There is very little difference between the three time periods.

The use of this indicator is strongly dependent on the availability of estimates of E_{MEY} (and possibly B_{MEY} and F_{MEY}). It should be noted that for other stocks and fisheries, reference to E_{MEY} may relate to alternative economic levels.

This indicator is one of two that are concerned with maximising economic yield (the other being indicator 3, Table 2). Although it is possible to calculate this indicator, considering effort and biomass may not be the best guide to maximising the economic yield and it is recommended that these are not considered further. An alternative indicator for maximising economic yield is to calculate the economic rent. This would have the additional advantage of reducing the number of candidate performance indicators.

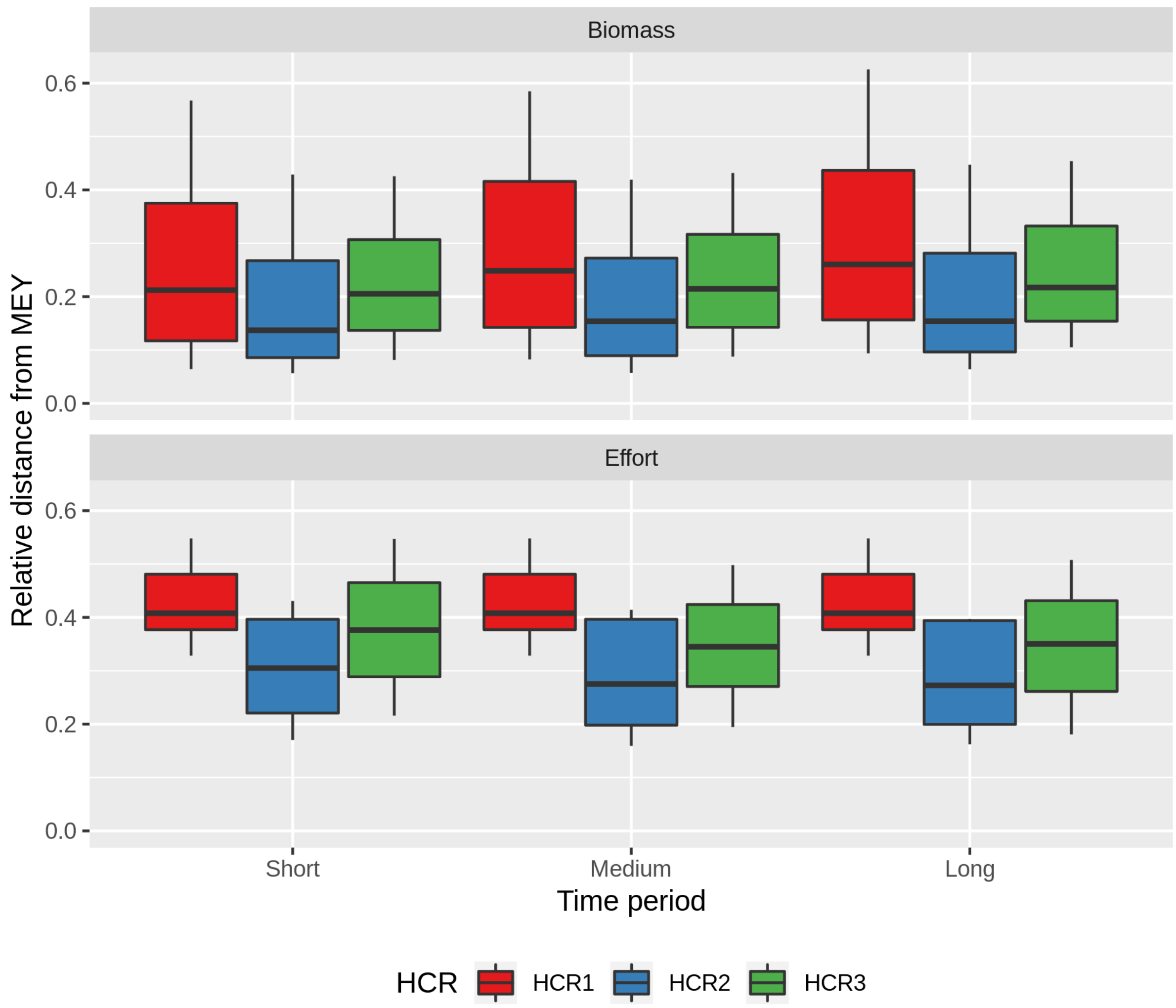

Figure 3: Distribution of performance indicator 2 (predicted E relative to E_{MEY} and B relative to B_{MEY}) in the three time periods (short-, medium- and long-). The higher the value, the worse the MP is considered to be performing in relation to this indicator. The boxes capture the 20-80th percentiles, the vertical lines capture the 5-95 percentile and the horizontal line is the median.

2.3 Indicator 3. Maximise economic yield from the fishery (average expected catch)

This indicator is based on the average expected catch. MULTIFAN-CL reports catches at the region and fishery level. It is therefore possible to calculate this indicator for each fishery and each region in the model. However, this would result in a large number of values for this indicator, making it challenging to understand the overall relative performances of the candidate MPs. Here, the indicator is calculated for the total expected catches across all regions and fisheries and also for the total of the expected catches from the tropical purse seine fisheries operating in regions 2, 3, and 5 (Figures 20 and 21).

The indicator is calculated by taking the average catch of each simulation over the desired year range (short-, medium- and long-term):

$$PI3_n = \frac{\sum_{y=y_1}^{y_2} C_{n,y}}{Y}$$

Where C is either the total catch of the whole region or of only the tropical purse seine fisheries in regions 2,3 and 5. This gives a distribution of N expected catches in each time period (Figure 4). The median value and 20th and 80th percentiles of the distribution are then calculated (Table 3).

The pattern of results are the same for the total catches and the catches from the purse seine fisheries. This is because the output of the HCR (an effort multiplier) applies equally to all fisheries and regions (see Section B). As the output of the HCR changes, the effort of all fisheries changes by the same proportion and so the relative pattern of fishery catches remains approximately the same. This suggests that under a system of WCPO-wide control, when developing a catch based indicator for evaluating the relative performance of MPs, it is only necessary to include an indicator based on the total expected catches. It is not necessary to include additional indicators that are based on the expected catches of sub-groups of fisheries. However, the catches of sub-groups of fisheries could still be considered as part of the monitoring strategy.

HCR2 produces the highest expected catches and *HCR1* produces the lowest. However, the expected catches from *HCR1* have a narrower distribution, i.e. there is less uncertainty in expected catches. This could be preferred to higher expected catches that are less certain. This illustrates the importance of considering the distribution of the indicator values as well as the median values. There is very little difference between the three time periods.

This indicator is presented in terms of absolute expected catches. It would be possible to calculate the indicator relative to the catches in a past period, e.g. in 2010. This could make it easier to compare the relative performances of the candidate MPs.

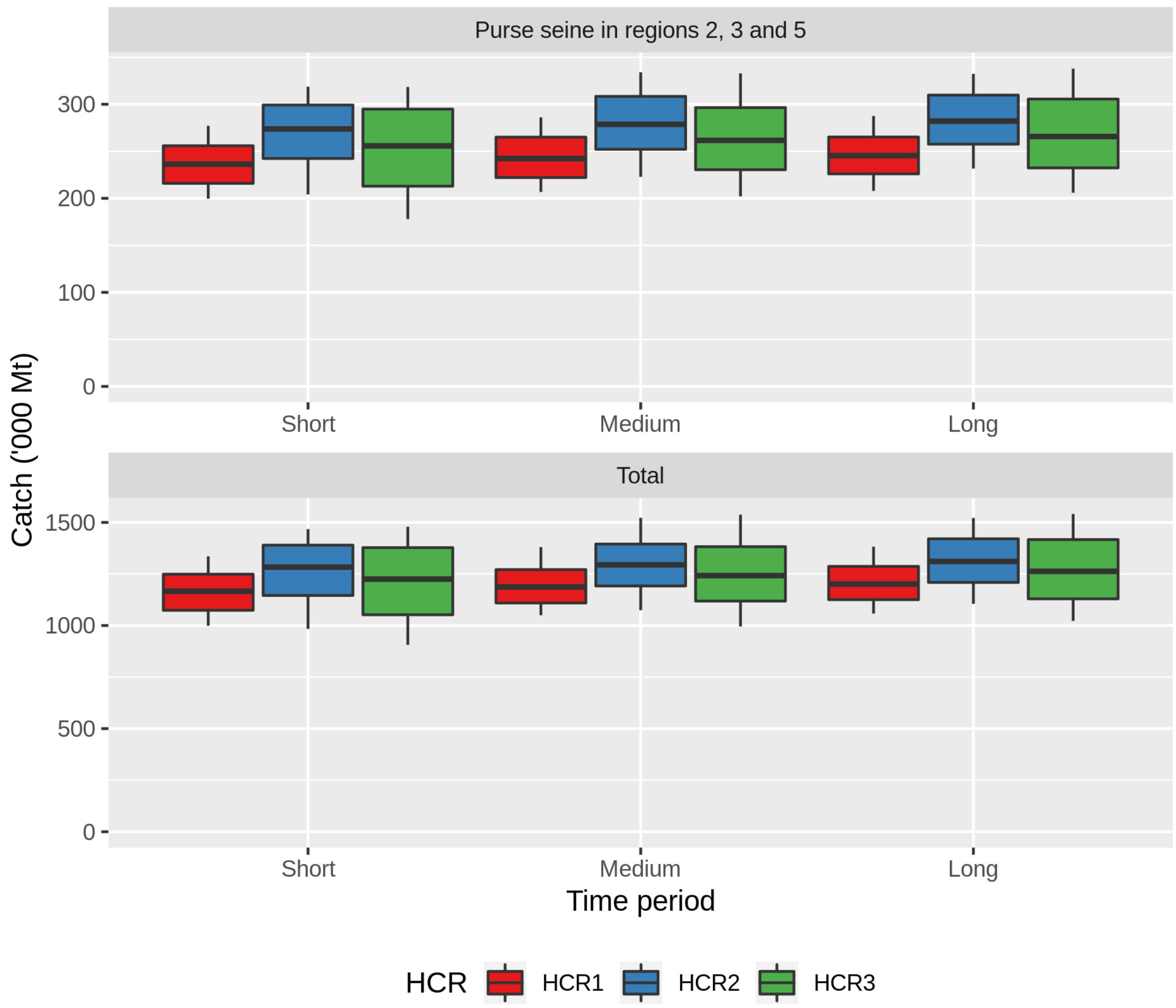

Figure 4: Distribution of performance indicator 3 (average expected catch) in the three time periods (short-, medium- and long-). The expected catch for the purse seine fisheries in regions 2,3 and 5 and the total catch are shown. The boxes capture the 20-80th percentiles, the vertical lines capture the 5-95 percentile and the horizontal line is the median.

2.4 Indicator 4. Maintain acceptable CPUE

This indicator is based on the average deviation of predicted skipjack CPUE from reference period levels. It is calculated as the CPUE relative to the CPUE in a reference period. Here the reference period is taken to be 2010.

As mentioned above, the OM reports effort and catch at the fishery level which means it could be possible to calculate the indicator at the fishery, region and total level. However, it is not always possible to combine the fishing effort from the different fisheries in the OM as the historical effort of each fishery may or may not have been standardised (this does not affect the simulations as internally the effort is scaled). This indicator is therefore calculated in each year for the purse seine fisheries operating in regions 2, 3 and 5 excluding the associated purse seine fishery in region 5 which has a standardised effort index (Figure 22):

$PI4_{n,y} = CPUE_{n,y}/CPUE_{ref}$

The average relative CPUE is calculated for each simulation over the desired year range (short-, medium- and long-term).

$$PI4_n = \frac{\sum_{y=y_1}^{y_2} PI4_{n,y}}{Y}$$

This gives a distribution of N values of relative CPUE in each time period (Figure 5). The median value and 20th and 80th percentiles of the distribution are then calculated (Table 3).

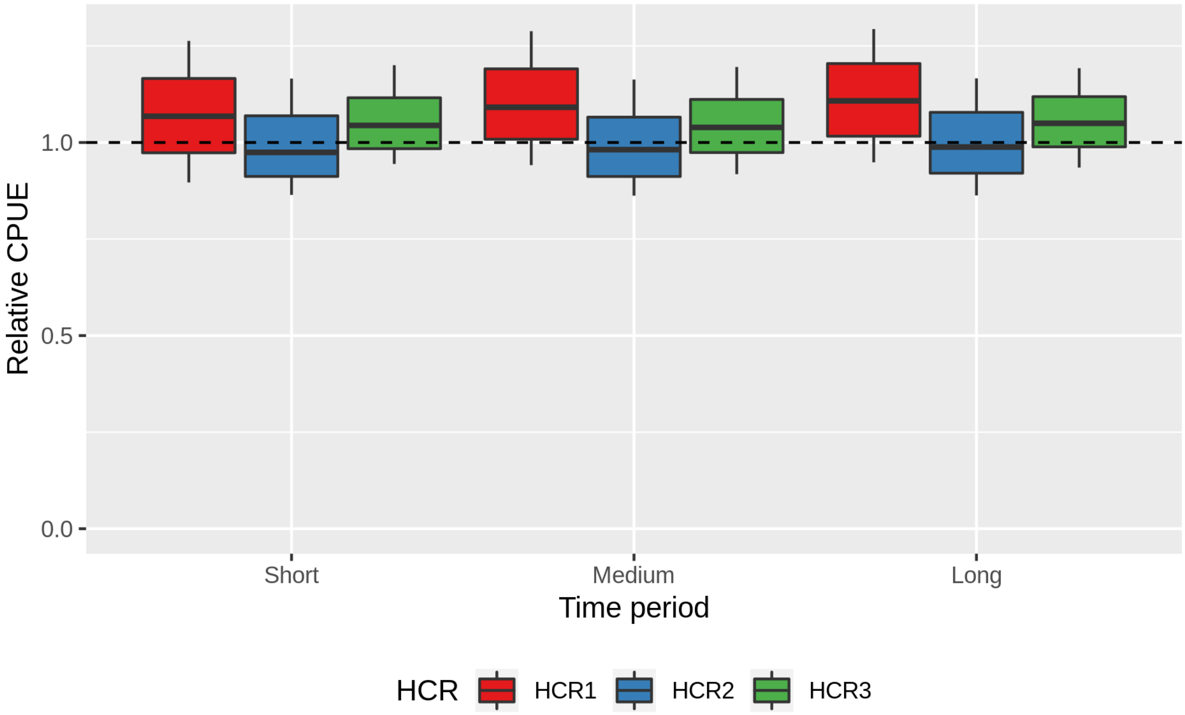

Figure 5: Distribution of performance indicator 4 (maintain acceptable CPUE relative to reference CPUE) in the three time periods (short-, medium- and long-). Only the CPUE of the purse seine fisheries in regions 2,3 and 5 are considered (excluding the associated purse seine fishery in region 5). The boxes capture the 20-80th percentiles, the vertical lines capture the 5-95 percentile and the horizontal line is the median.

HCR2 can be considered to perform worse than the other two HCRs as it has lower values. *HCR1* has the highest median value which would suggest that it performs the best. However, it has a slightly wider distribution of values for *HCR1* which mean that there is a higher degree of uncertainty. *HCR3* has a lower median value than *HCR1* but it has a smaller distribution, implying greater certainty, and this might be preferred. This illustrates the importance of considering the distribution of the indicator values and not just the median. There is no difference between each time period.

2.5 Indicator 6. Catch stability

This indicator is concerned with the average annual variation in catch. As with indicator 3, it is possible to calculate this indicator by fishery and region. Here it is calculated over the whole region and for the combined purse seine fisheries in regions 2, 3 and 5.

The indicator is calculated by taking the absolute annual difference of the catch for each simulation and in each year:

$$PI6_{y,n} = |C_{y+1,n} - C_{y,n}|$$

The absolute annual difference is then averaged over the desired year range (short-, medium- and long-term):

$$PI6_n = \frac{\sum_{y=y_1}^{y_2} PI6_{n,y}}{Y}$$

This gives a distribution of N values in each time period (Figure 6). The median value and 20th and 80th percentiles of the distribution are then calculated (Table 3).

This indicator measures the variation in the catch, i.e. the higher the value of the indicator, the less stable the catch is. It is assumed that an MP that produces a lower catch variation, i.e. a low value of the indicator, would be preferred. This indicator could be scaled by the mean catch so that it reflects the average stability in relative catch.

As with indicator 3, there is a similar pattern of results when looking at the total catch and only the catch of purse seine fisheries. *HCR1* produces the lowest values and also the narrowest distribution which means that it can be considered to perform better than the other two HCRs for this indicator. This means that the expected catches tend to be similar year-on-year. However, as seen in Section 2.3 *HCR1* tends to produce the lowest expected catches.

This is a good example of a trade-off between HCRs. *HCR1* tends to produce lower but stable catches, whereas *HCR2* tends to produce higher but less stable catches. Managers will need to decide which performance indicators and fishery objectives have a higher priority.

HCR3 has the highest variation in catch. As explained in Annex C, the threshold parameter of *HCR3* is relatively high compared to the other HCRs and is above the interim Target Reference Point (TRP). This means that even when SB is at or above the interim TRP, the fishery is on the sloping part of the HCR and it is likely that a new effort is set every management cycle (Figures 16 and 19). Consequently, under *HCR3* the fishery spends a lot of time moving up and down the slope of the HCR and the resulting effort and catches fluctuate. Conversely, *HCR1* has a low threshold parameter meaning that the effort and catches are less likely to change between management cycles.

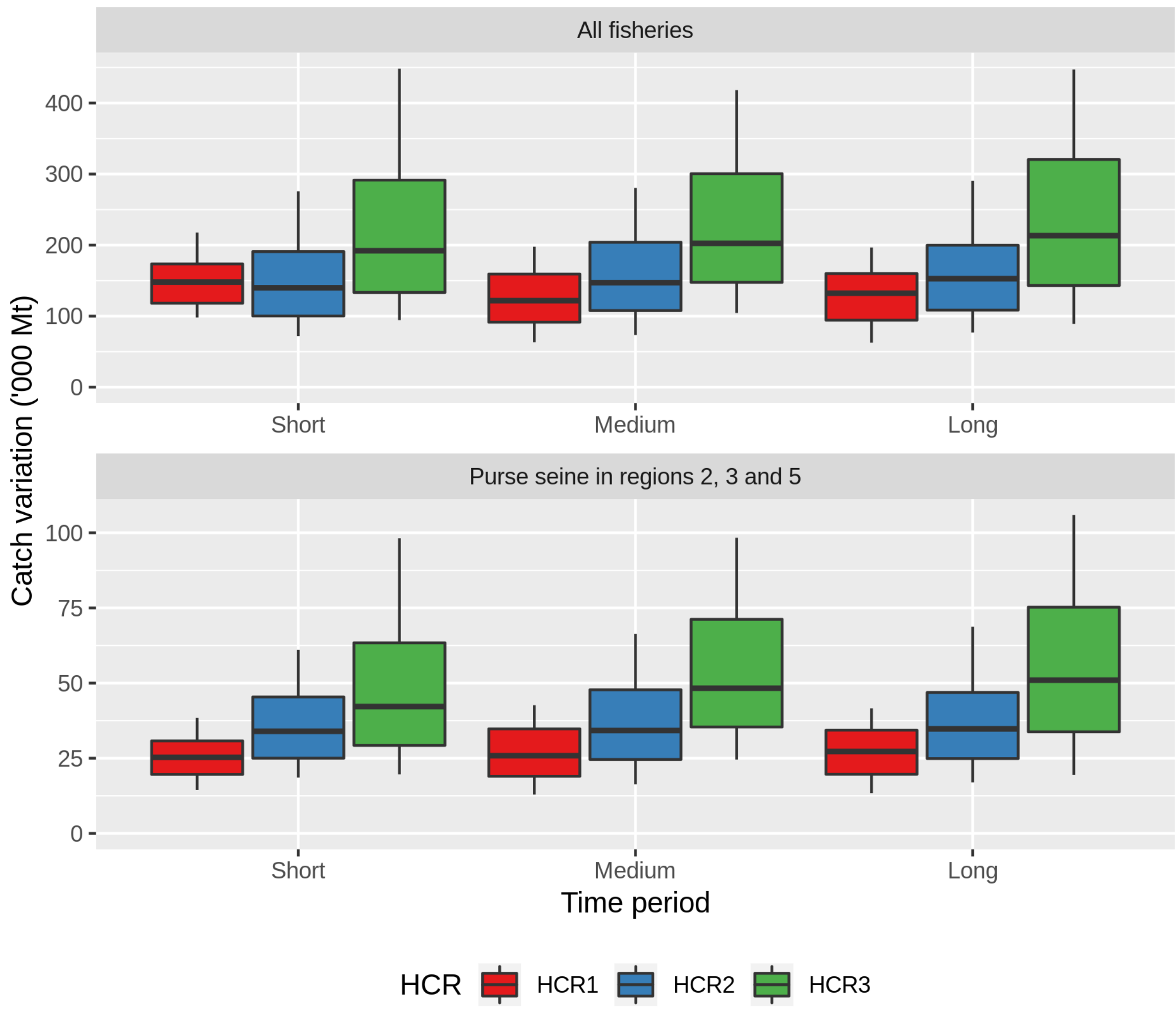

Figure 6: Distribution of performance indicator 6 (catch variation) in the three time periods (short-, medium- and long-). The catch variation for the purse seine fisheries in regions 2,3 and 5 and for all fisheries are shown. The boxes capture the 20-80th percentiles, the vertical lines capture the 5-95 percentile and the horizontal line is the median.

2.6 Indicator 7. Stability and continuity of market supply (effort variation relative to a reference period)

This indicator is concerned with effort variation relative to the effort in a reference period, i.e. stability of the relative effort. Here the reference period is taken to be 2010. As mentioned above, the OM reports effort at the fishery level which means this indicator could be calculated at the fishery, region and total level. Here, as with performance indicator 2 (see Section 2.2), this indicator is calculated for the purse seine fisheries operating in regions 2, 3 and 5 excluding the associated purse seine fishery in region 5 which has a standardised effort index.

This performance indicator is similar to indicator 6 (catch stability). However, it may still be worth considering as part of the suite of candidate indicators as it is closely related to the output of the HCR rather than the OM. This may be of interest when it comes comparing MPs.

The indicator is calculated in a similar way to performance indicator 6. The absolute annual difference of the effort relative to the base effort (in 2010) is calculated for each simulation in each year:

$$PI7_{y,n} = |E_{y+1,n}/E_{ref} - E_{y,n}/E_{ref}|$$

The absolute annual difference is averaged over the desired year range (short-, medium- and long-term):

$$PI7_n = \frac{\sum_{y=y_1}^{y_2} PI7_{n,y}}{Y}$$

This gives a distribution of the average absolute annual differences (Figure 7). The median value and 20th and 80th percentiles of the distribution are then calculated (Table 3).

Like indicator 6, this indicator measures the variation in the relative effort, i.e. the higher the value of the indicator, the less stable the relative effort is. It is assumed that MPs that produce a low value of this indicator are considered to be perform better. For *HCR1* the variation of relative effort is 0 with no distribution for the medium- and long-term time periods and almost 0 for the short-term time period. As explained in Section C, when this HCR is operating the fishery spends almost all of its time on the flat part of the HCR meaning that the effort level does not change between management cycles (Figures 16 and 19). *HCR3* has the highest variation in effort because the fishery spends most of the time moving up and down the slope of the HCR (as described in Section 2.5). There is a small change in performance over time.

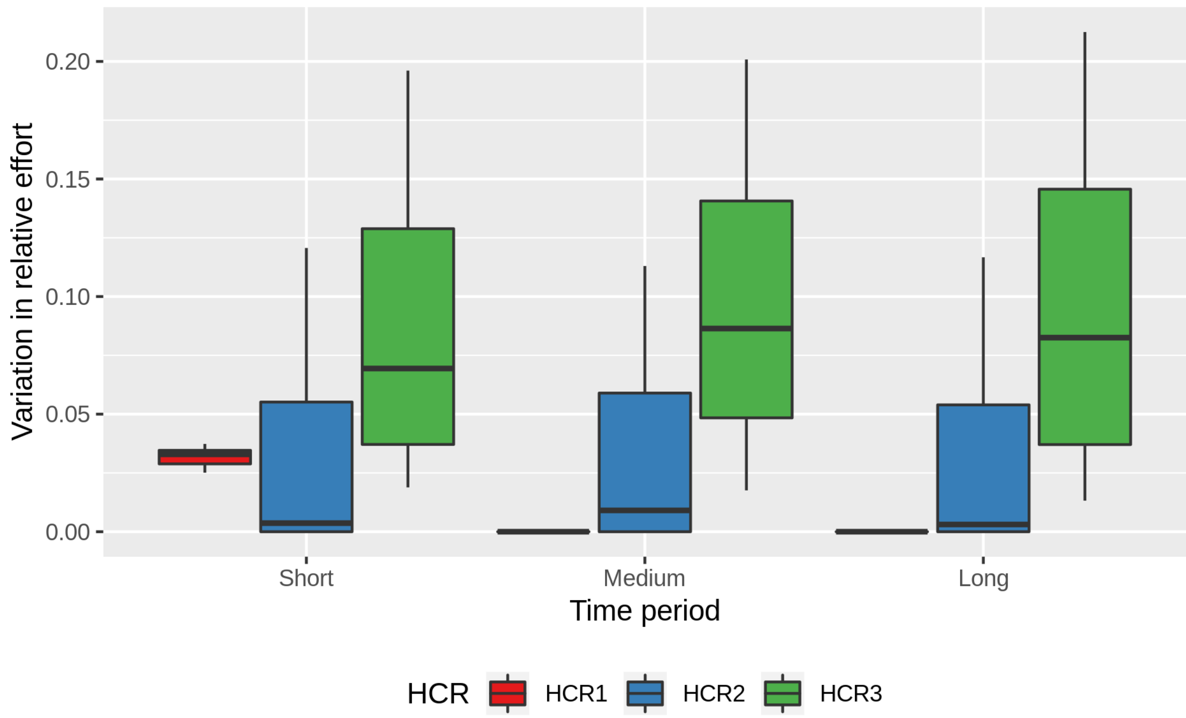

Figure 7: Distribution of performance indicator 7 (stability and continuity of market supply) in the three time periods (short-, medium- and long-). The indicator is based on the change in relative effort for purse seine fisheries operating in regions 2, 3 and 5 (excluding the associated purse seine fishery in region 5). The boxes capture the 20-80th percentiles, the vertical lines capture the 5-95 percentile and the horizontal line is the median.

2.7 Indicator 8. Stability and continuity of market supply (probability of and deviation from $SB/SB_{F=0} > 0.5$)

This indicator is concerned with maintaining the stock size around the TRP levels (where the interim TRP for skipjack is $SB/SB_{F=0} = 0.5$). As described in Table 2, the indicator has two values: the probability of and deviation from $SB/SB_{F=0} > 0.5$. It is assumed that the further away $SB/SB_{F=0}$ is from 0.5, the worse the MP can be thought of as performing, i.e. it is better to have $SB/SB_{F=0}$ close to 0.5 on average.

The probability of $SB/SB_{F=0} > 0.5$ (indicator 8a) is calculated in the same way as performance indicator 1 (Figure 23 and Table 3).

The deviation (indicator 8b) is calculated as the absolute difference between $SB/SB_{F=0}/0.5$ and 1, for each simulation in each year (similar to performance indicator 2), i.e. the further away $SB/SB_{F=0}$ is from the TRP (in either direction, higher or lower) the higher the value:

$$PI8_{n,y} = |SB/SB_{n,y,F=0}/0.5 - 1|$$

The average over the three time periods is then taken to give a distribution of N values for each time period (Figure 8):

$$PI8_n = \frac{\sum_{y=y_1}^{y_2} PI8_{n,y}}{Y}$$

The median value and 20th and 80th percentiles of the distribution are then calculated (Table 3). The lower the value the better the MP is considered to be performing at maintaining $SB/SB_{F=0}$ at the TRP, i.e. a value of 0 means that $SB/SB_{F=0}$ is always exactly equal to the TRP and never deviates from it. In Figure 8 it can be seen that *HCR1* tends to have the lowest values meaning that it tends to be closer to the TRP than the other HCRs and thereby performs better for this indicator.

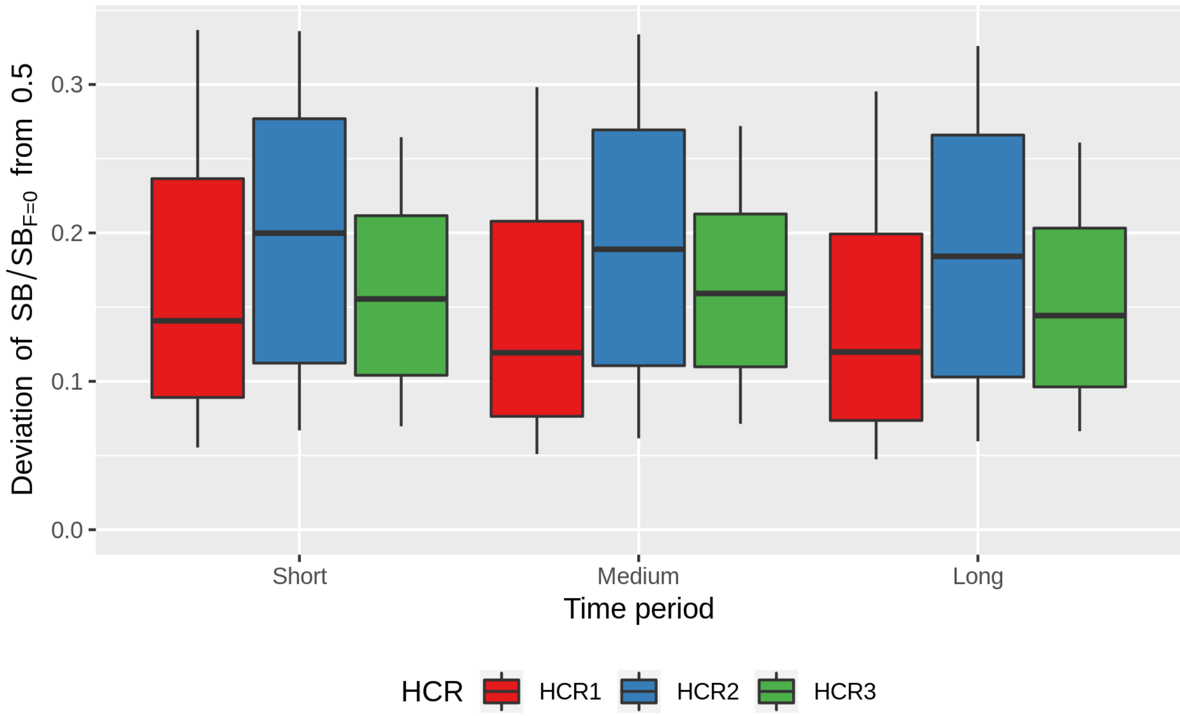

Figure 8: Distribution of performance indicator 8 (deviation from $SB/SB_{F=0} = 0.5$) in the three time periods (short-, medium- and long-). The boxes capture the 20-80th percentiles, the vertical lines capture the 5-95 percentile and the horizontal line is the median.

This pair of indicators is hard to interpret in the context of selecting between candidate MPs. The probability indicator is only concerned with the possibility of $SB/SB_{F=0}$ being above the TRP but has no information on how far away it is from the TRP. For example, if the probability indicator had a value of 0.5 it would mean that half the time $SB/SB_{F=0}$ is above the TRP and half the time below it, i.e. we are fluctuating around the TRP. This might seem like a good thing because it means that we are not always above or always below the TRP. However, we do not know how big the fluctuations are. This information is provided by deviation indicator which tells us how far away $SB/SB_{F=0}$ is from the TRP. However, the deviation indicator does not tell us the direction of the deviation. If the deviation indicator is high, meaning $SB/SB_{F=0}$ is far away from the TRP, we might want to know if $SB/SB_{F=0}$ is above or below the TRP or fluctuating around it. It is therefore necessary to consider the two indicators together to get a full understanding of $SB/SB_{F=0}$ relative to the TRP.

The status of $SB/SB_{F=0}$ relative to the TRP can provide important information on the status of the fishery. Information on the status of $SB/SB_{F=0}$ can be visualised by plotting its distribution (Figure 2). As mentioned in Section 2, an important quality of an indicator is that it is easy to communicate and understand. However, this is not the case for this pair of indicators and their inclusion in the suite of potential indicators should be reconsidered. Following discussions at WCPFC-SC14, an alternative indicator could be developed that considers $SB/SB_{F=0}$ relative to the TRP, for example the probability of $SB/SB_{F=0}$ being with a certain of range of the TRP (e.g. within $TRP \pm 10\%$).

2.8 Summary

Recommendation:

- Reconsider the use of predicted effort relative to E_{MEY} (and the B and F equivalents, indicator 2) as an indicator for maximising the economic yield and use an alternative indicator based on the economic rent;
- Further develop indicator 8 (probability of and deviation from $SB/SB_{F=0} > 0.5$) for its inclusion in the suite of potential performance indicators.

3 Performance indicators that were not calculated

Not all of the indicators listed in Table 2 can be calculated from the operating model (OM). Additionally, the information provided by an indicator may also be provided by other indicators. Following discussion at WCPFC-SC14, it is suggested that some of these indicators should not be considered further as part of the suite of potential performance indicators. Others should still be retained in the suite of potential indicators but require further development. Indicators that are no longer included in the suite of potential indicators may still be included as part of the monitoring strategy.

3.1 Indicator 5: Maximise SIDS revenues from resource rents

It is suggested that the proxy for this indicator is the ratio of the average value of SIDS to non-SIDS catch (see Table 2). Unfortunately, it is not possible to calculate this indicator in its current form from the current OMs. The fisheries in the biological and fishery model component of the OM are classified by region and gear type and not by country or state. This means that it is not possible to attribute the catches of the fisheries to SID or non-SID as this information is not available from the OM.

It may be possible to split projected future fishery catches into SIDS and non-SIDS components by using assumptions based on the historical catch distribution, e.g. assume that the future distribution of catches between SIDS and non-SIDS is the same as the mean over a particular year range. However, there are a number of problems with this approach. The main concern is that the value of the indicator would be very strongly influenced by the assumption that the future catch distribution is constant and related to the past. This is unlikely. As this assumption is hard to justify the indicator is potentially misleading. Additionally, this indicator will also be impacted by policy choices that cannot be predicted.

Following discussions at WCPFC-SC14, it is suggested that this indicator is retained in the suite of potential indicators but continues to be developed.

It is worth noting that if the total resource rent of the tuna fishery is maximised then that will probably be a positive outcome for the SIDS. This is covered by performance indicator 3 (maximise economic yield, see Section 2.3).

3.2 Indicator 9: Food security in developing states (import replacement)

It is suggested that the proxy for this indicator is the average proportion of CCMs-catch to total catch for fisheries operating in specific regions (see Table 2). As for indicator 5 above, it is not possible to calculate this indicator as the information for attributing the catches to CCMs is not available in the OM. It may be possible to make some assumptions about attributing catches to CCMs (e.g. based on historical distributions) but, as discussed above for performance indicator 5, the value of the indicator would be very strongly influenced by these assumptions, making the indicator potentially misleading.

There is an additional problem with the indicator in that it is calculated using a proportion. This may be not be appropriate and could give misleading results. For example, if the total catch for the fisheries strongly increases and the CCMs-catch also increases but not by as much, the proportion will show a negative outcome despite there being an increase in the CCMs-catch.

It is worth noting that there is overlap between this indicator and other indicators. Indicator 9 is specifically concerned with food security in developing states. If $SB/SB_{F=0}$ is sustainable (as measured by indicators 1 and 8) then it can be assumed that there should be no problem with food security for developing states.

This indicator should not be considered as part of the suite of potential indicators but should be retained as part of the monitoring strategy.

3.3 Indicator 10. Avoid adverse impacts on small scale fisheries

This indicator is concerned with how small scale fisheries may be affected by management plans, including information on other fisheries. However, there are a number of challenges in calculating this indicator.

One challenge is that it is not clear what is meant by 'small scale fisheries' in the indicator description. As mentioned above, individual fishing fleets are not modelled in the OM. If small scale means artisanal fisheries then it is not possible to calculate this indicator as this information is not available in the OM. An alternative interpretation is that small scale means non-purse seine fisheries for skipjack. This is possible to calculate but may be of limited value. As described in Section B, if the MP and HCR that will be developed are WCPO wide and apply equally to all fisheries then output of the HCR will affect the effort of all fisheries by the same proportion. This would mean that there is no need to investigate the performances of sub-groups of fisheries as the performance of all fisheries are likely to respond in a similar same way to the MP. For example, the expected total catches of all the fisheries show a similar pattern to the expected catches of only the purse seine fisheries in regions 2, 3 and 5 (Table 3).

Following dicussion at WCPFC-SC14 it was clarified that this indicator 10 also considers the MSY of the stocks. This could be calculated from the OM and included in the suite of potential indicators.

3.4 Indicator 11: Minimise bycatch

This indicator considers the number of FAD sets and the expected catch of other species. However, the number of FAD sets is not included in the OM (associated effort being in 'days' in the assessment model, though this might change in the future) and it is not possible to make robust assumptions about FAD sets that could result.

Concerning the expected catch of other species only the main tuna species are considered in the OMs at this time (for example, the OM for skipjack is a single species model) and so it is not possible to model the bycatch of other species. However, the bigeye and yellowfin tuna harvest strategy evaluation will use an OM that includes both species so that the catches of both are considered together and linked to that of the purse seine fishery.

It is not possible to calculate this indicator at present. However, it may be possible to in the future. Following discussions at WCPFC-SC14, it could still be retained in the suite of potential indicators.

3.5 Summary

Recommendation:

- Indicator 5 cannot be calculated in its current form but should be developed further and included as part of the suite of potential indicators;
- Do not include indicator 9 in the suite of potential performance indicators. It should still be considered for the monitoring strategy;
- By calculating MSY values for indicator 10 it could be considered as part of the suite of potential indicators;
- Indicator 11 cannot currently be calculated from the operating model. It may be possible to calculate it in the future and could still be retained in the suite of potential indicators.

4 Presenting indicators and trade-offs

Performance indicators are used to compare the relative performance of candidate MPs and measure how well each MP achieves the objectives of the fishery. An indicator is not considered in isolation but as part of a suite of indicators that collectively describe the performance of an MP. The MP which is judged to have the 'best' values overall for the indicators can be thought of as having the best overall performance in terms of meeting the management objectives. There may be trade-offs between the indicators in which case a decision will have to be made as to which indicators are more important, possibly using a pre-agreed weighting based on the prioritised objectives for the fishery.

In this section we compare the three candidate MPs (each with a different HCR, see Section B) by considering the indicators described in Section 2. Comparing the MPs through the indicators is not a trivial task. It is important the indicators are easily understood and can be presented in a way that allows an informed decision to be made. This can become more difficult as the number of indicators increases. It is good practice to present the results in a number of ways, as different managers will have different preferences. It was noted during the discussions at WCPFC-SC14 that examining the time series of dfferent metrics, such as those presented in Section C, can also provide information on the performance of an MP.

As mentioned above, performance indicators are used to evaluate the relative performance of a suite of MPs rather than the actual performance of an individual MP (Scott et al., 2016c). To compare the MPs it is not necessary to compare across indicators, e.g. it is not necessary to compare the value for indicator 1 to indicator 3. However, it is easier to visualise the relative performances of the MPs if the indicators are are on a similar scale. Additionally, for some of the indicators the MP is thought to be performing well if the value is high, e.g. a high probability of $SB/SB_{F=0} > LRP$ (indicator 1) is better than a low probability, whereas for other indicators the MP is thought to be performing well if the value is low e.g. variation in relative effort (indicator 7). It is easier to compare the indicators if they all 'point the same way', e.g. the higher the value of the indicator the better the MP is thought to be performing. This means that to compare the indicators it is helpful to first transform them. Care must be taken with the transformation method so that the relative differences within the indicators are preserved. There are a number of options for the transformation. For example, the indicator values can be scaled so that they are relative to the performance in a historical time period, relative to a 'status quo' management scenario or relative to one of MPs.

In Section 2 the distribution of values for each indicator for each MP was calculated. The distribution is from the 1000 simulations that were run to capture the different sources of uncertainty (see Section B). This means it is possible to consider the distributions of the indicator values, as well measures of the central tendency such as the median. Each indicator was calculated for three different time periods (short-, medium- and long-term).

4.1 Example plots

One method for exploring the results is a simple bar chart of the median values of each indicator for each HCR in each time period. To allow comparison the median values are transformed. For the indicators that 'point up' each indicator is scaled by the maximum value. For the indicators that 'point down' each indicator is scaled by the minimum value, and then further transformed by taking the reciprocal. This ensures that for each indicator, the values across the HCRs range between 0 and 1 where the best HCR has a value of 1 (the higher the value the better the HCR).

Figure 9 shows the transformed median values for the long-term time period only (the short- and medium- term values are similar). The value for indicator 1 (the probability of $SB/SB_{F=0} > 0.2$ is 1 for all three HCRs. For the other indicators there are clear trade-offs between the HCRs. *HCR2* has the highest value for indicator 2 (the predicted E and B relative to E_{MEY} and B_{MEY} and indicator 3 (the expected catch). However, it has the lowest value for indicator 4 (relative CPUE), indicator 7 (effort variation) and indicator 8 (deviation from TRP). *HCR1* has the highest value for indicators 4, 6 (catch variation), 7, and 8 but the lowest values for indicators 2 and 3. *HCR2* is in between the other two apart from indicator 6 for which it has the lowest value.

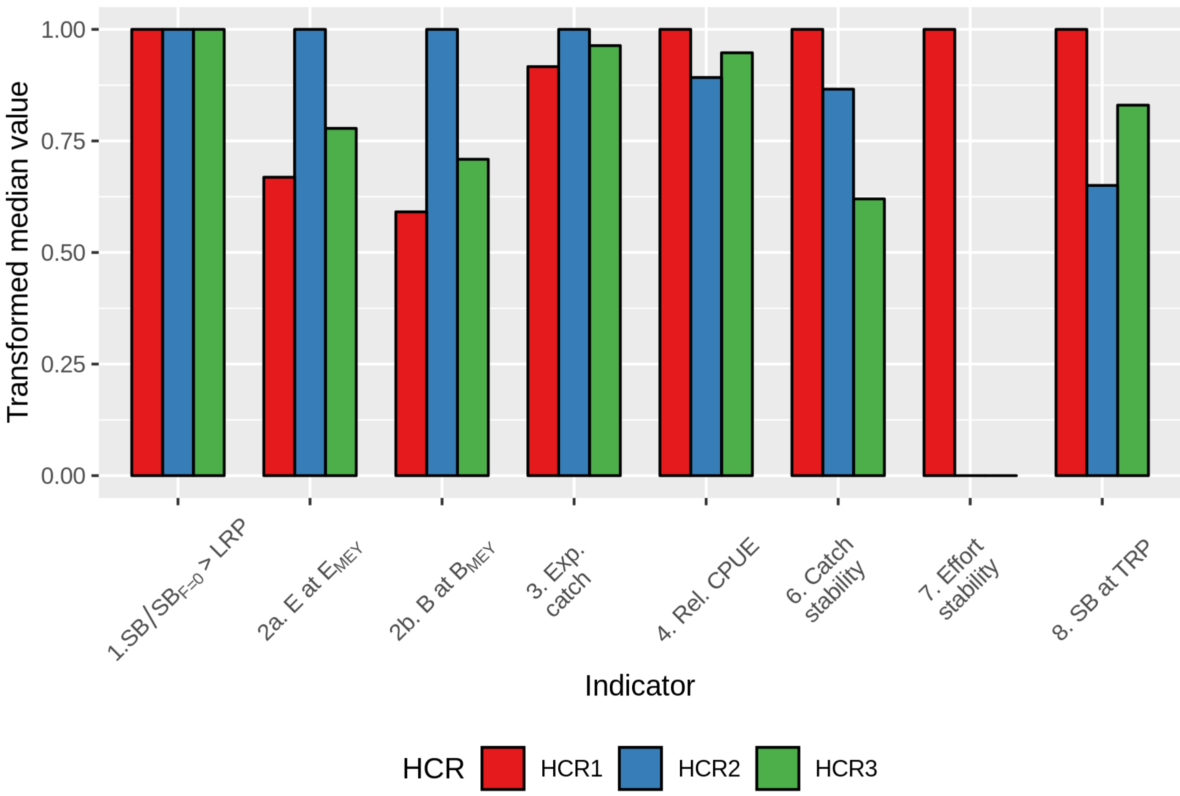

Figure 9: Transformed median values of the indicators in the long-term time period. The higher the value, the better the MP is thought to be performing for that indicator.

The bar chart in Figure 9 presents the transformed median of each indicator but does not include any information on the distribution of the indicator values. A box plot can summarise the median as well the distribution of values. An alternative transformation method is used to produce the box plot because we are now trying to plot a transformed distribution of values, not just the transformed medians. All of the indicators are scaled by the maximum value. For the indicators that 'point down' are further transformed by subtracting it from 1. The distribution of the transformed values range between 0 and 1, but the largest value on the plot is not necessarily 1 as we are only plotting a summary of the distribution.

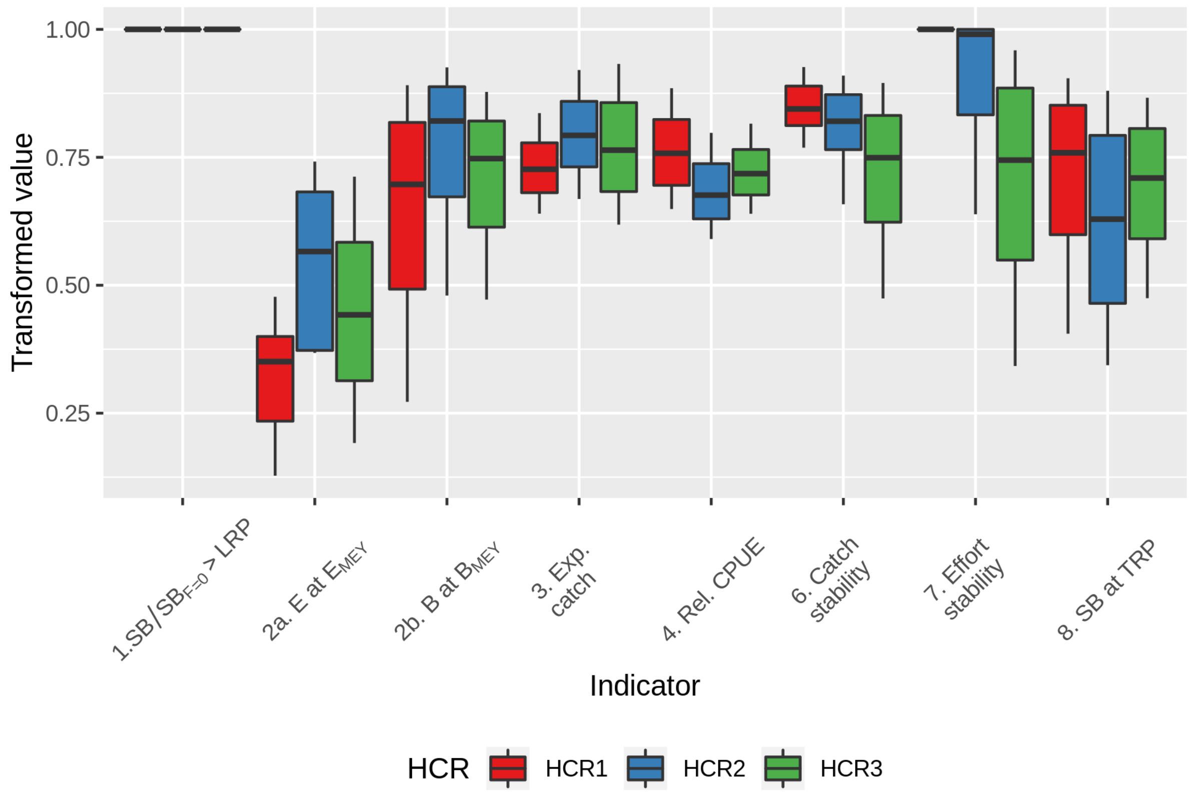

Figure 10: Distribution of the transformed indicators in the long-term time period. The higher the value, the better the MP is thought to be performing for that indicator. The boxes capture the 20-80th percentiles, the vertical lines capture the 5-95 percentile and the horizontal line is the median.

The ranges of the transformed indicators can be seen in Figure 10. The medians follow the same patterns as the bar chart but the distribution is also considered. This can be important. For example, although *HCR2* has the highest values for the effort based indicator 2 (E relative E_{MEY}) it also has the largest distribution of values. It may be preferable to trade-off a narrower distribution for a slightly lower median value for this indicator.

The box plot provides information on the distributions of the indicators but it can be hard to process all the information they contain. Alternative methods for exploring the distributions of the indicator values should be explored.

An alternative presentation method is a 'radar' or 'cobweb' plot. Figure 11 presents a radar plot of the transformed medians of the indicators for each HCR. These are the same values as presented in the bar chart in Figure 9 (only the long-term values are shown). If the indicators have equal weighting (they have the same priority), then the HCR that covers the largest area can be thought of as being the 'best'. These types of plots are sometimes preferred to bar charts. However, they can become confusing as the number of indicators and HCRs increases. Furthermore, like the bar charts, there is no consideration of the distribution of values for the indicators (unlike the box plot).

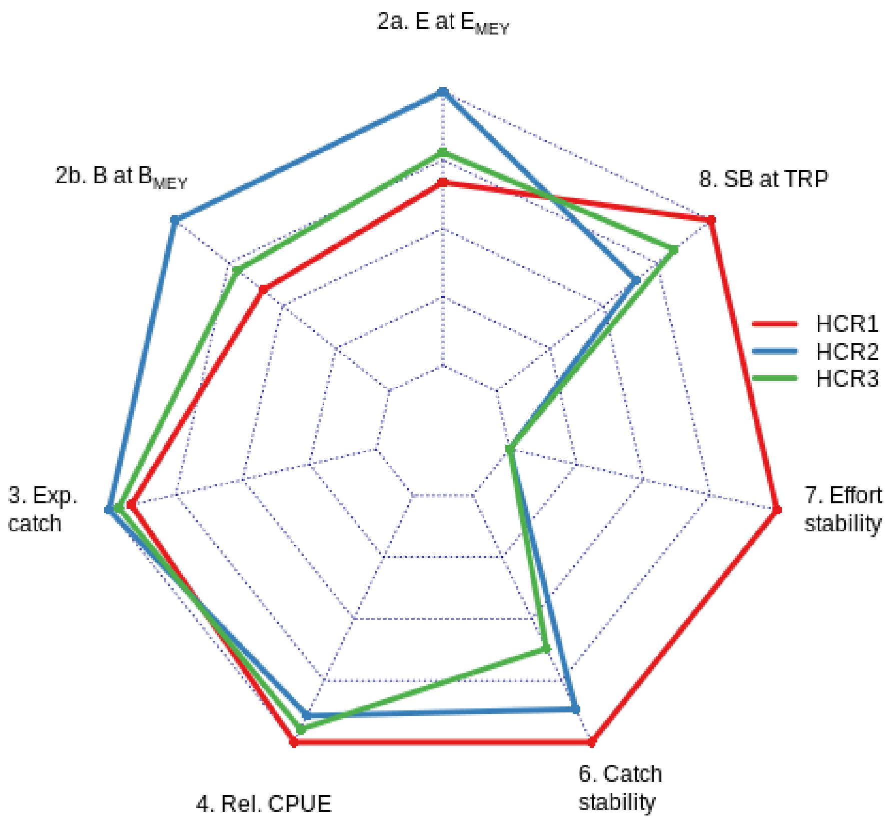

Figure 11: Radar plot of the transformed median values of indicators for each HCR over the long-term time period. These are the same values as presented in the bar chart (Figure 9). Indicator 1 (the probability of $SB/SB_{F=0} > 0.2$) is not shown as the value is 1 for all the HCRs.

An alternative transformation for the radar plot is to scale the values so that the HCR indicator with the highest median value has a value of 1, the lowest has a value of 0, and the others are scaled in between (Figure 12). This can visually enhance the differences between HCRs which can make it easier to select between them and is similar to ranking the values. However, the differences may be artificially increased which can distort the true underlying pattern. In Figure 12 the resulting areas have been shaded to further enhance the differences. Using this type of plot the trade-offs between indicators can be clearly seen. As described above, *HCR2* is the best performing HCR for some indicators (e.g. total expected catch) but the worst for other indicators (e.g. relative CPUE). *HCR1* performs well for the indicators that *HCR2* performs badly at, and vice versa. *HCR3* is neither the best nor the worst HCR and overlaps the other two in the plot.

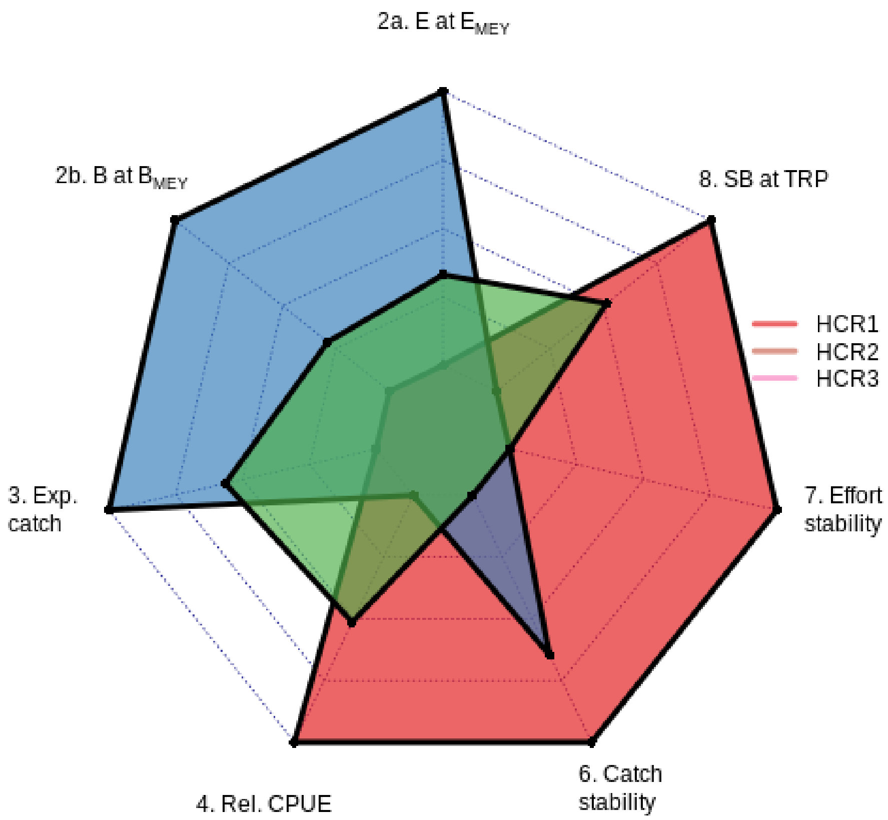

Figure 12: Alternative radar plot of the transformed median values of the indicators for each HCR over the long-term time period. The largest transformed median has a value of 1, the smallest a value of 0 and the other is scaled in between. Indicator 1 (the probability of $SB/SB_{F=0} > 0.2$) is not shown as the value is 1 for all the HCRs.

4.2 Discussion

Many other types of plots and presentation methods are available for communicating multi-variable information and there is no single best way of presenting the information contained in the performance indicators. For example, it is possible to simply rank the MPs by each indicator. This could be presented by colour coding the cells in Table 3. However, ranking ignores the magnitude of the differences between the MPs and can give a distorted impression of their relative performance. The preferred strategy is to continually communicate with stakeholders to develop a range of presentation methods. These presentation methods are likely to evolve during the harvest strategy process but it is good to develop some standard approaches as soon as possible.

Selecting the MP that will be adopted through stakeholder consensus is one of the main goals of the harvest strategy approach. To achieve this goal it will be necessary to synthesise the results from multiple performance indicators across a potentially large range of candidate MPs in such a way that an informed decision can be made. This can become more difficult as the number of performance indicators and candidate MPs increases. One approach is to agree weights for the indicators so that they can be combined into a single metric. However, agreeing weights among stakeholders will not necessarily be easy.

We invite SC14 to:

- Discuss approaches for transforming indicator values and displaying results to aid the comparison of the performances of MPs; and
- Discuss approaches for synthesising results across performance indicators to facilitate the selection of a single MPs.

Recommendation:

- Regularly work with stakeholders to develop a range of preferred presentation methods for communicating performance indicators.

5 Proposed other indicators

As mentioned above it has not been possible to calculate the ecosystem indicator for skipjack as the OMs do not include the required information. However, it is possible to calculate other indicators that reflect the biological status of the stock and which may link to economic indicators. For example, it is possible to calculate size-based indicators for the population. These kind of indicators are important because they can provide information on changes to the size structure of a population as a result of fishing and changes in environmental conditions (Blanchard et al., 2005; Orio et al., 2017).

Here we calculate the mean weight of an individual in the population. This is calculated as the sum of the number of individuals divided by the sum of the total biomass of the stock (Table 3 and Figure 13). Interpretation of this indicator must be done with care. A decrease in mean weight could be interpreted as a bad sign as it implies a reduction in the numbers of large individuals. However, the reduction could also have been caused by an increase in small individuals due to an increase in recruitment which is generally seen as a good thing. Generally, this indicator can be used to identify changes in stock structure which can be a signal for further investigation and management intervention.

There is almost no difference between the HCRs and the time periods for this indicator. Additionally, there is no change through the historical period of the stock (Figure 24). This suggests that the indicator is of limited value. However, this is not guaranteed to always be the case and different OMs and HCRs may show changes in the stock structure that would be detected by this indicator. An alternative approach is to monitor an upper percentile instead of the mean weight.

It is worth noting that a report of OM population size composition from MULTIFAN-CL at defined intervals of the projection period is a development proposed for 2018-19.

This type of indicator could also be included as part of the monitoring strategy as it may be useful for detecting population responses to environmental changes and ecosystem interactions etc.

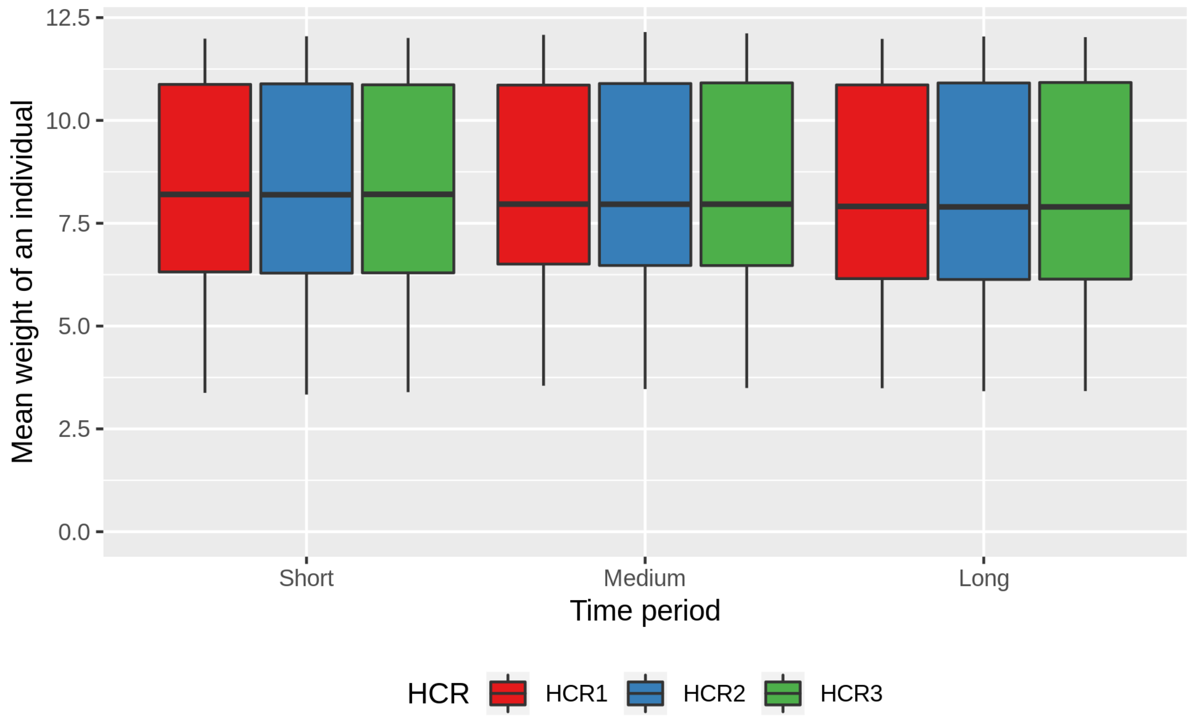

Figure 13: Distribution of the mean weight of an individual in the stock in the three time periods (short-, medium- and long-). The boxes capture the 20-80th percentiles, the vertical lines capture the 5-95 percentile and the horizontal line is the median.

Recommendation:

- Include a size-based indicator that can detect changes in stock structure as a performance indicator and as part of the monitoring strategy.

Acknowledgments

We gratefully acknowledge funding for this work from the New Zealand Ministry of Foreign Affairs and Trade (MFAT) funded project "Pacific Tuna Management Strategy Evaluation"; from the European Union through their funding support for Project 75 (Technical Support for the Development of Harvest Strategies". In addition we thank the Center for High Throughput Computing (CHTC UW-Madison) for generously providing access to their computing resources.

References

Blanchard, J. L., Dulvy, N. K., Jennings, S., Ellis, J. R., Pinnegar, J. K., Tidd, A., and Kell, L. T. (2005). Do climate and fishing influence size-based indicators of celtic sea fish community structure? *ICES Journal of Marine Science*, 62(3):405–411.

Davies, N., Fournier, D. A., Takeuchi, Y., Bouye, F., and Hampton, J. (2018). Developments in the MULTIFAN-CL software 2017-2018. WCPFC-SC14-2018/SA-IP-02, Busan, South Korea, 9–17 August 2018.

McKechnie, S., Hampton, J., Pilling, G. M., and Davies, N. (2016). Stock assessment of skipjack tuna in the western and central Pacific Ocean. WCPFC-SC12-2016/SA-WP-04, Bali, Indonesia, 3–11 August 2016.

Orio, A., Florin, A.-B., Bergström, U., Šics, I., Baranova, T., and Casini, M. (2017). Modelling indices of abundance and size-based indicators of cod and flounder stocks in the baltic sea using newly standardized trawl survey data. *ICES Journal of Marine Science*, 74(5):1322–1333.

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

Scott, F., Scott, R. D., Davies, N., Pilling, G., and Hampton, S. (2018a). Technical developments in the MSE modelling framework. WCPFC-SC14-2018/MI-IP-02, Busan, South Korea, 5–13 August 2018.

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

Scott, R., Pilling, G. M., and Hampton, J. (2016b). Performance statistics and monitoring strategies for skipjack and south Pacific albacore commensurate with: candidate management objectives for the Tropical Purse Seine and Southern Longline Fisheries. Technical Report WCPFC-SC12-2016/MI-WP-04, Bali, Indonesia, 3–11 August 2016.

Scott, R., Pilling, G. M., Hampton, J., Reid, C., and Davies, N. (2016c). Report of the expert consultation workshop on management strategy evaluation. WCPFC-SC12-2016/MI-WP-05, Bali, Indonesia, 3–11 August 2016.

Scott, R. D., Scott, F., Davies, N., Pilling, G., and Hampton, S. (2018b). Selecting, conditioning and weighting of operating models for WCPO skipjack. WCPFC-SC14-2018/MI-WP-03, Busan, South Korea, 5–13 August 2018.

WCPFC (2013). Report of the Expert Working Group on Management Objectives, Performance Indicators and Reference Points. Summary report, Commission for the Conservation and Management of Highly Migratory Fish Stocks in the Western and Central Pacific Ocean, Cairns, Australia, 28–29 November 2013.

WCPFC (2017). Report of the Scientific Committee Thirteenth Regular Session. Technical report, Commission for the Conservation and Management of Highly Migratory Fish Stocks in the Western and Central Pacific Ocean, Rarotonga, Cook Islands, 9–17 August 2017.

A Proposed initial suite of performance indicators

	Objective type	MOW4 Strawman	Performance Indicator (WP14)
1	Biological	Maintain SKJ (and YFT and BET) biomass at or above levels that provide fishery sustainability throughout their range	Probability of $SB/SB_{F=0} > 0.2$ as determined from MSE.
2	Economic	Maximise economic yield from the fishery	Predicted effort relative to E_{MEY} (to take account of multi-species considerations, SKJ, BET and YFT may be calculated at the individual fishery level). B_{MEY} and F_{MEY} may also be considered at a single species level.
3	Economic	Maximise economic yield from the fishery	Average expected catch (may also be calculated at the assessment region level)
4	Economic	Maintain acceptable CPUE	Average deviation of predicted SKJ CPUE from reference period levels
5	Economic	Maximise SIDS revenues from resource rents	Proxy: average value of SIDS / non-SIDS catch
6	Economic	Catch stability	Average annual variation in catch
7	Economic	Stability and continuity of market supply	Effort variation relative to reference period level (may also be calculated at the assessment region level)
8	Economic	Stability and continuity of market supply	Probability of and deviation from $SB/SB_{F=0} > 0.5$ (SKJ) in the short-, medium- and long-term as determined from MSE (may also be calculated at the assessment region level)
9	Social	Food security in developing states (import replacement)	As a proxy: average proportion of CCMs-catch to total catch for fisheries operating in specific regions
10	Social	Avoid adverse impacts on small scale fishers	• MSY of SKJ, BET, YFT • Possible information on other competing fisheries targeting SKJ (may also be calculated at the assessment region level) • Any additional information on other fisheries / species as possible
11	Ecosystem	Minimise bycatch	Number of FAD sets; Expected catch of other species

Table 2: Proposed performance indicators for the purse seine skipjack tuna fishery (WCPFC, 2017).

B Description of the demonstration MSE

The skipjack MSE that is under development (Scott et al., 2018a) has been run with a limited number of OMs and MPs to examine its current performance and to allow demonstration performance indicators to be calculated. The components of the demonstration MSE can be seen in Figure 14. It should be noted that the figures and indicators presented here do not represent results or even preliminary results. They are provided only to act as a 'proof of concept' and to illustrate how the performance indicators may be calculated.

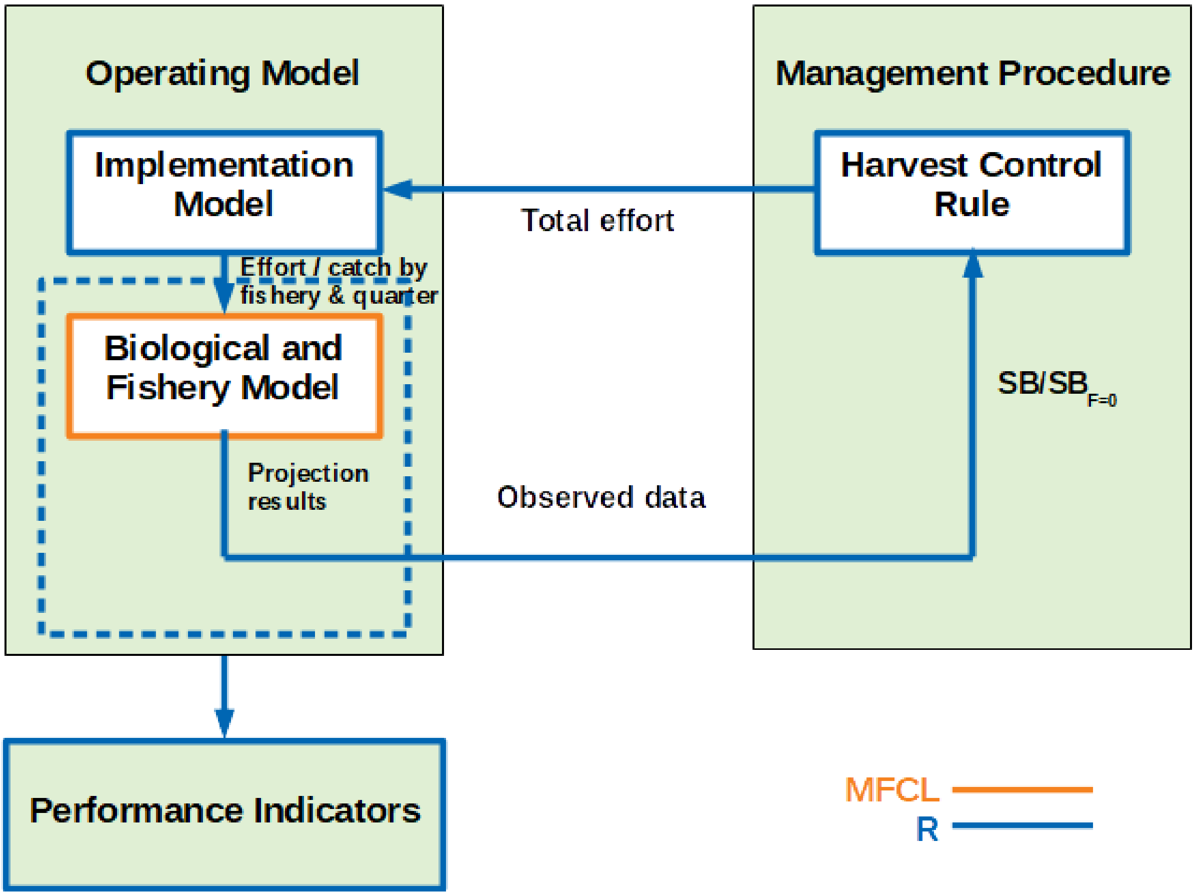

Figure 14: Implementation diagram of the proof of concept MSE framework (Scott et al., 2018a).

The management cycle of the fishery was set to three years, i.e. the MP is called every three years and the output applied during the 3 years. It should be noted that the harvest strategy that is being developed for skipjack is not concerned with modelling effort allocation and so the output of the MP will be applied equally to all fisheries in all regions.

B.1 The operating model

The biological and fisheries model was based on the 2016 skipjack stock assessment which has 23 fisheries and 5 regions (McKechnie et al., 2016). In each of the 5 regions, the fisheries are grouped by gear type. The model is implemented using MULTIFAN-CL (Scott et al., 2018a). It should be noted that the fisheries in the OM do not necessarily correspond to the fisheries of different states and countries. The five long-line fisheries were dropped from the projection as their recent historical catches are set to a low level for assessment purposes.

The skipjack stock assessment uncertainty grid in 2016 was based on four different options:

steepness of the stock recruitment relationship (3 levels), length composition weighting (3 levels), tag mixing (2 levels) and overdispersion (3 levels). The length composition weighting was considered to be the least important and so only the base level was included here (McKechnie et al., 2016). This gives a total of 18 different stock assessments on which to base the OM, covering a wide range of different types of uncertainty (Scott et al., 2018b).

It should be noted that the MPs that will be developed under the WCPFC harvest strategy approach for tuna stocks will be WCPO wide, excluding archipelagic waters. Managers will need to consider whether the MP will apply to all regions and fisheries in the WCPO, or whether there will be a HCR that specifically relates to sub-sets of fisheries such as purse seines, as in the previous modelling for skipjack tuna (Scott et al., 2016a). In this demonstration, all fisheries in the model are equally affected by the output of the MP. The fisheries in the biological and fisheries model are driven by setting their future fishing effort. The fishing effort is held constant for the length of the management cycle (3 years). There is no implementation error.

Future recruitment in the biological model is subject to process uncertainty by applying deviates from the fitted stock-recruitment relationship to the predicted recruitment. There was no uncertainty in the initial population abundance of the the projection period, i.e. numbers at age in the terminal year of the estimation model. This feature is currently under development for MULTIFAN-CL (Davies et al., 2018).

The implementation model and data generation components of the OM are the same in each of the OMs. Therefore, the difference between the OMs in this demonstration is based only on the differences in the biological and fishery model.

The future effort of the fishing fleets is set as the 'base' effort multiplied by the output of the HCR, i.e. all the fishing efforts of the fleets are modified by the same factor. The base fishing effort is taken as the estimated effort in the year 2015. The historical effort is estimated from the assessment results using a new feature of MULTIFAN-CL (Davies et al., 2018). The implementation is assumed to be perfect, i.e. each fishery implemented exactly the prescribed amount of effort.

To account for the process uncertainty in recruitment each OM is simulated multiple times. The simulations from all of the different OMs are combined to give a single set of results with multiple simulations. These simulations contain all of the uncertainties in the different OM assumptions (Scott et al., 2018b). The OMs are assumed to have equal weighting and 56 simulations were drawn from each OM giving a total of 1008 simulations.

B.2 The management procedure

There is no estimation model in the demonstration MSE. Instead, a simple approach to data generation is used where the 'true' value of $SB/SB_{F=0}$ from the operating model is multiplied by samples from a log-normal distribution whose logarithm has mean equal to 0.1 and standard deviation equal to 0.1 (so that the samples has a median of 1.11 and a standard deviation of 0.11) to generate an 'estimated' value of $SB/SB_{F=0}$. This estimated value is used as the input to the HCR. $SB_{F=0}$ is taken to be the mean $SB_{F=0}$ over the years 2005 to 2014. The estimation model is the same for each of the demonstrations which means that the difference between the management procedures (MPs) is only based on the difference between the HCRs that are tested.

Three HCRs are tested in the MP (Figure 15). The HCRs take an estimate of $SB/SB_{F=0}$ as the input and provide an effort multiplier as the output. As mentioned above, in this demonstration the HCR is WCPO wide and the output effort multiplier is applied to the base effort (the effort in 2015) of all the fisheries in the OM so that their relative efforts remain the same during the projections. The HCRs have the same shape (a typical 'threshold' type formulation) and differ only in their parameterisation. All three 'turn off' the fishery when estimated $SB/SB_{F=0}$ falls below the LRP of 0.2 and have an increasing effort multiplier as estimated $SB/SB_{F=0}$ increases until a threshold is reached. The difference between the HCRs is in the position of the threshold and the maximum effort multiplier.

It is important to note that the chosen HCRs do not represent actual candidate HCRs for skipjack. They have been deliberately chosen so that their relative performance differs sufficiently and they provide contrast in the performance indicators.

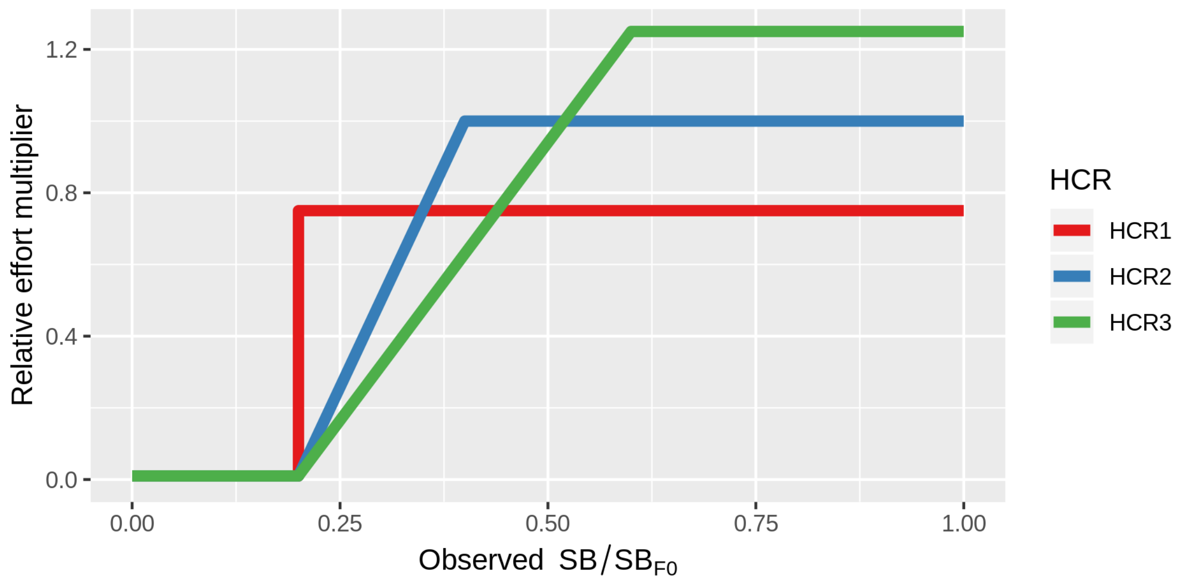

Figure 15: The three HCRs used to demonstrate the MSE framework.

C Outputs of the MSE simulations

The fishery performed differently under the three HCRs. The HCRs take the observed $SB/SB_{F=0}$ as an input and produce an effort multiplier as an output. The occurrence of the observed $SB/SB_{F=0}$ and corresponding effort multiplier of the three HCRs can be seen in Figure 16. The *HCR1* has a low value of the threshold parameter (0.2) which results in the fishery spending all of its time on the upper flat part of the HCR. This means that the effort does not change between management cycles. Conversely, *HCR3* has a relatively high value of the threshold parameter (0.6) which results in the fishery spending more time on the sloping part of the HCR.

This means that the effort and catches are likely to change between management cycles. This is reflected in the performance indicators that measure stability (indicators 6 and 7).

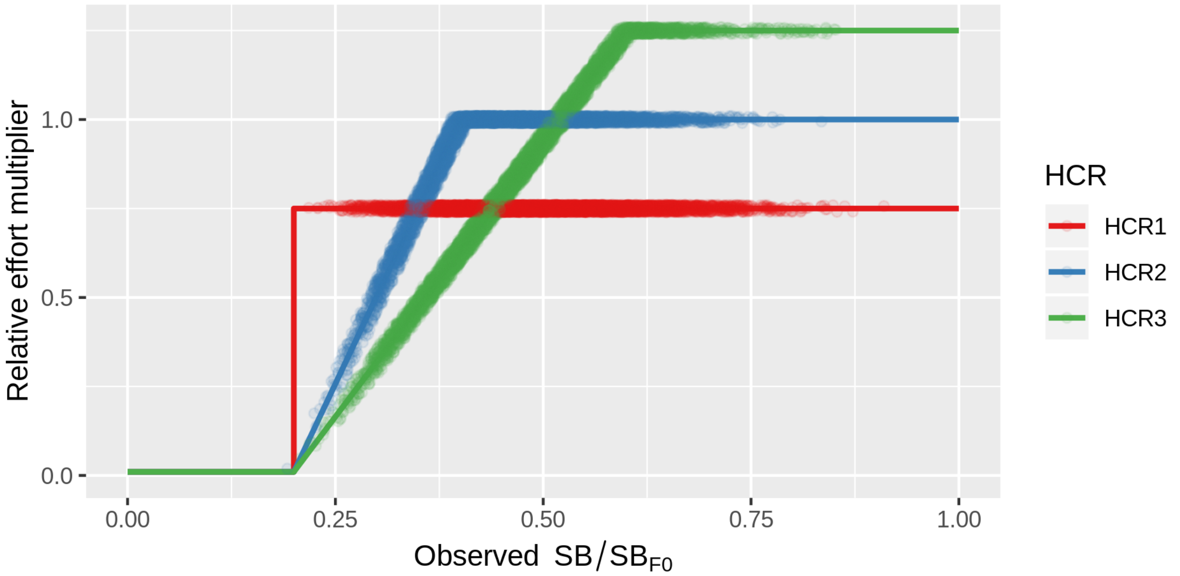

Figure 16: The occurrence of the observed $SB/SB_{F=0}$ and corresponding effort multiplier of the three HCRs. The points have been slightly jittered to better allow analysis.

Figure 17 shows $SB/SB_{F=0}$ over time respectively for the whole region. The variability in the historical period is a result of combining the different OMs, each of which has a different historical stock abundance. Performance indicator 1 is based on the probability of $SB/SB_{F=0} > 0.2$. The time series of the probability is shown in Figure 18.

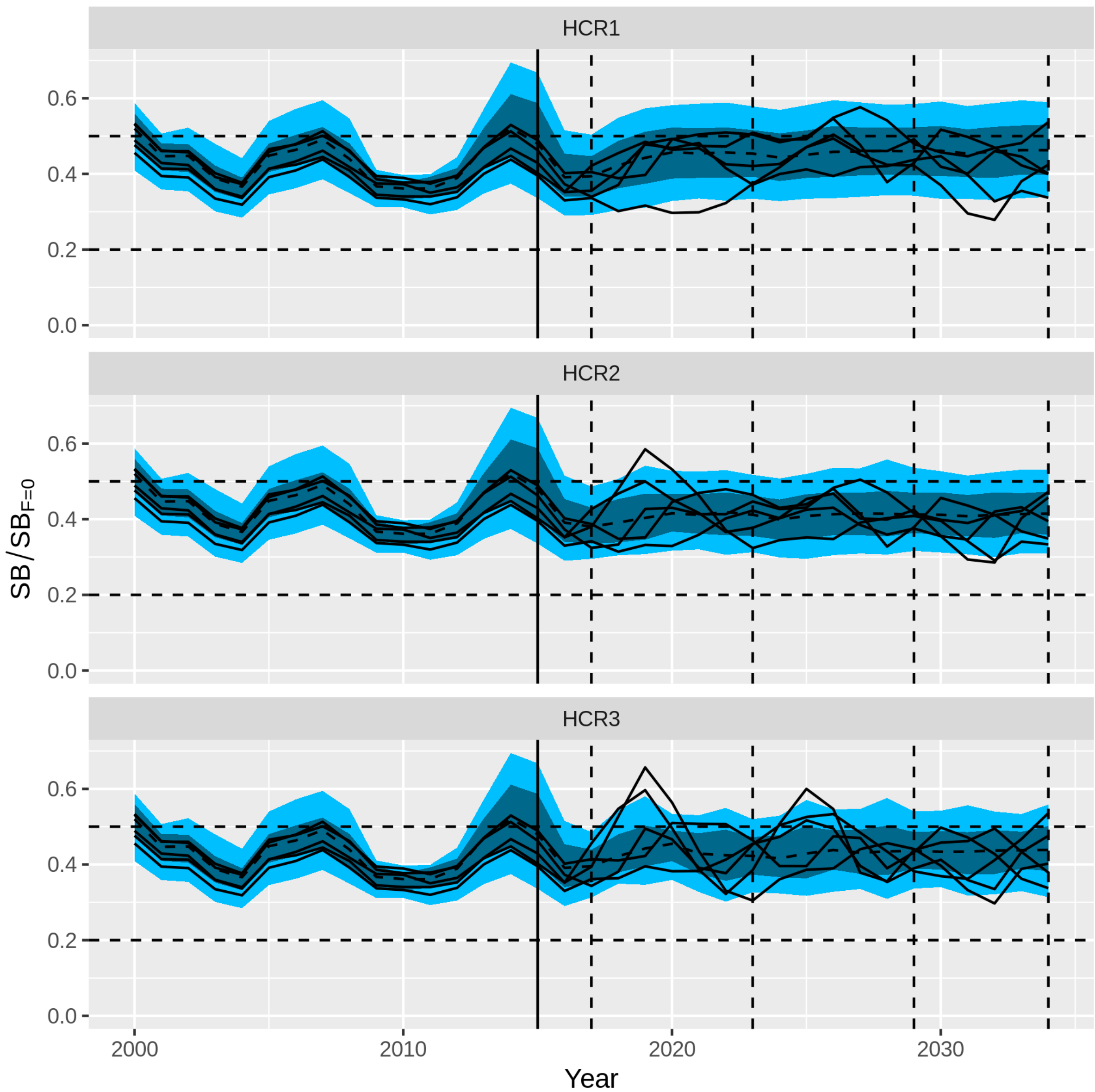

Figure 17: Predicted $SB/SB_{F=0}$ for the whole region for the three HCRs. Only the historical period from 2000 is shown. The ribbons show the 0.05-0.95 (light blue) and 0.20-0.80 (dark blue) percentiles. The dashed line is the median. The solid lines are example trajectories of 5 randomly selected simulations. The dashed vertical lines indicate the short-, medium- and long-term time periods over which the indicators are calculated. The solid vertical line at 2015 indicates the start of the projection. The dashed horizontal lines are the interim TRP (0.5) and LRP (0.2).

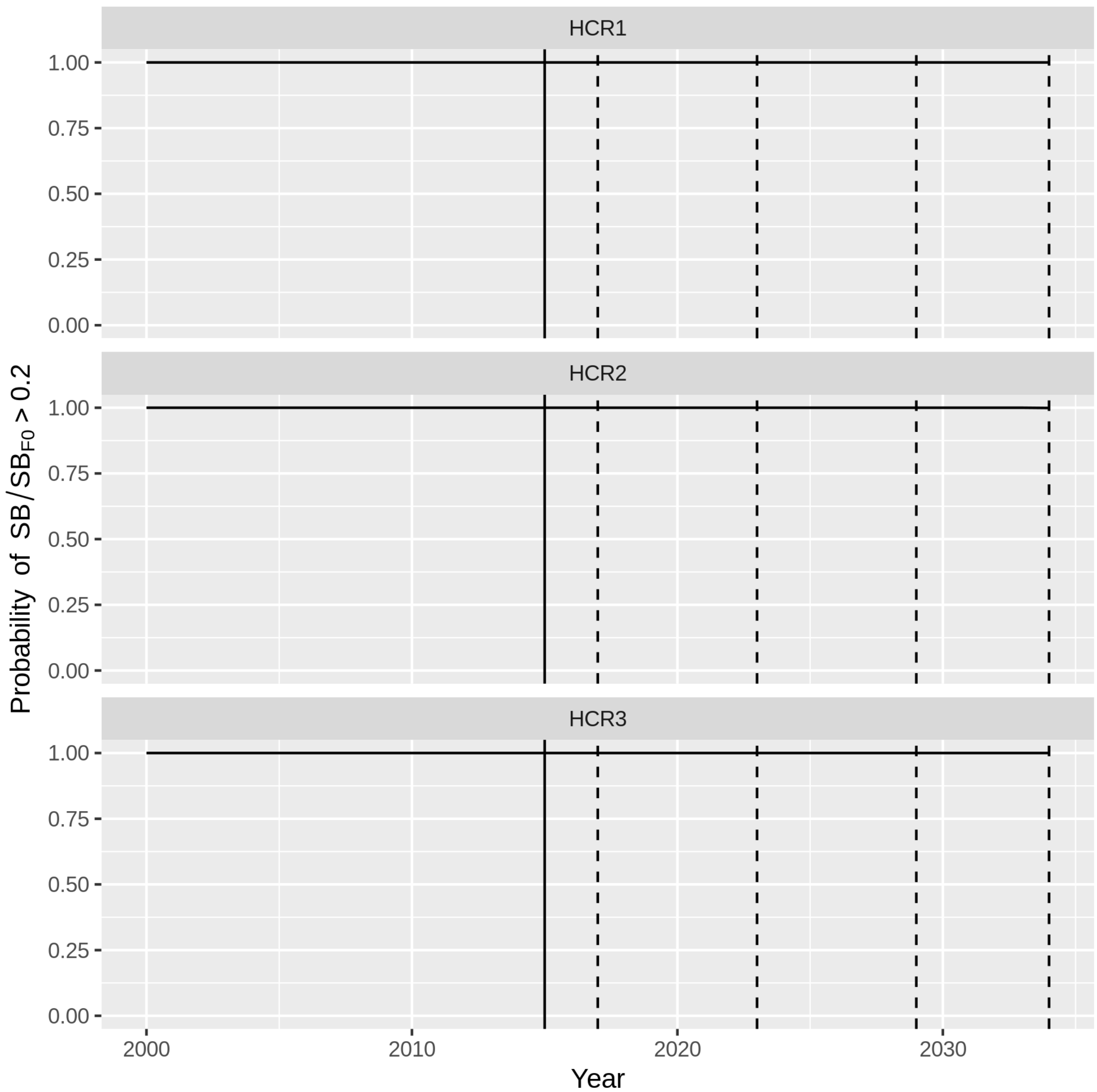

Figure 18: Probability of $SB/SB_{F=0} > 0.2$ for the whole region for the three HCRs. The solid vertical line at 2015 indicates the start of the projection. The dashed vertical lines indicate the short-, medium- and long-term time periods over which the indicators are calculated.

Figure 19 shows the effort relative to E_{MEY} for the purse seine fisheries operating in regions 2, 3 and 5 (excluding the associated purse seine fishery in region 5 due to the effort being standardised) for the three HCRs. Here E_{MEY} is taken to be the effort in 2010. The step-like pattern is the result of the effort being fixed during the management cycle (3 years), i.e. the effort only changes every 3 years as a result of the application of the HCR.

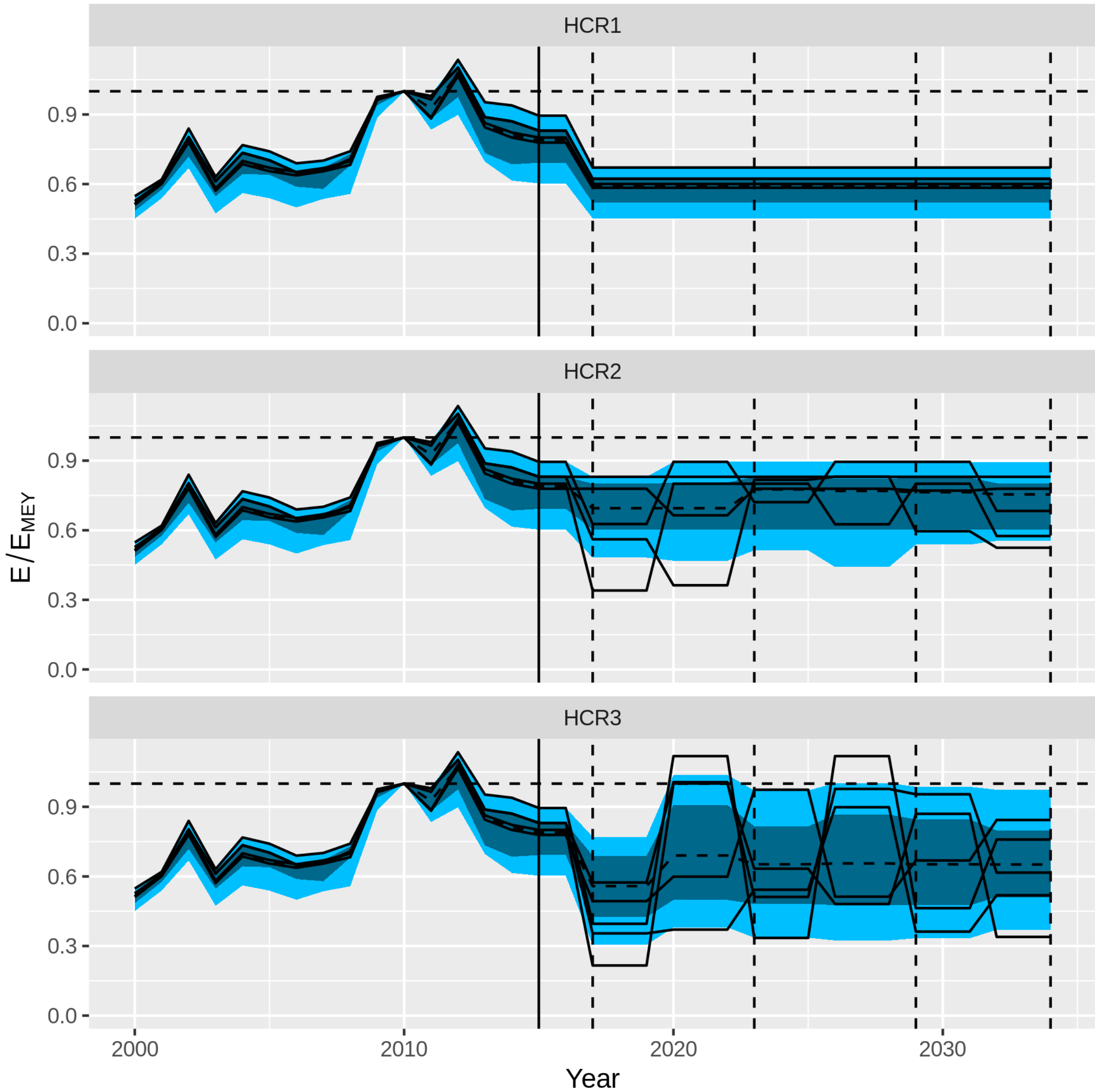

Figure 19: Effort relative to E_{MEY} for the purse seine fisheries operating in regions 2, 3 and 5 (excluding the associated purse seine fishery in region 5) for the three HCRs. E_{MEY} is taken to be the effort in 2010. The ribbons show the 0.05-0.95 (light blue) and 0.20-0.80 (dark blue) percentiles. The dashed line is the median. The solid lines are example trajectories of 5 randomly selected simulations. The dashed vertical lines indicate the short-, medium- and long-term time periods over which the indicators are calculated.

The expected total catches and the catches of the purse seine fisheries in regions 2, 3, and 5 are shown in Figures 20 and 21. The initial decrease in catch is the result of a period of low recruitment that was estimated in the last time steps of the 2016 stock assessment. This period of low recruitment is included in the conditioning of OMs.

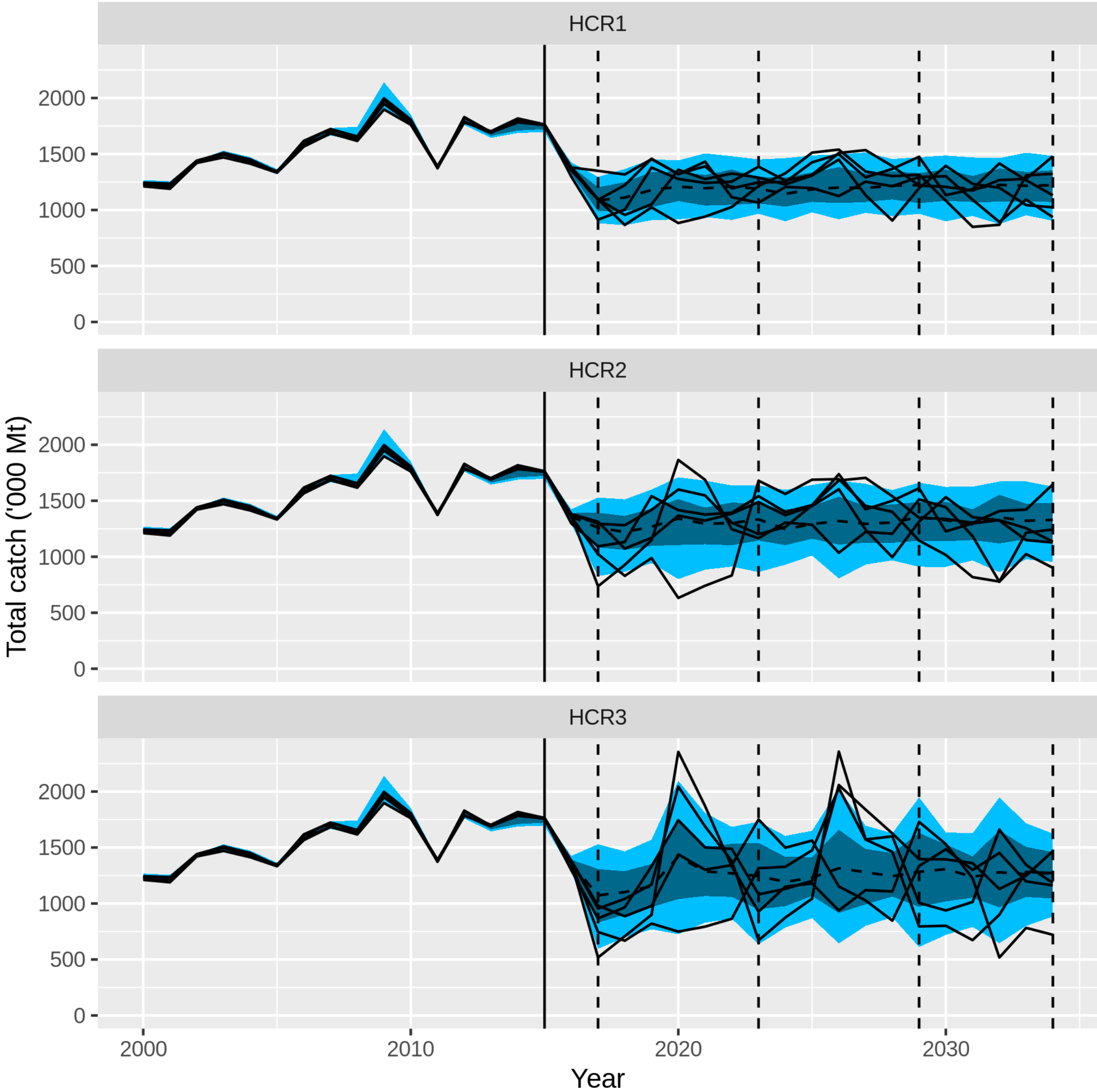

Figure 20: Expected total catch for the three HCRs. The ribbons show the 0.05-0.95 (light blue) and 0.20-0.80 (dark blue) percentiles. The dashed line is the median. The solid lines are example trajectories of 5 randomly selected simulations. The dashed vertical lines indicate the short-, medium- and long-term time periods over which the indicators are calculated.

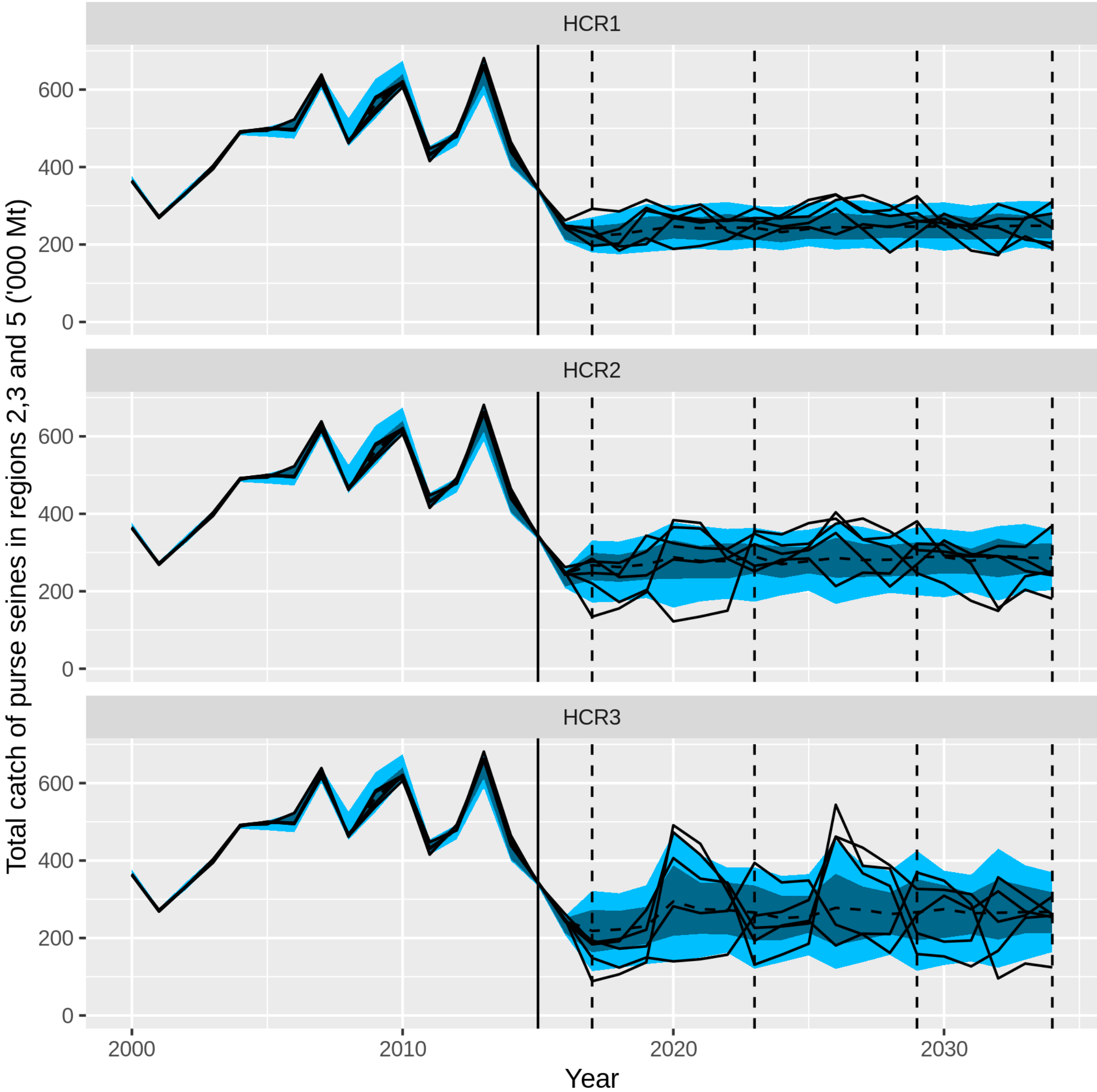

Figure 21: Expected catch of the purse seine fisheries in regions 2,3, and 5 for the three HCRs. The ribbons show the 0.05-0.95 (light blue) and 0.20-0.80 (dark blue) percentiles. The dashed line is the median. The solid lines are example trajectories of 5 randomly selected simulations. The dashed vertical lines indicate the short-, medium- and long-term time periods over which the indicators are calculated.

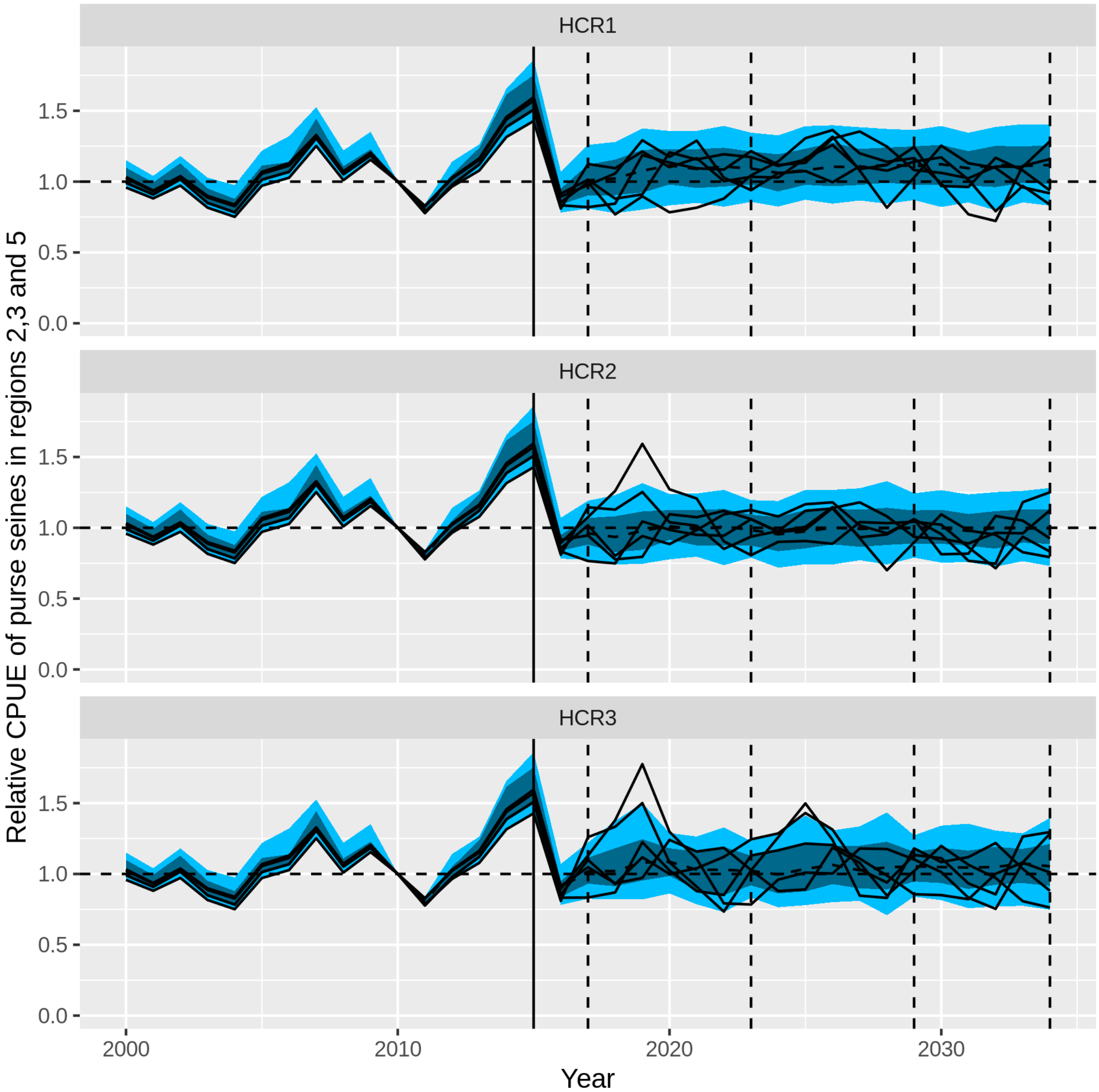

Figure 22: Expected CPUE relative to the CPUE in 2010 for the purse seine fisheries operating in regions 2, 3 and 5 (excluding the associated purse seine fishery in region 5) for the three HCRs. The ribbons show the 0.05-0.95 (light blue) and 0.20-0.80 (dark blue) percentiles. The dashed line is the median. The solid lines are example trajectories of 5 randomly selected simulations. The dashed vertical lines indicate the short-, medium- and long-term time periods over which the indicators are calculated.

Performance indicator 8 is based on the probability of $SB/SB_{F=0} > 0.2$. The time series of the probability is shown in Figure 23.

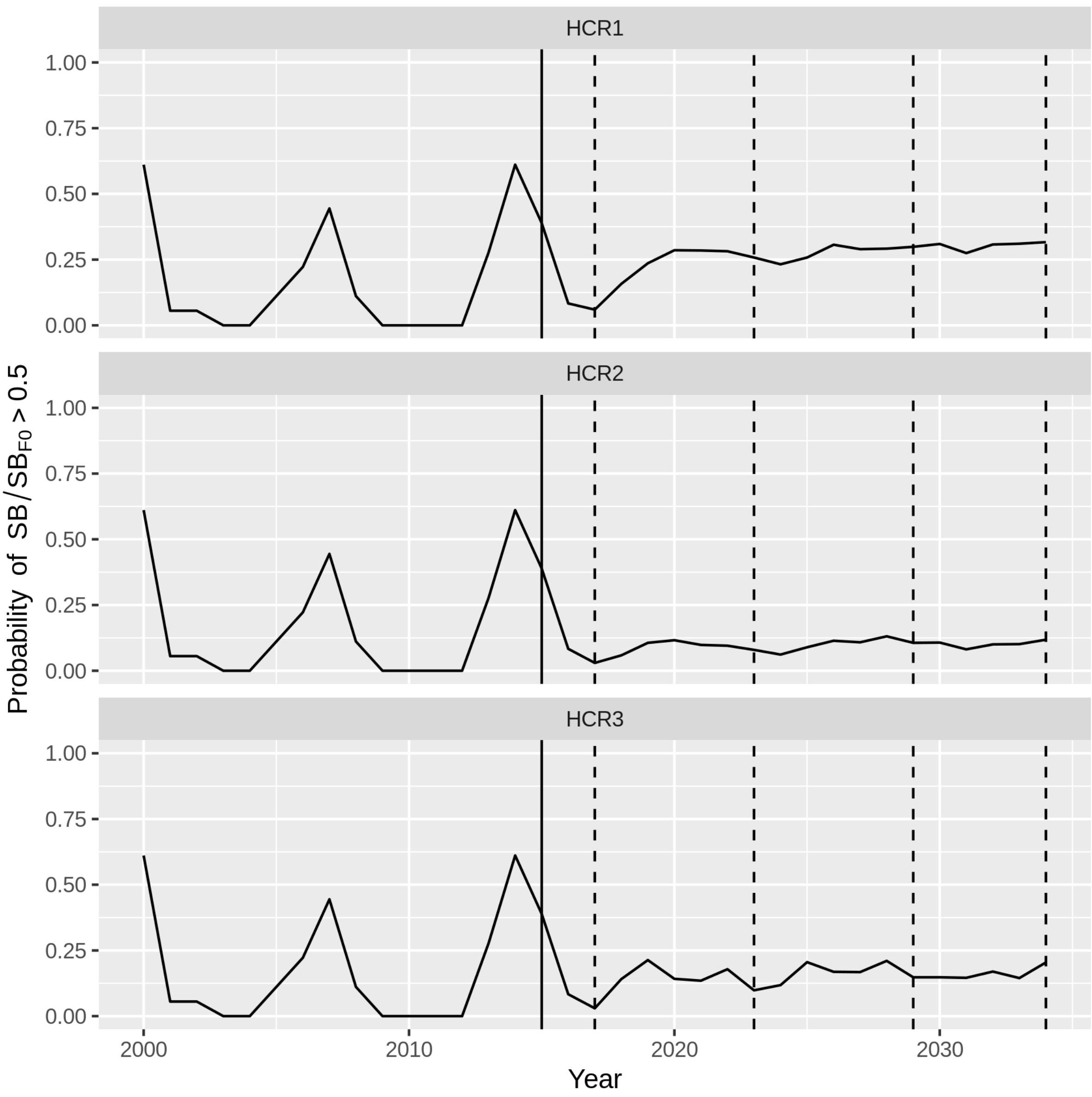

Figure 23: Probability of $SB/SB_{F=0} > 0.5$ for the whole region for the three HCRs. The solid vertical line at 2015 indicates the start of the projection. The dashed vertical lines indicate the short-, medium- and long-term time periods over which the indicators are calculated.

The mean weight of an individual is constant through time and there is almost no difference between the HCRs (Figure 24). However, this is not guaranteed to always be the case. This indicator can be used to identify changes in stock structure which could suggest the need to change the management of the fishery.

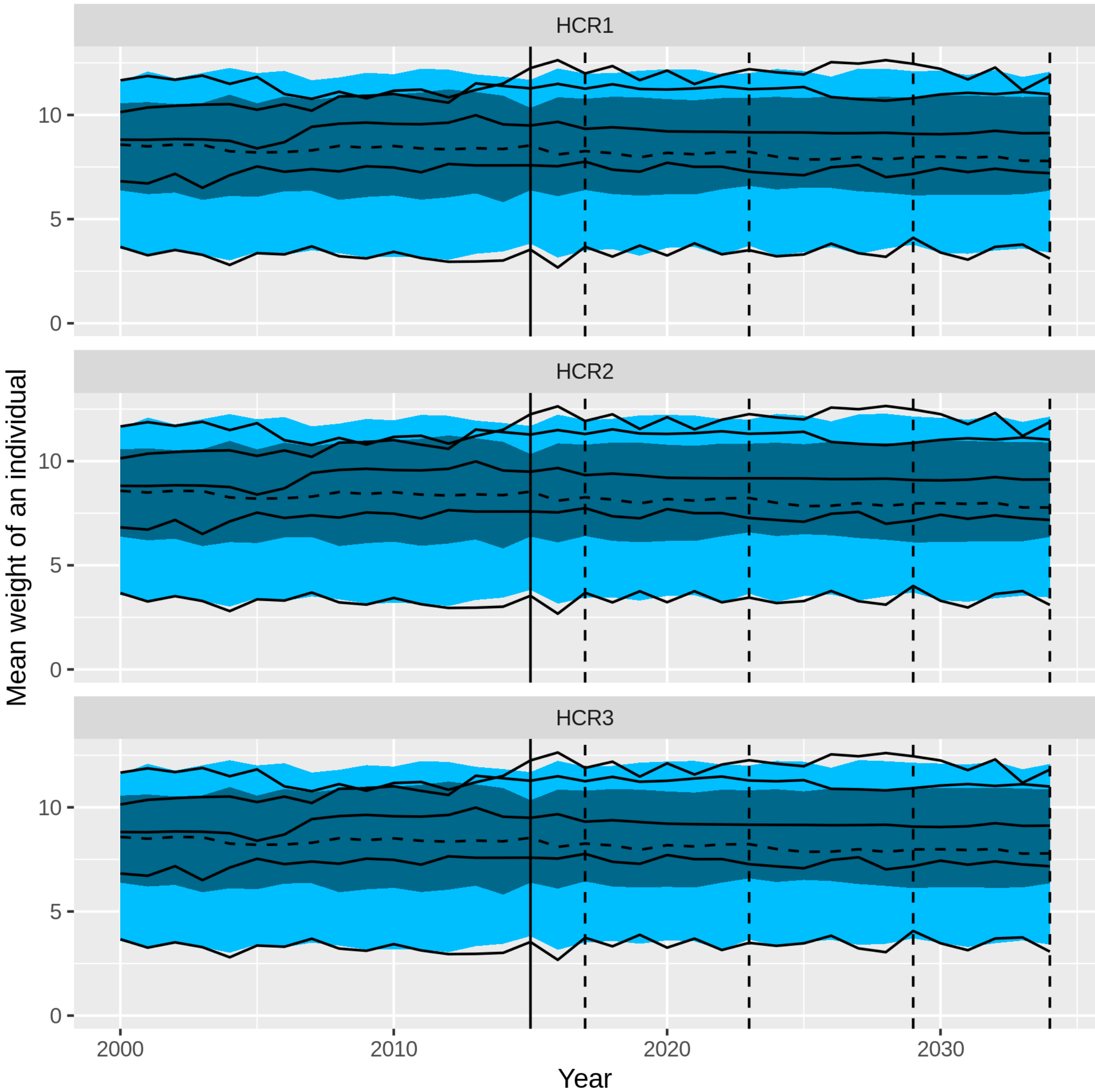

Figure 24: Mean weight of an individual across the total stock for the three HCRs. The ribbons show the 0.05-0.95 (light blue) and 0.20-0.80 (dark blue) percentiles. The dashed line is the median. The solid lines are example trajectories of 5 randomly selected simulations. The dashed vertical lines indicate the short-, medium- and long-term time periods over which the indicators are calculated.

D Performance indicator summary table

PI	Description	Period	HCR HCR 1	 HCR 2	 HCR 3
1	Prob. of $SB/SB_{F=0} > 0.2$	Short	1	1	1
1	Prob. of $SB/SB_{F=0} > 0.2$	Medium	1	1	1
1	Prob. of $SB/SB_{F=0} > 0.2$	Long	1	1	1
2a	$\lvert 1 - E/E_{MEY} \rvert$	Short	0.41 (0.38,0.48)	0.31 (0.22,0.4)	0.38 (0.29,0.47)
2a	$\lvert 1 - E/E_{MEY} \rvert$	Medium	0.41 (0.38,0.48)	0.28 (0.2,0.4)	0.34 (0.27,0.42)
2a	$\lvert 1 - E/E_{MEY} \rvert$	Long	0.41 (0.38,0.48)	0.27 (0.2,0.39)	0.35 (0.26,0.43)
2b	$\lvert 1 - B/B_{MEY} \rvert$	Short	0.21 (0.12,0.38)	0.14 (0.086,0.27)	0.21 (0.14,0.31)
2b	$\lvert 1 - B/B_{MEY} \rvert$	Medium	0.25 (0.14,0.42)	0.15 (0.089,0.27)	0.21 (0.14,0.32)
2b	$\lvert 1 - B/B_{MEY} \rvert$	Long	0.26 (0.16,0.44)	0.15 (0.096,0.28)	0.22 (0.15,0.33)
3a	Catch (total, ’000 Mt)	Short	1200 (1100,1200)	1300 (1100,1400)	1200 (1100,1400)
3a	Catch (total, ’000 Mt)	Medium	1200 (1100,1300)	1300 (1200,1400)	1200 (1100,1400)
3a	Catch (total, ’000 Mt)	Long	1200 (1100,1300)	1300 (1200,1400)	1300 (1100,1400)
3b	Catch (p. seine in 2,3,5, ’000 Mt)	Short	240 (220,260)	270 (240,300)	260 (210,290)
3b	Catch (p. seine in 2,3,5, ’000 Mt)	Medium	240 (220,270)	280 (250,310)	260 (230,300)
3b	Catch (p. seine in 2,3,5, ’000 Mt)	Long	250 (230,270)	280 (260,310)	270 (230,310)
4	Relative CPUE	Short	1.1 (0.97,1.2)	0.97 (0.91,1.1)	1 (0.98,1.1)
4	Relative CPUE	Medium	1.1 (1,1.2)	0.98 (0.91,1.1)	1 (0.97,1.1)
4	Relative CPUE	Long	1.1 (1,1.2)	0.99 (0.92,1.1)	1 (0.99,1.1)
7	Relative effort variation	Short	0.033 (0.029,0.035)	0.0036 (0,0.055)	0.069 (0.037,0.13)
7	Relative effort variation	Medium	0 (0,0)	0.0091 (0,0.059)	0.086 (0.048,0.14)
7	Relative effort variation	Long	0 (0,0)	0.0031 (0,0.054)	0.083 (0.037,0.15)
8	Prob. of $SB/SB_{F=0} > 0.5$	Short	0.22	0.084	0.14
8	Prob. of $SB/SB_{F=0} > 0.5$	Medium	0.27	0.097	0.16
8	Prob. of $SB/SB_{F=0} > 0.5$	Long	0.3	0.1	0.16
8	Deivation of SB from 0.5	Short	0.14 (0.089,0.24)	0.2 (0.11,0.28)	0.16 (0.1,0.21)
8	Deivation of SB from 0.5	Medium	0.12 (0.076,0.21)	0.19 (0.11,0.27)	0.16 (0.11,0.21)
8	Deivation of SB from 0.5	Long	0.12 (0.074,0.2)	0.18 (0.1,0.27)	0.14 (0.096,0.2)
	Mean weight in the stock	Short	8.2 (6.3,11)	8.2 (6.3,11)	8.2 (6.3,11)
	Mean weight in the stock	Medium	8 (6.5,11)	8 (6.5,11)	8 (6.5,11)
	Mean weight in the stock	Long	7.9 (6.2,11)	7.9 (6.1,11)	7.9 (6.1,11)

Table 3: Table of performance indicators over the time periods. The value is the median with the 20 and 80 percentiles in parenthesis.

E Performance indicator equations

	Performance Indicator	Equation
1	Probability of $SB/SB_{F=0} > 0.2$.	$PI1_y = \sum_{n=1}^{N} SB/SB_{F=0,n,y} > 0.2/N$ $PI1 = \sum_{y=y_1}^{y_2} PI1_y/Y$
2a 2b	Predicted effort relative to E_{MEY}. Predicted biomass relative to B_{MEY}.	$PI2a_{n,y} = \lvert E_{n,y}/E_{MEY} - 1\rvert$ $PI2b_{n,y} = \lvert B_{n,y}/B_{MEY} - 1\rvert$
3	Average expected catch	$PI3_n = \sum_{y=y_1}^{y_2} C_{n,y}/Y$
4	Average deviation of predicted SKJ CPUE from reference period levels	$PI4_{n,y} = CPUE_{n,y}/CPUE_{ref}$
6	Average annual variation in catch	$PI6_{y,n} = \lvert C_{y+1,n} - C_{y,n}\rvert$
7	Effort variation relative to reference period level	$PI7_{y,n} = \lvert E_{y+1,n}/E_{ref} - E_{y,n}/E_{ref}\rvert$
8	Deviation of $SB/SB_{F=0}$ from 0.5	$PI8_{n,y} = \lvert SB/SB_{n,y,F=0}/0.5 - 1\rvert$
PI 2,4,5,6,7,8 are summarised over the different time periods: $PIX_n = \sum_{y=y_1}^{y_2} PIX_{n,y}/Y$		

Table 4: Equations for calculating the the proposed performance indicators for the purse seine skipjack tuna fishery (WCPFC, 2017). SB is the adult biomass, E is the effort, C is catch, CPU is the catch per unit of effort, N is the number of simulations, n an individual simulation, y_1 and y_2 are the start and end years of the time period, Y is in the number of years in the time period, y is an individual year.