INTER-AMERICAN TROPICAL TUNA COMMISSION

SCIENTIFIC ADVISORY COMMITTEE

11TH MEETING

San Diego, California (USA)
11-15 May 2020[1]

DOCUMENT SAC-11 INF-J REV

RISK ANALYSIS FOR YELLOWFIN TUNA: REFERENCE MODELS AND THEIR RELATIVE WEIGHTS

Mark N. Maunder, Carolina Minte-Vera, Cleridy E. Lennert-Cody, Juan L. Valero, Alexandre Aires-da-Silva, Haikun Xu

CONTENTS

SUMMARY

Reference models are used in the IATTC staff's new 'risk analysis' approach to implement reference point-based fishery harvest control rules within a probabilistic framework (see SAC-11 INF-F) for yellowfin tuna in the EPO (SAC-11-08). The approach considers multiple hypotheses in a pragmatic way, as a compromise among computational demands, model complexity, and statistical rigor. It acknowledges the need to weight models based on information from the available data, recognizing that the complexity of fisheries stock assessment models prevents strict adherence to statistical rigor.

The five main steps involved in the implementation of this pragmatic approach are:

1. establish a hierarchy of hypotheses and models;
2. implement the weight system for hypotheses and models;
3. calculate the probability distributions for quantities of interest for a model;
4. combine probability distributions across models;
5. present the results in the form of a risk analysis.

This document addresses steps 1 and 2; steps 3-5 (the probability distributions for management quantities of interest and their application in the IATTC's harvest control rule) are addressed in SAC-11-08. A hierarchy of hypotheses and models was developed for yellowfin to address the unresolved issues with the assessment. Models and hypotheses were assigned weights by a panel of six experts among the IATTC staff following a set of predefined metrics.

---

[1] Postponed until a later date to be determined

1. BACKGROUND

The IATTC's **harvest control rule (HCR)** for the management of tropical tunas in the eastern Pacific Ocean (EPO), established in Resolution C-16-02, stipulates **limit and target reference points (RPs)** for both bigeye and yellowfin tunas, based on the **spawning biomass (*S*)** and **fishing mortality (*F*)** of the stock, and expressed as **probability statements**: the probability that a given management action will maintain the size (biomass) of that stock within certain specific limits **relative to its maximum sustainable yield (MSY)**[2]. Implementing this rule requires not only estimates of both *S* and *F*, as in previous assessments, but also a way of quantifying the probability that a management action will have its intended effect. Consequently, the IATTC staff has, as part of its work plan to improve the stock assessments of tropical tunas, developed a new 'risk analysis' approach (SAC-11 INF-F) to these assessments. This approach, which uses reference points within a probabilistic framework that considers multiple hypotheses, is a pragmatic compromise among computational demands, model complexity, and statistical rigor. In fisheries science, the large gaps in data, and the unavoidable complexity of stock assessment models, prevent strict adherence to statistical rigor, so in this approach, the practical need to use not only information in the available data but also from other sources, such as expert opinion, is accounted for.

The main features of the approach are:

1. Each hypothesis ('*state of nature*') is represented by a different combination of stock assessment model configuration, data, and parameters;
2. Hypotheses are grouped in a hierarchical framework, which highlights similarities among models, facilitating model development and weight assignment, and also preventing any one hypothesis from inadvertently becoming the 'overarching' hypothesis and dominating the results of the analysis;
3. Sub-hypotheses represent models with parameters that cannot be reliably estimated within the assessment model, and are therefore fixed in the models;
4. Multiple metrics are used to evaluate the reliability of the models and the plausibility of the hypotheses they represent;
5. Model fit plays only a limited role in metrics used to evaluate models;
6. Efficient elimination of unlikely hypotheses.

The five main steps for implementing the approach are:

1. Establish a hierarchy of hypotheses and models;
2. Implement the weighting system for hypotheses and models;
3. Calculate the probability distributions for quantities of interest for a model;
4. Combine probability distributions across models;
5. Present the results in the form of a risk analysis.

This approach was used to quantify, for the bigeye and yellowfin tuna stocks in the EPO (SAC-11-08), the probability statements in the IATTC HCR about:

a. current biomass (*S*) relative to target and limit reference points (S_{MSY} and S_{LIMIT}), and
b. fishing mortality (*F*) relative to target and limit reference points (F_{MSY} and F_{LIMIT}) under different

---

[2] **For *S*:** "*If the probability that the spawning biomass (S) is below the limit reference point (S_{LIMIT}) is greater than 10%, as soon as is practical management measures shall be established that have a probability of at least 50% of restoring S to the target level (dynamic S_{MSY}) or greater, and a probability of less than 10% that S will descend to below S_{LIMIT} in a period of two generations of the stock or five years, whichever is greater*"; and **for *F***: "*if the probability that fishing mortality (F) will exceed the limit reference point (F_{LIMIT}) is greater than 10%, as soon as is practical management measures shall be established that have a probability of at least 50% of reducing F to the target level (F_{MSY}) or less, and a probability of less than 10% that F will exceed F_{LIMIT}*".

management scenarios.

This approach is used to evaluate probability statements about:

- Current status: the current status relative to IATTC reference points, and
- Future management: probability statements about the fishing mortality relative to reference points under different management scenarios for bigeye and yellowfin tunas (SAC-11-08)

This document focuses on the application of steps 1 and 2 to yellowfin tuna. Section 2 provides a general overview of the methods and rationale, and Section 3 presents the hierarchy of hypotheses and models developed for yellowfin, and their weight assignments by a panel of six experts among the IATTC staff (see SAC-11 INF-F for a detailed description of the technical aspects). Results of the benchmark assessment of yellowfin used in the risk analysis are presented in SAC-11-07, and the results of the risk analysis and their interpretation (steps 3-5) are provided in SAC-11-08.

2. SETTING UP THE RISK ANALYSIS APPROACH

The hierarchy of hypotheses and models, and their weighting, form the foundation of the risk analysis. The two sections below briefly describe the methodology used to obtain these two components.

2.1. Hierarchy of hypotheses and models

In a risk analysis, hypotheses about the state of nature are represented by models. Typically, there will be multiple models that could be used to represent a single hypothesis; also, hypotheses may be nested, adding further complexity. In reality, there is an unlimited combination of hypotheses, and possible models to represent them, and care is needed when defining the models and combinations. Therefore, to facilitate the development of models that represent hypotheses about states of nature, a flow chart is used to represent a hierarchical structure of hypotheses with three levels: 1) overarching hypotheses; 2) hypotheses; 3) sub-hypotheses. These levels have different functions. The overarching hypotheses (Level 1) correspond to broad states of nature (*e.g.* number of stocks), and are represented by a variety of models of differing complexities that may use different data. Level 2 hypotheses are specific, can be represented by models, and can be differentiated by fitting the models to data; they may be divided into sub-levels (*e.g*. Level 2A and 2B), each addressing a particular issue in the assessment. Level 3 sub-hypotheses are evaluated differently than the Level 2 sub-levels to avoid the influence of data, reduce the number of analyses, or for convenience.

2.2. Weighting system for hypotheses and models

Once a hierarchy of hypotheses has been established (Step 1), various sources of information (both internal and external to each model) can be used to evaluate which hypotheses are considered more likely than others. The model weights must then be rescaled so that they represent probabilities and can be used in the risk analysis.

To assign weights to the various models and hypotheses, the following categories are used for all metrics except overall model fit (*W(Fit)*; see Section 2.2.2b of SAC-11 INF-F).

Category	Value
None	0
Low	0.25
Medium	0.5
High	1.0

2.2.1. Weight metrics

We use a set of metrics related to the reliability of a model, in addition to model fit, which will later be combined to assign weights to each model representing a hypothesis. These metrics include:

a. *W(Expert):* Expert opinion, assigned *a priori*, without consideration of model fit.
b. *W(Convergence):* Model convergence criteria of the estimation algorithm.
c. *W(Fit):* The fit of the model to the data.
d. *W(Plausible parameters):* The plausibility of the estimates of the parameters representing the hypothesis (*e.g*. estimates of the growth parameters).
e. *W(Plausible results):* The plausibility of the model results (*e.g*. estimates of fishing mortality).
f. *W(Diagnostics):* Reliability of the model based on diagnostics.
g. *W("Empirical" selectivity)*: How well the assumed selectivity curves represent the implied selectivity.

The metric *W("Empirical" selectivity)* was added to evaluate the misfit to the composition data for the fishery with asymptotic selectivity, which is one of the main issues in the yellowfin assessment.

The weight given to each model is then the product of the individual component weights, once each of those components has been rescaled (See Section 2c of SAC-11 INF-F):

W(Model) = W(Expert) x *W(Convergence)* x *W(Fit)* x *W(Plausible parameters)* x *W(Plausible results)* x *W(Diagnostics)* x *W("Empirical" selectivity)* **[Equation 1]**

The *W(Diagnostics)* component is calculated based on a variety of diagnostics. The weights of each set of diagnostics are added together (rather than multiplied) to ensure that the individual diagnostics are not overweighted in the calculation of *W(Model)*:

W(Diagnostics) = W(ASPM, likelihood profile R_0*, Catch curve) + W(Retrospective analysis) + W(Composition residuals) + W(Index residuals) + W(Recruitment residuals)* **[Equation 2]**

The *W(Fit*) and *W(Convergence)* metrics needed modification to address specific issues in the yellowfin tuna application. Those modifications are described in Section 3.

2.2.2. Assigning weights

Weights need to be assigned to each of the hypotheses and models using the weight categories and weight metrics described above. The hierarchy of hypotheses and models is used to assign weights, where hypotheses and models may be weighted considering branches of the hierarchy separately or they may be weighted across all branches of the hierarchy. The model weights must then be rescaled so that they represent relative probabilities and can be used in the risk analysis. There are several technical details to consider when assigning weights, which are not covered here but are discussed in SAC-11 INF-F.

3. APPLICATION TO YELLOWFIN TUNA

3.1. Hierarchy of hypotheses and models

There are several unresolved issues in the yellowfin tuna assessment that need to be accounted for in the management advice. These include 1) spatial structure, 2) inconsistencies between the index of abundance based on CPUE from the purse-seine fishery on yellowfin associated with dolphins and the index based on CPUE from the longline fishery, 3) inability of the model to fit the high values in the indices of abundance, and 4) misfit to the length composition data for the fishery that is assumed to have asymptotic selectivity. We use a hierarchical structure to present the hypotheses that may address these issues (Figure 1).

3.1.1. Details of the hierarchy

Level 1
The first level in the hierarchy addresses the spatial structure, which is an overarching set of hypotheses about the degree of mixing between stocks within the EPO (Figure 1a). It also addresses the

inconsistencies between the various indices. There is evidence of strong spatial structure of yellowfin in the EPO, but the stock boundaries are not clear, and mixing between the two potential northern and southern stocks may be episodic or there may be continual mixing, the magnitude of which varies from year to year. Therefore, we define three overarching hypotheses: *High mixing*, *Episodic/high-variability mixing*, and *Negligible mixing*. The *High mixing* hypothesis is represented by single-stock models similar to previous assessments. The *Episodic/high-variability mixing* hypothesis is represented by single-stock models that are driven by either the northern or the southern stock data. This means that the model is fit to data for the north (south) and only uses the catch for the south (north) while fixing the selectivity for these fisheries. The *Negligible mixing* hypothesis is represented by two independent assessments, one for the north and one for the south. The inconsistencies between the various indices may also be due to spatial structure since the indices from the main purse seine fisheries associated with dolphins are from the north and the longline indices are predominantly from the south. Many of these models were developed for the yellowfin tuna review, and this provided information that informed the decision to eliminate all hypotheses except *High mixing* from the risk analysis to make it practical to implement. At this point, the risk analysis is thus focused on the hypotheses and models nested within the overarching *High mixing* hypothesis.

Level 2

Figure 1b shows the flow chart representing the hierarchy of hypotheses and models under the *High mixing* overarching hypothesis used in the risk analysis. Models in Level 2A are implemented in combination with models in Level 2B. On Level 2A of the model sub-hierarchy are hypotheses that address the inability of the model to fit the high values in the indices of abundance. Level 2B of the model sub-hierarchy is based on hypotheses that attempt to address the misfit to the composition data for the fishery with asymptotic selectivity (the main purse-seine fishery on yellowfin associated with dolphins).

Level 3

Models representing different steepness scenarios are added as a third level in the hierarchy.

3.1.2. Model descriptions

The hierarchy of hypotheses in Figure 1b translates into a large number of models because many of these hypotheses are considered in combination. These models focus on the purse-seine fishery associated with dolphins because it targets yellowfin and produces the majority of the EPO yellowfin catch. To minimize issues with possible spatial variability in the growth rates of yellowfin, which are confounded with the longline fisheries catching larger yellowfin, the length composition for the longline fisheries and the southern dolphin-associated fisheries were not used in the analysis and their asymptotic selectivities were fixed. A description and a brief rationale of the hypotheses and associated models is provided below. Table 1 provides names and acronyms for the models.

Level 2A: Hypotheses that address the inability to fit the high index values

Index is proportional to abundance (BASE)

This model is most similar to previous models used to assess yellowfin tuna in the EPO (*e.g.* SAC-10-07) and is the basis for all the other models. It assumes, in contrast to the other models, that the index is proportional to abundance for the whole time period.

Density dependent catchability (DDQ)

This model allows the estimation of a coefficient that determines how catchability is influenced by abundance. It is hypothesized that during periods of high abundance the purse-seine fleet that fishes on yellowfin associated with dolphins can more efficiently catch yellowfin tuna and this will allow the model to better fit the high index observations.

Time block in the middle (TBM)
This model allows the estimation of catchability and selectivity of the index during the period where there is peak abundance. During this period, the fishery associated with dolphins catches smaller yellowfin on average. It is hypothesized that during the period of high abundance the purse-seine fleet that fishes on yellowfin associated with dolphins can more efficiently catch yellowfin tuna and this will allow the model to better fit the high index observations, but unlike the DDQ hypothesis, it assumes that at other times this is not the case. It assumes that if catchability changes, selectivity is also likely to change.

Time block at the end (TBE)
This model allows the estimation of catchability and selectivity for the survey and selectivity for some fisheries during the later period where the size of fish caught by the purse seine fishery associated with dolphins is higher on average. It is hypothesized that if selectivity changes, catchability is likely to change as well. This model does not directly address the inability of the model to fit the peak in the index of abundance, but was included in Level 2A because it includes a time block in the index catchability.

Level 2B: Hypotheses that address the misfit to the length composition data

Fixed parameters (Fixed)
This model is most similar to previous models used to assess yellowfin tuna in the EPO and is the basis for all the other models. In contrast to other models, it does not estimate growth and assumes that the selectivity is asymptotic for the main purse seine fishery on yellowfin associated with dolphins.

Estimate growth (EstGro)
This model estimates all the growth parameters of the Richards growth curve, assumes the variation of length at age to have a coefficient of variation fixed at 7.5%, and fits to the otolith age-length data. The fixed value used for asymptotic length is higher than the limited tagging data, but is somewhat consistent with the otolith data, although old fish cannot be aged using the daily increment method. The otolith data comes from before the model starts and the tagging data is limited in its spatial and temporal distribution. Therefore, estimation of growth within the stock assessment model may be appropriate and may allow a better fit to the length composition data for the fishery with asymptotic selectivity.

Dome selectivity (EstSel)
This model allows the selectivity to be dome-shaped for the main purse-seine fishery on yellowfin associated with dolphins and estimates the parameters of the double normal selectivity curve. This will allow the model to fit the length-composition data better.

3.2. Weight assignments

Weights were assigned by six experts among the IATTC staff. The results are presented for each expert but are randomized for each metric to obscure the expert. There was consensus on some weighting assignments, but due to the subjective nature of many of the weighting metrics, most cases had differences in the weighting assignments. Therefore, we took the approach where the weighting assignments were first discussed among all experts and then each expert provided their own weighting for each of the metrics. The average of the individual weighting assignments, first scaled to sum to one for each expert, were used as the final weighting assignments to compute the model weights. The numerical weights provided in the tables may not sum exactly to one due to rounding.

Stock mixing hypothesis (Level 1)
As mentioned above, only the *High mixing* overarching hypothesis is included in the risk analysis, reducing the number of models and making the analysis more practical.

W(Expert)
This weight represents the subjective judgment of the experts, taking into consideration past experience

with the specific assessment, other assessments of that species, assessments of other species, and other biological, ecological and related factors. Population dynamics theory is also taken into consideration. Based on expert opinion, assigned *a priori*, without consideration of model fit. The weights were considered on the same level of the model sub-hierarchy. Describing the rationale for the weights by each expert and each model would be too voluminous to include here, so the general support for a model is included above in their description. There was a wide range of weights for *W(Expert)* given to the models assigned by the different experts (Table 2).

W(Convergence)

It is not clear whether a large maximum gradient for *W(Convergence)* means that the model has not converged on the global MLE or whether the likelihood surface is flat and that it is appropriately reflected in the parameter uncertainty. Therefore, only weights based on the Hessian matrix being positive definite were assigned. All models had a positive definite Hessian and were given a *High* weight. It should be noted that the maximum gradient components were high for the TBM-EstGro and TBM-EstSel models (Table 3).

W(Fit)

The *W(Fit)* weighting procedure had to be modified because the models that estimated growth are the only ones to include the otolith data, and therefore the AIC for this model was based on the difference in AIC excluding the otolith data. The TBM model that estimates *Dome shape selectivity* fits the data the best based on the AIC score and is given a weight of *H*. The TBE model that estimates *Dome shape selectivity* fits the data the worst and is given a weight of *L*. The weights for the other models based on a linear relationship with AIC are presented in Table 4.

W(Plausible parameter)

There was a wide range of weights for *W(Plausible parameters)* given to the models assigned by the different experts (Table 5). Therefore, it is complicated to describe the rationale for all choices and only some generalizations are provided here. The estimates of asymptotic length were consistent with the limited tagging data, but were in conflict with the otolith data. The estimated dome shape selectivity curve for the main purse seine fishery on yellowfin associated with dolphins was very dome shaped and had essentially zero selectivity for the oldest yellowfin represented in the model. Other fisheries such as the dolphin associated fisheries in the south and the longline fisheries generally catch larger yellowfin and therefore, if all the fisheries operate on the same stock, then the extreme doming may be reasonable. On the other hand, if this variation in maximum lengths is due to stock structure, then the doming may be unreasonable. The time block models estimate only moderate changes in catchability, which appear plausible, but also estimate extreme doming in selectivity.

W(Plausible results)

The *W(Plausible results)* criterion was based on the fishing mortality and initial equilibrium catch estimates. The plausibility of biomass levels is hard to judge, and fishing mortality levels, which are related to the biomass level, may be easier to judge. The initial depletion level is also related to the estimated biomass level, so we compare the predicted initial equilibrium catch (and other associated parameters including the initial recruitment offset and recruitment deviates) with historical catch to make sure they are somewhat consistent.

The experts agreed that the fishing mortality and initial catch were plausible for all models.

W(Diagnostics)

The steepness runs (*h* = 0.7, 0.8, 0.9, 1.0) of Level 3 (Figure 1b), which were applied to all models, greatly increased the number of model runs. It was not feasible to run all the diagnostics for these models so we assumed that the diagnostics for all steepness values would be similar to the *h* = 1.0 runs.

The experts all agreed on the weights for the difference diagnostics. The index and recruitment residuals were similar for all models and the recruitment residuals are difficult to interpret, therefore these two diagnostics were not used in the weighting. All three models that estimated *Dome shape selectivity* for the main purse seine fishery associated with dolphins were given a *L* weighting for the *W*(R_0 *- ASPM)* diagnostic (Table 6) because the composition data had a higher influence on the estimates of absolute abundance and provided information that was inconsistent with the index data. The other models were given a *Medium* weight. The three TBE models fitted the composition data better for this period and were given a *H* weight for *W(Composition residuals)*, the other models were given a *Medium* weight (Table 7). There was some variability in the retrospective diagnostic among models that was most apparent in the change in relative scaling. The TBE and DDQ models , except the one that estimated *Dome shape selectivity*, had the least retrospective pattern and were given a *H* weight (Table 8).

W("Empirical" selectivity)

The experts all agreed on the weighting based on *W("Empirical" selectivity)*. The consistency between the assumed selectivity curve and the "empirical" selectivity was generally good, except for the purse seine fishery associated with dolphins that was assumed to have asymptotic selectivity. This fishery was the focus of this weighting criteria. The models that estimated a *Dome shape selectivity* for this fishery provided the best consistency and were given *H* weight, some of the models that *Estimate growth* were given a *Medium* weight, and the remaining models were given a *L* weight (Table 9).

Steepness

The weights for steepness were the same as used for bigeye tuna. There was large variability in the experts' opinions about weights for different values of the steepness of the Beverton-Holt stock-recruitment relationship. This parallels the general disagreement in the value of this parameter among fisheries scientists. Some experts gave *H* weight to high levels of steepness, which is consistent with the approach used in previous assessments of tropical tunas in the EPO, while others gave *H* weights to lower values of steepness, which is more consistent with other tuna RFMOs. One expert did not provide weights. The average of the weights across experts gave an almost linear positive relationship between the weight and the steepness value (Table 10).

Final model weights: W(Model)

The final model weights (Equation 1) are shown in Table 11. The highest weights were given to the *Time block in the middle*, *Time block at the end* and *Density dependence*, all when *Dome shape selectivity* was estimated. The lowest weights were given to models that have fixed growth and asymptotic selectivity.

4. DISCUSSION

The IATTC staff's risk analysis procedure to implement reference point-based fishery harvest control rules within a probabilistic framework that considers multiple hypotheses is applied to yellowfin tuna in the EPO. Here we presented the hierarchy of hypotheses and models developed to address the unresolved issues with the assessment and the weights for models and hypotheses. These were combined to provide model probabilities to use in the risk analysis. The probability distributions for management quantities of interest and their use in evaluating management quantities relative to IATTC's harvest control rule can be found in SAC-11-08. Technical details of the risk assessment procedure can be found in SAC-11 INF-F.

Of the three overarching hypotheses, only the *High Mixing* hypothesis can be evaluated at this time. To evaluate the other hypotheses, additional information may be needed to make it feasible to translate them into models. Tagging and biology studies are underway in the IATTC and new genomic tools provide a promising new source of information that could be considered. Moore *et al.* (2020a and 2020b) provide insights on how to address questions of tuna stock structure in the Pacific Ocean.

Although, the risk analysis for yellowfin tuna evaluates a wide range of hypothesis within only one overarching hypothesis, *High Mixing*, the spatial heterogeneity is accounted for, given the available information, by fitting to data from the core area where most of the catch is taken. This a pragmatic way to address the uncertainty. Once more information is available, the risk analysis could be expanded to encompass the other overarching hypotheses.

REFERENCES

More, B.R. *et al*. 2020a. Defining the stock structures of key commercial tunas in the Pacific Ocean I: Current knowledge and main uncertainties. Fisheries Research, x:xx-xx

More, B.R. *et al*. 2020b. Defining the stock structures of key commercial tunas in the Pacific Ocean II: Sampling considerations and future directions Fisheries Research, x:xx-xx

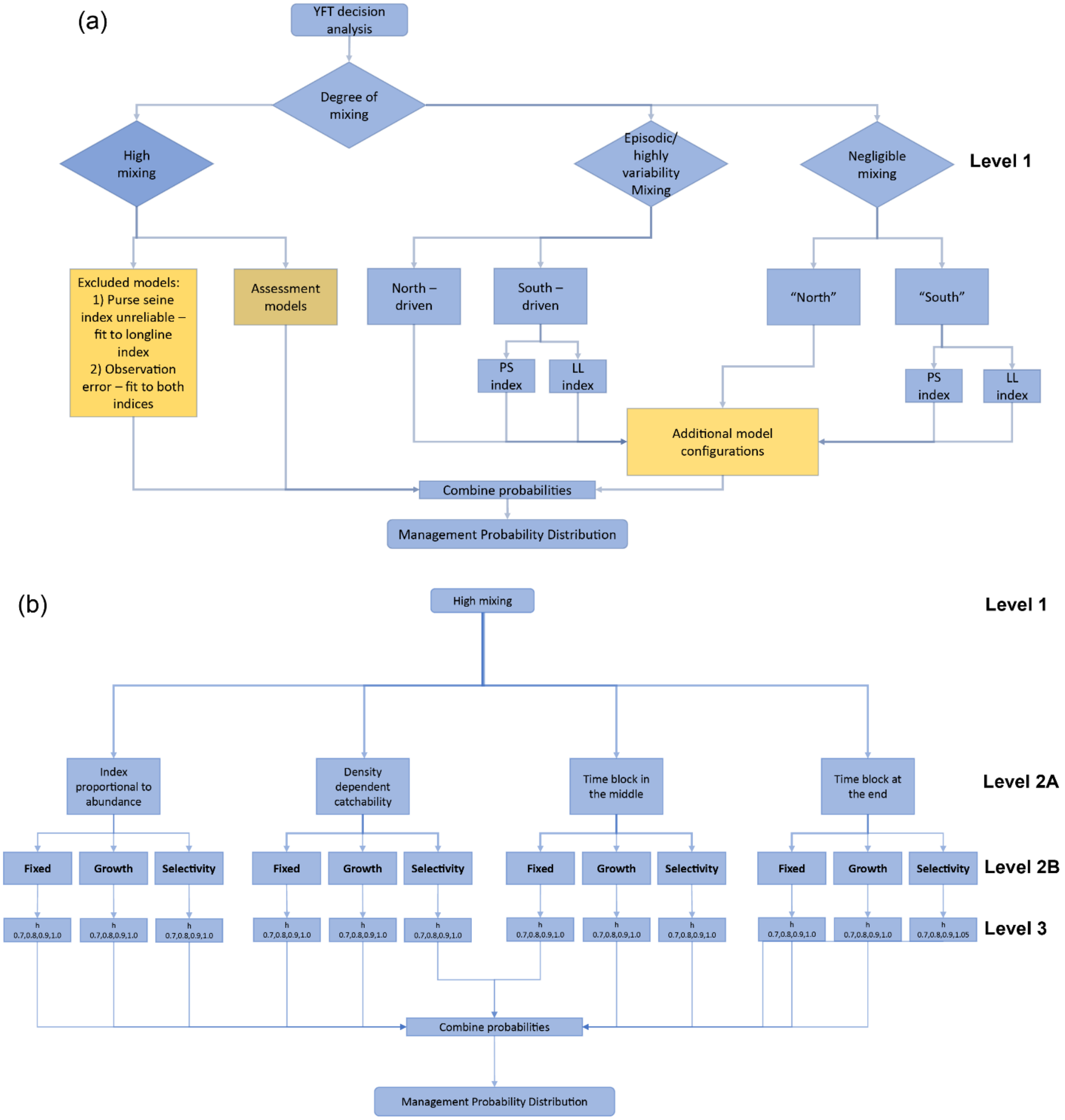

FIGURE 1. Flow charts of the full set of hypotheses considered for the yellowfin tuna risk analysis (a), and of the subset of hypotheses used in the analysis (b).

TABLE 1. Model configurations used to represent the Level 2 hypotheses (Figure 1b) addressing the main issues in the stock assessment.

Model		**Description**
Level 2A: ***Indices inconsistent*** **hypotheses**		
BASE	**Proportional**	Index of abundance proportional to abundance.
DDQ	**Density dependence**	Assumes index of abundance is non-linearly related to biomass, with parameters estimated.
TBM	**Time block middle**	Assumes a time block 2001.Q1-2003.Q2 for the index catchability (q) (to accommodate for large increase in the index) and a time block for selectivity 2002.Q3-2007.Q3 for the index, and the F18 and F19 fisheries. F19 selectivity assumed asymptotic, except during 2002.Q3-2007.Q3, when it was assumed dome-shaped.
TBE	**Time block end**	Assumes a time block beginning in 2015 for the index (both catchability and selectivity) and for F19 selectivity (to accommodate for increase in size in the index and asymptotic fishery).
Level 2B: ***Misfit to composition data*** **hypotheses**		
Fixed		Growth fixed; selectivity of fleets and survey time-invariant; F19 selectivity asymptotic; and catchability of the index is time invariant.
EstGro	**Estimate growth**	Growth estimated
EstSel	**Estimate selectivity**	Dome-shaped selectivity used for the fishery that assumes asymptotic selectivity and parameters estimated.

TABLE 2. Weights assigned by each expert to the alternative models: *W(Expert)*. H: high; M: medium; L: low. Rel. Wt. stands for relative weight and are the weights rescaled to sum to one.

2A	**Weight**						**Rel. Wt.**
BASE	M	L	M	L	L	L	0.16
DDQ	L	L	M	L	L	L	0.14
TBM	M	M	H	M	H	H	0.35
TBE	H	H	M	L	H	H	0.36
2B							
Fixed	L	L	M	L	L	L	0.18
EstGro	H	H	M	M	H	H	0.48
EstSel	H	H	M	M	M	L	0.35

TABLE 3. Maximum gradient components for the alternative models.

2A	2B	Steepness-Inclinación			
		h = 1	*h* = 0.9	*h* = 0.8	*h* = 0.7
BASE	Fixed	0.00128	<0.00001	<0.00001	0.000248
	EstGro	0.00014	0.000073	0.0132	0.000412
	EstSel	<0.00001	<0.00001	0.000137	<0.00001
DDQ	Fixed	0.000241	0.000552	0.000239	0.0285
	EstGro	<0.00001	0.0109	<0.00001	0.0013
	EstSel	0.000163	0.00302	0.00162	0.000618
TBM	Fixed	0.00224	0.000512	0.000158	0.00102
	EstGro	3.45	3.53	11.2	4.41
	EstSel	10.5	1.1	1.27	9.8
TBE	Fixed	0.000221	<0.00001	0.000451	0.00161
	EstGro	<0.00001	0.00413	<0.00001	0.000823
	EstSel	0.000142	<0.00001	0.000396	0.00141

TABLE 4. AIC values calculated using all the data, except the otolith age-length data, and the consequent weights for each model: *W(Fit)*.

2A	2B	dAIC	Weight	Rel. Wt.
BASE	Fixed	132.40	0.28	0.04
	EstGro	96.70	0.47	0.07
	EstSel	82.80	0.55	0.08
DDQ	Fixed	51.20	0.72	0.10
	EstGro	17.40	0.91	0.13
	EstSel	29.50	0.84	0.12
TBM	Fixed	53.00	0.71	0.10
	EstGro	25.80	0.86	0.12
	EstSel	0.00	1.00	0.14
TBE	Fixed	136.10	0.26	0.04
	EstGro	127.70	0.31	0.04
	EstSel	137.90	0.25	0.04

TABLE 5. Weights assigned by each expert to the alternative models, where appropriate, with respect to the plausibility of the estimates for the parameters that represent the hypotheses: *W(Plausible parameters)*. H: high; M: medium; L: low; NA: not applicable.

2A	2B	Weight						Rel. Wt.
BASE	Fixed	NA	NA	NA	NA	NA	NA	0.11
	EstGro	M	H	L	H	H	M	0.08
	EstSel	M	H	M	H	M	M	0.07
DDQ	Fixed	H	H	H	H	H	H	0.11
	EstGro	M	H	L	H	H	M	0.08
	EstSel	M	H	M	H	M	M	0.07
TBM	Fixed	H	H	H	H	M	H	0.1
	EstGro	M	H	L	H	M	M	0.06
	EstSel	M	H	M	H	M	M	0.07
TBE	Fixed	H	H	H	H	M	H	0.1
	EstGro	M	H	L	H	M	M	0.06
	EstSel	M	H	M	H	M	M	0.07

TABLE 6. Results of applying the algorithm to calculate the weights based on the likelihood profile of R_0 and the ASPM diagnostic: *W(R_0, ASPM)*. H: high; M: medium; L: low.

2A	2B	Weight	Rel. Wt.
BASE	Fixed	M	0.1
	EstGro	M	0.1
	EstSel	L	0.05
DDQ	Fixed	M	0.1
	EstGro	M	0.1
	EstSel	M	0.1
TBM	Fixed	M	0.1
	EstGro	M	0.1
	EstSel	L	0.05
TBE	Fixed	M	0.1
	EstGro	M	0.1
	EstSel	L	0.05

TABLE 7. Results of the weights for the composition residuals: *W(Composition residuals)*. H: high; M: medium; L: low.

2A	2B	Weight	Rel. Wt.
BASE	Fixed	M	0.07
	EstGro	M	0.07
	EstSel	M	0.07
DDQ	Fixed	M	0.07
	EstGro	M	0.07
	EstSel	M	0.07
TBM	Fixed	M	0.07
	EstGro	M	0.07
	EstSel	M	0.07
TBE	Fixed	H	0.13
	EstGro	H	0.13
	EstSel	H	0.13

TABLE 8. Results for the weights for the retrospective analysis: *W(Retrospective analysis)*. H: high; M: medium; L: low; N: none.

2A	2B	Weight	Rel. Wt.
BASE	Fixed	L	0.04
	EstGro	L	0.04
	EstSel	L	0.04
DDQ	Fixed	H	0.14
	EstGro	H	0.14
	EstSel	N	0
TBM	Fixed	M	0.07
	EstGro	M	0.07
	EstSel	L	0.04
TBE	Fixed	H	0.14
	EstGro	H	0.14
	EstSel	H	0.14

TABLE 9. Weights assigned based on the empirical selectivity: *W(Empirical selectivity)*. H: high; M: medium; L: low.

2A	2B	Weight	Rel. Wt.
BASE	Fixed	L	0.038
	EstGro	M	0.077
	EstSel	H	0.154
DDQ	Fixed	L	0.038
	EstGro	M	0.077
	EstSel	H	0.154
TBM	Fixed	L	0.038
	EstGro	L	0.038
	EstSel	H	0.154
TBE	Fixed	L	0.038
	EstGro	L	0.038
	EstSel	H	0.154

TABLE 10. Weights assigned by each expert to the alternative values of steepness. H: high; M: medium; L: low; N: none.

h	Weight						Rel. Wt.
0.7	L	N	L	-	N	N	0.04
0.8	M	L	H	-	L	L	0.21
0.9	H	L	H	-	M	M	0.31
1	M	H	M	-	H	H	0.44

TABLE 11. Rescaled weights assigned to each model, *W(Model)* (from Equation 1, rescaled; values corresponding to "P(Model)" in Table 1 of SAC-11-08).

Model		Rel. Wt.
BASE	Fixed	0.01
	EstGro	0.05
	EstSel	0.06
DDQ	Fixed	0.03
	EstGro	0.13
	EstSel	0.09
TBM	Fixed	0.05
	EstGro	0.10
	EstSel	0.24
TBE	Fixed	0.03
	EstGro	0.06
	EstSel	0.14