Workshop on ICES reference points (WKREF2)

Lordan, Colm ; Orio, Alessandro; Kempf, Alexander; Pierucci, Andrea; Kuparinen, Anna; Rindorf, Anna; Peyronnet, Arnaud; Wilson, Ashley; Albertsen, Christoffer Moesgaard; Konrad, Christoph

Total number of authors:
54

Link to article, DOI:
10.17895/ices.pub.20557008

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

Citation (APA):
Lordan, C. (Ed.), Orio, A., Kempf, A., Pierucci, A., Kuparinen, A., Rindorf, A., Peyronnet, A., Wilson, A., Albertsen, C. M., Konrad, C., Sparrevohn, C. R., Minto, C., Howell, D., Gilljam, D., Miller, D., Garcia, D., Armelloni, E., Abad, E., Masnadi, F., ... Bartolino, V. (2022). *Workshop on ICES reference points (WKREF2)*. International Council for the Exploration of the Sea (ICES). ICES Scientific Report Vol. 4 No. 68 https://doi.org/10.17895/ices.pub.20557008

WORKSHOP ON ICES REFERENCE POINTS (WKREF2)

VOLUME 4 | ISSUE 68

ICES SCIENTIFIC REPORTS

RAPPORTS SCIENTIFIQUES DU CIEM

ICES INTERNATIONAL COUNCIL FOR THE EXPLORATION OF THE SEA

CIEM CONSEIL INTERNATIONAL POUR L'EXPLORATION DE LA MER

International Council for the Exploration of the Sea
Conseil International pour l'Exploration de la Mer

H.C. Andersens Boulevard 44-46
DK-1553 Copenhagen V
Denmark
Telephone (+45) 33 38 67 00
Telefax (+45) 33 93 42 15
www.ices.dk
info@ices.dk

ISSN number: 2618-1371

ICES Scientific Reports

Volume 4 | Issue 68

WORKSHOP ON ICES REFERENCE POINTS (WKREF2)

Recommended format for purpose of citation:

ICES. 2022. Workshop on ICES reference points (WKREF2).
ICES Scientific Reports. 4:68. 96 pp. http://doi.org/10.17895/ices.pub.20557008

Editor

Colm Lordan

Authors

Alessandro Orio • Alexander Kempf • Andrea Pierucci • Anna Kuparinen • Anna Rindorf • Arnaud Peyronnet • Ashley Wilson • Christoffer Moesgaard Albertsen • Christoph Konrad • Claus Reedtz Sparrevohn • Cóilín Minto • Daniel Howell • David Gilljam • David Miller • Dorleta Garcia • Enrico Armelloni • Esther Abad • Francesco Masnadi • Giuseppe Scarcella • Gjert Endre Dingsør • Henning Winker • Henrik Sparholt • Inês Farias • Jan Horbowy • Jean-Baptiste Lecomte • Jeffrey A. Hutchings Johanna Fall • Johan Lövgren • John Simmonds • Jonathan Shrives • José De Oliveira • Kristiina Hommik Laurence Kell • Lies Vansteenbrugge • Lisa Borges • Luke Batts • Marc Taylor • Martin Pastoors Martina Scanu • Massimiliano Cardinale • Michaël Gras • Mikael van Deurs • Nicolas Goñi • Norman Graham • Paula Silvar Viladomiu • Paz Sampedro • Ray Hilborn • Rishi Sharma • Sarah Millar • Sofie Nimmegeers • Tanja Miethe • Tommi Perälä • Valerio Bartolino

ICES CIEM
International Council for the Exploration of the Sea
Conseil International pour l'Exploration de la Mer

Contents

i Executive summary

The ICES Workshop on ICES reference points (WKREF2) was tasked review the WKREF1 report and based on the outcome develop updated guidelines for the ICES reference points system and recommendations for ACOM consideration. The WKREF1 report has suggested 5 key recommendations to simplify and harmonise the ICES reference points framework representing a major change to the current guidelines. At WKREF2, we detailed discussions and four key concerns were raised about the proposed approach.

The first related to the simplification of rules to define B_{lim}. Around two thirds of category 1 stocks would end up as WKREF1 "B_{lim} Type 2" where B_{lim} would be set as a fraction of B0. The Allee effect or "depensation" maybe more important than previously thought and should be furthered explored for ICES stocks since it has important consequences for B_{lim}. A number of challenges and issues around defining B_{lim} using the current guidelines were documented. Some suggestions on improvement criteria were discussed including using classifiers to define spasmodic stocks and using change point algorithms to address non-stationary productivity regimes. However, further work is need to make these approaches operational and there was no consensus that the WKREF1 B_{lim} types should replace the current guidelines.

WKREF1 recommended that the F_{MSY} proxy should be based on a biological proxies and should be less than the deterministic F_{MSY}. It was pointed out that the stochastic F_{MSY} estimated in EqSim for example, is lower than the deterministic F_{MSY} and that the current guidelines ensure that the F_{MSY} should not pose a more than 5% risk to B_{lim}. A large amount of work described in WD 1 was carried out to develop an MSE framework to consistency and robustness test a candidate reference point system for North East Atlantic stocks. However, WKREF2 recommended that further work needs to be carried out to condition and test the simulation framework before the conclusions could be adopted by ICES and incorporated into the guidelines.

A number of considerations for defining MSY related reference points were discussed including using model validation and prediction skill to ensure that ICES provide robust and credible advice. There is evidence that density dependence (DD) is important in the majority of ICES stocks (68% in recruitment and 54% in growth). The correct prediction of the shape and strength of density-dependence in productivity is key to predicting future stock development and providing the best possible long-term fisheries management advice. A suggested approach to use surplus production models (SPMs) to account for DD in F_{MSY} was suggested and discussed but there was no consensus on whether that approach was appropriate. There was consensus that the FECO approach as a means of adapting target fishing mortality to medium-term changes in productivity should be included in the guidelines subject to a benchmark and ACOM approval.

While WKREF1 and 2 focused mainly on Category 1 stocks ToR c) called for a "simplified and harmonised set of guidelines for estimating MSY and precautionary reference points applicable in the advice framework across various ICES stock categories." Ideally the ICES assessment categories should provide equivalent risk across all stocks. This issue was discussed but no recommendations emerged.

There was no consensus a revised reference point framework was proposed at WKREF2. However, it was agreed that it should be presented here for further discussion at ACOM and other fora. The key feature of the suggested approach is that the stock status evaluation is treated independent of the Advice Rule (AR). The main feature of the system is that the biomass trigger is not linked to a stock status evaluation, it is linked to the expected biomass when fishing at the target fishing mortality, in contrast to the current ICES approach. It also entailed that F_{MSY} would also become an upper limit of fishing mortality and that the advised fishing mortality would be

set at or lower than that level. WKREF2 did not discuss what to do in situations where SSB< B_{lim} or alternative forms of HCR for the advice rule. Building community understanding and consensus around simplified and harmonised guidelines has yet to be achieved. A further workshop WKREF3 will be required to achieve that aim. The report includes 6 recommendations for ACOM consideration.

ii Expert group information

Expert group name	Workshop on ICES reference points (WKREF2)
Expert group cycle	Annual
Year cycle started	2021
Reporting year in cycle	1/1
Chair(s)	Colm Lordan, Ireland, Rishi Sharma, FAO, Italy.
Meeting venue(s) and dates	11-13 January 2022, Online meeting (58 participants)

This report is dedicated to the memory of Professor Jeffrey Hutchings who through his career contributed huge insights in the fields of fisheries science and advice. He made an important presentation to WKREF on the “Allee effect” shortly before his untimely death on 30 January 2022.

Opening of the meeting

The ICES Workshop on ICES reference points (WKREF2) was held online, on 11–13 January 2022. The list of participants and contact details are given in Annex 2. The chairs, Colm Lordan (Ireland) and Rishi Sharma (FAO, Italy) welcomed the participants and highlighted the variety of Terms of References (ToRs). The draft agenda was presented and ToRs for the meeting (Section 1) were discussed. The agenda was agreed and the online meeting proceeded.

1 Terms of Reference

The **Workshop on guidelines for reference points** (WKREF2) chaired by Colm Lordan, Ireland and Rishi Sharma, Italy, will meet as a hybrid meeting online and in ICES, 11-13 January 2022 to:

1. Review the outcome of the Workshop on ICES reference points (WKREF1).
2. Based on the outcome of WKREF1, develop best practice guidelines on the estimation of reference points with worked examples.
3. Develop recommendations for ACOM on a simplified and harmonised set of guidelines for estimating MSY and precautionary reference points applicable in the advice framework across various ICES stock categories.

WGREF2 will report by 15/02/2022 for the attention of the Advisory Committee.

Supporting information

Priority	High
Scientific justification	WKREF1 will propose a range of candidate methods to define and estimate reference points based on best available science which are appropriate to the ICES advisory framework and end user needs. WKREF2 will explore these methods in more detail by applying them to a range of ICES stocks and where possible also simulation testing the methods. Based on these worked examples the WK will make recommendations to ACOM on reference points guidelines. In relation to b) the worked examples will need to be clearly documented in TAF for the community to use in the future.
Resource requirements	One meeting room at ICES HQ with at least one breakout room and facilities for online participation.
Participants	Scientists with experience and interest in reference points definition and estimation procedures from inside and also from outside the ICES area.
Secretariat facilities	Secretariat administrative, scientific and TAF support.
Financial	No financial implications.
Linkages to advisory committees	The results of this work will directly feed the ICES advisory process.
Linkages to other committees or groups	HAWG, WKGMSE3, WGWIDE, WGBFAS, WGCSE, WGNSSK, NWWG, AFWG, WGHANSA
Linkages to other organizations	All advice recipients having an interest in ICES reference points.

2 Scope of the workshop

The main objective of the workshop was to review the recommendations of WKREF1 and consider how these might feed into a new reference points framework and guidelines for ICES. There were a number of presentations on the wider issues of best practice for reference points, the Allee effect, density dependence and the WKIRISH approach. The starting point was to try and develop a set of simplified and harmonised guidelines based on the WKREF1 report rather than evolving the current guidelines to include the WKREF1 conclusions. A key aspect of the meeting was to allow for discussions in order to build a shared understanding of the strengths and weakness of the current framework and of the new framework emerging from WKREF1.

3 Review of WKREF1

WKREF1 was tasked to provide a thorough review of the ICES reference points system as a basis to re-evaluate the process for estimating, updating and communicating reference points in the context of the ICES advice (ICES, 2022). The key recommendations of WKREF1 were to:

i) revise and simplify how B_{lim} is derived, including the possibility to determine B_{lim} as a fraction of B_0 based on biological principles and international best practice;

ii) $F_{P.05}$ should be calculated without $B_{trigger}$

iii) to use biological proxies for deriving F_{MSY}, and the F_{MSY} proxy must not exceed $F_{P.05}$ consistent with current ICES guidelines

iv) to report a biomass target (B_{trg}) that corresponds to the F_{MSY} proxy

v) to set $B_{trigger}$ as either a fraction of B_{trg} or multiplier of B_{lim}

This constituted quite a major change compared to the current guidelines (ICES, 2021). Four key concerns were raised in response to the initial presentation on the WKREF1 findings. The first related to the three options for B_{lim} proposed, this is discussed further in section 4.

The second related to the fact that the simulations carried out at WKREF1 used a Beverton and Holt SR model, whereas in the current ICES framework different functional forms of SR are used and the experts consider the most appropriate one(s) at benchmarks. WKREF1 recommended that the F_{MSY} proxy should be based on biological proxies and should be less than the deterministic F_{MSY}. WKREF2 concluded that if either a Beverton-Holt or hockey-stick S-R relationship are used, or maybe other functional forms as well, then a precautionary F_{MSY} would be needed, i.e the F_{MSY} should not pose a more than 5% risk to B_{lim}, consistent with the current ICES guidelines. However, it was pointed out that the stochastic F_{MSY} estimated in EqSim for example, is lower than the deterministic F_{MSY} (also estimated in EqSim).

The third issue related to how well B_0 could be estimated given the long history of exploitation of stocks in the ICES area compared to other areas. For New Zealand and US northwest coast rockfish stocks, for example, the fisheries have recently developed and plausible levels of B_0 are known, whereas for North Sea cod, it is difficult to know what B_0 might be. Mace and Sissenwine (1993) only investigated heavily exploited stocks in Europe and North America to estimate replacement %SPR. This was explored further during the meeting through the use of sanity checks.

The final concern that surfaced was around communication and governance considerations, and the risk of scientist overstepping their remit and taking management decisions. WKREF1 reviewed the differences between what is done inside and outside ICES with reference points. There was not much consideration on how any changes suggested might fit with the governance system in the ICES area. WKREF1 recommended having a biomass target (B_{trg}) corresponding to the equilibrium biomass when fishing at the F target (F_{MSY} or proxy). It was noted that any changes to the ICES reference points framework will require very clear terminology and rational.

3.1 Consistency and Robustness testing of candidate reference point systems for North East Atlantic stocks

Working Document 1 took the key recommendations of WKREF1 and conducted a large-scale simulation testing experiment with feedback control for 64 ICES Category 1 stocks, with the aim to evaluate the consistency and robustness of candidate reference point systems. Based on the objectives of ICES advice framework (ICES, 2022), the evaluation criteria for testing consistency are based on the following objects:

(1) to not exceed a 5% probability of SSB falling below B_{lim},

(2) to achieve high long-term median yields that correspond to at least 95% of the median yield at constant F_{MSY} (MSY),

(3) to attain a high probability that SSB is above the FAO threshold of 80% of the B_{trg} proxy for B_{MSY}.

By considering stock-specific productivity and taxonomic grouping, WD1 then put forward the recommended candidate reference point systems for further robustness testing under alternative misspecifications of the stock recruitment relationship. Based on the simulation results, WD1 presents straightforward and transparent guidelines for setting optimal reference points depending on the stock's productivity characteristics. WD1 aligns this new reference point system with a status classification system that is intended to facilitate clear and unambiguous interpretation of the stock status.

Key results and conclusions

The results of both the self-test and robustness test clearly highlight the need to consider the stock's biology and productivity for setting reference points. Based on these results the following guidelines for setting reference points for category 1 stocks assessed by ICES are proposed according to productivity category (Table 3.1).

Table 3.1 Guidelines for deriving target and trigger reference points in the newly proposed ICES system. The Type 1 and 2 approaches (see Section 4 for types) can be used for all stocks to derive B_{lim}. SRR: Stock-recruitment relationship; BH: Beverton- Holt; HS: Hockey-Stick.

	Productivity	F_{trg}	B_{trg}	$B_{trigger}$	SRR
SPR%	Low	F_{spr40}	B_{spr40}	0.8 x B_{trg}	BH/HS
	Medium	F_{spr40}	B_{spr40}	0.8 x B_{trg}	BH/HS
	High	F_{spr50}	B_{spr50}	1x B_{trg}	BH/HS
B%	Low	F_{B35}	B_{35}	0.8x B_{trg}	BH
	Medium	F_{B35}	B_{35}	0.8x B_{trg}	BH
	High	F_{B40}	B_{40}	1x B_{trg}	BH

Low and medium productive species - $F_{SPR40\%}$ with stock and recruitment modelled as BH or HS fulfils both the PA and the MSY criteria and is proposed as candidate for the future ICES system to derive a target reference point (TRP). $F_{B35\%}$ with stock and recruitment modelled as BH fulfils both the PA and the MSY criteria and is proposed as candidate for the future ICES system to derive a target reference point TRP. B_{lim} can be derived as either Type 1 or Type 2 (see section 4 for types), B_{trg} (the median biomass when fishing at TRP) is the SSB that corresponds to $F_{SPR40\%}$ or $F_{B35\%}$ and $B_{trigger}$ is set at 0.8 B_{trg}.

High productive species - $F_{SPR50\%}$ with stock and recruitment modelled as BH or HS fulfils both the PA and the MSY criteria and is proposed as candidate for the future ICES system to derive TRP. $F_{B40\%}$ with stock and recruitment modelled as BH fulfils both the PA and the MSY criteria and is proposed as candidate for the future ICES system to derive TRP. B_{lim} can be derived as either Type 1 or Type 2, B_{trg} is the SSB that corresponds to $F_{SPR50\%}$ or $F_{B40\%}$ and $B_{trigger}$ is set equal to B_{trg} or higher.

WKREF2 recognised the large amount of work described in WD 1 and the framework developed. However, similar concerns to those already outlined above were raised regarding the need to condition the simulation for the specificities of individual stocks (e.g. stock specific SR functions) rather than using a generic approach, and the requirement to sanity-check the results carefully before making far-reaching changes to the current framework and guidelines. WKREF2 recommended that further work is carried out on the simulation framework before the conclusions could be adopted by ICES.

4 Considerations for defining B_{lim}

WKREF1 proposed a simplified framework with 3 different types when defining B_{lim}.

B_{lim} Type 1: Consider an empirical Hockey-Stick for deriving B_{lim} only if the data show contrast and a break point is clearly defined

B_{lim} Type 2: Determine a plausible B_{lim}/B_0 ratio based on biological principles and life history of the stock (for instance, 10% to 25% of B_0 depending on the type of stock)

B_{lim} Type 3: For stocks where the stock development is dominated by occasional good year-classes (i.e. spasmodic recruitment), the lowest observed SSB(s) that gave rise to a good year class can be used as basis for B_{lim}

Alternative approximations (i.e. current type associated with subjective decisions) should be discouraged. Biological plausibility checks (e.g. $B_{lim} > 0.1B_0$) to ensure there is a sufficient safety margin when setting B_{lim}.

In practice around two thirds of category 1 ICES stocks would end up as Type 2 according to the analysis presented at WKREF1. This raises two fundamental questions: how well B_0 is estimated in practice for those stocks, and what % of B_0 is appropriate. There was no consensus that the WKREF1 B_{lim} types should replace the current guidelines. Below are some of the considerations discussed by WKREF2.

4.1 Allee Effects, Allee-Effect Thresholds, and Their Potential Utility in Setting Limit Reference Points

Interest in Allee effects, or 'depensation', in marine fishes has increased since the first meta-analysis in the mid-1990s (Myers *et al.*, 1995); examples include Liermann and Hilborn (1997, 2001), Keith and Hutchings (2012), Hilborn *et al.* (2014), Hutchings (2014, 2015), and Perälä and Kuparinen (2017). One stimulus for this interest is the observation that cessation of overfishing has not always resulted in stock recovery and rebuilding, raising the question of whether population-size thresholds exist below which recovery is significantly impaired. A second stimulus lies in the establishment of reference points in support of sustainable fisheries management, particularly limit reference points for stock biomass, such as B_{lim} (ICES, 2022).

Allee effects describe a positive association between population size (e.g. SSB) and realized (as opposed to maximum, rmax) per capita population growth rate (realized or simply r), a metric of the average individual fitness in a population. An Allee effect describes a pattern, not causal mechanisms. Classic stock-recruitment (S-R) models implicitly assume that compensatory dynamics persist as populations decline, meaning that r continually increases as SSB declines. However, if a declining population reaches a size below which the strength of negative dependence weakens, or r begins to decline, with declining SSB, this pattern can be indicative of an Allee effect.

The SSB at which r begins to decline (relative to the negatively density dependent pattern exhibited at larger SSB) is termed the Allee-effect threshold (Hutchings, 2015; Perälä *et al.*, 2022). That is, Allee-effect thresholds identify the SSB below which negative density-dependence weakens and below which stock recovery is increasingly impaired and uncertain.

The most common metric of r in the fisheries literature on depensation is recruits per spawner (e.g. R/SSB). However, it is important to note that r may well be more sensitive to changes in natural mortality (M) than changes in R/S. If M increases as SSB declines (e.g. Swain & Benoît, 2015), this will almost always reflect depensation, irrespective of the pattern in R/S (Kuparinen *et al.*, 2014; Hutchings, 2015). Increased variance in M with declining SSB (Minto *et al.*, 2008) is also likely to cause an Allee effect. Lack of awareness of the importance of M to r, coupled with empirical challenges in estimating M, may mean that a weakening of the strength of negative density dependence below the Allee-effect threshold (Keith and Hutchings, 2012) is more common than the literature would suggest.

The greater the magnitude of population decline, the greater the likelihood that Allee effects will be manifest. However, empirical determination of where these population-size or Allee-effect thresholds are in relation to parameters such as B_0 or SSB_{MSY} remains a challenge. This provided impetus to compare compensatory and depensatory S-R models for Atlantic herring, *Clupea harengus*, and Atlantic cod, *Gadus morhua*. For example, by addressing methodological issues associated with previous analyses, Perälä and Kuparinen's (2017) Bayesian statistical approach documented depensation in 4 of 9 herring stocks in the Northeast Atlantic. In 2022, the Bayesian inference approach was extended by applying four S-R models (Beverton-Holt, Ricker, Sigmoidal Beverton-Holt, Saila-Lorda) to study depensation in the southern Newfoundland cod stock in NAFO Subdivision 3Ps (Perälä *et al.*, 2022). In addition to finding strong evidence of Allee effects, the SSB Allee-effect threshold (below which recovery is impaired) was inferred, and determined the years during which the population dynamics of 3Ps cod switched from negative to positive density dependence.

As noted by ICES (2022), a key challenge lies in determining where Allee-effect thresholds are in relation to B_0. For 3Ps cod, Perälä *et al.* (2022) estimated that a weakening of compensation was first evident at 44-46% of SSB_{max}; this delineates the Allee-effect threshold. If the SSB of 3Ps cod in the late 1950s (when SSB_{max} was experienced) was approximately half of B_0 (a not unreasonable supposition; 3Ps cod have likely been fished since the late 1400s; Castañeda *et al.*, 2020), this would yield an Allee-effect threshold of ~0.20-0.25 B_0. This is similar to the 0.2B_0 threshold suggested by Hutchings (2014), based on previous work (Keith and Hutchings, 2012).

Key considerations:

1. The relevance of Allee effects, or depensation, to the population dynamics of marine fishes at low abundance may be more important than previously thought.
2. The most commonly applied metric of r — recruits per spawner (R/S) — may not be as sensitive a proxy for r as natural mortality, M; reductions in M with declining SSB can cause Allee effects, irrespective of patterns in R/S.
3. Using Bayesian inference and methods, recent work provides: (i) evidence of depensation in Northeast Atlantic herring and Northwest Atlantic cod; (ii) estimates of the SSB threshold at which Allee effects, or a weakening of compensation, are manifest; and (iii) methods for determining the SSBs and time frame during which the population dynamics of depleted stocks might switch from negative to positive density dependence.

WKREF2 recommends that the evidence for Allee effects in ICES stocks should be explored further using the latest methods that have been developed.

4.2 Challenges with the current approach to define B_{lim}

A number of challenges were mentioned in connection with the application of the current framework (ICES S-R type specific guidance in ICES 2021) to define B_{lim} in ICES benchmark processes. A general observation was that the guidelines provided much space for subjective interpretation and this made the process of estimating B_{lim} non-reproducible, not transparent to outsiders and possibly prone to inconsistencies. A number of specific issues where this is seen as a problem were mentioned, including the following list:

- WKREF1 found that the most common approach was to set B_{lim} at B_{loss} (32 out of 79, S-R Type 5 stocks) or to derive B_{lim} from B_{loss} (10/79, Type 6 stocks). This calls into question the biological relevance of B_{lim} for those stocks;
- WKREF1 also found for 16 out of 79 stocks B_{lim} was set at the lowest SSB where recruitment is good/high/not impaired across various stock types, which is more in line with the definition of B_{lim}, but is not part of the current guidelines (ICES, 2001);
- Defining B_{lim} with short or truncated time-series. There has been an increasing tendency to truncate time-series due to concerns about changing productivity regimes (e.g. North Sea herring her.27.3a47d and North Sea cod cod.27.47d);
- Unclear criteria to define a stock as spasmodic (ICES S-R type 1) and inconsistent selection of B_{lim} when S-R Type 1 is chosen;
- Differences in hockey-stick/segmented regression breakpoints from different methods (including different windows and different estimation methods) (ICES S-R type 2);
- Differences in B_{lim} when recruitment is ever increasing when stock sizes increases depending on the approach taken (ICES S-R type 3);
- Differences in derivation of B_{lim} when S-R shows no pattern (ICES types 5 and 6);
- Differences in the estimation of B_{lim} or B_{pa} based on output from biomass models (e.g. SPICT).

A further complication was cases where large natural variability in the stock recruitment relationship may lead to a greater than 5% risk of falling below B_{lim} even under no fishing-scenarios.

In addition to this, concerns were raised that the use of B_{pa} estimated as 1.4 x B_{lim} (or $B_{lim} \times \exp(1.645 \times \sigma)$) as MSY $B_{trigger}$ would not ensure MSY. WKREBUILD pointed out that if B_{lim} and MSY $B_{trigger}$ are too close to each other, small reductions in biomass below MSY $B_{trigger}$ can lead to large changes in F with little time for the stock to adapt/respond (ICES, 2020).

4.3 Lack of confidence in the accuracy of the estimated recruitment and SSB

The assessment models occasionally provide highly uncertain estimates of recruitment and SSB. In this case, a different model can be investigated. If this does not solve the issue, the stock should be moved to stock assessment category 3.

4.4 Spasmodic stocks

Stocks with spasmodic recruitment are common for many fish species and their management is particularly challenging (Licandeo *et al.*, 2020). In ICES, spasmodic stocks (SR type 1) are defined as "stocks with occasional large year classes" (ICES 2021b). The Spencer and Collie (1997) classification identified spasmodic stocks as those having the highest variation in their study, with low-frequency components without clear periodicities. Spasmodic recruitment might have long periods of weak recruitment with infrequent or irregular strong recruitment.

An approach to objectively classify stocks as spasmodic was presented and discussed at WKREF2. This involved using objective measures to define spasmodic stocks, such as CDFs (Figure 4.1). The cumulative distribution function (CDF) can be used to identify high variance and infrequent strong recruitment, but the order of the time-series is not preserved and long periods of low recruitment (e.g. some haddock stocks) or infrequent strong recruitments (as seen in some redfish) cannot readily be identified. One promising approach would be to used CDF intervals and sigma after removing low variability.

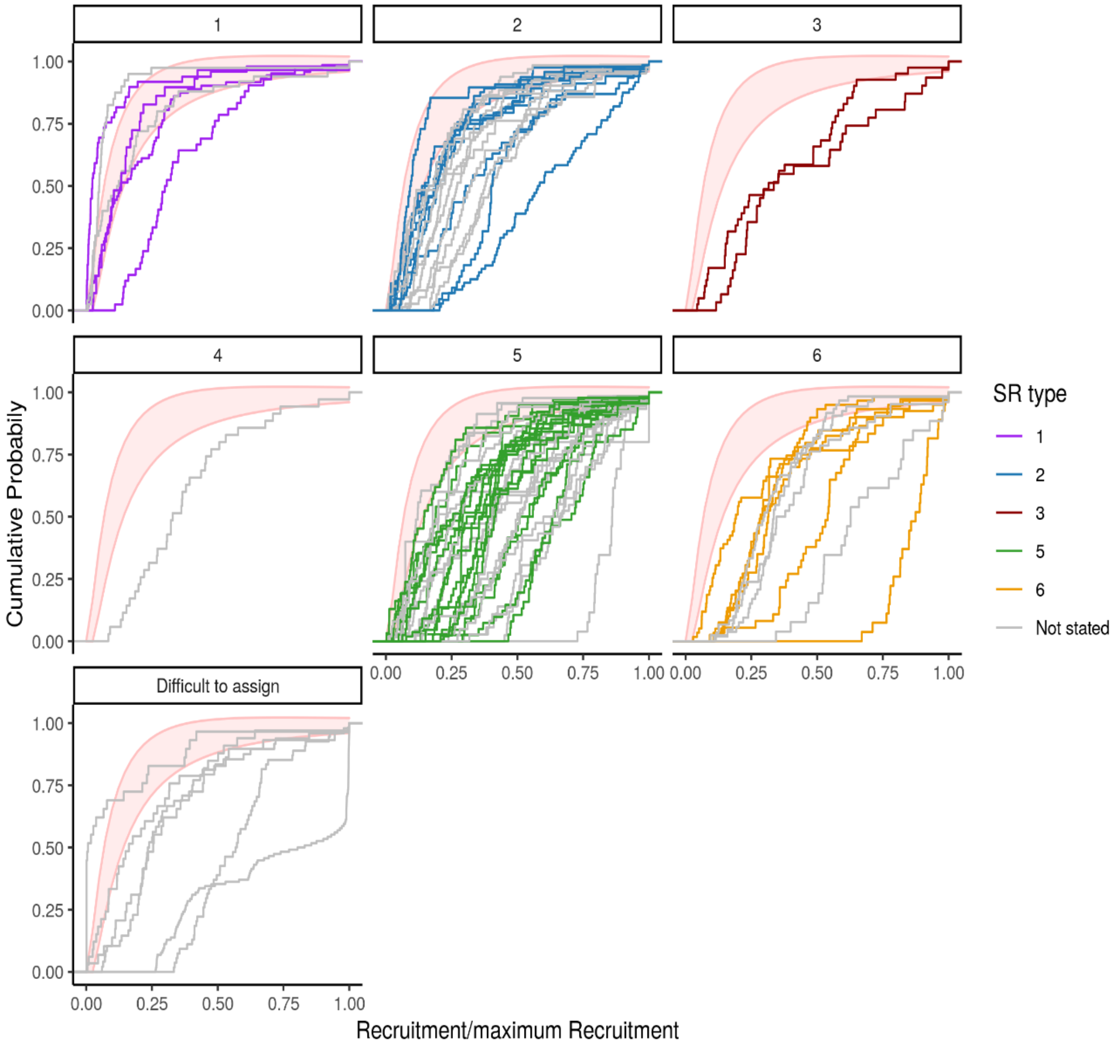


Figure 4.1. Empirical cumulative distribution function (CDF) of recruitment relative to maximum recruitment by inferred stock-recruitment type. Colour shows stated stock-recruit type. Pink area shows the theoretical expected 80% interval for CDFs of time series (length = 42) of lognormal variance = 1.

More research is needed to define spasmodic criteria, as well as simulation evaluation on how to define reference points and manage this type of stock. Nature of recruitment time series variation could be used in developing a control system. Another important effort would be to perform simulations to understand how a spasmodic recruitment links with the productivity of the stock and thus the reference points.

Stocks identified as spasmodic use approaches such as ICES S-R type 1, 2 or 3. Often single data points are highly influential on the outcome of these methods (Figure 4.2). It is possible that the sensitivity to single points can be reduced by estimating B_{lim} after simulating estimates from stock recruitment relationships including parametric (bimodal or heavy tailed) bootstrap of residuals. No solutions to the issue were presented.

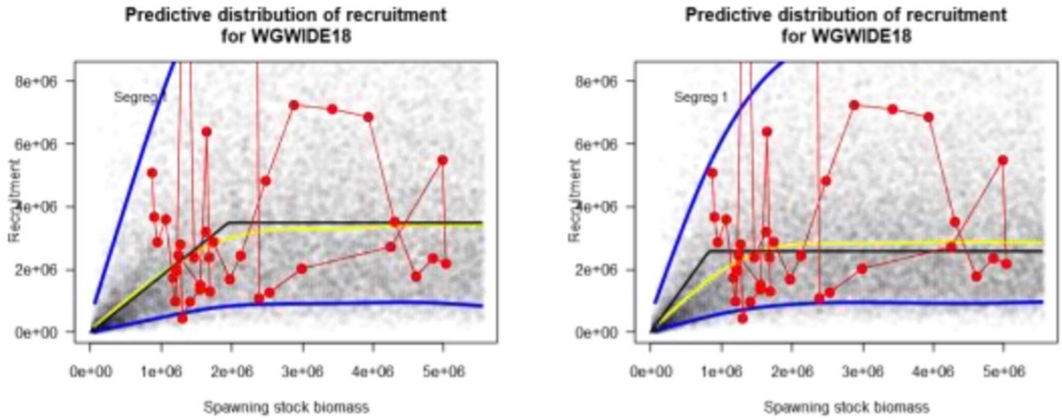


Figure 4.2. Segmented regression of Western horse mackerel including all data points (left) and removing the 1982 data point (right) (ICES IBHM).

4.5 B_{lim} when S-R is ever increasing

As mentioned above in section 4.2, estimation of hockey-stick (HS) break-points is sensitive to the method used, and the grid-search method recommended by Barrowman and Myers (2000) is computationally intensive. Ever increasing relationships are frequent among stocks (32%, Rindorf *et al.*, 2021) and present specific problems, because the B_{lim} is not clearly defined. For example, the estimated B_{lim} of herring west of Scotland ranges from the lowest observed to almost the highest observed SSB depending on the method used (Figure 4.3). Another example is North Sea cod, where HS estimation results in a breakpoint right of the middle of the plot which does not fully match the breakpoint as imagined by eye-balling the figure (Figure 4.4). It was suggested that more appropriately accounting for differences between different productivity periods may help narrow the range of possible B_{lim} estimates. However, this should not be done by truncating the time series (see next section).

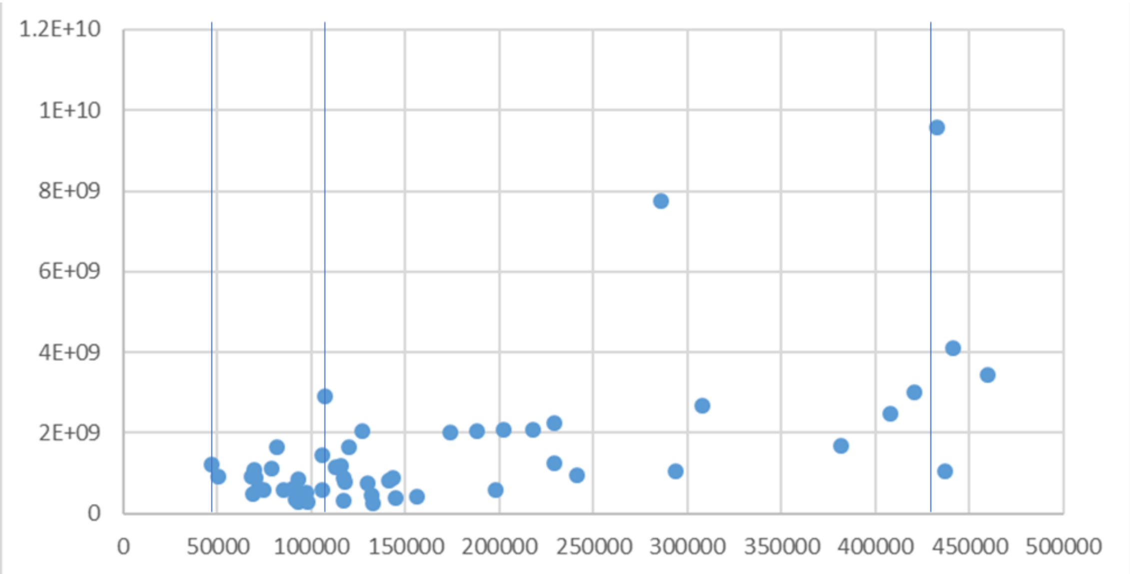


Figure 4.3. Stock recruitment relationship of herring west of Scotland. Vertical lines from left to right: P50 = 46 600, P80 = 107 000 and HS = 430 770 (method from van Deurs *et al*., 2020).

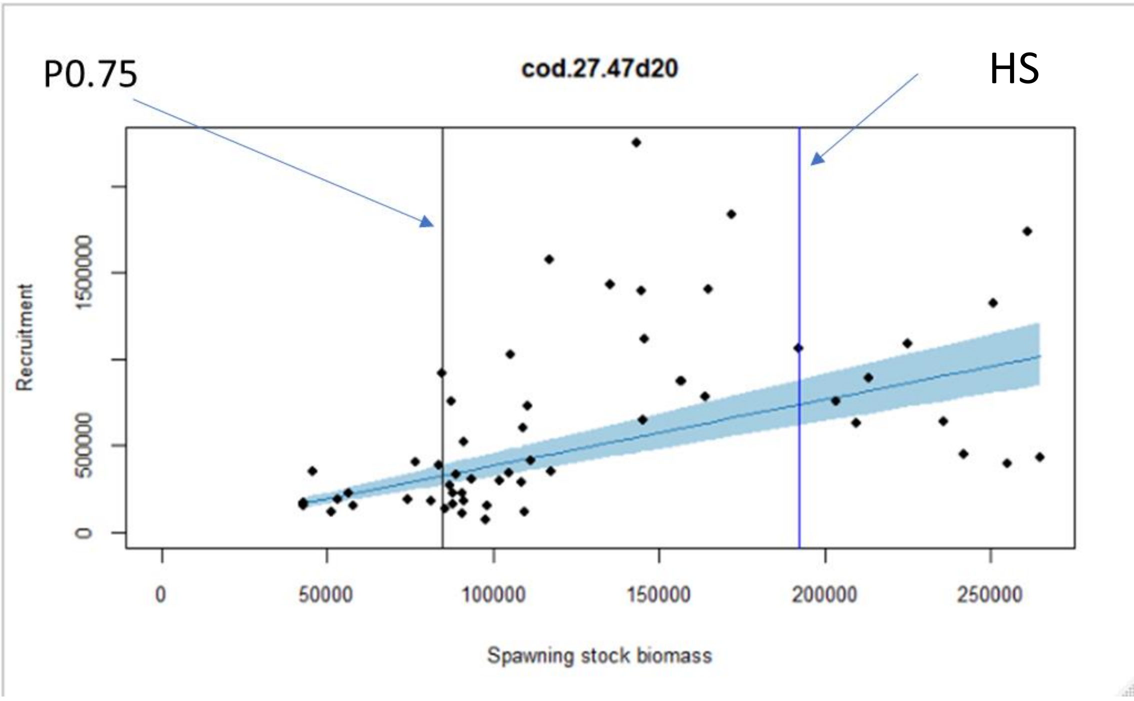


Figure 4.4. Stock recruitment relationship of North Sea cod.

4.6 New approaches to account for changing regimes

Non-stationarity in productivity has implications for reference points, including B_{lim}. The Bayesian online change-point detection (BOCPD) algorithm applied to stock-recruitment relationships (SRRs) could be used to partition the recruitment and spawning stock biomass time-series (Perälä *et al.*, 2016). The output of BOCPD can be used to segment the data into different regimes, and SRR can be fitted for each segment separately. This will result in a different set of SRR parameters for each regime based on which RPs can be then derived. Furthermore, the full posterior and posterior predictive distributions of BOCPD representing the current state of the SRR and used to predict the future recruitment, respectively, automatically incorporate the uncertainty about the current regime. Thus, less weight is given to older observations in estimating the parameters of SRR if there is evidence of a regime shift in the data. The probability of a regime shift is a combination of the modeler's prior belief about the frequency of such shifts and the discrepancy between the posterior predictive distribution of the model and the most recent observations. To ensure that the information in the data is used most efficiently, the model should specifically investigate whether only the level of recruitment is changing or also the shape of the relationship.

Re-calculating B_{lim} on truncated time-series when moving between regimes should only be allowed if this does not increase the risk of getting into a non-precautionary situation, i.e. B_{lim} is reduced when moving into a low productivity regime. This requires a good process understanding i.e. what is driving recruitment and how strong is the impact of SSB on recruitment compared to environmental factors. Forecasting changes in productivity regimes would also be needed.

4.7 B_{lim} when S-R shows no pattern (ICES types 5 and 6)

The current guidance suggests that B_{lim} in this case is either defined as B_{loss} or undefined. If B_{lim} is undefined, B_{pa} is set to B_{loss}. WKREF1 pointed out that B_{lim}, defined in this way, is simply a consequence of the history of the exploitation of the stock, and has no biological underpinning. If a stock is already depleted, it could encourage further stock depletion. WKREF1 recommended that B_{lim} would more appropriately be defined as some fraction of B_0 in these cases. This fraction would be stock dependent, and there is a lot of variability between stocks, which implies that in this case a "least bad" estimate (e.g. B_{lim} always larger than 10% of B_0) is going to be the best we can get (ICES, 2022). Studies during the workshop of $0.1B_0$ showed that this was unrelated to B_{lim} for the stocks where both were available (Winker, *et al*. presentation). Furthermore, in some cases, the estimated $0.1B_0$ was very high compared to the observations of SSB apparently providing high recruitment. An additional observation was that a change in individual growth (weight-at-age) or natural mortality would change B_0 though the stock recruitment plot, and hence other estimates of B_{lim} would not be affected. No change to the current guidelines was therefore agreed.

4.8 B_{lim} as a fraction of B_0 in integrated models

An alternative for setting B_{lim} is to use a fraction of B_0, as suggested by WKREF1. This might constitute a viable alternative to current methodology, especially for stocks where there is no data in the plausible B_{lim} range or where there is little contrast in the SR data. In WKREF2 it was pointed out that, in integrated models where the S-R relationship is included in the assessment and the B_0 estimate must be consistent with the data and estimated stock dynamics, it may be appropriate to base B_{lim} on a fraction of B_0 (although in some cases the B_0 estimate appears relatively high compared to the historical development of the stock).

WKREF2 recommended that the guidelines should allow for benchmarks to set B_{lim} as a fraction of B_0, particularly within integrated models and ensembles, where it is internally consistent to do so. By "integrated models" we mean models like SS3 where a stock and recruitment function is included in the stock assessment and reference points estimation process.

In fact, this method is already implemented in ICES for spurdog (ICES, 2020). It would be particularly important for the development of ensemble assessments, where biomass and F in absolute terms will vary depending on the model configuration, but the ratio may stay fairly stable.

Where a benchmark decides to adopt this methodology, the suitability of using a fraction of B_0 to derive B_{lim} should be scrutinised, and the fraction to be used should be defined. It is important to analyse how well B_0 is estimated by the model in respect to historical values of SSB and also to consider environmental impacts on productivity. If stocks are influenced by environmental factors to a larger extent just looking at historic SSBs can give biased/unrealistic B_0 estimates. Comparing how B_{lim} (as derived as a fraction of B_0) relates to the SSB range observed within integrated models generally more suited for providing robust estimate of B_0.

4.9 B_{lim} or B_{pa} based on output from biomass models

The interpretation of B_{lim} and B_{pa} when estimated from biomass models is inherently different from cases where the values are based on the stock recruitment relationship. The definition that B_{lim} is the biomass below which recruitment is impaired no longer applies when using outputs from biomass models. However, to ensure that F_{MSY} estimates are precautionary, a biomass limit reference point is necessary. Estimates of unacceptable biomass levels in the literature are generally set as fractions of the median SSB when fishing at F_{MSY} (B_{MSY}) or no fishing (B_0). The latter of these depends on stock-recruitment relationships, growth, natural mortality and maturity in the unfished state. The former depends on all of these as well as the selection pattern for a stock fished for prolonged periods at F_{MSY}. Concern was raised that for many stocks, there is very limited information on the productivity of unfished stocks, and that this may lead to a highly uncertain estimate of $0.1B_0$. In general, more information is likely to be available for stocks fished at F_{MSY}, leading to a lower uncertainty of B_{MSY}, though variability in selection pattern (which is not estimated in biomass models) may act to make B_{MSY} estimates more variable.

Values of $0.1B_0$ and $0.2B_{MSY}$ were discussed and considered broadly acceptable. However, they should be provided together with estimates of their precision.

4.10 Stocks with a greater than 5% risk of falling below B_{lim} under no fishing-scenarios

Some stocks exhibit high variability even in the absence of fishing. This may lead to cases where $F_{P.05}$ is estimated to be zero, and hence F_{MSY} would be zero. The occurrence of this is likely to increase as models are expanded to fully account for all sources of uncertainty (e.g. different shape of the stock recruitment relationship or variability in natural mortality). WKMSE3 has suggested that this can be addressed by estimating the 5% additional risk relative to the risk estimated in the unfished stock.

4.11 Estimation of MSY $B_{trigger}$

The current guidelines states that B_{PA} is defined as B_{lim}*(1.645*σ) where σ is the estimated standard error of the log-transformed SSB in the final assessment year. In the absence of an estimated standard error or where this standard error is considered highly uncertain or unrealistically low, a default of $B_{PA} = B_{lim}$*1.4 is often used, which corresponds approximately with $\sigma = 0.2$. MSY $B_{trigger}$ is estimated as the maximum of B_{PA} and the 5th percentile of SSB when fishing at F_{MSY}, unless SSB has historically been so low that estimates of the percentiles of SSB when fishing at F_{MSY} are considered highly uncertain. Because many ICES stocks have been fished above F_{MSY}, the guidelines envisage an adaptive approach where MSY $B_{trigger}$ would be redefined as stocks are fished at F_{MSY} levels for 5 or more years (see flow chart in the ICES Technical Guidelines, ICES 2021).

The above approach has led to confusion about the basis of and use of MSY $B_{trigger}$ as a stock status classifier when compared to other management systems. For example, overfishing is often defined at biomass levels higher than the 5% percentile of B_{MSY}, e.g. 50% of B_{MSY} in the US or 80% of B_{MSY} in Canada (see Hilborn, 2020). In addition, exceptions to the guidelines have also emerged. The recent benchmark for North Sea herring found that the 5th percentile of B_{MSY} was less than B_{lim}, so for that stock, MSY $B_{trigger}$ was set at 50% B_{MSY} (ICES, 2021).

The MSE work carried out by ICES for mackerel (ICES, 2020) and several of North Sea stocks (ICES, 2019) explored a range of biomass triggers and target fishing mortalities. What this work illustrates is that quite a wide range of combinations of target Fs and trigger biomasses could optimise long-term yield with relatively low risk and varying levels of inter-annual catch variation. The harvest control rules for Icelandic stocks are specifically designed to have a low target fishing mortality and low biomass triggers in order to avoid the slope of the harvest control rule. This is purposefully done to minimise inter-annual catch variations (IAV) (ICES, 2022).

The current reference point framework in ICES has mixed biological reference points with harvest control rule parameters. The generic harvest control rule within the MSY approach has advantages (e.g. inherently reduces risk) but there are also some downsides (e.g. increases complexity of reference points, may not optimise performance i.e. yield and IAV) which needs to be considered carefully. An example of this complexity is the discussion over $F_{P.05}$ and whether it should be estimated with or without the Advice Rule (AR). ACOM decided that $F_{P.05}$ with AR should be used to cap F_{MSY}; WKREF1 recommended $F_{P.05}$ should be estimated without the AR, which might be considered as being overly precautionary, since ICES advice is always given according to the MSY approach (i.e. with the advice rule).

5 Considerations for defining MSY-related reference points

5.1 Validation and Plausibility

The ICES framework relies on stock recruitment relationships (SRRs) to include density dependence, estimate reference points, and conduct projections to determine management action. However, the quantities of interest, SSB and recruitment, are not observable as they are latent variables estimated by models under a variety of assumptions. There are two main approaches, for estimating the SRR, i.e. either when fitting the assessment model or post-hoc. However, there is often a lack of information in stock assessment datasets on system processes such as the SRR, and therefore a key question is whether a fitted relationship is plausible. A definition of a highly plausible model scenario is one that fits prior knowledge well, with many different sources of corroboration, without the complexity of explanation, and with minimal conjecture (Connell *et al.,* 2006).

Currently, the primary diagnostics used to evaluate plausibility, and select and reject models when conducting stock assessment, are to examine residuals to check goodness-of-fit and to conduct retrospective analysis to check stability. However, residual patterns can be removed by adding more parameters than justified by the data, and retrospective patterns by ignoring data. Therefore, models must be validated if they are to provide robust and credible advice, which requires assessing whether it is plausible that a system equivalent to the model generated the data (Thygesen *et al.*, 2017). An alternative to residual and retrospective analysis is to perform cross-validation by omitting recent observations and then predicting their out-of-sample values. This allows the estimation of prediction skill, a measure of the accuracy of a predicted value unknown by the model relative to its observed value (Weigel, *et al.*, 2008). Prediction skill can be used to explore model misspecification and data conflicts, help to compare alternative hypotheses, weight operating models when conducting MSE, or weight models in an ensemble approach.

WKREF2 recommends therefore that stock assessment models are validated using prediction skill to ensure that they provide robust and credible advice (e.g. Saltelli *et al.*, 2020).

5.2 Compensation and overcompensation in stock recruitment relationships and somatic growth

The correct prediction of the shape and strength of density-dependence in productivity is key to predicting future stock development and providing the best possible long-term fisheries management advice (Rindorf *et al.*, 2022). Working on from WKRPCHANGE, unbiased estimators of the relationship between somatic growth, recruitment and density were identified and applied to 80 stocks in the Northeast Atlantic (one stock had data for the growth analysis only). The analyses revealed density-dependent recruitment in 68% of the stocks, while 32% were best fitted by a proportional (ever-increasing) relationship between SSB and recruitment (corresponding to ICES type 3). Excluding pelagic stocks exhibiting significant trends in spawning stock biomass, the probability of significant density dependence was even higher at 78% as the proportion of stocks best fitted by a proportional relationship decreased to 22%. The relationships demonstrated that at 0.2 times maximum observed spawning stock size, considered a proxy for either B_0 or B_{MSY}, depending on the stock, 32% of the stocks attained 75% of maximum recruitment. Hence, for 68% of the stocks, a biomass limit of 0.2 times B_{MSY} (corresponding to $0.1B_0$ if

$B_{MSY} = 0.5B_0$) would not be sufficient to avoid decreased recruitment. The estimated recruitments at 0.2 times maximum observed spawning stock size were scrutinised to determine if a unimodal distribution with a clear median could be determined for any of the species groups, indicating that the median would be a useful predictor of recruitment at 0.2 times maximum observed spawning stock size for stocks where the stock recruitment relationship is unknown. The distribution of benthic and demersal show several peaks along the whole domain, indicating that a specific fraction of maximum observed spawning stock size cannot be used as a high probability estimate of when recruitment is impaired (Figure 5.1). For pelagic stocks, the distribution was unimodal with a median of 0.4, indicating that the pelagic stocks had a 50% probability of producing less than half the maximum recruitment at 0.2 times B_{MSY}.

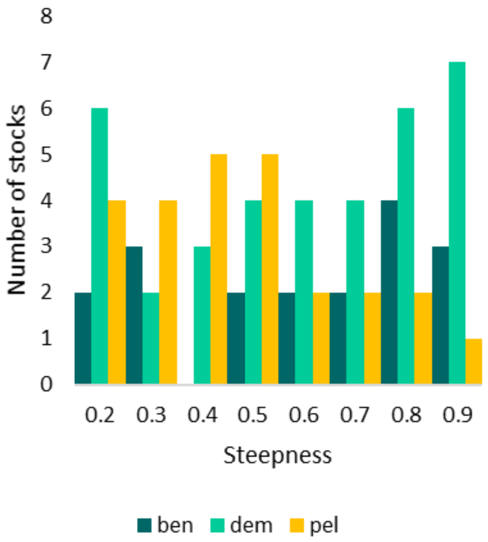


Figure 5.1. Frequency plot of steepness = recruitment (as a proportion of maximum recruitment) attained at 0.2 times maximum observed spawning stock size for 79 stocks in the North Atlantic.

Significantly lower recruitment at high stock size than at intermediate stock size (overcompensation, corresponding to Ricker-like relationships) was seen in 38% of the stocks, indicating this to be common among stocks. Pelagic stocks were less likely to exhibit density dependence in recruitment than demersal and benthic stocks. Density dependent decreases in growth after recruitment occurred in 54% of the stocks.

A direct comparison of SSB reference points, estimated as break points in a hockey stick relationship, was made with 0.2 times maximum observed spawning stock size, as well as other approaches, as conducted for pelagic stocks by van Deurs *et al*. (2021). The definitions of the reference points are illustrated in Figure 5.2, which also shows the relationship between the breakpoint of a hockey stock relationship (ICES type 2) and 0.2 times the maximum observed biomass in Figure. 5.3. Similar to the previous study, there is no relationship between the hockey stock breakpoint and 0.2 times the maximum observed biomass.

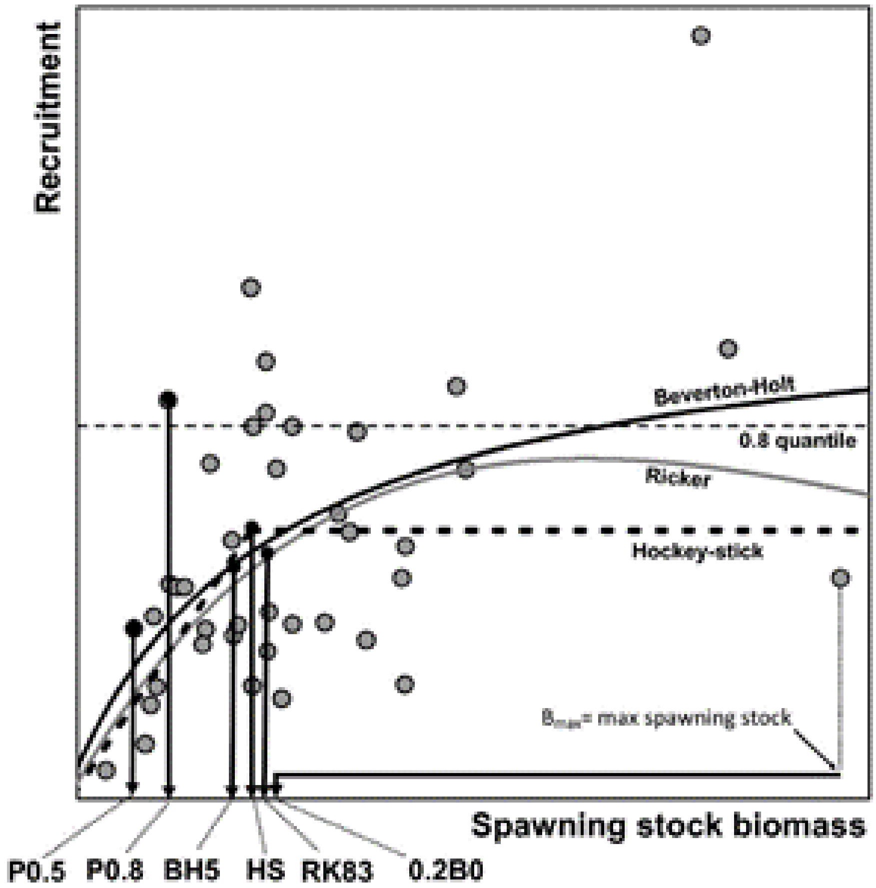


Figure 5.2. Illustration of different methods (P0.5, P0.8, HS, RK83, BH51, 0.2Bmax) for estimating biomass thresholds (BTs). Hockey stick (black dotted), Ricker (grey) and Beverton–Holt (black) curves fitted to SR data from a hypothetical stock (grey dots). The dashed horizontal line represents the 0.8 quantile of recruitment, and the maximum spawning stock biomass (B_{max}) used in 0.2Bmax approach is indicated by a bent arrow. The vertical arrows point to the spawning stock biomasses representing the BT derived from each of the methods. Reproduced from van Deurs *et al*. (2021), fig. 1.

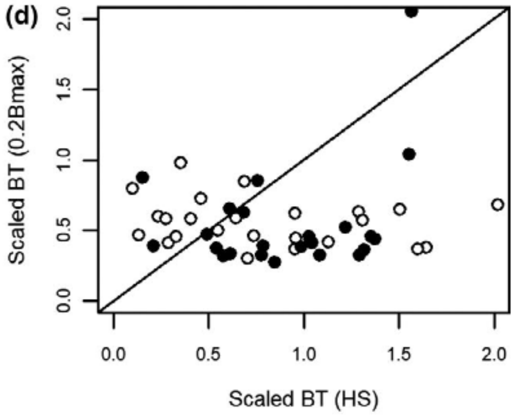


Figure 5.3. Comparison of biomass thresholds (BTs) of 51 small-bodied pelagic stocks scaled to geometric mean spawning stock biomass calculated using HS and 0.2Bmax. Stocks with relatively well-defined hockey stick breakpoints (black dots, CV_{HS}< 0.3) and stocks with poorly defined breakpoint estimates (white dots, CV_{HS}> 0.3). Axes were cut off at 2.0. Hence, outlier values twice the geometric mean spawning stock biomass were not included in the plots (amounting to 5 data points). Reproduced from van Deurs et al. (2021), fig. 4.

In summary, there is no correlation between biomass reference points determined from the stock recruitment relationship and biomass reference points determined as a fraction of maximum observed biomass. This conclusion is not related to where the reference points are located relative to median biomass (i.e. whether the stock has shown high contrast in SSB or not).

5.3 Density dependence and F_{MSY}

The Fmsy-project (www.fmsyproject.net , with condensed results in Sparholt *et al*. 2020) has postulated that there is a substantial systematic bias in ICES current F_{MSY} calculations due to missing 3 out of 4 density dependent mechanisms. The Fmsy-project also suggests a way forward. Further work is done in a new, ongoing MSE-project (www.mseproject.org), which works in close cooperation with ICES benchmark and methods workshops. The following builds on a presentation of the relevant results from these projects and the discussion at WKREF2.

It was suggested that the current ICES approach for data-rich stocks underestimates F_{MSY} because density dependence (DD) in growth, maturity and natural mortality are ignored. This is a mathematical fact based on the mechanics in the calculations and shown by many case studies see e.g. Gislason (1999); Collie, J. S., and Gislason, H. (2001), Pope *et al.* (2006), ICES WGSAM (2008), ICES Advice Report (2012 and 2013), Froese et al. (2016), Andersen *et al.* (2017), Szuwalski *et al.* (2017), Zhou *et al*. (2019) and Sparholt *et al.* (2020). This means that the statement made in WKREF1: "…as long as benchmarks are conducted every 3-5 years …, explicit modelling of the density dependence is not required ....", is problematic. For instance, even if the stock has been around B_{MSY} for a long time and weight-at-age have stabilised, if DD in growth is not modelled dynamically, F_{MSY} will be underestimated. However, to model density dependence you need a basis to do so and extrapolating beyond the range of observations for something like DD might entail considerable risk.

Density dependence did not matter much for fisheries management for decades because the stocks were overfished, and direction of advice was easy: managers should reduce F. Now that F in general has reduced significantly for ICES stocks, the issue is getting important as many ICES fish stocks are rebuilt to levels where density dependent effects may matter.

A meta-analysis of 53 data-rich ICES stocks showed that on average F_{MSY} is 50% higher than ICES current values (Sparholt *et al.* 2020). This was based on an ensemble approach using ICES multispecies models for some stocks, age-based models including sub-models of DD for all four parameters for a few stocks, and Surplus Production Models (which by design include all DD effects although not in a disentangled way) for all stocks.

DD is how ecosystem works – without it, all stocks would increase indefinitely (F_{MSY} would be zero – sustainable fisheries would not be possible). DD is a proxy for multispecies interactions, food limitation, diseases, etc. It is a useful simplification. S-R models are based on it. However, often we do not see DD in our data because, stock size variation is too small or the noise in the data too big. This should, however, not mislead us to assume DD doesn't exists for the stock we are working with, just because we do not see a statistically significant manifestation of it.

Recently, meta-analysis has been conducted to see how often we see DD in the ICES stocks. Zimmermann *et al.* (2018) looked at 70 ICES stocks and found significant DD in R in 70% and DD in weight-at-age in 69%. Rindorf *et al.* (2022) looked at 80 ICES stocks and found DD in R in 68% and in growth after recruitment in 54%. There are several documentations of DD in smaller sets of stocks (e.g., Morgan *et al*. (2016), Lorenzen (2016), Sparholt and Cook (2009), Horbowy and Luzeńczyk (2017), Kovalev and Bogstad 2005)). Thus, even with the noisy data and a limited dynamic range in stock size we have experienced in the past, we quite often see DD significantly manifested.

Two solutions to the problem were suggested at WKREF2. One is to produce DD sub-models for all four parameters and do the normal age-based forecast models including these sub-models to determine biological reference points (BRPs). This has been done for NEA-cod by Kovalev and Bogstad (2005), and by several ICES Benchmark WKs since then. This has also been done by Sparholt *et al*. (2020) for North Sea cod and Northeast Atlantic mackerel, by Danielsson *et al*. (1997) for cod at Iceland, and by Horbowy and Luzeńczyk (2017) for Baltic sprat. However, for all other stocks these DD sub-models belong to the category of "known unknowns". It will be difficult to develop the needed sub-models and they are not likely to be available for most stocks until a decade or so from now.

The other solution suggested is to use Surplus Production Models (SPM) for estimating BRP. These models include, implicitly, all density dependent elements by design. The basic idea is to use the total stock biomass (TB) from ICES routine age-based stock assessment as the biomass index (assumed to be an absolute metric of exploitable stock biomass) and the catch. The annual surplus production (SP) is then the catch plus the change in TB from one year to the next. SP is plotted against TB and an SPM curve fitted. A worked example for mackerel is provided in Annex 2 Working Document 4.

In conclusion, it was proposed that using a Surplus Production Models (SPM) in a combination with the current age-based assessment models may circumvents the problem of known unknowns of density dependence in growth, maturity and natural mortality ----and gives an F_{MSY} estimate that has no (documented) bias. Although using the age-based assessment model provides the stock biomass time series to the SPM model may introduce some biases. It was proposed that the SPM model could provide the long-term optimal F (F_{MSY}) and that value can be applied to the current age-based short-term forecast to give the TAC advice for the coming year. There was no consensus at WKREF2 on whether that approach was appropriate.

Plenary discussion

Following the presentation, the issues listed below were discussed.

It has sometimes been suggested that one should also use an SPM model for the historical assessment of the stock if SPMs are used for BRP calculations to be consistent. However, the current ICES age-based assessments are probably the best reflection of the history of the stock. It has been developed over decades and involves sophisticated tunings with fisheries independent tuning time series from surveys and much more. The historic assessment also includes DD in growth, generally measured every year, and best available maturity data. Sometimes also DD in natural mortality (for cannibalistic species) is included from multispecies models via a simple relationship to stock size or estimates from multispecies models (WGSAM 2021). Age-based historic assessments furthermore provide information about recruitment and selectivity which may vary significantly over time. This can be useful for identifying regime shifts in the ecosystem relevant for the stock in question. To try to do all that in SPMs seems not prudent. Also, for short-term projections, the current age-based approach seems superior to SPMs, because it can take account of cohort sizes and recruitment and DD changes to weight-at-age.

How SPMs reacts when we have a regime shift was also questioned. SPMs ignoring regime shifts is probably not any better than the current age-based approach ignoring this as well. However, SPMs might be easier to deal with in this regard because it is simpler. It is very important to select a SPM configuration which reflects ecosystem situation for the coming say about 5 years accurately. This should be revisited every 5 years or so, to see if the ecosystem is still as it was. Also, changes in exploitation pattern in the fisheries could be important if that changes substantially. A new F_{MSY} values might be developed, but maybe a new SPM is not needed, only the translation from the SPM F-metric (yield in biomass divided by stock size in biomass) to the ICES age-based F-metric (typically mean F over some ages). If a regime shift is identified, one should consider shortening the time series of estimation. The MSE-project has worked on North Sea cod, Baltic sprat and Northeast Atlantic mackerel in this regard and presented this type of sensitivity analysis at various ICES workshops.

Some approaches using SPMs, assume equilibrium in the sense that a given F has been applied for many years and the stock has stabilised at that level of F. In those cases, SP (Surplus Production) can be plotted against F in order to fit an SPM. This is not the approach suggested by the Fmsy-project. Here SPMs are based on SP plotted against the stock biomass (exploitable stock biomass). Such an approach does not require equilibrium. This is because the SP is only dependent on the stock size.

The use of the meta-analysis of Thorson *et al.* (2012) of 141 fish stocks to get the shape of the SPM curve was questioned. The shape varies quite a bit if one uses the parameters from a specific taxonomic group or from the overall average. In the cases for mackerel, North Sea cod and Baltic sprat (WK Documents available on the SharePoint site) these shape parameters were scrutinised carefully, as they were in the Fmsy-project for all 53 data rich stocks. This is something which, for a given stock, should be explored and tested, and for instance, for Baltic sprat, it seems that an ordinary Schaefer curve was more consistent with the data than Thorson *et al.* (2012) taxonomic group shape for Clupeiformes stocks (which were quite asymmetric, with B_{MSY}/K being only 0.2649 compared to the Schaefer value of 0.50).

The two alternative approaches suggested, one using the traditional age-based approach including DD sub-models and the other using the SPM approach, might not be mutually exclusive, but could supplement each other. It was mentioned at WKREF 2 that it is unrealistic to assume you get the best picture of the stock population dynamics by just using one model. It might be better using several structurally different models and then averaging the results afterwards, resulting in an ensemble approach. Tests of the model performance could be accomplished using the approaches in section 5.1.

It seems prudent when using ICES traditional age-based long-term forecast models for obtaining BRPs, to only include DD when it is important for the stock, and this is mainly at higher stock sizes. The experience so far is that including DD in maturity is often only increasing the F_{MSY} estimate by very little (a few percentages) and can be ignored. However, this might be difficult to know without doing the calculations, and when the calculations are done anyway, one might as well include it in the final long-term forecast calculations (i.e., in the operating model used).

5.4 F_{eco} as a means of adapting target fishing mortality to medium-term changes in productivity

F_{eco} is an approach to allow ecosystem information or outputs of ecosystem models to be used to tune the F_{target} to account for medium term ecosystem driven variability in productivity. Assessment models are tuned to as long a time series of data as possible, and there is good evidence that curtailing these time series imposes errors in the assessment. Obviously, the ecosystem rarely remains unchanged over time periods measured in multiple decades. In some cases, the variability can be accounted for directly in the assessment model and potentially used directly in the calculation of the fishing target reference point. However, in many cases, this medium-term variability is not accounted for in the fisheries target reference point, meaning that the fishing pressure is out of step with the current state of the ecosystem.

Ecosystem models are generally not suitable for setting annual quota advice, but they do provide the best ecosystem overview available. F_{eco} is a method to allow for information from the ecosystem models to enter the quota advice without directly transferring values between different models. F_{eco} entails identifying indicators (either physical or synthetic model outputs) which track stock productivity, and then using these indicators to scale up or down the pre-defined single species F_{target}, while not exceeding the pre-defined limit reference points (F_{lim}, B_{lim}). This approach allows for some influence of the ecosystem information, while retaining the advantages of the current single species workflow. This approach is outlined in WKIRISH6 (ICES 2020) and Howell et al. (2021). This process gives a large degree of flexibility in accounting for ecosystem variability, with similar approaches being used to account for predator needs (Chagaris et al., 2021), variable stock productivity (Bentley et al., 2021), and the use of risk assessment to potentially reduce catch if required to remain precautionary (Dorn and Zador, 2020).

Recommendation

WKREF2 recommends that ICES guidelines include the possibility to use an F_{eco} approach to adjust the F based on ecosystem model information (e.g. WKIRISH 6 ICES 2020)). If such an approach is desired in a particular benchmark, then the following criteria should be applied:

- The revised F should not exceed $F_{P.05}$
- The ecosystem model to be used should have been reviewed as a Key Run by WGSAM
- The implementation should be evaluated and reviewed at a benchmark process.

6 Risk Equivalence

WKREF1 was focused on Category 1 stocks, but it is important that they are not considered in isolation. ToR c) calls for a "simplified and harmonised set of guidelines for estimating MSY and precautionary reference points applicable in the advice framework across various ICES stock categories." Ideally the ICES assessment categories should provide equivalent risk across all stocks (Fischer et al., in press). A definition of risk is the probability of a stock being depleted below a limit reference point or not being maintained at a target reference point, and risk equivalence requires that this should be the same, irrespective of the stock assessment method used to provide management advice, or the amount of data and knowledge available (Fulton *et al.*, 2016). This requires standardised metrics for the calculation of risk.

Risk equivalence can help provide robust and accountable management decision-making in the absence of perfect knowledge, and provide an incentive to evaluate the value-of-information and the development of robust feedback control (Roux *et al.*, 2021). For example, when there is large uncertainty around the estimated stock size, fishing rates must be lower than when uncertainty is small to ensure the same risk to the stock in the two cases. There is therefore a positive relationship between information and utilisation, and so the value of information to the fishery is positive (Cooke, 1999).

An implicit assumption of the ICES guidelines is that, within stock assessment categories, the assessment and rule are linked. Geromont and Butterworth (2015) showed that for category 1 stocks, simple catch control rules based upon survey indices (i.e. a category 3 rule) can be developed to achieve virtually equivalent catch and risk performance as for category 1 advice. Whether such approaches could be acceptable to advice recipients and stakeholders remains to be seen given that there is usually a demand for better process and dynamic understanding from science.

Often there is insufficient data in stock assessment data sets to describe and parameterise key processes, and multiple models may explain the data equally well, for example on the relationship between stock recruitment, multi-species effects and density dependence. Therefore, there may be more uncertainty than admitted in a single "best assessment". Model ensembles or MSE can be used to develop test advice rules that are robust to uncertainty and de-risking even data limited situations. For example, Fischer *et al.* (2020) developed advice based on indicators conditioned on life history traits and ecological knowledge that were able to meet precautionary and MSY objectives (Fischer *et al.*, 2021). However, a generic rule may not work for all stocks. Therefore, the option should be available to develop case-specific advice if appropriate sources of uncertainty have been included in model ensembles or the conditioning of the Operating Models.

7 Possible revised reference point framework

To make a reference point system operational requires general guidelines on how to specify the reference points in practice. A guiding principle for developing these guidelines is that reference points, such as the F_{MSY} proxies, should be stock-specific by considering a stock's biology, productivity and ecology, and the nature of the fisheries, following international best practice. The reference point system should be based on understandable and transparent rules and should provide a clear and unambiguous interpretation of the stock status. A possible candidate reference point system, which builds on the key recommendations by ICES WKREF1 (2021), was discussed and define as follows (Figure 7.1).

There was no consensus at WKREF2 on this suggested below; however, it was agreed that it should be presented here for further discussion. The key feature of the suggested approach is that the stock status evaluation is treated independent of the Advice Rule (AR). The main feature of the system is that the biomass trigger is not linked to a stock status evaluation, it is linked to the expected biomass when fishing at the target fishing mortality, in contrast to the current ICES approach. The F_{MSY} would also become an upper limit of fishing mortality.

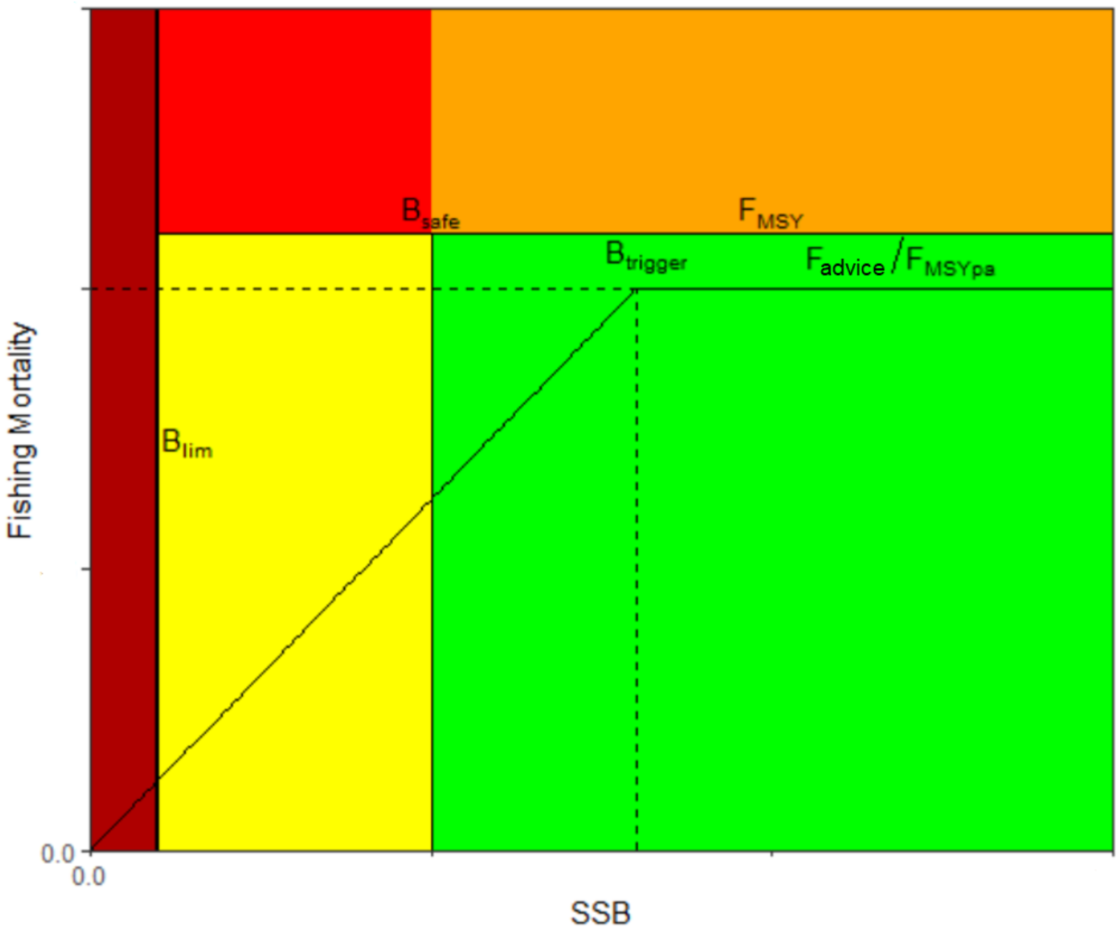


Figure 7.1. Proposed ICES Reference points system with integrated Harvest Control Rule. (source `FLRef` function plot-WKREF(); https://github.com/henning-winker/FLRef)

- $F_{P.05}$ is a probabilistically derived quantity the fishing mortality that is associated with a 5% risk that SSB falls below B_{lim} as derived using stochastic long-term projections.
- **F_{MMY}:** The maximum medium yield F_{MMY} denotes the fishing mortality that corresponds to the peak of the median landings yield curve derived from stochastic forward projections as is typically derived from the EqSim software (i.e. "FMSYmedianL"). Within the ICES advice framework, the quantity F_{MMY} is typically referred to as F_{MSY}. However, for F_{MMY} to directly translate into F_{MSY} as reported on the advice sheet, F_{MMY} first requires meeting the condition that $F_{MMY} \leq F_{P.05}$ in accordance with precautionary principle. For the purposes of this report, a clearer definition was therefore needed to separate the initial estimate of F_{MSY}, here F_{MMY}, from the final advice for F_{MSY}.
- **F_{MSY}:** Remains $F_{MSY} = min(F_{MMY}, F_{P05})$, but becomes the overfishing threshold (It should be noted that this definition is potentially not consistent with current EU management nor with the new MSFD guidelines which refer to 6-year averages of F/F_{MSY} as defining the limit, not annual values).
- **B_{MSY}:** is the median biomass corresponding to fishing at F_{MSY} (noting that B_{MSY} in this definition has also changed as it may in fact be median B at F_{p05}), which can be computed in deterministic or stochastic projections (EqSim/Assessment Model, etc.).
- **F_{advice}** or **F_{MSYpa}** or **F_{target}** is the fishing mortality advice of the Harvest Control Rule, which could be set at or lower than F_{MSY}, e.g. F_{95} that corresponds to 95% MSY (i.e. MMY) or a biological proxy for F_{MSY} ($F_{SPR\%}$, $F_{0.1}$, $F_{B\%}$) that must not exceed F_{MSY}.
- **B_{lim}** is the biomass limit below which there is evidence of impaired recruitment or if no such evidence exists set according to new ICES guidelines addressing the issue outlined in Section 4.2.
- **B_{safe}** is the lower of range of stochastic fluctuations around B_{MSY}, the default could be tentatively set to $0.5B_{MSY}$. This could be adjusted for stocks with more or less natural variability. B_{safe} would be used as a stock status classifier.
- **B_{trg}** is the median biomass corresponding to fishing at F_{advice}, which can be computed in deterministic or stochastic projections (EqSim/Assessment Model, etc.). Noting that if agreed management plan opts for a lower F_{advice} in which case this B_{trg} would exceed B_{MSY}.
- **$B_{trigger}$** is the operationalised biomass trigger point for tuning of the harvest control rule (not a reference point). If biomass falls below $B_{trigger}$, F_{advice} is decreased linearly toward minimum biomass (default is zero) at which the fishery may be closed. The $B_{trigger}$ is a generalization of the MSY $B_{trigger}$. Two options remain to considered for specifying the $B_{trigger}$ value:
    - (1) as a fraction of B_{trg} (here: 0.6 - 1.0 of B_{MSY})
    - (2) as multiplier of B_{lim} (here: 2 x B_{lim})

WKREF2 did not discuss what to do in situations where SSB <B_{lim} or alternative forms of HCR (e.g. steeper linear declines to zero at B_{lim}).

Conclusions

Biological and MSY target and limit reference points are an essential part of normative fisheries management. The definition of reference points is both complex and variable worldwide, and best practice has yet to be achieved. In ICES, guidelines for reference points are a key element of achieving quality assurance, consistency and transparency across stocks. However, in general, setting reference points according to the guidelines remains extremely challenging. Despite the challenge of discussing such a complex issue in relatively short and intense virtual meetings, WKREF 1 and WKREF 2 have made significant progress to advance the ICES approach to setting reference points. Building community understanding and consensus around simplified and harmonised guidelines has yet to be achieved. A further workshop WKREF 3 will be required to achieve that aim.

8 Recommendations

- WKREF2 recommended that further work is carried out on the simulation framework before the conclusions could be adopted by ICES.
- WKREF2 recommended that stock assessment models are validated using prediction skill to ensure that they provide robust and credible advice
- WKREF2 recommended that the evidence for Allee effects in ICES stocks should be explored further using the latest methods that have been developed.
- WKREF2 recommended that the guidelines should allow for benchmarks to set B_{lim} as a fraction of B_0, particularly within integrated models and ensembles, where it is internally consistent to do so.
- WKREF2 recommended that ICES guidelines include the possibility to use an F_{eco} approach to adjust F based on ecosystem model information (e.g., WKIRISH 6 ICES 2020)). If such an approach is desired in a particular benchmark, then the following criteria should be applied:
    - The revised F should not exceed $F_{P.05}$
    - The ecosystem model to be used should have been reviewed as a Key Run by WGSAM
    - The implementation should be evaluated and reviewed at a benchmark process.
- WKREF2 recommended that **a further workshop WKREF 3 be established to provide the evidence base and build** community understanding and consensus around simplified and harmonised guidelines**.**

References

Andersen, K. H. 2017. Fish and Fisheries: The Size – and Trait –Based Approach. Technical University of Denmark, National Institute of Aquatic Resources. 284 pp.

Castañeda RA, Burliuk CMM, Casselman JM, Cooke SJ, Dunmall KM, Forbes LS, Hasler CT, Howland KL, Hutchings JA, Klein GM, Nguyen VM, Price MHH, Reid AJ, Reist JD, Reynolds JD, Van Nynatten A, Mandrak NE. 2020 A brief history of fisheries in Canada. Fisheries 45, 303-318.

Collie, J. S., and Gislason, H. 2001. Biological reference points for fish stocks in a multispecies context. Canadian Journal of Fisheries and Aquatic Sciences, 58: 2167–2176.

Danielsson, A., G. Stefansson, F.M. Baldursson, K. Thorarinsson. 1997. Utilization of the Icelandic Cod Stock in a Multispecies Context. Marine Resource Economics, 12, 329–344.

Fischer, S. H., De Oliveira, J. A. A., and Kell, L. T. (2020). Linking the performance of a data-limited empirical catch rule to life-history traits. ICES Journal of Marine Science 77, 1914–1926

Fischer, S.H., De Oliveira, J.A.A., Mumford, J.D. and Kell, L.T. (2021). Application of explicit precautionary principles in data-limited fisheries management. ICES Journal of Marine Science, 78(8), pp.2931-2942.

Fischer, S.H., De Oliveira, J.A.A., Mumford, J.D. and Kell, L.T. (in press). Risk equivalence in data-limited and data-rich fisheries management: an example based on the ICES advice framework. Fish and Fisheries.

Froese, R., Garilao, C., Winker, H., Coro, G., Demirel, N., Tsikliras, A., Dimarchopoulou, D., et al. 2016. Exploitation and status of European stocks. Updated version. World Wide Web electronic publication, http://oceanrep.geomar.de/34476/.

Fulton, E. A., Punt, A. E., Dichmont, C. M., Gorton, R., Sporcic, M., Dowling, N., et al. (2016). Developing risk equivalent data-rich and data-limited harvest strategies. Fisheries Research 183, 402 574–587. doi:10.1016/j.fishres.2016.07.004

Geromont, H.F. and Butterworth, D.S., 2015. Generic management procedures for data-poor fisheries: forecasting with few data. ICES Journal of Marine Science, 72(1), pp.251-261.

Gislason, H. 1999. Single and multispecies reference points for Baltic fish stocks. ICES Journal of Marine Science, 56: 571–583.

Hilborn R, Hively DJ, Jensen OP, Branch TA. 2014 The dynamics of fish populations at low abundance and prospects for rebuilding and recovery. ICES Journal of Marine Science 71, 2141–2151.

Hilborn R, Hively DJ, Loke NB, et al. 2021. Global status of groundfish stocks. Fish Fish. 22:911–928. https://doi.org/10.1111/faf.12560

Horbowy, J., L. A. Luzeńczyk. 2017. Effects of multispecies and density-dependent factors on MSY reference points: example of the Baltic Sea sprat. Can. J. Fish. Aquat. Sci. 74, 864–870. dx.doi.org/10.1139/cjfas-2016-0220.

Hutchings JA. 2014 Renaissance of a caveat: Allee effects in marine fishes. ICES Journal of Marine Science 71, 2152-2157.

Hutchings JA. 2015 Thresholds for impaired species recovery. Proceedings of the Royal Society B 282, 20150654.

ICES. 2008. Report of the Working Group on Multispecies Assessment Methods (WGSAM), 6–10 October 2008, ICES Headquarters, Copenhagen. ICES Document CM 2008/RMC: 06. 113 pp.

ICES. 2012. Report of the ICES Advisory Committee 2012. ICES Advice, 2012. Book 9. 389 pp.

ICES. 2013. Report of the ICES Advisory Committee 2013. ICES Advice, 2013. Books 1–11, 2095 pp.

ICES. 2021. Inter-Benchmark Protocol on North Sea Herring (IBPNSHerring 2021). ICES Scientific Reports. 3:98. 168 pp. https://doi.org/10.17895/ices.pub.8398

ICES. 2022. Cod (Gadus morhua) in Division 5.a (Iceland grounds). In Report of the ICES Advisory Committee, 2022. ICES Advice 2022, cod.27.5a, https://doi.org/10.17895/ices.advice.19447886

ICES. 2022 Workshop on ICES reference points. ICES Scientific Reports 4:2. International Council for the Exploration of the Sea. doi.org/10.17895/ices.pub.9822.

Keith DM, Hutchings JA. 2012 Population dynamics of marine fishes at low abundance. Canadian Journal of Fisheries and Aquatic Sciences 69, 1150-1163.

Kovalev, Y. A., B. Bogstad. 2005. Evaluation of maximum long-term yield for Northeast Arctic cod. In Ecosystem Dynamics and Optimal Long-Term Harvest in the Barents Sea Fisheries. Proceedings of the 11th Russian–Norwegian Symposium, Murmansk, Russia, 15–17 August 2005, pp. 138– 157. IMR-PINRO Joint Report Series 2/2005.

Kuparinen A, Hutchings JA. 2014 Increased natural mortality at low abundance can generate a demographic Allee effect. Royal Society Open Science 1, 140075.

Licandeo R, Duplisea DE, Senay C, et al (2020) Management strategies for spasmodic stocks: A canadian atlantic redfish fishery case study. Can J Fish Aquat Sci. https://doi.org/10.1139/cjfas-2019-0210

Liermann M, Hilborn R. 1997 Depensation in fish stocks: a hierarchical Bayesian meta-analysis. Canadian Journal of Fisheries and Aquatic Sciences 54, 1976-1984.

Liermann M, Hilborn R. 2001 Depensation: evidence, models and implications. Fish and Fisheries 2, 33-58.

Lorenzen, K. 2016. Toward a new paradigm for growth modeling in fisheries stock assessments: Embracing plasticity and its consequences. Fisheries Research 180 4–22. https://doi.org/10.1016/j.fishres.2016.01.006.

Minto C, Myers RA, Blanchard W. 2008 Survival variability and population density in fish populations. Nature 452, 344-347.

Morgan, M. J., P. A. Shelton, Fernando González-Costas, Diana González-Troncoso. 2016. Compensation potential in six depleted groundfish stocks from the Northwest Atlantic. Can. J. Fish. Aquat. Sci. Vol. 73, 2016.

Myers RA, Barrowman NJ, Hutchings JA, Rosenberg AA. 1995 Population Dynamics of Exploited Fish Stocks at Low Population Levels. Science 269, 1106–1108.

Perälä, T. A., Swain, D. P., & Kuparinen, A. 2017. Examining nonstationarity in the recruitment dynamics of fishes using Bayesian change point analysis. Canadian Journal of Fisheries and Aquatic Sciences, 74(5), 751-765.

Perälä T, Kuparinen A. 2017 Detection of Allee effects in marine fishes: analytical biases generated by data availability and model selection. Proceedings of the Royal Society B 284, 20171284.

Perälä T, Hutchings JA, Kuparinen A. 2022 Allee effects and the Allee-effect zone in Northwest Atlantic cod. Biology Letters. 182021043920210439 http://doi.org/10.1098/rsbl.2021.0439

Pope, J. G., Rice, J. C., Daan, N., Jennings, S., and Gislason, H. 2006. Modelling an exploited marine fish community with 15 parameters—results from a simple size-based model. ICES Journal of Marine Science, 63: 1029–1044.

Rindorf, A., van Deurs, M., Howell, D., Andonegi, E., Berger, A., Bogstad, B., Cadigan, N., Elvarsson, B. Þ., Hintzen, N., Savina Roland, M., Taylor, M., Trijoulet, V., van Kooten, T., Zhang, F., & Collie, J. (2022). Strength and consistency of density dependence in marine fish productivity. Fish and Fisheries, 00, 1– 17. https://doi.org/10.1111/faf.12650

Roux, M.-J., Duplisea, D., Hunter, K., and Rice, J. (2021). Consistent risk management in a changing world: risk equivalence in fisheries and other human activities affecting marine resources and ecosystems

Saltelli, A. et al. 2020. Five ways to ensure that models serve society: a manifesto. Nature 582, 482-484 (2020), doi: https://doi.org/10.1038/d41586-020-01812-9

Sparholt, H., R. M. Cook. 2010. Sustainable Exploitation of Temperate Fish Stocks. Biol. Lett. 6(1):124-127 (2010).

Sparholt, H., B. Bogstad, V. Christensen, J. Collie, R. van Gemert, R. Hilborn, J. Horbowy, D. Howell, M. C. Melnychuk, S. A. Pedersen, C. R. Sparrevohn, G. Stefansson and P. Steingrund. 2020. Estimating Fmsy from an ensemble of data sources to account for density-dependence in Northeast Atlantic fish stocks. ICES Journal of Marine Science. ICES Journal of Marine Science, Volume 78, Issue 1, January-February 2021, Pages 55–69, https://doi.org/10.1093/icesjms/fsaa175. First published online in October 2020.

Swain DP, Benoît HP. 2015 Extreme increases in natural mortality prevent recovery of collapsed fish populations in a Northwest Atlantic ecosystem. Marine Ecology Progress Series 519, 165-182.

Thorson, J. T., J. M. Cope, T. A. Branch, O. P. Jensen. 2012. Spawning biomass reference points for exploited marine fishes, incorporating taxonomic and body size information. Canadian Journal of Fisheries and Aquatic Sciences, 69: 1556-1568.

Zimmermann, F., Ricard. D., Heino, M. 2018. Density regulation in Northeast Atlantic fish populations: Density dependence is stronger in recruitment than in somatic growth. J. Anim. Ecol. May;87(3):672-681. doi: 10.1111/1365-2656.12800

Annex 1: List of participants

Member	Dept/Institute	Email
Alessandro Orio	Swedish University of Agricultural Sciences	alessandro.orio@slu.se
Alexander Kempf	Thünen-Institute of Sea Fisheries	alexander.kempf@thuenen.de
Andrea Pierucci	COISPA Tecnologia & Ricerca	pierucci@coispa.eu
Ane Iriondo	AZTI Sukarrieta	airiondo@azti.es
Anna Rindorf	DTU Aqua, National Institute of Aquatic Resources	ar@aqua.dtu.dk
Arnaud Peyronnet	European Commission Directorate-General for Maritime Affairs and Fisheries	Arnaud.PEYRONNET@ec.europa.eu
Ashley Wilson	Pew Charitable Trusts London Office	awilson@pewtrusts.org
Cecilie Kvamme	Institute of Marine Research	cecilie.kvamme@hi.no
Christoffer Moesgaard Albertsen	DTU Aqua, National Institute of Aquatic Resources	cmoe@aqua.dtu.dk
Christoph Konrad	Joint Research Centre	christoph.konrad@ec.europa.eu
Christoph Priebe	EC DGMARE	Christoph.Priebe@ext.ec.europa.eu
Claus Reedtz Sparrevohn	Danish Pelagic Producers' Organisation	crs@pelagisk.dk
Cóilín Minto	Galway-Mayo Institute of Technology	Coilin.Minto@gmit.ie
Colm Lordan (chair)	International Council for the Exploration of the Sea	colm.lordan@ices.dk
Daniel Howell	Institute of Marine Research	daniel.howell@hi.no
David Gilljam	SLU Department of Aquatic Resources Institute of Marine Research	david.gilljam@slu.se
David Miller	International Council for the Exploration of the Sea	david.miller@ices.dk
Dorleta Garcia	AZTI-Tecnalia	Dorleta.Garcia@ices.dk
Enrico Armelloni	University of Bologna	enrico.armelloni@irbim.cnr.it
Esther Abad	Centro Oceanográfico de Vigo	esther.abad@ieo.es

Member	Dept/Institute	Email
Francesco Masnadi	University of Bologna	francesco.masnadi@studio.unibo.it
Giuseppe Scarcella	Institute of Marine Sciences	giuseppe.scarcella@cnr.it
Gjert Endre Dingsør	Norwegian Fishermen's Association	gjert@fiskebat.no
Gonçalo Ferreira de Carvalho	Marine Conservation and Cooperation Association	gcarvalho@sciaena.org
Graham Norman	EC DGMARE	Norman.Graham@ec.europa.eu
Henning Winker	Joint Research Centre	henning.WINKER@ec.europa.eu
Henrik Sparholt	University of Copenhagen	henrik.sparholt@gmail.com
Inês Farias	Portuguese Institute for the Sea and the Atmosphere	ifarias@ipma.pt
Jan Horbowy	National Marine Fisheries Research Institute	horbowy@mir.gdynia.pl
Jean-Baptiste Lecomte	Centre Atlantique	jean.baptiste.lecomte@ifremer.fr
Jeffrey A. Hutchings	Department of Biology	Jeff.Hutchings@Dal.Ca
Johan Lövgren	SLU Department of Aquatic Resources-SLU Aqua	johan.lovgren@slu.se
Johanna Fall	Institute of Marine Research	johanna.fall@hi.no
John Simmonds	Unaffiliated	e.j.simmonds1@gmail.com
Jonathan Shrives	European Commission Directorate-General for Maritime Affairs and Fisheries	jonathan.shrives@ec.europa.eu
José De Oliveira	Cefas Lowestoft Laboratory	jose.deoliveira@cefas.co.uk
Kristiina Hommik	University of Tartu	kristiina.hommik@ut.ee
Laurence Kell	SEA++	laurie@seaplusplus.co.uk
Lies Vansteenbrugge	The Flanders Research Institute for Agriculture, Fisheries and Food	lies.vansteenbrugge@ilvo.vlaanderen.be
Lisa Borges	FishFix	info@fishfix.eu
Luke Batts	Galway-Mayo Institute of Technology, Ireland	luke.batts@research.gmit.ie
Marc Taylor	Thuenen Institute	marc.taylor@thuenen.de

Member	Dept/Institute	Email
Mark Dickey-Collas	ACOM Chair	Mark.dickey-collas@ices.dk
Martin Pastoors	Pelagic Freezer-Trawler Association	mpastoors@pelagicfish.eu
Martina Scanu	Institute for Marine Biological Resources and Biotechnologies	martina.scanu@irbim.cnr.it
Massimiliano Cardinale	SLU Department of Aquatic Resources-SLU Aqua	massimiliano.cardinale@slu.se
Michaël Gras	Joint Research Centre	michael.gras@ec.europa.eu
Mikael van Deurs	DTU Aqua, National Institute of Aquatic Resources	mvd@aqua.dtu.dk
Mikaela Bergenius Nord	SLU Department of Aquatic Resources-SLU Aqua	mikaela.bergenius.nord@slu.se
Nicolas Goñi	Natural Resources Institute Finland	nicolas.goni@luke.fi
Paula Silvar Viladomiu	Galway-Mayo Institute of Technology, Ireland	paula.silvarviladomiu@research.gmit.ie
Paz Sampedro	Centro Oceanográfico de A Coruña	paz.sampedro@ieo.es
Pekka Jounela	Natural Resources Institute Finland	pekka.jounela@luke.fi
Ray Hilborn	School of Aquatic and Fishery Sciences	rayh@u.washington.edu
Rishi Sharma (chair)	The Food and Agriculture Organization of the United Nations	Rishi.Sharma@fao.org
Sarah Millar	International Council for the Exploration of the Sea	sarah-louise.millar@ices.dk
Sofie Nimmegeers	The Flanders Research Institute for Agriculture, Fisheries and Food	sofie.nimmegeers@ilvo.vlaanderen.be
Tanja Miethe	Marine Laboratory	t.miethe@marlab.ac.uk
Tommi Perälä	University of Jyväskylä	tommi.a.perala@jyu.fi
Valerio Bartolino	SLU Department of Aquatic Resources-SLU Aqua	valerio.bartolino@slu.se

Annex 2: Working documents

WD1 Consistency and Robustness testing of candidate reference point systems for North East Atlantic stocks

Henning Winker1*, Massimiliano Cardinale2 , Iago Mosqueira3 , Laurence Kell4 , Christoph Konrad1 , Michael Gras1 , Rishi Sharma5 & Colm Lordan6

Working Document

DRAFT:

Consistency and Robustness testing of candidate reference point systems for North East Atlantic stocks

Henning Winker[1], Massimiliano Cardinale[2], Iago Mosqueira[3], Laurence Kell[4], Christoph Konrad[1], Michael Gras[1], Rishi Sharma[5] & Colm Lordan[6]*

[1]*Joint Research Centre (JRC), European Commission, TP 051, Via Enrico Fermi 2749, 21027 Ispra (VA), Italy*

[2]*Swedish University of Agricultural Sciences, Department of Aquatic Resources, Institute of Marine Research, Lysekil, Sweden*

[3]*Wageningen Marine Research, Haringkade 1, 1976CP IJmuiden, the Netherlands*

[1]*Centre for Environmental Policy, Imperial College London, Weeks Building, 16-18 Princes Gardens, London SW7 1NE, UK*

[5]*Food and Agricultural Organization, Fishery and Aquaculture Policy and Resources Division, Rome, Lazio 00153, Italy*

[6]*Marine Institute, Oranmore, Galway,Ireland*

Abstract

Recently, the ICES Workshop on ICES reference points (WKREF1, 2021) was tasked to provide a thorough review of the ICES reference points system as a basis to re-evaluate the process for estimating, updating and communicating reference points in the context of the ICES advice. The key recommendations of WKREF1 were to: i) revise and simplify how B_{lim} is derived, including the possibility to determine B_{lim} as a fraction of B_0 based on biological principles and international best practice; ii) $F_{P.05}$ should be calculated without $B_{trigger}$; iii) to use biological proxies for deriving F_{MSY}, and the F_{MSY} proxy must not exceed $F_{P.05}$ consistent with ICES Precautionary Approach (PA) ; iv) to report a biomass target (B_{trg}) that corresponds to the F_{MSY} proxy; and v) to set $B_{trigger}$ as either a fraction of B_{trg} or multiplier of B_{lim}. In this paper, we conduct a large-scale simulation testing experiment with feedback control for 64 ICES Category 1 stocks, with the aim to evaluate the consistency and robustness of candidate reference point systems. In accordance with the objectives of ICES advice framework, the evaluation criteria for testing consistency are based on the following objects: (1) to not exceed a 5% probability of SSB falling below B_{lim} , (2) to achieve high long-term yields that correspond to at least 95% of the median yield at constant F_{MSY} (MSY), (3) to attain a high probability that SSB is above the FAO threshold of 80% of the B_{trg} proxy for B_{MSY}. By considering stock-specific productivity and taxonomic grouping, we then put forward the best performing candidate reference point systems for further robustness testing under alternative misspecifications of the stock recruitment relationship . Based

on our simulation results, we present straightforward and transparent guidelines for setting optimal reference points depending on the stock's productivity characteristics. We align this new reference point system with a status classification system that is intended to facilitate clear and unambiguous interpretation of the stock status.

Keywords: North East Atlantic stocks; Reference points; simulation testing; harvest control rule; shortcut MSE

1. Introduction

Central to fisheries advice worldwide are reference points, which are used to classify and communicate current resource status relative to sustainability limits and to provide targets for determining future fishing opportunities, e.g. to set the total allowable catch (TAC) in quota managed fisheries. Stock assessment models are generally considered the basis for scientific advice. In practice, however, the process starts with the processing of fishery dependent and independent observations used by the assessment model, which are typically associated with large and systematic sampling errors (Carruthers et al. 2017). The assessment model itself relies on many assumptions about model structure, in the form of the underlying deterministic relationships (e.g. between stock and recruitment, SR) and population parameters (e.g. natural mortality, M). These contribute to structural and estimation uncertainties associated with stock assessment results (Patterson et al., 2001), where uncertainty is the difference between the model and reality. Accounting for these uncertainties is one of the key challenges for operationalising reference point systems able to provide consistent and robust scientific advice on fishing opportunities (Ralston et al. 2011). However, despite common commitments to maintain or restore stocks at levels capable of producing maximum sustainable yield (MSY) and the Precautionary Approach (PA) to fisheries (UN 1995; FAO 1995), international advice standards vary widely in how this challenge is addressed in particular regarding specifying and estimating the corresponding target- (TRPs) and limit reference points (LRPs), as well as setting the operational trigger points used in harvest control rules.

A main objective of reference points is to prevent overfishing, e.g. growth, recruitment, economic and target overfishing. Growth and recruitment overfishing are generally associated with limit reference points, while economic overfishing may be expressed in terms of either targets or limits. The difference between targets and limits is that indicators may fluctuate around targets, but in general limits should not be crossed. Target overfishing occurs when a target is overshot, although variations around a target are not necessarily considered of serious concern unless a consistent bias becomes apparent. In contrast, even a single violation of the LRP may indicate the need for immediate action in order to be consistent with the PA. On the other hand, triggers are intended to implement action before limits are reached.

In age-structured assessments, MSY based reference points can be either estimated in the model, i.e. when the SR is fitted internally in the assessment model, or derived post-hoc from the model results, using yield and spawner per recruit assumption combined with a SR relationship. These reference points typically assume equilibrium, or an alternative approach is to run long-term stochastic projections. Benefits of the latter approach are that reference points can account for structural uncertainties and estimation errors (e.g. required for ensembles). A problem, however, is that as reference points estimation procedures become more complicated and computationally demanding, they become less transparent and difficult to verify and validate; where verification is the provision of objective evidence that a given procedure meets the specified requirements, and validation is ensuring that management objectives are actually met. This is complicated by the fact that the quantities used to compute reference points are model-based estimated latent quantities, such as numbers-at-age, spawning stock biomass (SSB) and fishing selectivity, which can therefore not be validated by observations (Kell et al. 2021). Thus, verification and validation of reference point systems need to be based on simulation-testing.

Simulation-testing allows verifying consistency of a reference point system in meeting the quantifiable management objectives (e.g. thresholds of TRPs and LRPs) and validating the system's robustness of achieving the underlying goals (e.g. biomass levels at MSY). The consistency of a reference point system relies on the setting TRPs, LRPs and trigger points so that target thresholds are exceeded and the limit thresholds are not breached. By contrast, a reference point system would be internally inconsistent if, for example, the rules for setting the target fishing mortality (F_{trg}) would fail systematically to exceed the corresponding target biomass threshold. Evaluating consistency does not need knowledge of the "true" quantities and can therefore be simulation-tested using "self-tests". The term self-test is used because the assumptions for simulating the stock dynamics are the same as the assumptions for computing biological reference point proxies. Thus, the reference point estimator is correctly specified with respect to the operating model (OM) simulator (Deroba *et al.*, 2015). In contrast to consistency, the term robustness refers in statistics to a model that provides correct inference despite its assumptions being violated; whereas robustness in engineering means that a system functions correctly in presence of uncertainty (Kell *et al.*, 2016). In the context of fisheries advice both meanings are interrelated and highly relevant. Evaluating the robustness of a reference point system therefore requires testing if it can also produce desired outcomes in situations where the reality (OM) differs in assumptions from the reference point estimator (Deroba *et al.*, 2015). Using simulations for robustness testing provides an additional scope beyond a self-test because it can be used to validate that if by meeting management objectives, the desired yet latent state of the stock (e.g. biomass at or above the "true" B_{MSY}) is achieved with high probability despite imperfect knowledge of the true population dynamics.

In the International Council for the Exploration of the Sea (ICES), the PA to fishing (UN 1995; FAO 1995) was introduced first in 1998 without consideration of MSY, and the ICES MSY approach was subsequently integrated into the PA framework in 2009 (ICES, 2012). The ICES reference point system has since evolved and undergone several revisions (Lassen *et al.*, 2014; Silvar-Viladomiu *et al.*, 2021). Part of this evolution is driven by the scientific advances in accounting for risk and uncertainty, but also by policy requirements to implement the ecosystem-based approach to fisheries management, among which multi-species interactions, impacts on bycatch species, adaptation to environmental change and socio-economic considerations, are important drivers. Fisheries advice is therefore becoming increasingly more sophisticated and also more complex. However, sequentially adding more elements, rules and exceptions can also result in ambiguities, inconsistencies and conflicts in achieving multiple objectives. Various ambiguities and potential inconsistencies related to reference points have recently been identified by ICES workshops WKREBUILD (ICES, 2021a), WKGMSE3 (ICES, 2020) and WKRPCHANGE (ICES, 2021b) and concerns have been raised that current reference point estimation procedures are complex and difficult to communicate both internally, among the scientific community, and externally to stakeholders and clients.

For age-structured data rich Category 1 assessments in ICES, F_{MSY} is mostly derived through stochastic forward simulations that are externally implemented in the EQSIM software (Simmonds *et al.*, 2010). The fishing mortality (F) at MSY (F_{MSY}) is in the first instance determined as the F that achieves maximum median long-term yield (F_{MMY}). These projections are commonly run with an HCR, in which the ICES MSY $B_{trigger}$ point instantly reduces fishing mortality linearly if biomass falls below it. Therefore, F_{MMY} in conjunction with MSY $B_{trigger}$ can lead to higher F_{MSY} estimates in comparison to values from projections run at constant fishing mortality (WKREF1, 2021). Both lower and upper ranges for F_{MSY} are provided, but these are bound on the condition to not reduce the long-term yield corresponding to F_{MMY} by more than 5%. However, to be consistent with ICES PA the final estimate of F_{MSY} must not exceed the fishing mortality $F_{P.05,}$ where $F_{P.05}$ is associated with a 5% probability for biomass to fall below B_{lim} (i.e. F_{MSY} is the minimum of $F_{P0.5}$ and F_{MMY}). A, perhaps unique, feature is that ICES MSY approach does not have a formal biomass TRP, with the B_{MSY} estimate corresponding to F_{MSY} being neither used nor reported. Strictly speaking there is one exception, however, in that MSY $B_{trigger}$ can be specified as the lower 5th percentile of the B_{MSY} estimate. This has the seemingly risk-prone property that the higher the uncertainty, the lower biomass has to fall to reduce fishing. In practice, however, this MSY $B_{trigger}$ rule is used rarely, and instead the precautionary biomass (B_{pa}) is normally set equal to MSY $B_{trigger}$, which is approximated by a multiplier of B_{lim} (typically ~ 1.4). Without a biomass TRP, the MSY $B_{trigger}$ not only serves as an operationalized trigger in the ICES Advice Rule, but also as a threshold to classify the stock status to be within 'safe biological' limits if biomass is above it. Without a biomass TRP, the ICES MSY approach is strongly "bottom-up" dependent on B_{lim}. For age-structured models, B_{lim} is a derived deterministic value of absolute spawning stock biomass (SSB) that is independent from any other biomass reference point (e.g. B_{MSY} or B_0). In ICES, there are currently 6 typologies of SR data patterns for determining B_{lim}. Of these, one approach is based on fitting a segmented regression to the SR data to quantify its breakpoint as B_{lim}, but this was only used for 14% of 77 stocks analysed (WKREF1, 2021). Of the other five rule-based approaches, setting B_{lim} to the lowest observed SSB (B_{loss}) was the most common (41%). A meaningful comparison of B_{lim} ranges across stocks or advisory bodies is challenging because the common reference values of B_{MSY} and B_0 are not reported. A recent analysis on 69 ICES stocks, which used a default segmented regression with B_{lim} benchmark estimates as its break-point, indicated a wide variation of B_{lim} value relative to B_0, as estimated by the EQSIM procedure, ranging from 1.3 to 38% of B_0 with a median of just under 10% (WKREF1 2021).

Recently, the ICES Workshop on ICES reference points (WKREF1, 2021) was tasked to provide a thorough review of the ICES reference points system as a basis to re-evaluate the process for estimating, updating and communicating reference points in the context of the ICES advice. The key recommendations of WKREF1 were to: i) revise and simplify how B_{lim} is derived, including the possibility to determine B_{lim} as a fraction of B_0 based on biological principles and international best practice; ii) $F_{P.05}$ should be calculated without $B_{trigger}$; iii) to use biological proxies for deriving F_{MSY}, and the F_{MSY} proxy must not exceed $F_{P.05}$ consistent with ICES PA; iv) to report a biomass target (B_{trg}) that corresponds to the F_{MSY} proxy; and v) to set $B_{trigger}$ as either a fraction of B_{trg} or multiplier of B_{lim}.

In this paper, we first present an overview of international reference point systems. This is to provide the conceptual basis for conducting a large-scale simulation testing experiment with feedback control for 64 ICES Category 1 stocks, with the aim to evaluate the consistency and robustness of candidate reference point systems in accordance with the recommendations made by WKREF1 (ICES, 2021). The evaluation criteria for testing consistency are based on the following three main objectives: (1) to not exceed a 5% probability of SSB falling below B_{lim} (ICES, 2021c), (2) to achieve high long-term yields that correspond to at least 95%

of the median yield at constant F_{MSY} (MSY) (ICES, 2021c), (3) to attain a high probability that SSB is above the FAO threshold of 80% of the B_{trg} proxy for B_{MSY} (DFO, 2009; FAO, 2020; Sharma *et al.*, 2021). By considering stock-specific productivity and taxonomic grouping, we then select the best performing candidate reference point systems for robustness testing. Robustness testing based on simulations enables comparison against 'true' quantities of B_{lim}, MSY and B_{MSY} as derived from the OM, which can differ to various extent from the reference estimator. For example, it allows us to evaluate if a median SSB close to the "true" B_{MSY} is indeed achieved in cases where SSB is above the lower threshold set for B_{trg}. Based on this simulation testing framework, we provide best practice guidelines on the estimation of reference points that are simplified, yet robust, data driven and consistent with the criteria of ICES advice framework.

2. Overview of international reference point systems

Direct estimates for fishing mortality (F_{MSY}) and biomass (B_{MSY}) that correspond to the maximum surplus production MSY are the default TRPs in tuna Regional Management Organizations (RMFOs), such as the International Commission for the Conservation of Atlantic Tunas (ICCAT) and the Indian Ocean Tuna Commission (IOTC). However, there can be exceptions of using a ratio relative to the unfished biomass (B_0) for the biomass and the corresponding TRP if there is high uncertainty about the stock recruitment relationship (e.g. SKJ). For several tuna and billfish stocks, Management Strategy Evaluations (MSEs) and harvest control rules are under development by ICCAT and IOTC but implementation of interim Management Procedures (NAtl. Albio, ICCAT Laurie) or harvest control rules (IO SKJ) are limited. In the absence of a harvest control rule, catch advice is typically based on the Kobe-2-Strategy Matrix, which depicts the probabilities of biomass exceeding B_{MSY} and F remaining below F_{MSY} as derived from medium to long-term projections (7-15 years) over a range of constant catch scenarios. In tune RFMOs, the total allowable catch (TAC) advice has generally to fulfil the minimum requirement that $B > B_{MSY}$ and $F < F_{MSY}$ with 50% probability at the end of the projection horizon. Like harvest control rules, formal implementation LRPs are pending for most stocks, but interim LRPs are increasingly put forward (Refs, Rishi). For example, in the IOTC interim LRPs were specified as a biomass limit at $B_{lim} = 0.4B_{MSY}$ and F_{lim} =1.4-1.5 F_{MSY} for tunas and swordfish, pending further updates as part of the ongoing MSE development process. By contrast, MSE has already been successfully implemented since 2012 by the Commission for the Conservation of Southern Bluefin Tuna (CCSBT) to provide rigid TAC advice for Southern bluefin tuna (Hillary *et al.*, 2015). Here, the management procedure specifies the interim rebuilding objective to achieve spawning stock biomass (SSB) levels above a LRP of 20% B_0 with a least 70% probability and a TRP of 30% B0 to be achieved with at least 50% by 2035.

In Canada, the maximum acceptable harvest removal reference point is determined analytically as the best estimate of F_{MSY} from the stock assessment model (DFO, 2009). However, the advised fishing mortality (F_{trg}) can be at or below F_{MSY}, but must not exceed it, i.e. $F_{trg} \leq F_{MSY}$. The value for F_{trg} can be set smaller than F_{MSY} by factoring in the impact on other stocks ecosystem considerations and precaution in light of uncertainty. The stock status zones are defined as the Limit Reference Point (LRP) at the *Critical-Cautious* zone boundary, and an Upper Stock Reference Point (USR) at the *Cautious-Healthy* zone boundary. In absence of a pre-agreed harvest rule developed in the context of the PA, DFO (2019; Appendix 1b) provides provisional guidance for specifying the LRP and USR. The stock is considered to be in the *Critical Zone*, if the mature biomass, or its index, is less than or equal to 40% of the B_{MSY} estimate (i.e. $B_{lim} = 0.4\ B_{MSY}$), where B_{MSY} is the expected biomass corresponding to F_{MSY}. The stock is considered to be in the *Cautious Zone* if the biomass, or its index, is higher than 40% of B_{MSY} but lower than 80% of B_{MSY} ($0.4\ B_{MSY} < B < 0.8\ B_{MSY}$). F_{trg} is linearly reduced between the URP and the LRP. The stock is considered to be "healthy" if the biomass, or its index, is higher than 80% of B_{MSY} ($B > 0.8\ B_{MSY}$), with $F_{trg} \leq F_{MSY}$. In this case, the URP therefore serves the purpose of both reference point for stock status classification and an operationalised $B_{trigger}$ point that is bound to B_{trg} (i.e. $B_{trigger} = 0.8B_{trg}$).

In New Zealand, Australia and the USA, biological proxies for F_{MSY} and B_{MSY} are predominantly used (Punt *et al.*, 2013). For the New Zealand Harvest Strategy Standard (New Zealand Ministry of Fisheries, 2008), detailed guidelines (New Zealand Ministry of Fisheries, 2011) on selecting proxies, so called "MSY-compatible reference points", are specified for B_{MSY} as ratios to B_0 ($B_{\%}$) and F_{MSY} based on the per-recruit spawning potential ratio ($F_{SPR\%}$). The ratios are specified according to biological classifications into very low, low, medium and high productivity species (Musick, 1999; FAO, 2001), where lower productivity is associated with more conservative ratios (e.g. F_{SPR45} and SB_{40}). The default target is to achieve B_{MSY} (or its proxy) with at least a 50% probability. LRPs comprise a "soft-limit" at 0.5 B_{MSY} or 0.2 B_0, whichever is higher, and "hard-limit" at 0.25 B_{MSY} or 0.1 B_0, whichever is higher. The soft-limit is considered breached and the stock classified as depleted if there is a more than 50% probability that the biomass falls below the soft limit, whereas the hard-limit is considered breached and stock classified as collapsed if there is more than 50% that the biomass is below the hard-limit. Catch advice is implemented via a HCR. If biomass falls below the biomass trigger point ($B_{trigger}$) located between the biomass target (B_{trg}) and the soft-limit, fishing mortality is reduced linearly to keep the stock close to the target and away from the soft-limit, where $B_{trigger}$ is typically set relative to B_{trg} (Restrepo *et al.*, 1998). Harvest strategies based on MSE are advocated and tuning criteria are designed to be fully compatible with the minimum requirements of the Harvest Strategy Standard. The default performance criteria for MSEs are therefore specified to ensure that: (1) the probability of achieving the biomass target is at least 50%, (2) the probability of breaching the soft limit does not exceed 10%, (3) and the probability of breaching the hard limit does not exceed 2%.

In the USA, the choices of proxies for F_{MSY} and B_{MSY} vary widely, but those based on $F_{SPR\%}$ (typically $F_{SPR30\%}$ to $F_{SPR45\%}$) and its corresponding $B_{SPR\%}$ or $B_{\%}$ (e.g. B_{40}) are most frequently used. F_{MSY} or its proxy determines the Maximum Fishing Mortality Threshold (MFMT), where F > MFMT invokes a condition of overfishing and associated management interventions. The target fishing mortality F_{trg} is set lower than F_{MSY} so that the probability of overfishing is reduced below 50% according to the degree of scientific uncertainty, which is referred to as P* approach for data-rich assessments (Shertzer *et al.*, 2008). The LRP is referred to as Minimum Stock Size Threshold (MSST) below which the stock is considered to be overfished and invokes requirement for a rebuilding plan. The MSST is explicitly linked to the B_{MSY} or its proxy that is often specified to be larger or equal to $0.5B_{MSY}$.

Horbowy and Luzenzzyk (2012) interpreted the use of more conservative biological proxies for F_{MSY} to be consistent with the guidelines for applying a PA within an MSY framework in Annex 2 of the UN Fish Stocks Agreement (1995), which states that fishing mortality that produces the MSY should be considered as a fishing mortality limit rather than a management target. The basis for this is also well founded in the scientific literature, which frequently found that more conservative biological proxies for F_{MSY} are more robust to asymmetric risk associated with fishing below or above the 'true' unknown F_{MSY} (Mace, 2001; Horbowy and Luzeńczyk, 2012; Hordyk *et al.*, 2019), where asymmetric risk describes the phenomenon that one direction of bias for an estimate leads to disproportionately higher risk than if the bias would occur in the other direction (Hordyk et al. 2019).

The consequence of fishing above F_{MSY} is that the biomass will decrease relative to B_{MSY}, so that yield levels close to MSY cannot be maintained. Subsequent rebuilding requires fishing mortalities lower than F_{MSY} which may come at high costs of reduced catches and long recovery time. Fishing below F_{MSY} can result in short-term yield loss but in contrast to overshooting F_{MSY} the catch opportunity still exists at higher biomass levels. In comparison to the substantial biomass increase at F < F_{MSY}, the long-term loss in yield is relatively small (Hordyk *et al.*, 2019). For example, Beverton (1998) noted that instead of striving for F_{max} "a simple management system based on careful monitoring of fishing effort, biological targets such as F_{95} (i.e. a lower fishing mortality the results in 95% of the maximum yield), and exploitation of a diversity of fish resources may suffice to avert further disaster and hedge against uncertainty." Restrepo et al. (1998) showed that fishing at just 75% F_{MSY} would still yield an average 0.949 - 0.989 of MSY based on deterministic age-structured models that was parameterized with 600 combination of variations of life history parameters (Mace, 1994). Hilborn's (2010) concept of 'Pretty Good Yield' is also founded on the principle that fishing near but not at the maximum yield will reduce risk of overfishing and increase robustness to uncertainties with little long-term yield loss. Horbowy and Luzenzzyk (2012) and Punt et al. (2013) showed that fishing mortality corresponding to a biomass at 40% B_0 as a proxy for B_{MSY} leads to high yield and safe biomass levels irrespective of the steepness value of the stock recruitment function. Even fishing under a harvest control rule at F_{MSY} can still be associated with high risk of a stochastic collapse below $0.5B_{MSY}$ as a result of recruitment variability, while this risk can be significantly reduced by fishing somewhat below F_{MSY} (Thorson *et al.*, 2015). Recently, Hordyk et al. (2019) demonstrated by way of simulations with stock assessment feedback-loop that there is much higher risk to long-term yields and sustainable stock biomass levels when positively biased stock parameter (*e.g. M,* steepness and historical catches) lead to an overoptimistic F_{MSY} than with the equivalent negative bias.

3. *Proposed candidate reference point system*

To make a reference point system operational requires general guidelines on how to specify the reference points in practice. A guiding principle for developing these guidelines is that reference points, such as the F_{MSY} proxies, should be set stock-specific by considering its biology, productivity and ecology, and the nature of the fisheries, following international best practice. The reference point system should be based on tangible and transparent rules and should provide a clear and unambiguous interpretation of the stock status. The proposed candidate reference point system builds on the key recommendations by ICES WKREF1 (2021), which interpret and define as follows (Figure 1):

- **B_{lim}** is the deterministic biomass limit below which a stock is considered to have reduced reproductive capacity, or productivity. For stocks where quantitative information is available, a reference point B_{lim} may be identified as the stock size below which there is a high risk of reduced recruitment. In this study, we consider Type 1 and Type 2 of the three newly proposed typologies to derive B_{lim} made by WKREF1 (2021):

- **Type 1**: Consider an empirical Hockey-Stick for deriving B_{lim} only if there is a clear relationship between stock and recruitment, the data show contrast and a breakpoint is clearly defined
- **Type 2**: Determine a plausible B_{lim}/B_0 ratio based on biological principles and life history of the stock (e.g. 10% to 25% of B_0 depending on the type of stocks)
- **Type 3:** It meant for those stocks where recruitment is dominated by occasional good year-classes (i.e. spasmodic recruitment), e.g. dynamics are process error driven, the lowest observed SSB(s) that gave rise to a good year class can be used as basis for B_{lim}

- $\mathbf{F_{P.05}}$ is the fishing mortality that is associated with a 5% risk that SSB falls below B_{lim} as derived using stochastic long-term projections.
- $\mathbf{F_{brp}}$ is the biological reference point proxy for F_{MSY} which can be computed at equilibrium or derived from long-term projections to incorporate additional structural uncertainties and estimation errors (e.g. required for ensembles). Here, we consider two type of F_{brp} estimators:
    - $\mathbf{F_{SPR\%}}$: The fishing mortality at which the spawner-biomass-per-recruit (SPR), e.g. 40%, of its unexploited level of SPR_0 at F = 0. $F_{SPR\%}$ requires no assumption about SR. We consider a range is F_{spr35} -F_{spr50} for evaluations using simulation.
    - $\mathbf{F_{B\%}}$**:** The fishing mortality at which the spawning stock biomass (SSB) is e.g. 40% of its unexploited level at B_0, i.e. F_{B40}. Computation of $F_{B\%}$ relies on a SSR assumption. For a Beverton-Holt SRR $F_{B\%}$ is smaller to the equivalent $F_{spr\%}$ (i.e. $F_{B40} < F_{spr40}$). A specific property of the segmented regression SSR (Hockey-Stick) is that $F_{B\%}$ is equal to the equivalent $F_{spr\%}$ if the corresponding biomass is larger than B_{lim}. Here, we therefore consider a lower range of F_{B30} -F_{B45} for simulation-testing.
- $\mathbf{B_{trg}}$ is the **biomass target** ($\mathbf{B_{trg}}$), i.e. the expected average biomass that corresponds to F_{brp}, which can be computed at equilibrium quantity or derived from long-term projections.
- $\mathbf{F_{trg}}$ is the fishing mortality used in the advice rule. In accordance with the ICES PA, F_{trg} must not exceed $F_{P.05}$, such that $F_{trg} = min(F_{brp}, F_{P0.5})$. The definition of F_{trg} is used here as the equivalent of F_{MSY} as defined in the current ICES advice framework to ensure a clear distinction between F_{trg} and "true" F_{MSY} of the simulated stock. Note that if $F_{P0.5} < F_{brp}$, $\mathbf{B_{trg}}$ is not adjusted upward to correspond to F_{trg}. The reasoning is that $F_{P.05}$ is thought to act as precautionary safeguard to prevent biomass to fall below B_{lim} which has no obvious implications for the need of changing B_{trg}.

- **B_0** is not directly used in the advice rule, but included here because it can be considered for specifying **F_{brp}** and **B_{trg}** values based on **$F_{B\%}$** as well as B_{lim} (Type 2). In age-structured models, B_0 is the unfished spawning biomass that is given by the product of recruitment R_0 of an unfished stock (implicit to the SR relationship) and the unfished spawning biomass-per-recruit (SPR_0) being a function of weight-at-age, maturity-at-age and natural mortality under current conditions (e.g. average of the last 3 years). If the biology is time-varying, B_0 will therefore differ from the virgin biomass that is assumed to be representative of historical conditions prior to fishing.
- Like B_{MSY}, it is therefore an implicit property of any age-structured model for which a stock recruitment relationship is estimated or assumed.
- **$B_{trigger}$** is the operationalised biomass trigger point for tuning of the harvest control rule (not a reference point). If biomass falls below $B_{trigger}$, F_{trg} is decreased linearly toward minimum biomass (default is zero) at which the fishery may be closed. The $B_{trigger}$ is a generalization of the MSY $B_{trigger}$. Two options are considered for specifying the $B_{trigger}$ value:
    - (1) as a fraction of B_{trg} (here: 0.6 - 1.0 of B_{trg})
    - (2) as multiplier of B_{lim} (here: 2 x B_{lim})

A new element to the ICES reference point system is the introduction of an explicit B_{trg} reference point that corresponds to the F_{brp} proxy for F_{MSY}. Therefore, guidelines are needed on how to quantify stock status relative to B_{trg}, for example, by specifying the level probability being close or above B_{trg} at biomass levels capable of producing MSY. In real-world stochastic systems, the biomass will fluctuate around B_{trg} when fishing at the corresponding F_{brp}. The extent of biomass fluctuation depends on the variability and autocorrelation of recruitment, as well as time-varying biological processes (e.g. somatic growth, maturation and survival). An important property of stochastic stock dynamics to consider is that the probability of biomass being below the B_{trg} tends to be larger than being above it due to the lognormal nature of the system (Thorson et al. 2015). This is aggravated for species that exhibit high recruitment variability and short generation turn-over, such as many small pelagic foraging species (Thorson et al. 2015; Mildenberger et al. 2021). One option to attain probabilities above 50% of being above B_{trg} is to reduce F_{trg} relative to F_{MSY} or its proxy (Mildenberger et al. 2021). However, considering that F_{brp} proxies tend to be more conservative than F_{MSY}, this could result in increased risk of reduced fishing opportunities by reducing a conservative a F_{brp} to an even more conservative F_{trg}. As an alternative we therefore adopted a target threshold (**B_{thresh}**) at 80% B_{trg}

to be achieved with probability of at least 50%, as used by FAO (e.g. Sharma et al., 2021) and Canada (DFO, 2019) for classifying stock status as "sustainably fished" and within "Healthy Zone (Green)", respectively.

We seek to align the reference point system with a status classification system that facilitates clear and unambiguous interpretation. A clear definition of sustainability is important to make the reference point system operational and useful, so that the achievement of sustainability can be assessed against quantitative objectives and effectively communicated to stakeholders (Quinn and Collie, 2005). Currently, ICES uses pictograms (i.e. green, yellow, red) to represent the status of the stocks and their exploitation, relative to management objectives as defined by separate categories for the ICES MSY and PA reference points. Stocks are classified by "green" and "red" symbols with respect to the reference points for fishing pressure and stock size. Separating the PA and the ICES MSY approach into different categories of reference points results in the current classification system being complex and difficult to illustrate. With the aim to unify the MSY and Precautionary approach within a single reference point system, we integrate the four colour classification system of the Kobe MSY framework used in tuna RFMOs (de Bruyn et al., 2013) with key elements for the PA frameworks drawn from ICES (ICES, 2020), the New Zealand Harvest Standard (New Zealand Ministry of Fisheries, 2008) and the Canadian Harvest Strategy (DFO, 2009).

The harvest control rule is embedded in the stock classification system and is shown together with governing reference points in **Figure 1**. The reference point system includes five stock status zones delineated for stock size by B_{lim} and $\mathbf{B_{thresh}}$, in this example set to 80% B_{trg}, and for fishing pressure by F_{trg}. Stock size for age-structured assessment is usually represented by stock spawning biomass. The $B_{trigger}$ location may vary relative to B_{trg} or B_{lim}, depending on stock's biology, and is therefore explicitly not considered for stock status classification.

The stock status zone below F_{trg} and above $\mathbf{B_{thresh}}$ is the "Sustainable" zone illustrated in green (B > $\mathbf{B_{thresh}}$ and F < F_{trg}). The orange "Overfishing" zone demarcates sustainable biomass levels above $\mathbf{B_{thresh}}$, but unsustainable fishing pressure (B > $\mathbf{B_{thresh}}$ and F > F_{trg}). The stock is classified to be in the yellow rebuilding zone if biomass is below $\mathbf{B_{thresh}}$ but fishing pressure is below F_{trg} so that biomass is predicted to increase (B > B_{tresh} and F > F_{trg}). The stock falls in the red "overfished" zone if fishing mortality is above F_{trg} and biomass falls below $\mathbf{B_{thresh}}$. However, to be consistent with the principles of the PA, a fifth "Critical" zone is introduced, if spawning stock biomass is below B_{lim}. The classification of this zone is conceptually independent of fishing pressure relative to F_{trg}, but the dark red shading that overlays the yellow "Rebuilding" zone still allows depicting if a status stock is "Critical" and under "Rebuilding" (Figure 1).

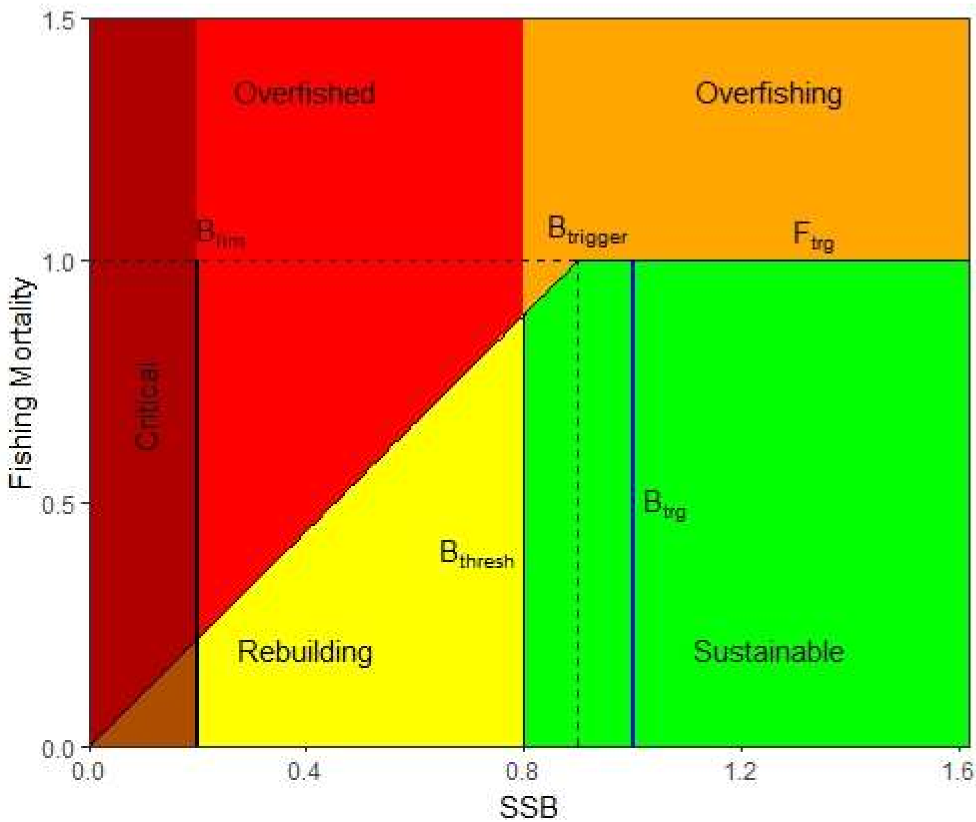


Figure 1: Proposed ICES Reference points system with integrated Harvest Control Rule. (source `FLRef` function plotWKREF(); https://github.com/henning-winker/FLRef)

4. Simulation-test framework

We develop our simulation-testing framework using the tools available in the Fisheries Library for R (FLR; Kell et al., 2007; https://flr-project.org/). The simulation framework was implemented in the FLR library `mse` (https://github.com/flr/mse) with `FLasher` (https://github.com/flr/FLasher) being used to carry out the forward projections. All Stock Recruitment relationships were conditioned using the FLR package `FLSRTMB` (https://github.com/flr/FLSRTMB). Reference points at equilibrium were calculated with `FLBRP` (https://github.com/flr/FLSRTMB). To facilitate customised reference point estimation and visualisation of F_{brp} ($F_{S\%}$ and $F_{B\%}$), B_{lim}, $F_{p0.5}$, B_{trg}, F_{trg}, we developed the FLR package `FLRef` (https://github.com/henning-winker/FLRef). `FLRef` makes use of the new fast forward projection 'ffwd()` in `FLasher` together with the bisection function `bisect()` in 'mse' to efficiently derive precise values of $F_{P0.5}$ based stochastic simulations. All data and R code used in this analysis will be made available in the github repository of `FLRef`.

4.1 *Stock assessment data*

For simulation-testing, we use a unique dataset of detailed stock assessment outputs for 64 stocks that cover the entire ICES region across the North-East Atlantic, which were collated in the form of objects of the `FLStock` class as defined in FLR. The 64 stocks were sourced from a database that include 77 stocks which are assessed as Category 1 by ICES in 2020 and 2021. In the following, stocks are referred to by ICES stock IDs, with details on the assessment outputs of all stocks in the form of a Shiny Application (https://michaelgras.shinyapps.io/WKREF1), which also comes with a range plots visualising various aspects of each stock's population dynamic and productivity characteristics .

Of the 77 stocks, eight stocks were excluded as MSY reference points are undefined (i.e. cod.27.1-2coastN, cod.27.24-32, san.sa.1r, san.sa.2r, san.sa.3r, san.sa.4, spr.27.3a4 and reb.27.1-2). Further five stocks (sol.27.7e, sol.27.8ab, cod.27.7e–k, her.27.20–24, whg.27.47d) were excluded due to challenges of estimating realistic B_0 or F_{MSY} values (i.e. $F_{MSY} < 1$) within plausible limits during Operating Model conditioning (see below) for the given assessment assumptions, such as natural mortality and selectivity. The final set of 64 `FLStock` objects represent the unified assessment outputs of 12 different age-structured assessments platforms, of which SAM (n = 27; Nielsen and Berg, 2014), Stock Synthesis (n = 9; Methot and Wetzel, 2013) and XSA (n = 8) were the most common. The 64 stocks comprised 23 bony fish species representative of nine taxonomic orders as well as one crustacean, *Pandalus borealis* (pra.27.3a4a). The majority of stocks belonged to the following three taxonomic orders Gadiformes (n = 27), Pleuronectiformes (n = 14) and Clupeiformes (n = 11). Note that there is only one chondrichthyes species (North East Atlantic spurdog) assessed as category 1 by ICES, but the assessment is not included in our database.

We characterised the stocks into low, medium and high productivity categories in accordance with the classification scheme proposed by FAO (2001), using the intrinsic rate of population increase r and mean generation time G (Table 1). In cases where r and G resulted in different categories, the lower productivity class was chosen. Productivity is a function of somatic growth, reproduction, survival and longevity. More productive species tend to have high somatic growth, early maturation and short generation times. These life history traits are typically associated with high resilience to growth overfishing and fast rebuilding potential if conditions are favourable. High productivity is therefore often perceived as highly resilient to fishing pressure based on their "ability to rebound after perturbation" (Holling 1973). On the other hand, these traits are often associated with high variability in recruit and fewer mature fish to buffer against sequential recruitment failure, which can make them more vulnerable to recruitment overfishing and risk stochastic depletion, even under light fishing pressure (Thorson *et al.*, 2015).

A direct indicator for productivity is *r,* which summarizes several key life history traits into a single metric (Musick, 1999) (Musick, 1999). FAO (2001) suggested the mean generation time *G* as an additional indicator, which quantifies the turnover time of generations and is widely considered for setting targets for rebuilding plans. Both *r* and *G* can be directly derived from a Leslie Matrix (McAllister *et al.*, 2001; Thorson, 2020), which requires weight, maturity, and *M*-at-age from the `FLStock` objects as well as an estimate of recruitment compensation in the form of the steepness *s* of the stock recruitment relationship. We implemented the Leslie matrix generic tool in the R package FLSRTMB (https://github.com/flr/FLSRTMB) and provide details on methods in Appendix B. Stock specific steepness *s* were derived as the expected means from the hierarchical taxonomic FishLife model (Thorson, 2020; https://github.com/James-Thorson/FishLife), which are summarised in **Table B1**. Most stocks fell into the medium productivity category (n = 37), followed by low productivity stocks (n = 17), and high productivity stocks (n = 10).

Table 1: Guidelines used for categorising productivity levels for exploited fish species. Criteria for intrinsic rate population increase *r* are from Musick (1999) and value of Generation Time *G* are adopted from FAO (2001). In cases where *r* and *G* resulted in different categories, the lower productivity class is chosen.

Parameter	Productivity		
	Low	Medium	High
Intrinsic population Growth *r*	< 0.15	0.15 - 0.5	> 0.5
Generation Time *G*	> 10	5 -10	< 5

4.2 Conditioning of Operating models

Operating Models were implemented as single sex and single fleet models with an annual time step. Future projections were run over 60 years (i.e. 2021-2080) with 250 iterations and based on the 3-years average of the most recent data years for weight-at-age (w_a), maturity-at-age (mat_a), natural mortality-at-age (M_a) and the F_a pattern determining the selectivity-at-age (s_a). This choice was made to account for non-stationary processes in these quantities. The performance evaluations were based on the last 10 years of the 60-year projection horizon (i.e. 2071-2080).

For the simulation testing, a generic Beverton-Holt model (BH-SRR) was assumed for all stocks. The recruitment deviation is assumed to be associated with a first-order autocorrelation (AR1) process and a function of recruitment standard deviation σ_r and the AR1 coefficient ρ (Johnson et al. 2016). To ensure an objective and unified approach representative over the wide range of life histories across the 64 stocks, species-specific predictive distributions for steepness *s* were and expected means for σ_r and ρ were sourced from the hierarchical taxonomic FishLife model to fit a Beverton-Holt (BH) to the SR data and generate the recruitment deviations, respectively (Thorson, 2020; https://github.com/James-Thorson/FishLife).

The parameters of stock-recruit curves are notoriously difficult to estimate, and often little inference can be made from a single stock-recruit fit, but meta-analysis and the use of distributions as a Bayesian prior can provide a useful starting point from which meaningful updates could occur. This approach of using prior information to condition the SR to the data, is consistent with discussions and suggestions for future work in WKMSYREF2 (ICES, 2014). Instead of assuming that nothing is known, other than the information that is contained in the stock data alone, this approach assumes that at least within taxonomic groupings (family, species) information from one stock can provide some useful prior information about the recruitment compensation for another (Myers *et al.*, 1999; Thorson, 2020). For stocks with few years of SR data, or where the observations appear uninformative, priors can assist in making less spurious inference about the SR, whereas if the SR data are informative, so that the priors are effectively updated by the data.

The Beverton-Holt SSRs were fitted to S-R data using the R package FLSRTMB (https://github.com/flr/FLSRTMB), which implements a re-parameterised of the BH SR as a function of steepness s and annual unfished spawning biomass per-recruit SPR_0 to accommodate the integration of priors for s (Thorson, 2020). A notable difference to the conventional parameterization is that SPR_{0_y} is treated as non-stationary, being a function of annual quantities of $W_{a,y}$, $Mat_{a,y}$ and $M_{a,y}$. By way of using time-varying $SPR_{0,y}$, it also takes into consideration the recent criticism by Miller and Brooks (2021) that specifying a set biological parameters to define a single time-invariant SPR_0 can be highly sensitive to reference estimation when using steepness values from meta-analysis (See Appendix A for details).

4.3 Implementation system

To facilitate comparability of the tested reference point systems, all considered harvest control rules (HCRs) are kept generic and in the same form of the conventional ICES Advice Rule (ICES, 2021d), where the F

advice decreases from F_{trg} to zero and $B_{trigger}$ and zero SSB (Figure 1). Variations of the tested HCRs are therefore determined by the parameters F_{trg} and $B_{trigger}$.

The HCRs were implemented using a simulated feedback control loop between the implementation system and the operating model, where the implementation system translates the assessment outcome via the HRC into the Total Allowable Catch (TAC) advice (Figure 2). The key difference to a simple stochastic risk simulation, such as EQSIM, is the simulated feedback control loop between the implementation system and the operating model allows accounting for the lag between the last of year data used in the assessment and the implementation year of TAC advice. In ICES, the implementation system of the harvest control rule is based on the assumption that advice is given for year y+1 based on an assessment completed in year y, which is typically fitted to data up until last data year y-1 (ICES, 2020b). Therefore, implementation of the TAC derived through HCR requires projection of the stock dynamics by way of a short-term forecast (Mildenberger *et al.*, 2021). To do this, numbers-at-age were projected through the year of assessment. Status quo recruitment, M_a, w_a and mat_a were set as the mean of the last 3 years. A projection based on a fixed fishing mortality-at-age to the last year (y-1) in the assessment is then made through to the implementation year (y+1).

In contrast to a full Management Strategy Evaluation (MSE) simulation design (Punt et al. 2017), this MSE 'shortcut' approach (e.g. ICES, 2020 WKGMSE3), omits the step of the annual updating of the estimation model (assessment) in the feedback control. Instead, it passes the 'true' age-structured dynamics from the OM (or with assumed some error) to the HCR implementation. For testing the robustness of reference point systems across a large number of stocks the merits of a short-cut MSE approach include: (1) the straight-forward implementation using the tools available in 'FLR' (Kell et al., 2007), i.e. 'mse' and 'FLasher', (2) reduced computation time, (3) data requirements are limited the available assessment outputs (FLStock class object) without the need of sourcing auxiliary data to recondition the assessment models, and (4) the incorporation of the lag effect between data, assessment and management implementation.

The limitations of the MSE short-cut approach are that it cannot fully account for uncertainties resulting from imperfect sampling of the full age-structure (e.g. poorly sampled recruits), observation error, model estimation error, misspecified assumptions about the biology (M_a, w_a or mat_a) and selectivity. Therefore, robustness testing is limited here to the structural uncertainty about the externally fitted SR, which determines the stock's recruitment compensation and the absolute scale of R_0, with direct impacts on reference points such as F_{MSY}, B_{MSY}, MSY, B_0, B_{trg} or $B_{trigger}$.

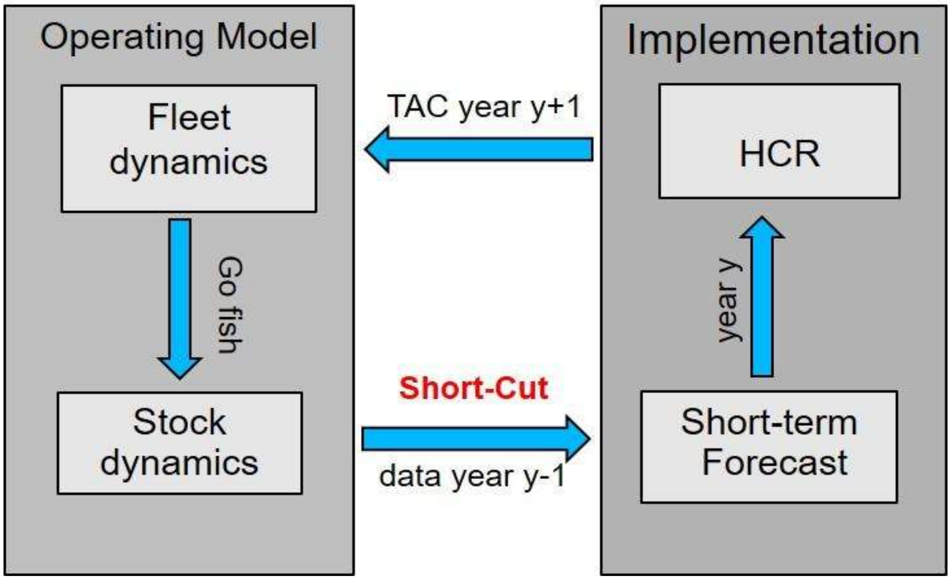


Figure 2: Schematic illustrating the key processes of the short-cut approach to MSE, showing the Operating model that simulates the fishery and stock dynamics on the left and Implementation System including the short-term forecast on the right. The short-cut denotes the omission of the estimation (stock assessment) model that updates to new observations (with estimation error) in conventional MSE implementations with full feedback control loop.

4.3 Performance Evaluation Criteria

The consistency tests were designed to identify the generic rules for specifying F_{brp}, B_{trg} and $B_{trigger}$ according to stock-specific productivity that provide the optimal trade-offs among the following three main objectives: (1) to not exceed a 5% probability of SSB falling below B_{lim} in any single year (2) to achieve high long-term yields that correspond to at least 95% of the median long-term yield attained by fishing at F_{MSY} (MSY), (3) to attain at least 50% probability that SSB is above B_{thresh} set at 80% of B_{trg}. Consistent with the objectives of ICES advice framework (ICES, 2020d), the three objectives are interpreted hierarchically in that objective (1) is the overriding criteria of maintaining stock size above B_{lim} with at least 95% probability to be compliant with the ICES PA. Conditional on objective (1), objective (2) is based on the ICES definition for using plausible values around F_{MSY} in the advice rule, which are derived so that they lead to no more than a 5% reduction of MSY obtained by fishing at F_{MSY} in the long term. The B_{tresh} in objective (3) replaces the current MSY $B_{trigger}$ threshold (which is normally set to $1.4B_{lim}$) for classifying stock size to be at biomass levels that can produce MSY (green).

For this performance evaluation, F_{trg} was set to F_{brp}, but we also analysed how often F_{P05} would be invoked based on the specifications of the OM (see Section 6.3). To set B_{lim}, we considered both estimators for Type

1 and Type 2. To derive Type 1 B_{lim}, a generic continuous Hockey-Stick (HS) model was fitted to the SR data (Appendix B). In absence of contrast in a large proportion of S-R dataset, the HS was constrained to assure that B_{lim} falls within a range of $0.1B_0 < B_{lim} < 0.3B_0$ to ensure that Type 1 B_{lim} was estimated within plausible biological limits. Within these constraints B_{lim} is estimated by the breakpoint $b = B_{lim}$, while R_0 is given by the product of the slope *a* and *b* (see details in Appendix B). Type 2 B_{lim} was derived as the 10% of B_0, where B_0 is the equilibrium estimate under F = 0 based on the “true” SR of the OM and average stock biology over the most recent three years. Regressing the so derived Type 1 and Type 2 B_{lim} values against each other on log-scale showed notable variation (CV = 40%) among the 64 stocks but indicated no systematic divergence from a 1:1 relationship. Type “P3” probability was applied to compute the risk for the biomass limits as the maximum of annual probabilities. The performance statistic for MSY was quantified as long term median yield obtained when fixing F_{trg} of HCR to the “true” F_{MSY} of the OM.

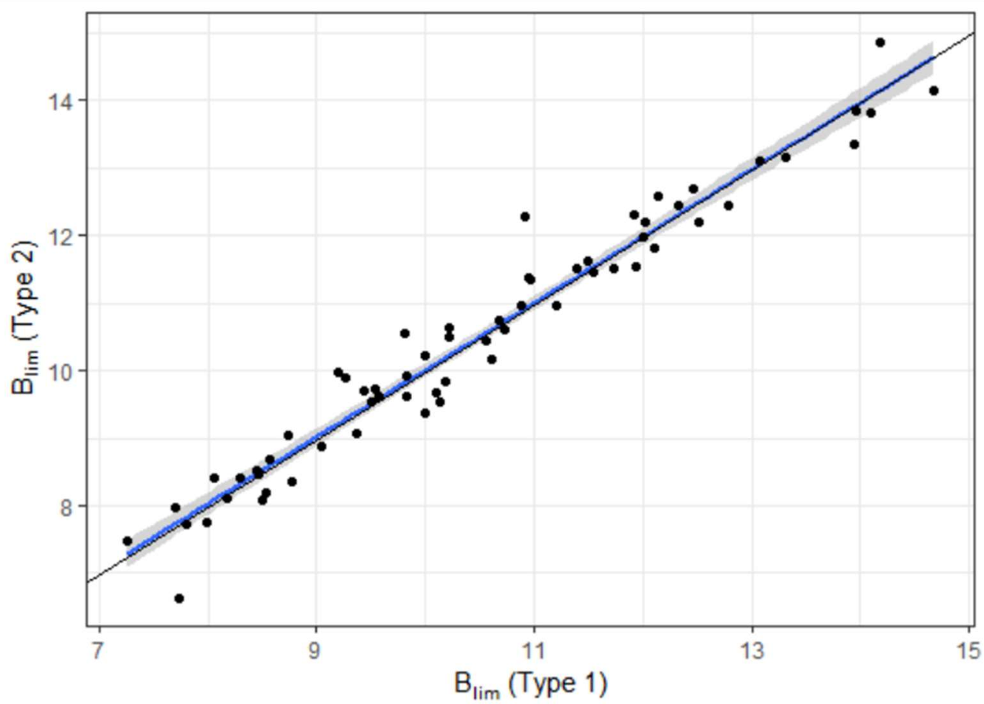


Figure 3. Relationship between Type 1 and Type 2 on a log-scale (CV=40%)

For the robustness evaluation, we retain objectives (1) and (2) as performance criteria, but instead of B_{tresh} from objective 3, we used the “true” B_{MSY} from the OM as the third performance criteria. This allowed us to evaluate if the underlying goal to restore and maintain stocks above average levels that can produce MSY is achieved by the selected candidates reference point systems that were most consistent in meeting the objectives. To test the robustness of the selected "WKREF" candidate reference point systems, we considered two scenarios for violating the assumptions about the SSR with respect to the "true" functional form of the OM. These were: (1) a Beverton Holt SRR, but fitted without informative priors about *s* and (2) the continuous Hockey-Stick SSR (Appendix B). This effectively achieved various extents of misspecification of the SRR and the associated production function across the 64 stocks (Supplement 1). For reference, we also compare the performance of "WKREF" candidate reference systems to: (1) the current ICES advice rule, by setting the official 2021 ICES benchmarks of F_{MSY} as F_{trg} and MSY $B_{trigger}$ as $B_{trigger}$, (2) the New Zealand Standard, and using directly the estimates of F_{MSY} and B_{MSY} to specify F_{trg} and B_{trg}, respectively (See Table 3).

5. *Results*

5.1 Results of Self-test consistency

A total of 32 scenarios in two 4 x 4 grids were tested. The first grid comprises F_{SPR} that ranges from 35 to 50% and the second grid ranges of F_B ranged from 30 to 45%. These F_{brp} ranges were tested in both grids with alternative $B_{trigger}$ set equal to 0.6, 0.8 and 1 × B_{trg} and 2 × $B_{lim,}$ where type 1 B_{lim} was in this case used to estimate $B_{trigger}$.

For low productivity stocks, all tested F_{brp} proxies for F_{MSY} are precautionary with a less than 5% probability of SSB falling below B_{lim} (Figure 4). This is irrespective of how B_{lim} is set (i.e. Type 1 or Type 2) or which fraction of B_{trg} is used to determine $B_{trigger}$. Medium productive species showed higher risk of falling below Type 1 than Type 2 B_{lim}. In accordance with the PA, F_{trg} needs to be set at $F_{B35\%}$ or F_{SPR40} (or larger) in combination with a trigger of at least 80% of B_{trg} or 2 × B_{lim}. In contrast to $F_{B\%}$ proxies, the 5% risk threshold for Type 1 B_{lim} was exceeded for some medium productivity stocks at $F_{SPR40\%}$. In contrast to Type 1 B_{lim}, all F_{brp} proxies consistently met the precautionary objective for low and medium productive species, with the exception of $F_{SPR35\%}$ in combination with the $B_{trigger}$ set at 0.6 B_{trg}. High productivity species were associated with substantially higher risk to fall below B_{lim}, with comparable levels of $F_{SRP\%}$ being substantially more risk prone than $F_{B\%}$. Like for medium productivity, Type 1 B_{lim} was associated with a higher risk than Type 2 B_{lim}. Consistency with the precautionary objective, was only achieved for $F_{B\%40}$ or combinations F_{SRP50} and $B_{trigger}$ set to B_{trg} or higher.

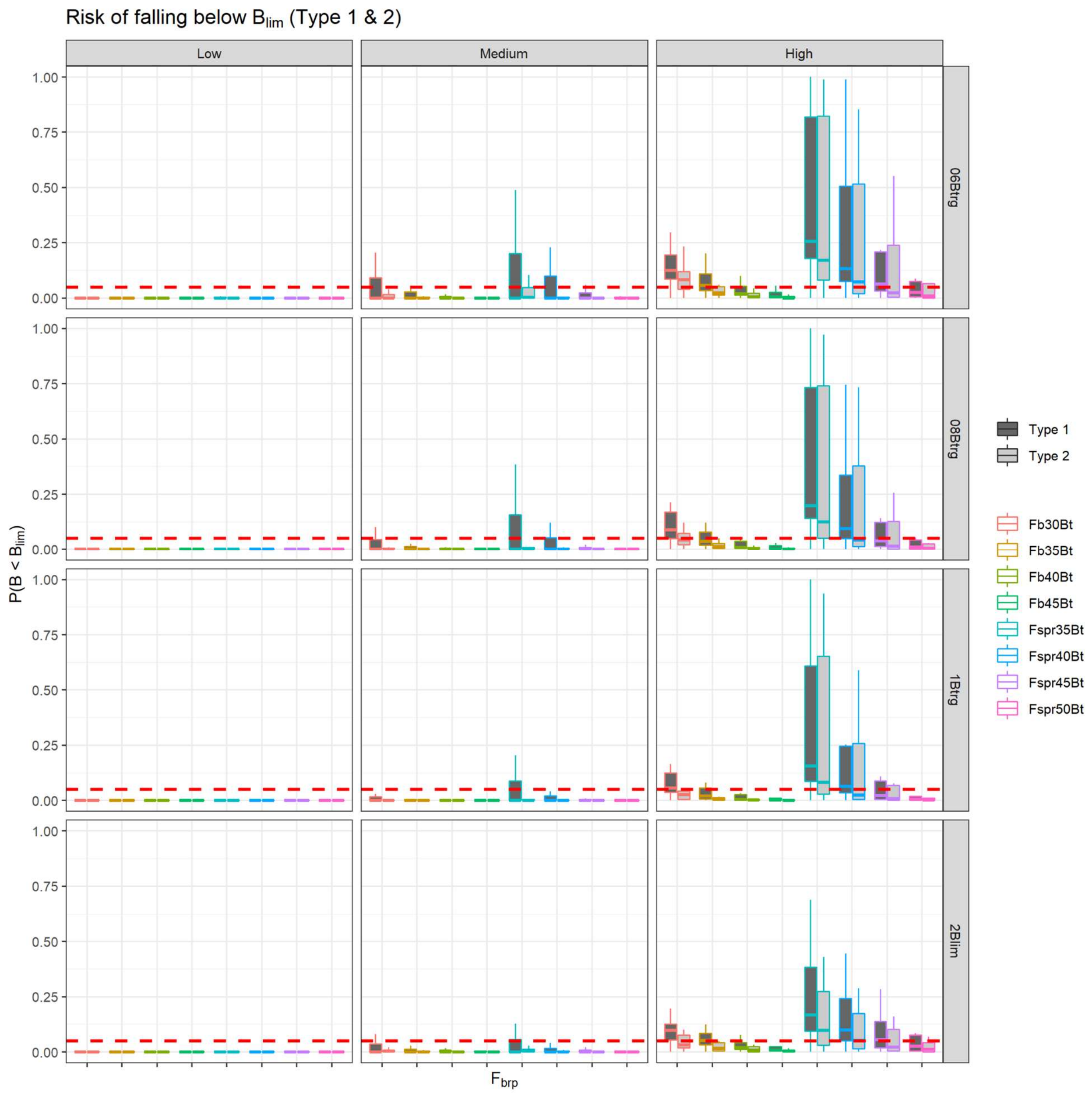


Figure 4: Self-test consistency evaluations of the type 3 risk probability (P3) that SSB falls below B_{lim} shown for low, medium and high productivity stocks (columns) across colour-coded ranges for $F_{SPR\%}$ of 35-50% and $F_{B\%}$ of 30-45 in combinations with alternative $B_{trigger}$ values of fractions of 0.6, 0.8, 1 B_{trg} and a multiplier of 2 x B_{lim} (rows). The red dashed line denotes the limit threshold of a 5% probability in accordance with ICES Precautionary Approach.

For low and medium productivity stocks, highest long term catches in excess of 95% MSY are obtained with F_{brp} proxies specified at levels of 30 - 35% for $F_{B\%}$ and 40 - 45% for $F_{SPR\%}$ in combination with $B_{trigger}$ between 0.8 and 1.0 B_{trg} or at 2 × B_{lim} (Figure 5). The situation is very different for high productivity stocks (e.g. sardine, sprat), for which more conservative proxies of $F_{B\%}$ and $F_{SPR\%}$ lead to increased yield. Here, highest long term catches in excess of 95% MSY are obtained with F_{brp} equal to F_B 40 to 45% and F_{SPR} 45 to 50% in combination with $B_{trigger}$ between 0.8 and 1.0 B_{trg} or equal to 2 × B_{lim}.

The results of the self-test showed that the probability of exceeding B_{tresh} (at 80% B_{trg}) increases by setting $B_{trigger}$ closer to B_{trg}. However, for low and medium productivity stock high $B_{trigger}$ values indicate yield loss and thus creates a conflict with the objective to optimise long yield. High productivity stocks, by contrast, indicated no conflicts among the objects of optimising yield, exceeding B_{tresh} and minimizing the risk of falling below B_{lim}, with optimal trade-off being achievable with more conservative combinations F_{brp} and $B_{trigger}$. Setting $B_{trigger}$ equal to 2 × B_{lim} performs generally similar in terms of the yield and risk objectives when compared to optimal setting of $B_{trigger}$ to 0.8 B_{trg} for low and medium productivity stocks and equal to B_{trg} for high productivity stocks. However, in particular for medium and high productivity species, the probability to exceed B_{tresh} is notably lower when $B_{trigger}$ is set 2 × B_{lim} associated with large variations among species. Setting $B_{trigger}$ to 2 × B_{lim} is therefore associated with increased risk of inconsistent stock status classification, which can be minimised by setting $B_{trigger}$ relative B_{trg}.

Based on these results, we chose to specify the candidate reference points for further robustness testing using $B_{trigger}$ equal to 0.8B_{trg} for low and medium productivity stocks and $B_{trigger}$ equal to B_{trg} for high productivity stocks (Table 3)

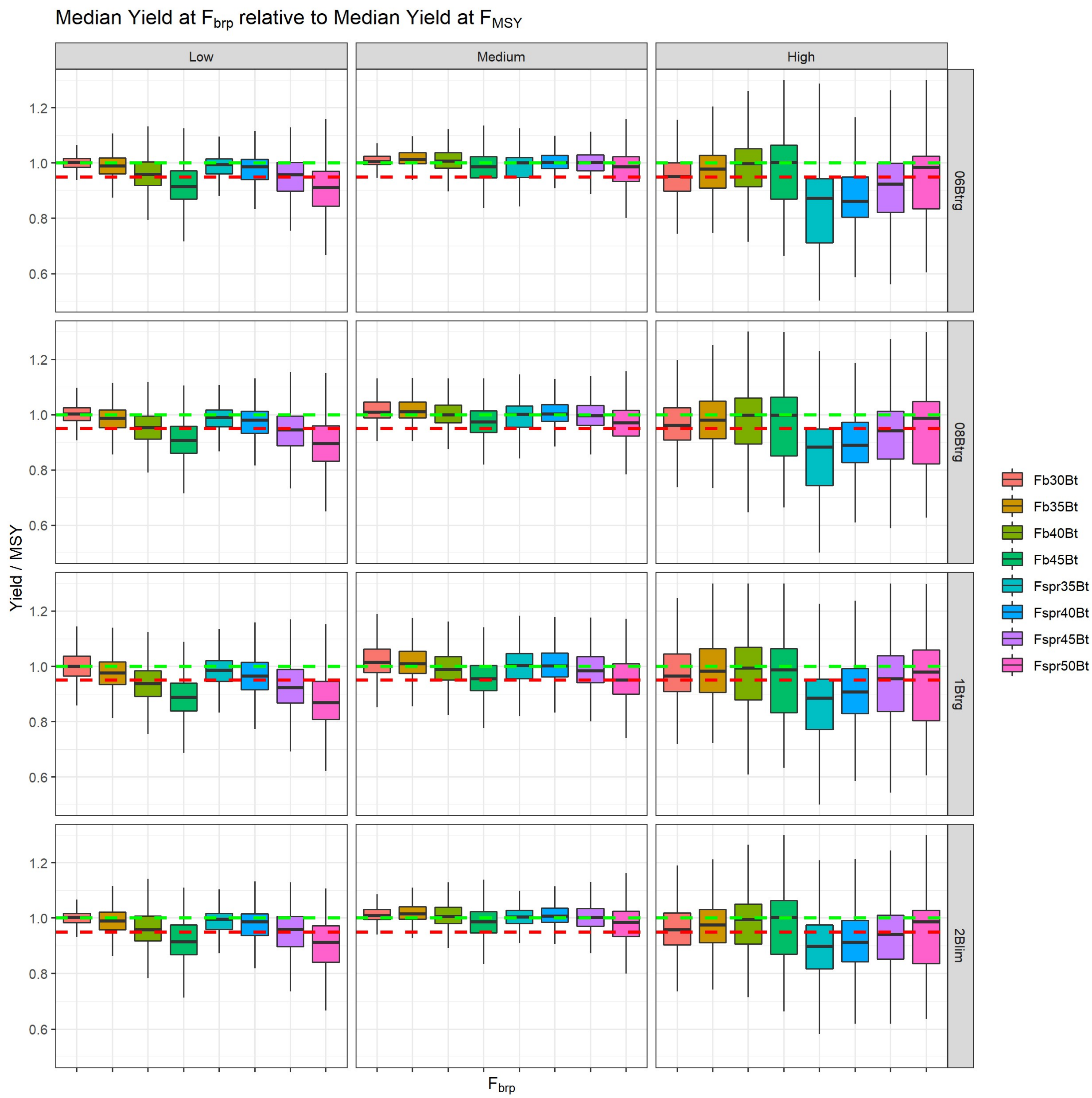


Figure 5: Self-test consistency evaluations of the median long term yield relative the median long-term obtained at fixed "true" F_{MSY} (MSY) shown for low, medium and high productivity stocks (columns) across colour coded ranges for $F_{SPR\%}$ of 35-50% and $F_{B\%}$ of 30-45 in combinations with alternative $B_{trigger}$ values of fractions of 0.6, 0.8, 1 B_{trg} and a multiplier of 2 x B_{lim} (rows). The green dashed line denotes a 1:1 ratio of long term Yield/MSY and the red dashed line denotes the yield threshold at 95% MSY.

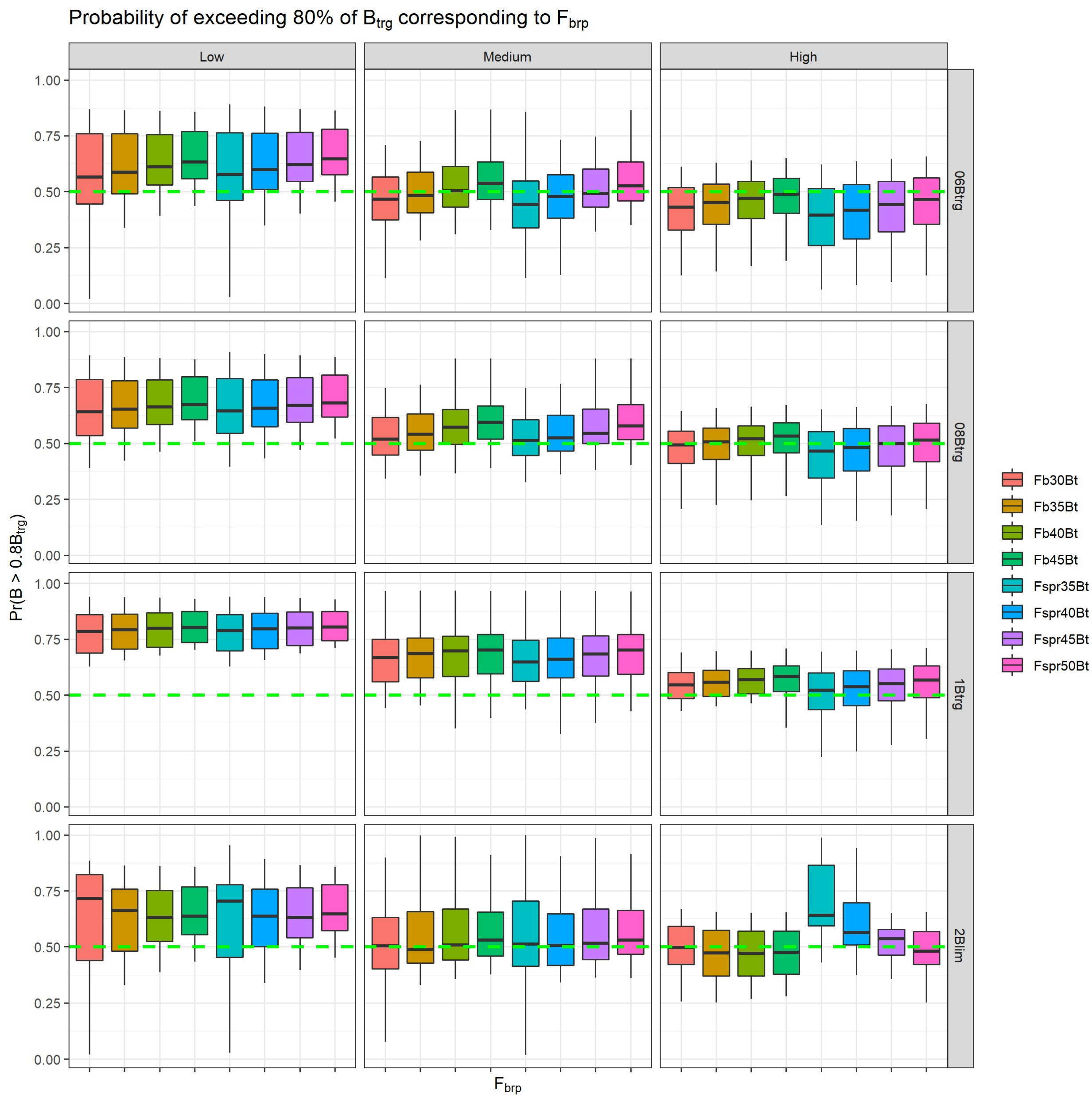


Figure 6: Self-test consistency evaluations of the probabilities that SSB exceeds B_{tresh} at 80% B_{trg} shown for low, medium and high productivity stocks (columns) across colour coded ranges for $F_{SPR\%}$ of 35-50% and $F_{B\%}$ of 30-45 in combinations with alternative $B_{trigger}$ values of fractions of 0.6, 0.8, 1 B_{trg} and a multiplier of 2 x B_{lim} (rows). The green dashed line denotes a 50% probability threshold of exceeding B_{tresh}.

5.2 Results from robustness tests

Details on the specifications of the reference point systems considered for robustness tested are presented in Table 3, together acronyms used hereafter. The candidate reference point systems that showed the best performance in the self-tests are referred to as "WKREF".

Table 3: Specifications of alternative reference point system evaluated by robustness testing. SRR: Stock-recruitment recruitment relationship; BH: Beverton and Holt; HS: Hockey-Stick.

Advice Rule	Productivity	F_{trg}	B_{trg}	$B_{trigger}$	SRR	Acronyms
ICES	-	Advice	-	Advice	N/A	ices.ar
F_{MSY}	All	F_{MSY}	B_{MSY}	0.8 *x* B_{trg}	BH	fmsy.bh
New Zealand	Low	F_{SPR45}	B_{40}	min(1-*M*, 0.5)	BH	
	Medium	F_{SPR40}	B_{35}	min(*1*-*M*, 0.5)	BH	nz.bh
	High	F_{SPR35}	B_{30}	min(*1*-*M*, 0.5)	BH	
WKREF1 SPR%	Low	F_{SPR40}	B_{SPR40}	0.8 *x* B_{trg}	BH / HS	fspr.bh / fspr.hs
	Medium	F_{SPR40}	B_{SPR40}	0.8 *x* B_{trg}	BH / HS	
	High	F_{SPR50}	B_{SPR50}	1 *x* B_{trg}	BH / HS	
B%	Low	F_{B35}	B_{35}	0.8 *x* B_{trg}	BH / HS	fb.bh / fb.hs
	Medium	F_{B35}	B_{35}	0.8 *x* B_{trg}	BH / HS	
	High	F_{B40}	B_{40}	1 *x* B_{trg}	BH / HS	

The ices.ar was found to be the least robust compared to any other tested reference point systems (Figure 7). For low and medium productivity stocks, the risk of falling below either of the two B_{lim} types was substantially higher, yield and SSB were on average lower. By contrast, the ices.ar was among the more precautionary reference point systems for high productivity stock. Similarly, the use of direct estimates F_{MSY} as F_{trg} in fmsy.bh performed generally poorer than the F_{brp} proxies in the WKREF candidates for low and medium productivity stocks, but also improved notably for high productivity stocks. Direct estimates fail in particular to achieve SSB levels at or below B_{MSY} for low and medium productivity species and the risk of

SSB falling below B_{lim} is above the 5% threshold is relatively high, in particular for medium productivity stocks. The fmsy.bh system performs comparably better for high productivity stocks.

Except for the ices.ar, all tested systems were robust to risk of SSB falling below B_{lim} for low productivity species (Figure 7). For medium productivity species, the WKREF fsb.bh was the only candidate system that was fully compliant with the PA for Type 1 B_{lim}, whereas for Type 2 B_{lim}, this also included the nz.bh and WKREF systems. For high productivity stocks, the best performing systems in terms of risk are fspr.bh, fspr.hs and fb.bh, while the nz.bh performed poorly with respected to the PA. This can be explained in that $F_{B\%}$ tends to be notable smaller than its equivalent $F_{spr\%}$ when the production function is based on Beverton-Holt SRR but equal for a Hockey-Stock. Therefore, the specifications fb.hs led to consistently higher F_{brp} (i.e. proxies (i.e. $F_{B30} = F_{SPR30}$), which then led poorer performance in the robustness tests.

Among the WKREF candidates, differences in long term yields are small for low and medium productivity species, with all medians exceeding the 95% MSY threshold and generally low yield variation among stocks. For high productive species largest median yields are attained with fspr.hs. The results indicate that fspr.bh, fsb.bh and fspr.hs lead to median SBB levels at or above $B_{MSY.}$ The exception is fspr.bh, which was generally the least robust of the tested WKREF candidates (Figure 7).

With respect to taxonomic orders, the WKREF candidates performed particular well for pleuronectiformes (flatfishes), which fells within medium productivity group (Figure 8). Pleuronectiformes showed negligible risk of SSB falling below B_{lim}, long-term yields at or above MSY and median SSB at B_{MSY}. With respect to yield and B_{MSY}, similar good performance was achieved for gadoids although some stocks are associated with higher risk to fall B_{lim}. Stocks of the order Clupeiformes showed a similar high risk profile as the high productivity stocks, but generally better performance in terms yield. Maintaining stock levels close to B_{MSY} was only achieved with fb.bh, fmsy.bh and current the ices.ar, albeit the latter with larger variation.

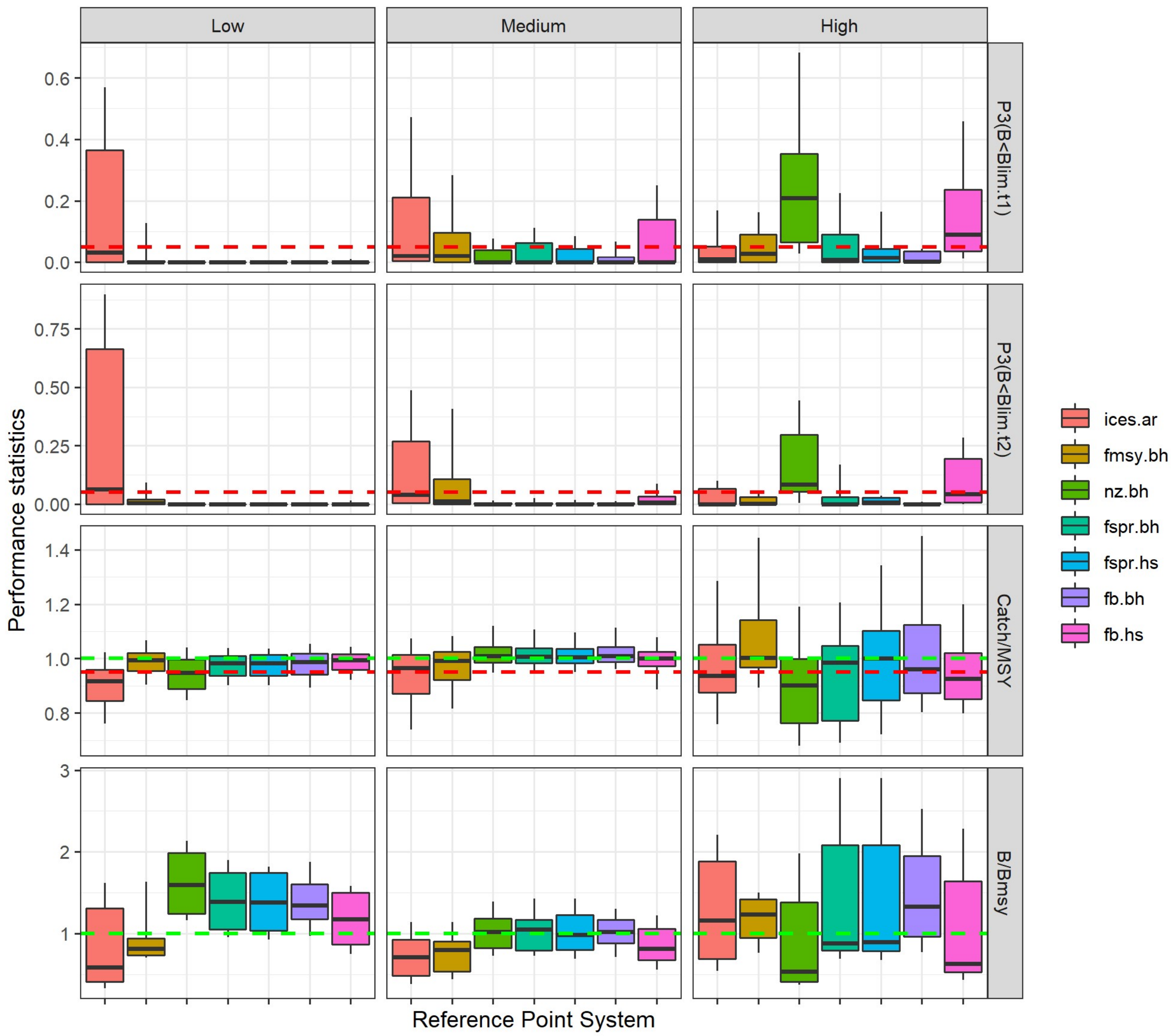


Figure 7: Results of robustness tests of evaluate reference point systems, showing the type 3 risk probabilities (P3) of SSB falling below B_{lim} of Type 1 (top row) and Type 2 (2nd row), the median long term yield relative the median long-term obtained at fixed "true" F_{MSY} (MSY) (3rd row) and the probabilities of SSB exceeding B_{tresh} at 80% B_{trg} (bottom row) for low, medium and high productivity stocks (columns). Green and red dashed lines denoting the target and limit thresholds, respectively. *ices.ar*: ICES Advice Rule; *fmsy.bh*: HCR with $F_{trg} = F_{MSY}$ and $B_{trigger} = 0.8\ B_{MSY}$; *nz.bh*: New Zealand Harvest Standard; *fspr.bh*/*.hs*: WKREF1 candidate based on $F_{SPR\%}$ and *fspr.bh*/*.hs:* WKREF1 candidate based on $F_{B\%}$, where *.bh* and *.hs* denotes if a fitted Beverton-Holt or Hockey-Stick was used, respectively (See Table 3 for details).

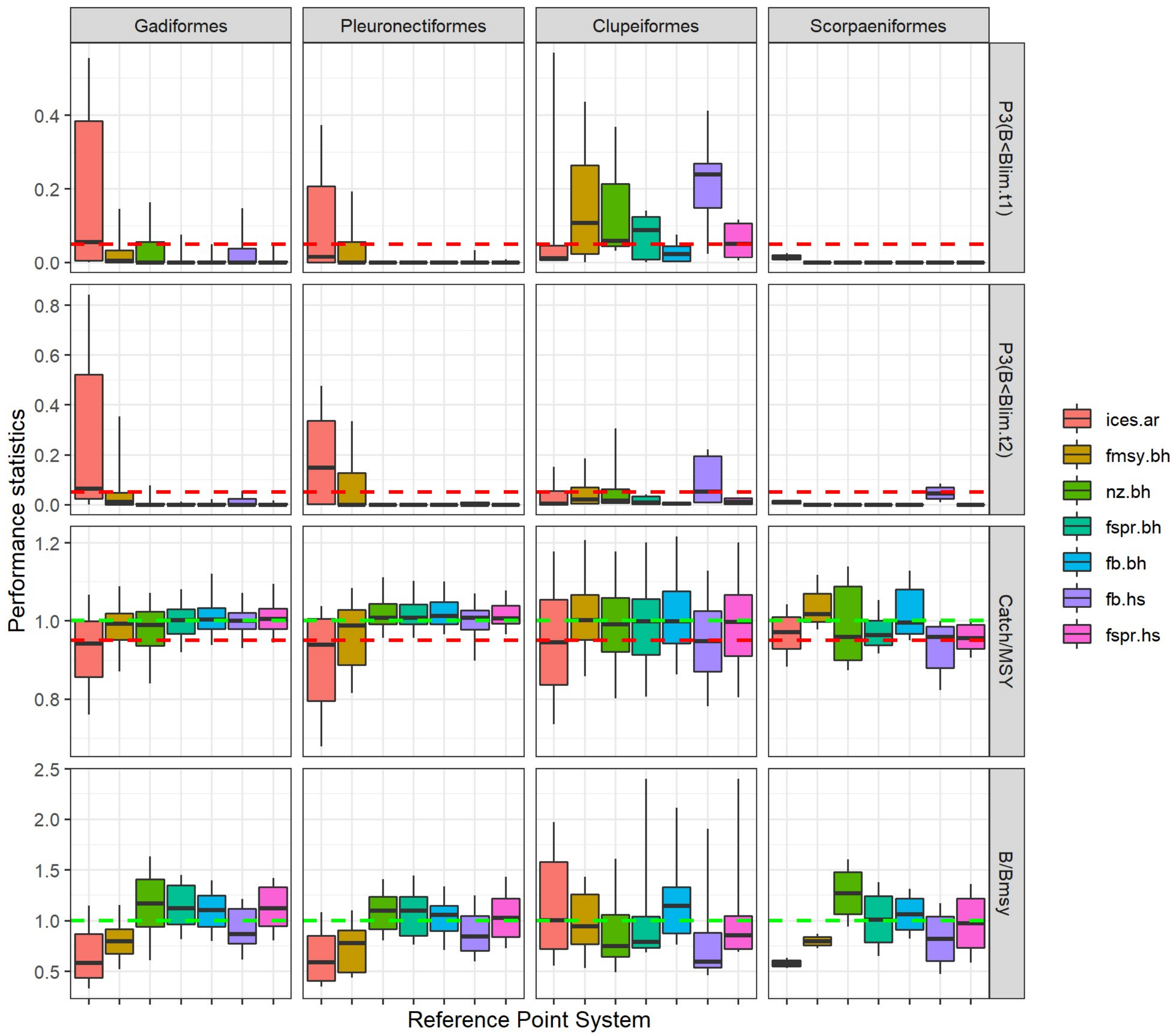


Figure 8: Results of robustness tests of evaluated reference point systems, showing the type 3 risk probabilities (P3) of SSB falling below B_{lim} of Type 1 (top row) and Type 2 (2nd row), the median long term yield relative the median long-term obtained at fixed "true" F_{MSY} (MSY) (3rd row) and the probabilities of SSB exceeding B_{tresh} at 80% B_{trg} (bottom row) for stock of four selected taxonomic orders, (columns). Green and red dashed lines denoting the target and limit thresholds, respectively. *ices.ar*: ICES Advice Rule; *fmsy.bh*: HCR with $F_{trg} = F_{MSY}$ and $B_{trigger} = 0.8\ B_{MSY}$; *nz.bh*: New Zealand Harvest Standard; *Fspr.bh*/*.hs*: WKREF1 candidate based on $F_{SPR\%}$ and *Fspr.bh*/*.hs:* WKREF1 candidate based on $F_{B\%}$, where *.bh* and *.hs* denotes if a fitted Beverton-Holt or Hockey-Stick was used, respectively (See Table 3 for details)

5.3. Invoking the precautionary fishing mortality target $F_{P0.5}$

Based on the SRR of OM we estimated B_{lim} for Type 1 and 2 and used the bisection function in FLFlasher to determine $\mathbf{F_{P.05}}$. As shown in Figure 7, the $\mathbf{F_{P.05}} < F_{SPR\%}$ was invoked for 16% for Type 1 B_{lim} and 6% for Type 2 B_{lim}. $\mathbf{F_{P.05}} < F_{B\%}$ was invoked for 8% of the stocks when using Type 1 B_{lim} but it was never invoked for Type

2 B_{lim}. High and medium productivity stocks were similarly likely to invoke $\mathbf{F_{P.05}}$ rule for Type 1 B_{lim}, whereas this was reduced for medium productivity stocks when Type 2 B_{lim} was used. In total only 10 stocks invoked $\mathbf{F_{P.05}}$ for any of the B_{lim} and F_{brp} combinations. These included six of the herring stocks.

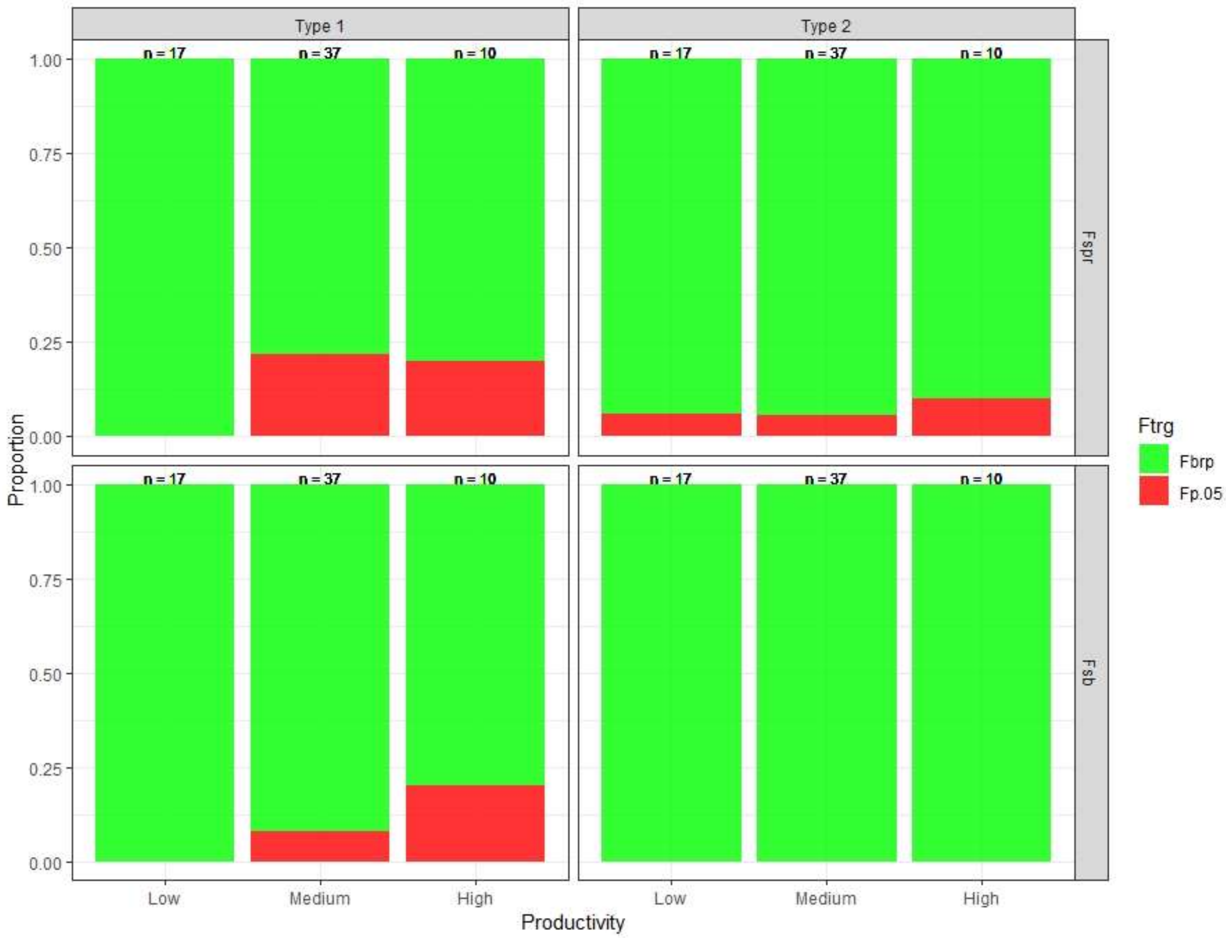


Figure 9: Proportion of stocks triggering $F_{P.05}$ for the different category of productivity when using $F_{SPR\%}$ or $F_{B\%}$

6. Recommendations

The results of both the self-test and robustness test clearly highlights the need to consider the stock's biological and productivity for setting reference points. Based on these results the following guidelines for setting reference points for category 1 stocks assessed by ICES are proposed according to productivity category (Table 4):

Table 4: Guidelines for deriving target and trigger reference points in the newly proposed ICES system. The Type 1 and 2 approaches can be used for all stocks to derive B_{lim}. SRR: Stock-recruitment recruitment relationship; BH: Beverton and Holt; HS: Hockey-Stick.

	Productivity	F_{trg}	B_{trg}	$B_{trigger}$	SRR
SPR%	Low	F_{spr40}	B_{spr40}	0.8 x B_{trg}	BH/HS
	Medium	F_{spr40}	B_{spr40}	0.8 x B_{trg}	BH/HS
	High	F_{spr50}	B_{spr50}	1x B_{trg}	BH/HS
B%	Low	F_{B35}	B_{35}	0.8x B_{trg}	BH
	Medium	F_{B35}	B_{35}	0.8x B_{trg}	BH
	High	F_{B40}	B_{40}	1x B_{trg}	BH

Low productive species $F_{SPR40\%}$ with stock and recruitment modelled as BH or HS fulfils both the PA and the MSY criteria and is proposed as candidate for the future ICES system to derive TRP. $F_{B35\%}$ with stock and recruitment modelled as BH fulfils both the PA and the MSY criteria and is proposed as candidates for the future ICES system to derive TRP. B_{lim} can be derived as the newly proposed Type 1 or Type 2, B_{trg} is the SSB that corresponds to $F_{SPR40\%}$ or $F_{B35\%}$ and $B_{trigger}$ is set at 0.8 B_{trg}.

Medium productive species $F_{SPR40\%}$ with stock and recruitment modelled as Beverton-Holt or Hockey-Stick SRR fulfils both the PA and the MSY criteria and is proposed as candidates for the future ICES system to derive TRP. $F_{B35\%}$ in combination with a Beverton-Holt SRR fulfils both the PA and the MSY criteria and is proposed as candidates for the future ICES system to derive TRP. B_{lim} can be derived as the newly proposed Type 1 or Type 2, B_{trg} is the SSB that corresponds to $F_{SPR40\%}$ or $F_{B35\%}$ and $B_{trigger}$ is set at 0.8 B_{trg}.

High productive species $F_{SPR50\%}$ with stock and recruitment modelled as BH or HS fulfils both the PA and the MSY criteria and is proposed as candidates for the future ICES system to derive TRP. $F_{B40\%}$ with stock and recruitment modelled as BH fulfils both the PA and the MSY criteria and is proposed as candidates for the future ICES system to derive TRP. B_{lim} can be derived as the newly proposed Type 1 or Type 2, B_{trg} is the SSB that corresponds to $F_{SPR50\%}$ or $F_{B40\%}$ and $B_{trigger}$ is set equal to B_{trg} or higher.

The Type 1 and 2 approaches can be used for all stocks to derive $\mathbf{B_{lim}}$ where Type 1 relies on the existence of a discernible relationship between stock and recruitment in that the data show contrast and a breakpoint is clearly defined. The $F_{B\%}$ guidelines should not be used in combination with Hockey-Stick SRR. In all cases it is recommended to estimate $\mathbf{F_{P.05}}$ although with the exception of herring, the newly proposed set of reference points should very rarely trigger $\mathbf{F_{P.05}}$.

References

Beverton, R. 1998. Fish, Fact and Fantasy: a Long View. Reviews in Fish Biology and Fisheries, 8: 229–249. https://doi.org/10.1023/A:1008888411100.

Carruthers, T., Kell, L., and Palma, C. 2017. Accounting for uncertainty due to data processing in virtual population analysis using Bayesian multiple imputation. Canadian Journal of Fisheries and Aquatic Sciences, 75: 883–896. NRC Research Press. https://doi.org/10.1139/cjfas-2017-0165.

de Bruyn, P., Murua, H., Aranda, M., 2013. The precautionary approach to fisheries management: how this is taken into account by Tuna regional fisheries management organisations (RFMOs). Mar. Policy 38, 397–406. http://dx.doi.org/10.1016/j.marpol.2012.06.019.

Deroba, J. J., Butterworth, D. S., Methot Jr, R. D., De Oliveira, J. A. A., Fernandez, C., Nielsen, A., Cadrin, S. X., *et al.* 2015. Simulation testing the robustness of stock assessment models to error: some results from the ICES strategic initiative on stock assessment methods. ICES Journal of Marine Science, 72: 19–30. https://doi.org/10.1093/icesjms/fst237.

DFO. 2009. A fishery decision-making framework incorporating the precautionary approach. https://www.dfo-mpo.gc.ca/reports-rapports/regs/sff-cpd/precaution-eng.htm.

FAO. 1995. Code of Conduct for Responsible Fisheries. FAO Technical Guidelines for Responsible Fisheries, 4 (Suppl.: 1–112. Rome.

FAO. 2001. Second technical consultation on the suitability of the CITES criteria for listing commercially exploited aquatic species. Windhoek, Namibia. https://www.fao.org/3/y1455e/y1455e.htm.

FAO. 2020. The State of World Fisheries and Aquaculture (SOFIA). Rome. https://www.fao.org/documents/card/en/c/ca9229en/.

Hilborn, R. 2010. Pretty good yield and exploited fishes. Marine Policy, 34: 193–196. http://www.sciencedirect.com/science/article/B6VCD-4W99NT8-

1/2/06cc5f8a2afa7621460cea7fd3a5a898.

Hillary, R. M., Preece, A. L., Davies, C. R., Kurota, H., Sakai, O., Itoh, T., Parma, A. M., *et al.* 2015. A scientific alternative to moratoria for rebuilding depleted international tuna stocks. Fish and Fisheries, 17: 469–482. Wiley/Blackwell (10.1111). https://doi.org/10.1111/faf.12121.

Horbowy, J., and Luzeńczyk, A. 2012. The estimation and robustness of F MSY and alternative fishing mortality reference points associated with high long-term yield. Canadian Journal of Fisheries and Aquatic Sciences, 69: 1468–1480.

Hordyk, A. R., Huynh, Q. C., and Carruthers, T. R. 2019. Misspecification in stock assessments: Common uncertainties and asymmetric risks. Fish and Fisheries, 20: 888–902.

ICES. 2014. Report of the Workshop to consider reference points for all stocks (WKMSYREF2). ICESCM 2014/ACOM, 47: 99.

ICES. 2020. The third Workshop on Guidelines for Management Strategy Evaluations (WKGMSE3). ICES Scientific Reports, 2: 1–112.

ICES. 2021a. Workshop on guidelines and methods for the evaluation of rebuilding plans (WKREBUILD). ICES Scientific Reports, 2: 1–79.

ICES. 2021b. Workshop of Fisheries Management Reference Points in a Changing Environment (WKRPChange). ICES Scientific Reports, 3: 39.

ICES. 2021c. ICES fisheries management reference points for category 1 and 2 stocks; Technical Guidelines.

ICES. 2021d. Benchmark Workshop on the development of MSY advice for category 3 stocks using Surplus Production Model in Continuous Time; SPiCT (WKMSYSPiCT). ICES Scientific Reports, 3: 1–317.

Kell, L.T., Mosqueira, I., Grosjean, P., Fromentin, J., Garcia, D., Hillary, R., Jardim, E., Mardle, S., Pastoors, M.A., Poos, J.J., Scott, F., Scott, R.D., 2007b. FLR : an open-source framework for the evaluation and development of management strategies. ICES J. Mar. Sci. 64, 640–646.

Kell, L. T., Nash, R. D. M., Dickey-Collas, M., Mosqueira, I., and Szuwalski, C. 2016. Is spawning stock biomass a robust proxy for reproductive potential? Fish and Fisheries, 17: 596–616.

Lassen, H., Kelly, C., and Sissenwine, M. 2014. ICES advisory framework 1977–2012: from Fmax to precautionary approach and beyond. ICES Journal of Marine Science, 71: 166–172. https://doi.org/10.1093/icesjms/fst146.

Mace, P. M. 1994. Relationships between common biological reference points used as thresholds and targets of fisheries management strategies. Canadian Journal of Fisheries and Aquatic Sciences, 51: 110–122.

Mace, P. M. 2001. A new role for MSY in single-species and ecosystem approaches to fisheries stock assessment and management. Fish and Fisheries, 2: 2–32.

McAllister, M. K., Pikitch, E. K., and Babcock, E. A. 2001. Using demographic methods to construct Bayesian priors for the intrinsic rate of increase in the Schaefer model and implications for stock rebuilding. Canadian Journal of Fisheries and Aquatic Sciences, 58: 1871–1890. http://www.nrcresearchpress.com/doi/abs/10.1139/f01-114.

Methot, R. D., and Wetzel, C. R. 2013. Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management. Fisheries Research, 142: 86–99. http://www.sciencedirect.com/science/article/pii/S0165783612003293.

Mildenberger, T. K., Berg, C. W., Kokkalis, A., Hordyk, A. R., Wetzel, C., Jacobsen, N. S., Punt, A. E., *et al.* 2021. Implementing the precautionary approach into fisheries management: Biomass reference points and uncertainty buffers. Fish and Fisheries: 1–20.

Miller, T. J., and Brooks, E. N. 2021. Steepness is a slippery slope. Fish and Fisheries, 22: 634–645.

Musick, J. A. 1999. Criteria to Define Extinction Risk in Marine Fishes. Fisheries, 24: 6–14.

Myers, R. A., Bowen, K. G., and Barrowman, N. J. 1999. Maximum reproductive rate of fish at low population sizes. Canadian Journal of Fisheries and Aquatic Sciences, 56: 2404–2419.

New Zealand Ministry of Fisheries. 2008. Harvest Strategy Standard for New Zealand Fisheries. https://fs.fish.govt.nz/Page.aspx?pk=113&dk=16543.

New Zealand Ministry of Fisheries. 2011. Operational Guidelines for New Zealand's Harvest Strategy Standard. New Zealand Government. 80 pp. https://fs.fish.govt.nz/Doc/22847/Operational_Guidelines_for_HSS_rev_1_Jun_2011.pdf.ashx.

Nielsen, A., and Berg, C. W. 2014. Estimation of time-varying selectivity in stock assessments using state-space models. Fisheries Research, 158: 96–101. http://www.sciencedirect.com/science/article/pii/S0165783614000228.

Patterson, K., Cook, R., Darby, C., Gavaris, S., Kell, L., Lewy, P., Mesnil, B., *et al.* 2001. Estimating uncertainty in fish stock assessment and forecasting. Fish and Fisheries, 2: 125–157.

Punt, A. E., Smith, A. D. M., Smith, D. C., Tuck, G. N., and Klaer, N. L. 2013. Selecting relative abundance proxies for BMSY and BMEY. ICES Journal of Marine Science, 71: 469–483. http://icesjms.oxfordjournals.org/content/71/3/469.short (Accessed 30 May 2014).

Ralston, S., Punt, A. E., Hamel, O. S., Devore, J. D., and Conser, R. J. 2011. A meta-analytic approach to quantifying scientific uncertainty in stock assessments. Fishery Bulletin, 109: 217–231.

Restrepo, V. R., Thompson, G. G., Mace, P. M., Gabriel, W. L., Low, L. L., MacCall, A. D., Methot, R. D., *et al.* 1998. Technical Guidance On the Use of Precautionary Approaches to Implementing National Standard 1 of the Magnuson-Stevens Fishery Conservation and Management Act. 56 pp. http://www.st.nmfs.noaa.gov/StockAssessment/documents/Tech-Guidelines.pdf.

Sharma, R., Winker, H., Levontin, P., Kell, L., Ovando, D., Palomares, M. L. D., Pinto, C., *et al.* 2021. Assessing the potential of catch-only models to inform on the state of global fisheries and the UN's SDGs. Sustainability (Switzerland), 13.

Shertzer, K. W., Prager, M. H., and Williams, E. H. 2008. A probability-based approach to setting annual catch levels. Fishery Bulletin, 106: 225–232.

Silvar-Viladomiu, P., Minto, C., Halouani, G., Batts, L., Brophy, D., Lordan, C., and Reid, D. G. 2021. Moving reference point goalposts and implications for fisheries sustainability. Fish and Fisheries, 22: 1345–1358. John Wiley & Sons, Ltd. https://doi.org/10.1111/faf.12591.

Simmonds, E. J., Portilla, E., Skagen, D., Beare, D., and Reid, D. G. 2010. Investigating agreement between different data sources using Bayesian state-space models: an application to estimating NE Atlantic mackerel catch and stock abundance. ICES Journal of Marine Science, 67: 1138–1153. http://icesjms.oxfordjournals.org/cgi/doi/10.1093/icesjms/fsq013.

Thorson, J. T., Jensen, O. P., and Hilborn, R. 2015. Probability of stochastic depletion: an easily interpreted diagnostic for assessment modelling and fisheries management. ICES Journal of Marine Science, 72: 428–435.

Thorson, J. T. 2020. Predicting recruitment density dependence and intrinsic growth rate for all fishes worldwide using a data-integrated life-history model. Fish and Fisheries, 21: 237–251. John Wiley & Sons, Ltd. https://doi.org/10.1111/faf.12427.

UN. 1995. United Nations Conference on Straddling Fish Stocks and Highly Migratory Fish Stocks. New York.

Appendix A:

FLSRTMB: Characterising stock productivity in FLR

Demographic information from FLStock objects can be used to construct an age-structured Leslie matrix **A** of the form:

$$\mathbf{A} = \begin{pmatrix} \phi_1 & \phi_2 & \phi_3 & \cdots & \phi_A \\ \theta_1 & 0 & 0 & 0 & 0 \\ 0 & \theta_2 & 0 & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & 0 & \theta_{T-1} & 0 \end{pmatrix} \quad \text{(B1)}$$

where ϕ_a is the average number of recruits expected to be produced by an adult female at age a and θ_a is the fraction of survivors at age, with T denoting the maximum age (plus group). The value of *r* is obtained from $\lambda = \exp(r)$, where λ is the dominant eigenvalue of **A** and *G*

Age-dependent survival calculated as $\theta_a = \exp(-M_a)$, where M_a is age-dependent natural mortality. The average number of recruits expected to be produced by an adult female at age t is expressed as:

$$\phi_t = \alpha w_a mat_a \quad \text{(B2)}$$

where α denotes the slope of the origin of the spawner-recruitment relationship (i.e. the ratio of recruits to spawner biomass at very low abundance) , w_a is the weight at age a, $mat_a$is the fraction of females that are mature at age a. For the calculation of the annual reproductive rate a first consider the BH-SSR of the form:

$$R = \frac{\alpha SB}{1+\beta SB} \quad \text{(B3)}$$

spr_0

where R is the number of recruits, S is the spawner biomass and β is the scaling parameter (Hilborn and Walters, 1992) . In contrast to alternative formulations of the BH-SSR, the parameter α can be directly interpreted as the slope in the origin of the S-R curve. We re-parameterized α as function of unfished spawner-biomass per recruit SPR_0 and the steepness parameter h of the spawner-recruitment relationship (Myers et al., 1999), such that:

$$\alpha = \frac{4h}{(1-h)} SPR_0^{-1} \quad \text{(B4)}$$

In cases where the quantities $W_{a,y}$, $Mat_{a,y}$ and $M_{a,y}$ varied annually, the averages of *r* and *G* across all years.

Appendix B:

FLSRTMB: Fitting conditioned Stock Recruitment Relationships (SRR) in FLR

Beverton-Holt SSR conditioning with prior information for steepness

The stock-recruitment relationship (SRR) was assumed to follow a Beverton and Holt model (BH-SRR) of the form

$$R_y = \frac{aSB_{y-a_{min}}}{b + SBB_{y-a_{min}}} e^{\epsilon_y - 0.5\sigma_r^2}$$

where R_y is the number of recruits in year y, SSB_{y-a_r} is the spawning biomass in year y minus minimum age a_{min} defined for the stock (typically age-0 or age-1). The recruitment deviation ϵ_t is assumed to be associated with a a first-order autocorrelation (AR1) process (Johnson et al. 2016; Simmonds et al. 2019), such that

$$\epsilon_y = \rho\epsilon_{y-1} + \delta_y\sqrt{1-\rho^2}$$

where ρ is the AR1 coeffient and $\delta_y \sim N(0, \sigma_r)$ determines variation in recruitment as a function of the recruitment standard deviation $sigma_r$.

The BH-SRR was fitted the recruitment R and SSB from FLStock objects using the FLR library FLSRTMB (Winker and Mosquiera; https://github.com/flr/FLSRTMB), which enables straight-forward integration of available prior information on the steepness s of the SSR from a recent meta-analysis (Thorson 2020).

For this purpose, the Beverton-Holt equation in FLSRTMB is re-parameterised as function of steepness s and annual unfished spawning biomass per-recruit SPR_0 (Mace and Doonan, 1988),

$$R_y = \frac{4sSB_{y-a_{min}}R0}{R_0 SPR_{0_y}(1-s) + SBB_{y-a_{min}}(5s-1)}$$

where steepness s is defined as the ratio of recruitment when SSB equals 20% of the unfished SSB_0 to the virgin recruitment R_0 at $SSB0$. A notable difference to the conventional parameterization is that SPR_{0_y} is treated as non-stationary, being function of annual quantities of $W_{a,y}$, $Mat_{a,y}$ and $M_{a,y}$. By way of using time-varying $SPR_{0,y}$, also takes into consideration the recent criticism by Miller and Brooks (2021) that specifying a set biological parameters to define a single time-invariant SPR_0 can be highly sensitive to reference estimation when using steepness values from meta-analysis.

The prior distribution for s is generated from truncated logit distributions ($TrunkLogit$) of the form

$$s = 0.2001 + 0.7999/\left(1 + exp\left(-s_{logit}\right)\right)$$

$$s \sim TrunkLogit\left(s_{logit}, \sigma_{logit}\right)$$

where s_{logit} and $sigma_{logit}$ correspond to the input of species-specific predictions for the distribution of s from the hierarchical taxonomic FishLife model (Thorson, 2020, https://github.com/James-Thorson-NOAA/FishLife), summarized in Table A1. The default prior is assuming an approximately uniform prior between 0.3 – 0.9, with a decreasing density (soft bounds) to the limits 0.2 and 1.0 (Figure. A1)

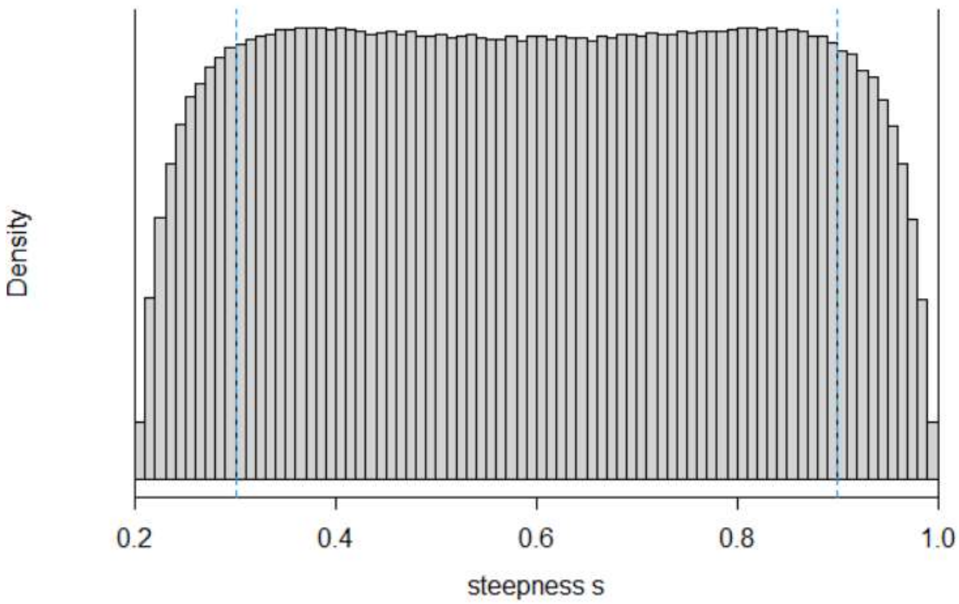


Figure A.1.1 Graphical illustration of default prior for estimating steepness s, with a mean of 0.6 and logit.sd = 1.5

The FLSRTMB estimates of $R0$ and s are then converted into the parameters a and b of the Beverton-Holt formulation in FLR, such that

$a = \frac{4sR_0SPR_0}{5sSPR_0-1}$ and $b = \frac{R_0SPR_0(1-s)}{5s-1}$

where the reference for SPR_0 to predict a and b was taken the average SPR_{0_y} across all years in the case of the OM.

A conditioned, continuous hockey-stick SSR

A new conditional Hockey-Stick formulation was developed and implemented in ’FLSRTMB`. The new Hockey-Stick is based on a continuous, quadratic hockey-stick (c.f. Barrowman and Myers), which is re-parameterised as a function of SPR_{0_y} and a “re-purposed” steepness parameter s^* given by

$$R_y = \frac{s^*}{2P_{lim}SPR0_y}\left(SSB_y + P_{lim}R_0SPR_{0_y}/s^* - \sqrt{\left(SSB_y - P_{lim}R_0SPR_{0,y}/s^*\right)^2}\right)$$

In addition, the parameter P_{lim} is introduced, which then determines the lower of the ratio B_{lim}/SSB_{0_y}, where $Blim$ corresponds to break point b of the segmented regression and SSB_{0_y} is allowed to be treated as non-stationary being a function of $SSB_{0_y} = R_0SPR_{0_y}$.

The break point b (B_{lim}) and slope a are given by

$b = P_{lim} * R_0SPR_{0,y}/s$ and $a = R_0/b$

In the chosen setting for FLSRTMB, the parameter s^* was bounded by a mostly uniform distribution between $0.2 > s^* \leq 1$, with soft bounds towards the limits to invoke a conditioned B_{lim} range of $0.1B_0 < B_{lim} < 0.3B_0$.

Table B1. List Species arranged by taxonomic order with FishLife (Thorson 2020) predictions for the recruitment standard deviation (σ_R), the auto-correlation coefficient (ρ), steepness (s) and the associated standard error (σ_S) on logit scale.

Species	Order	σ_R	ρ	s	σ_s
Argentina silus	Argentiniformes	0.69	0.38	0.52	1.14
Clupea harengus	Clupeiformes	0.67	0.32	0.58	0.26
Sardina pilchardus	Clupeiformes	0.49	0.50	0.77	0.60
Sprattus sprattus	Clupeiformes	0.70	0.31	0.80	0.67
Brosme brosme	Gadiformes	0.42	0.56	0.57	1.30
Gadus morhua	Gadiformes	0.53	0.39	0.79	0.22
Melanogrammus aeglefinus	Gadiformes	0.80	0.24	0.66	0.34
Merlangius merlangus	Gadiformes	0.64	0.31	0.71	0.43
Merluccius merluccius	Gadiformes	0.23	0.67	0.56	1.20
Micromesistius poutassou	Gadiformes	0.60	0.34	0.55	0.73
Molva molva	Gadiformes	0.38	0.56	0.53	1.33
Pollachius virens	Gadiformes	0.46	0.57	0.79	0.40
Pandalus borealis	Crustacian	0.28	0.27	0.84	0.30
Lophius piscatorius	Lophiiformes	0.30	0.88	0.92	1.28
Dicentrarchus labrax	Perciformes	0.34	0.75	0.90	1.93
Trachurus trachurus	Perciformes	0.53	0.47	0.75	0.87
Glyptocephalus cynoglossus	Pleuronectiformes	0.53	0.47	0.63	1.04
Lepidorhombus boscii	Pleuronectiformes	0.37	0.68	0.87	1.23
Lepidorhombus whiffiagonis	Pleuronectiformes	0.38	0.66	0.84	1.29
Pleuronectes platessa	Pleuronectiformes	0.48	0.58	0.82	0.40
Scophthalmus maximus	Pleuronectiformes	0.60	0.48	0.86	1.15
Solea solea	Pleuronectiformes	0.54	0.34	0.61	0.42
Scomber scombrus	Scombriformes	0.78	0.28	0.64	0.58
Sebastes norvegicus	Scorpaeniformes	0.56	0.61	0.58	0.96

WD2 A quick retrospective analysis on the estimation of reference points

Martin Pastoors, 12/1/2021

Aim: investigate the impacts of the length of the time series on estimated reference points.

Data: WKREF1 database

Two approaches:

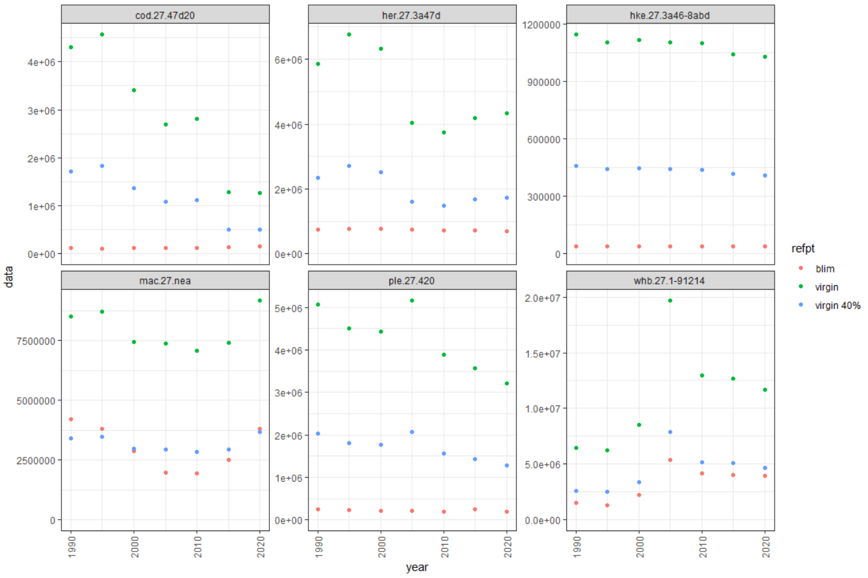
1. Just change the final year of the data from 1990 to 2020 in steps of 5 years
2. Just use a series of 20 years, also with final years from 1990 to 2020 in steps of 5 years

Examples: North Sea cod, North Sea herring, Northern hake, Mackerel, North Sea plaice, Blue whiting

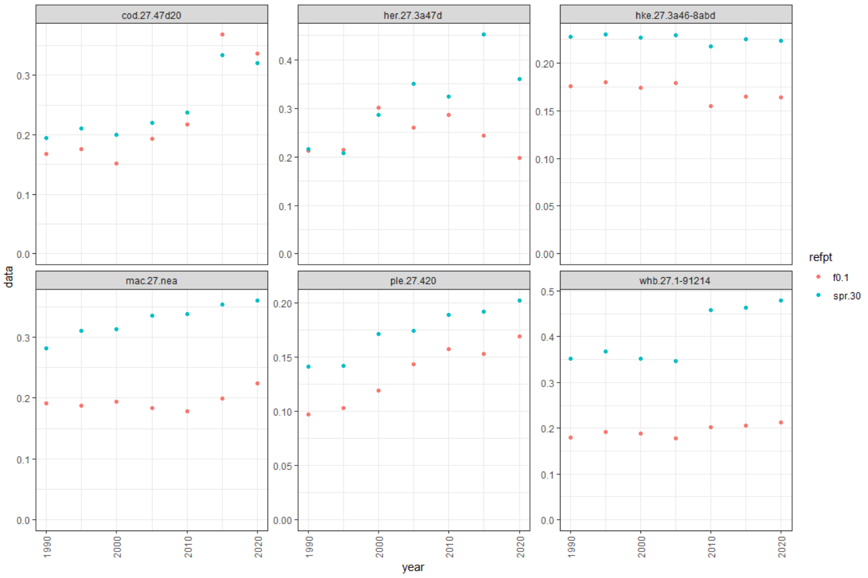
Reference points calculation based on segmented regression (see code in annex); also trial with Beverton-Holt

RETRO1; just changing the final year

Biomass

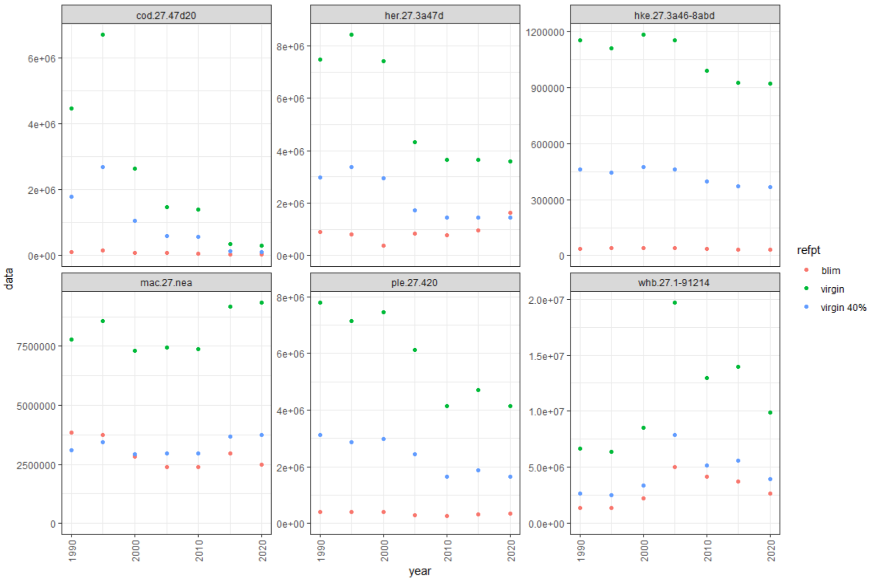


F (harvest)

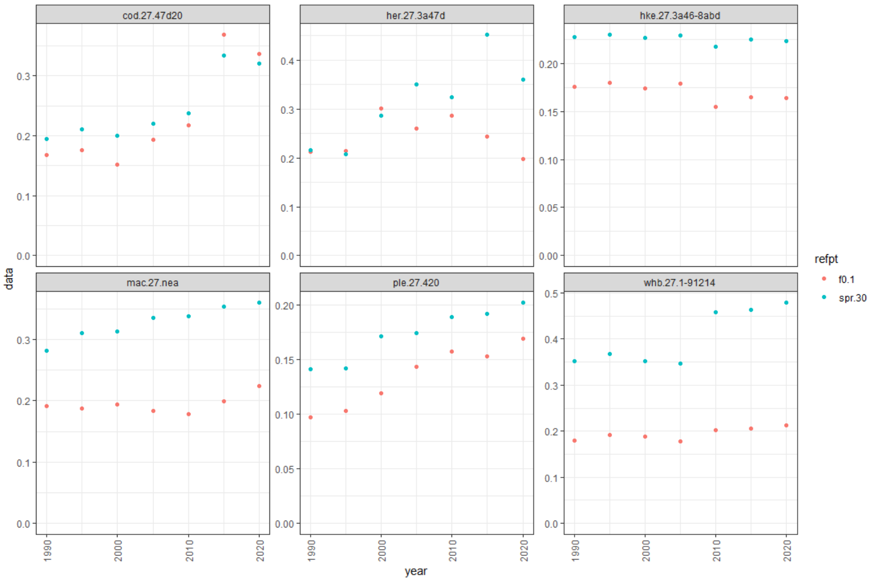


Retro 2: take 20 year time series only with the specified end years

Biomass (SSB)



F (harvest)



Conclusions:

- Final year and length of time series highly influential on estimates of biomass and F reference points
- Large variations in estimates of Bzero (and B40%) between different retro runs.
- E.g. North Sea cod between 1.2 Mt and 4.5 Mt; Blue whiting between 6 and 20 Mt.
- (Obviously) difference between using segmented regression and Beverton Holt SRR in estimated reference points.
- Overall: variability in reference points estimation due to specific set of data points included in the SRR data. Bzero very sensitive.

Retro1.r (only changing the last data year)

```
library(tidyverse)
library(FLCore)
library(FLBRP)

rm(list=ls())

load("bootstrap/data/ices.stks.n78.Rdata", verbose=T)

mystocks <- c("her.27.3a47d", "mac.27.nea", "whb.27.1-91214", "cod.27.47d20", "ple.27.420", "hke.27.3a46-8abd")

# 0. Create FLStocks objects with retro features; variability in final year
mystks <- FLStocks()
i <- 0
for (mystock in mystocks) {
 # for (myyear in 2020:1990) {
 for (myyear in c(2020,2015, 2010, 2005, 2000, 1995, 1990)) {
  i <- i+1
 mystks[[i]]  <- window(stks[[mystock]], end=myyear)
 mystks[[i]]@name <- paste (mystock, myyear)
 }
}

# 1. Fit segmented regression to derive breakpoint
mysr <- FLSRs(lapply(mystks, function(x) {
 return(fmle(as.FLSR(x,model=segreg)))
}))

# 2. create ices ref point objects with FLBRP
brp.ices = Map(function(x, y){
 brp = brp(FLBRP(x, y))
 brp@name = y@name
 brp
}, mystks, mysr)

# 3. Create data frame with reference points
df <- data.frame()
for (i in 1:length(brp.ices)) {
 df <-
 df %>%
 bind_rows(
  as.data.frame(brp.ices[[i]]@refpts) %>%
  bind_rows(data.frame(refpt="blim", quant="ssb", iter=factor(1), data=as.numeric(mysr[[i]]@params["b"]),
        stringsAsFactors = FALSE)) %>%
  mutate(stock = brp.ices[[i]]@name)
 )
}

# Add 40% Bvirgin
df <-
 bind_rows(
 df,
```

```
 df %>%
  filter(refpt == "virgin" & quant == "ssb") %>%
  mutate(
  data = 0.4*data,
  refpt = "virgin 40%")
 ) %>%
 tidyr::separate(stock, into=c("stock","endyear"), sep=" ") %>%
 mutate(year = as.numeric(endyear))


# 4. Plot data frame
df %>%
 filter(
 (refpt == "spr.30" & quant=="harvest") |
 (refpt == "f0.1" & quant=="harvest") ) %>%

 ggplot(aes(x=year, y=data)) +
 theme_bw() +
 theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
 geom_point(aes(colour=refpt)) +
 expand_limits(y=0) +
 facet_wrap(~stock, scales="free_y")


df %>%
 mutate(year = as.numeric(year)) %>%
 filter(
 (refpt == "blim" & quant == "ssb") |
 (refpt == "virgin 40%" & quant == "ssb") |
 (refpt == "virgin" & quant == "ssb")) %>%

 ggplot(aes(x=year, y=data)) +
 theme_bw() +
 theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
 geom_point(aes(colour=refpt)) +
 expand_limits(y=0) +
 facet_wrap(~stock, scales="free_y")
```

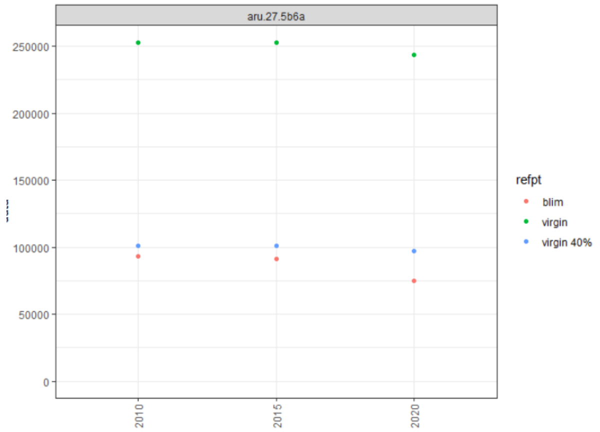
Retro2.r (only using 20 years in combination with a last data year)

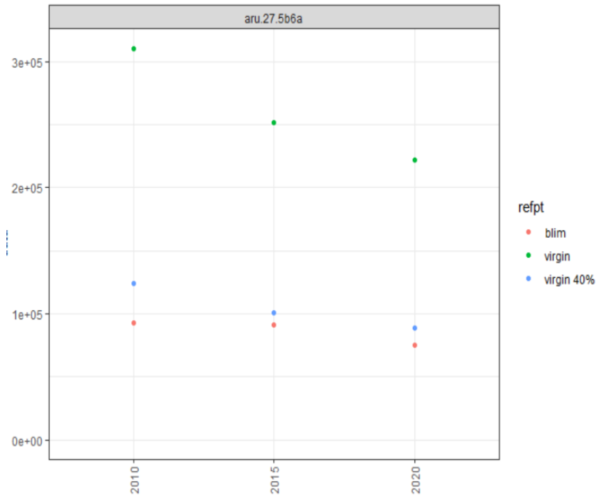
```
library(tidyverse)
library(FLCore)
library(FLBRP)

rm(list=ls())

load("bootstrap/data/ices.stks.n78.Rdata", verbose=T)

mystocks <- c("her.27.3a47d", "mac.27.nea", "whb.27.1-91214", "cod.27.47d20", "ple.27.420", "hke.27.3a46-8abd")

# 0. Create FLStocks objects with retro features; variability in final year
mystks <- FLStocks()
i <- 0
for (mystock in mystocks) {
```

```
 # for (myyear in 2020:1990) {
 for (myyear in c(2020,2015, 2010, 2005, 2000, 1995, 1990)) {
  i <- i+1
 mystks[[i]]  <- window(stks[[mystock]], end=myyear)
 mystks[[i]]@name <- paste (mystock, myyear)
 }
}

# 1. Fit segmented regression to derive breakpoint
mysr <- FLSRs(lapply(mystks, function(x) {
 return(fmle(as.FLSR(x,model=segreg)))
}))

# 2. create ices ref point objects with FLBRP
brp.ices = Map(function(x, y){
 brp = brp(FLBRP(x, y))
 brp@name = y@name
 brp
}, mystks, mysr)

# 3. Create data frame with reference points
df <- data.frame()
for (i in 1:length(brp.ices)) {
 df <-
 df %>%
 bind_rows(
  as.data.frame(brp.ices[[i]]@refpts) %>%
  bind_rows(data.frame(refpt="blim", quant="ssb", iter=factor(1), data=as.numeric(mysr[[i]]@params["b"]),
        stringsAsFactors = FALSE)) %>%
  mutate(stock = brp.ices[[i]]@name)
 )
}

# Add 40% Bvirgin
df <-
 bind_rows(
 df,
 df %>%
  filter(refpt == "virgin" & quant == "ssb") %>%
  mutate(
  data = 0.4*data,
  refpt = "virgin 40%")
 ) %>%
 tidyr::separate(stock, into=c("stock","endyear"), sep=" ") %>%
 mutate(year = as.numeric(endyear))


# 4. Plot data frame
df %>%
 filter(
 (refpt == "spr.30" & quant=="harvest") |
 (refpt == "f0.1" & quant=="harvest") ) %>%
```

```
 ggplot(aes(x=year, y=data)) +
 theme_bw() +
 theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
 geom_point(aes(colour=refpt)) +
 expand_limits(y=0) +
 facet_wrap(~stock, scales="free_y")


df %>%
 mutate(year = as.numeric(year)) %>%
 filter(
 (refpt == "blim" & quant == "ssb") |
 (refpt == "virgin 40%" & quant == "ssb") |
 (refpt == "virgin" & quant == "ssb")) %>%

 ggplot(aes(x=year, y=data)) +
 theme_bw() +
 theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
 geom_point(aes(colour=refpt)) +
 expand_limits(y=0) +
 facet_wrap(~stock, scales="free_y")
```

Comparison of Segmented regression (left) and BH (right)





WD 3 First approach on deriving biological reference points for black scabbardfish NE Atlantic stock components

Inês Farias, Isabel Natário, Lucília Carvalho and Ivone Figueiredo

First approach on deriving biological reference points for black scabbardfish NE Atlantic stock components

Inês Farias, Isabel Natário, Lucília Carvalho and Ivone Figueiredo

Preamble

The black scabbardfish, *Aphanopus carbo*, is a widely distributed species with high economic importance for some European fleets, being commercially exploited in several areas.

This species presents complex spatial dynamics along the Northeast Atlantic. The current understanding of the species' life history is that one single population undergoes an ontogenic migration along the Northeast Atlantic basin (Farias et al., 2013; ICES, 2021b). The only known spawning grounds are located in Madeira and Canary Islands. It is admitted that juveniles are born in those areas and later migrate northwards to Iceland, the Faroe Islands, and the West of the British Isles. After spending a few years in these grounds, individuals move southwards, namely to mainland Portugal (ICES 27.9.a) where they remain a few more years, after which they migrate to the spawning areas.

Given the complex spatial dynamics of the species along the Northeast Atlantic, the modelling procedure adopted and benchmarked by ICES (ICES, 2015) to assess its abundance takes into account the spatial aspects of the species dynamics, particularly the fact that ontogenic stages occur in different geographical areas linked through migration (Figure 1).

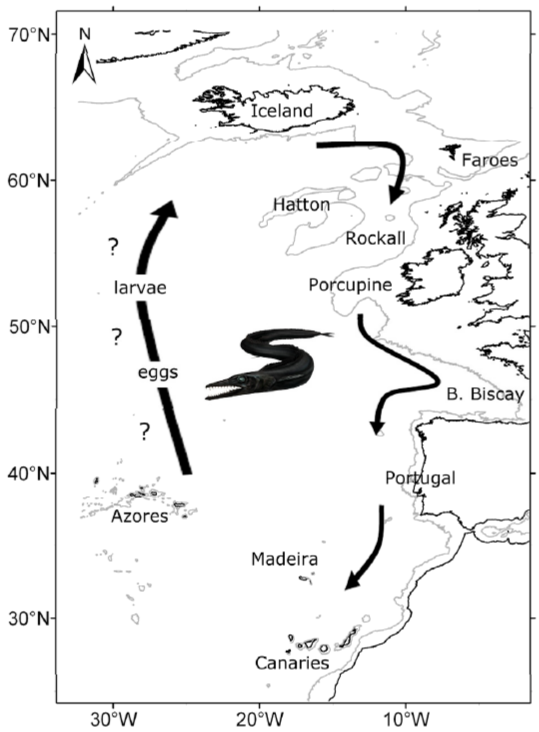


Figure 1. Map of the NE Atlantic representing the black scabbardfish's hypothetical migratory cycle. The 100 m depth contour is shown (Drawing of the species adapted from MARPROF, www.marprof.org). Source: Farias et al. (2013).

Brief overview of the assessment model

A unique stock is admitted for this species within the ICES area. Furthermore, based on the demographic structure of the population in NE Atlantic, two components are considered, encompassing the following areas: (i) the continental slope to the west of the British Isles, which is named the Northern component (BI); and (ii) the continental slope off mainland Portugal (P), which is the Southern component.

The scientific advice is provided based on outputs of the assessment model obtained for each component. The modelling procedure captures the spatial complexity of the species through the use of a stage-based model for each component and by including a migration process linking the two stock components. The model was benchmarked by ICES in 2014 (ICES, 2015) and since then the advice is being given using the ratio rule adopted for Category 3 ICES stocks.

Based on the known demographic structure of the population in the two ICES components, the population is partitioned into three broad size classes: (i) class C1 that is represented by juveniles with total length (TL) smaller than 70 cm; (ii) class C2 that encompasses the immature pre-adults (TL > 70 cm and < 103 cm); and (iii) class C3 that includes the immature adults (TL > 103 cm). The individuals in classes C2 and C3 are those exploited by the European fishing fleets and together they represent the exploitable fraction of the stock. No reliable information (e.g., surveys) are available to inform on the abundance of juveniles/recruits.

According to the methodology adopted, the abundance of the stock evolves at discrete time intervals. The observed seasonality on the monthly index of biomass (CPUE) of the stock components that reflects the underlying ontogenic migration process between the two components, lead to consider a six-month period as time unit. The periods considered are: first semester (corresponding to the months with lower CPUE: March to August) and second semester (corresponding to the months with higher CPUE: September to February).

The fitting of the model is done under the Bayesian paradigm. The model´s outputs are the posterior distributions of abundance by stock component (BI and P) and life stage (C2 and C3) and the posteriors of parameters ruling the life cycle and fishing activity.

Main objective

Recent simulation studies stressed the importance of considering the spatial structure of the stock on the development of management procedures that are more robust to those spatial complexities. In ICES, there is an increased interest in estimating reference points within the Management Strategy Evaluation (MSE) framework, which has the advantage of being tailored to the uncertainties of the stock (ICES, 2022).

The present contribution is a proposal to derive management reference points for the black scabbardfish NE Atlantic stock in terms of fishing mortality rate (F), considering the stochastic processes that describe the species spatio-temporal dynamics and its fishery under a Bayesian inference framework. The number of individuals in each stage and the state-space models for each component with the intrinsic migration process adjusted during the latest assessment of black scabbardfish (ICES, 2020b) will be used to project the stock status forward assuming a range of fishing mortality rates for the Northern component (BI) and for the Southern component (P).

Reference points - methodology

The reference points for black scabbardfish in NE Atlantic is derived from a self-consistent MSE simulation and following ICES technical guidelines from 1 March 2021 (ICES, 2021a). According to these guidelines, two types of reference points are defined for providing fisheries advice for category 1 stocks:

- precautionary approach (PA) reference points, used when assessing the state of stocks and their exploitation relative to the precautionary approach objectives;
- maximum sustainable yield (MSY) reference points, used in the advice rule applied by ICES and aimed at producing advice consistent with the objective of achieving MSY.

To derive the PA and MSY reference points, the status of black scabbardfish stock in NE Atlantic (bsf.27.nea) will be stochastically projected by considering different pairs of BI and P fishing mortality rates, each ranging from 0 to 1 (in steps of 0.05 and 0.1). Based on the adopted state-space models and migration linkage between BI and P, the inputs considered to initiate the simulation were:

- the latest output of the population status;
- the posterior distribution of the parameters from the latest assessment was used as the priors for the projection;
- the recruitment prior distribution in the projection is equal to the average of the entrance distribution medians for the last three years with a 20% CV.

The ICES guidelines suggest that the derivation of the MSY and PA reference points are to be calculated using two sources of error: stochastic variability in biology and assessment error. The present analysis only considers variability in recruitment as vital processes are estimated within the model.

The methodology adopted by ICES for deriving PA reference points was used (ICES, 2021a). This includes the following steps:

1. Identifying appropriate data;
2. Identifying stock type;
3. Estimating biomass limit reference points;
4. Deriving PA reference points from limit reference points;
5. Estimate MSY reference points.

1. Identifying appropriate data

In the model used for black scabbardfish no formal stock-recruitment relationship is considered. The annual recruitment, during the observation period, is estimated from the model without any consideration of the size of the spawning stock biomass.

Recruitment is equal to estimated annual entrances in BI of new individuals in class C2 and the spawning stock is the number of C3 individuals in P in the 2nd semester (spawning season occurs during the winter months). The timing of the recruitment is the 2nd semester and to accommodate the time required for juveniles to migrate from spawning to recruitment grounds a period of three years is considered. This time interval is consistent with assigned age of smallest individuals entering in BI. As a result, the recruitment is related with the projected spawning stock biomass from the previous three years.

The time-series of the stock and recruitment used includes the model estimates (median of the posteriori distribution) of S–R pairs covering the period from 2006-2015 (2nd semester) for Spawning biomass and 2009-2018 (2nd semester) for recruitment.

2. Identifying stock type

The relationship between the estimated spawning stock three years before the year y and recruitment in year y is shown in Figure 2.

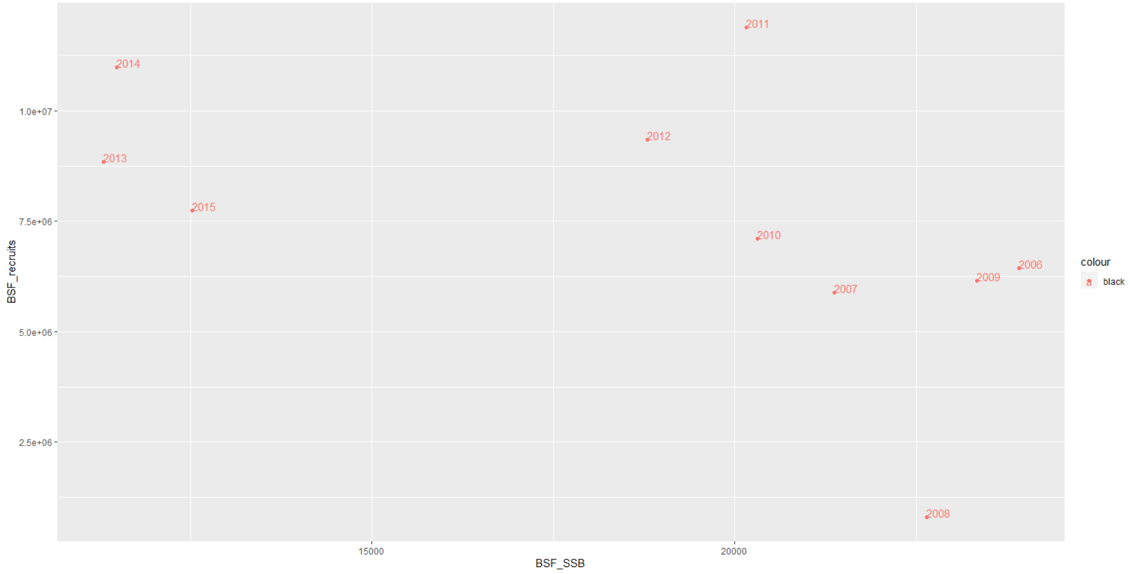


Figure 2. Black scabbardfish relationship between the estimated spawning stock (BSF_SSB, in tonnes) and recruitment (BSF_recruits, in number). Point labels are the years that correspond to SSB.

3. Estimating biomass limit reference points

The stock-recruitment (S-R) plot for black scabbardfish shows that there is relatively narrow dynamic range of SSB and there is no evidence of impaired recruitment (Figure 2). Following ICES guidelines this stock is categorized as Type 6 and, as a consequence, B_{lim} cannot be estimated from black scabbardfish S-R data.

4. Deriving PA reference points from limit reference points

B_{loss} is the lowest observed SSB from the base-line model during the period under analysis. For black scabbardfish, B_{loss} = SSB(2013) = 11.3 kt (Figure 2). According to ICES guidelines B_{loss} is suggested as a candidate for B_{pa}, because fishing pressure has been low since HR has historically fluctuated between 0.06 and 0.13 (Figure 3).

Furthermore, looking at the temporal evolution of the biomass trend (Figure 4) the selected year for B_{loss} is in agreement with ICES Guidelines requirements, according to which B_{loss} *should be taken from a stable part of the assessment and should not be from recent years if SSB is declining, since this could lead to a declining B_{lim} as the stock declines*.

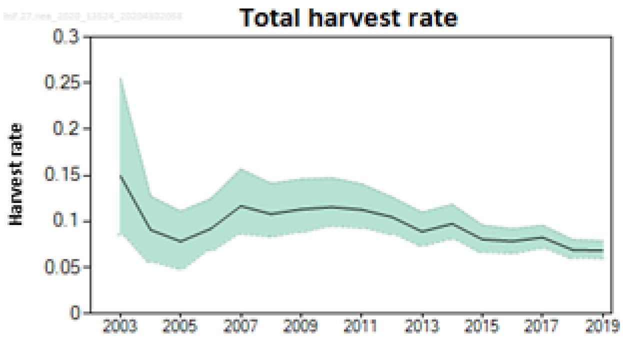


Figure 3. Black scabbardfish estimates of harvest rate (HR) between 2003 and 2019. Source: ICES, 2020a.

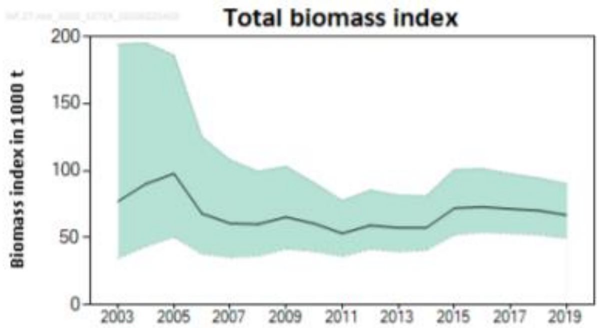


Figure 4. Black scabbardfish estimates of Total biomass (in thousand tonnes) between 2003 and 2019. Source: ICES, 2020a.

B_{lim}

According to ICES guidelines, given the fact that black scabbardfish S-R can be categorized as type 6, B_{pa} can be estimated but B_{lim} cannot. In such cases, also following ICES guidelines, a proxy for B_{lim} can be considered based on the inverse of the standard factor for calculating B_{pa} from B_{lim}, i.e., B_{lim} proxy equal to $B_{pa}/1.4$.

For black scabbardfish, B_{lim} is then set as $B_{pa}/1.4 = 8.07$ kt.

F_{lim}

The basis for defining F_{lim} (exploitation rate which leads SSB to B_{lim}) is that it corresponds to the fishing mortality rate (F) that in stochastic equilibrium will result in median SSB = B_{lim}, i.e. 50% probability of SSB being above or below B_{lim}.

To determine F_{lim}, the stock status of each component is projected forward during 50 semesters. In the present case, a series of pairs of BI and P fishing mortality rates (0 - 1 in steps of 0.05 and 0.1) are essayed by projecting the population forward 50 semesters.

The strategy followed to project the population forward under different pairs of fishing mortality rates uses the population model adopted for black scabbardfish in NE Atlantic and the recruitment values randomly selected from the posterior distribution of the parameters as the posteriori distribution for the projection. The recruitment distribution has a mean that is the average of the three previous recruitment posterior distributions and a CV of 20%. For each component and pair of BI and P fishing mortality rates, a total of 2 million recruitment particles are randomly selected in each semester.

F_{lim} is set as the pair of BI and P fishing mortality rates that in equilibrium, which corresponds to the *posteriori* distributions of the projected population, gives a 50% probability of SSB > B_{lim}. The results from the long-term simulations (50 semesters) are shown in Figure 5.

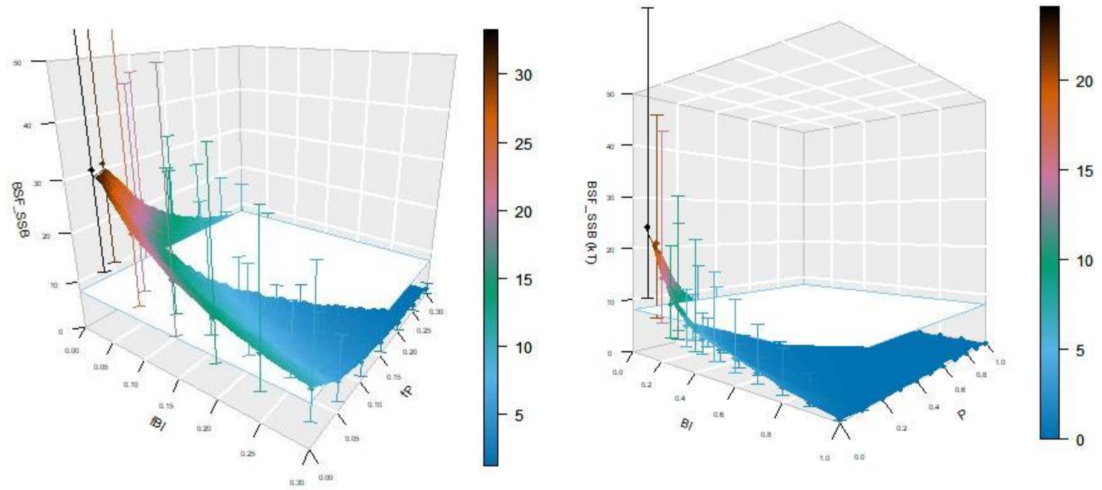


Figure 5. Black scabbardfish SSB (median and 95% credible interval) as function of the pair of BI and P fishing mortality rates (fBI and fP, respectively). Left panel represents fishing mortality rates varying from 0 to 0.3 in steps of 0.05; right panel represents fishing mortality rates varying from 0 to 1 in steps of 0.1. The white rectangle represents the B_{lim} threshold (B_{lim} = 8.07 kt)

The pairs of BI and P fishing mortality rates (fBI and fP, respectively) corresponding to a probability equal or greater than 50% of SSB > B_{lim}, i.e., the estimated pairs of fishing mortality that correspond to median SSB above the B_{lim} threshold, given B_{lim} = 8.07 kt, are presented in Table 1 (excluding the pairs where one of the fishing mortality rates is null).

Table 1. Black scabbardfish estimated pairs of fishing mortality rate that correspond to median SSB above the Blim threshold, and corresponding 5 % and 95 % quantiles.

fBI	fP	Median BSF_SSB (kt)	Q5	Q95
0.05	0.05	19.407741	13.77	25.551
0.10	0.05	13.855113	9.807	18.296
0.05	0.10	12.863898	9.901	15.578
0.15	0.05	10.621562	7.509	14.056
0.05	0.15	9.214900	7.468	10.671
0.10	0.10	9.112051	7.01	11.031
0.20	0.05	8.495535	6	11.265

The pair with median SSB above B_{lim} and highest concentration of points around that median is (fBI, fP) = (0.05, 0.15).

5. Estimate MSY reference points

F_{MSY}

F_{MSY}, in the present case, corresponds to the pair of Bl and P fishing mortality rates expected to give maximum sustainable yield in the long term. ICES defines yield as the catch above the minimum catch/conservation size. According to ICES guidelines, F_{MSY} calculation should be based on the median of the yield distribution. Median is a robust estimator that avoids undesirable properties of the mean for highly skewed yield distributions.

The F_{MSY} for black scabbardfish stock is calculated based on a yield surface, which is constructed using the simulated annual total catches (sum of catches in Bl and P) that correspond to the posterior distribution at the 50th iteration and considering the pairs of fishing mortality rate in Bl and in P (Figure 6).

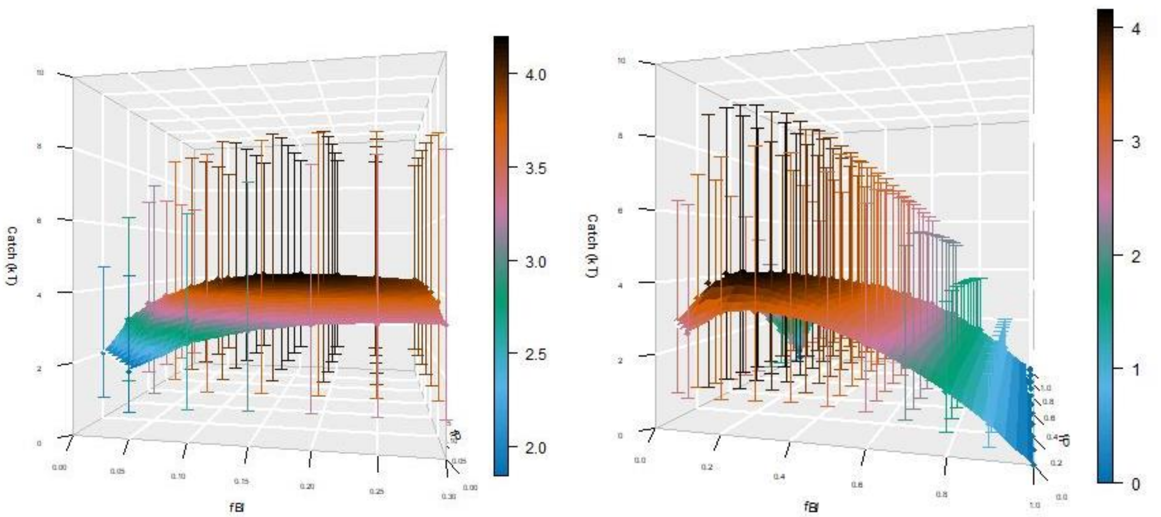


Figure 6. Black scabbardfish catch curve (median and 95% credible interval) under different pairs of Bl and P fishing mortality rates (fBl and fP, respectively). Left panel represents fishing mortality rates varying from 0 to 0.3 in steps of 0.05; right panel represents fishing mortality rates varying from 0 to 1 in steps of 0.1.

The F_{MSY} corresponding to the pairs of Bl and P fishing mortality rates (fBl and fP, respectively) that maximize the yield surface of the median, which is median = 4.20 kt (Q10 = 3.599 and Q90 = 4.802), are fBl = 0.15 and fP = 0.25.

MSY $B_{trigger}$

MSY $B_{trigger}$ corresponds to a lower bound of the expected range of SSB when the stock is fished at F_{MSY}. However, as black scabbardfish has not been managed according to F_{MSY} for more than 5 years, following ICES guidelines it is recommended that MSY $B_{trigger}$ be set as B_{pa}.
In the present case, as mentioned above, B_{loss} is a candidate for B_{pa}. Hence, $B_{pa} = B_{loss}$ = SSB(2013) = 11.3 kt.

F_{pa}

F_{pa} is interpreted as an exploitation rate reference point below which exploitation is considered to be sustainable, having accounted for estimation uncertainty. Important to remark that to ensure consistency between the precautionary and the MSY frameworks, ICES guidelines state that F_{MSY} is not allowed to be above F_{pa}; therefore, if the F_{MSY} value is above F_{pa}, F_{MSY} is reduced to F_{pa}.

According to ICES technical guidelines, F_{MSY} is set such that the annual risk of SSB falling below B_{lim} does not exceed 5%. As before, the population is projected forward under different pairs of fishing mortality rates and simulations are done with MSY $B_{trigger}$ (which, in this case, corresponds to B_{pa}) implemented. F_{pa} (F_{pa} also designated as $F_{0.05}$) corresponds to the maximum F that leads to 5% annual risk of SSB < B_{lim} with both assessment error and the MSY $B_{trigger}$ implemented. If the previous precautionary criterion is not met, i.e., $F_{0.05}$ < F_{MSY}, F_{MSY} should be reduced to F_{pa}.

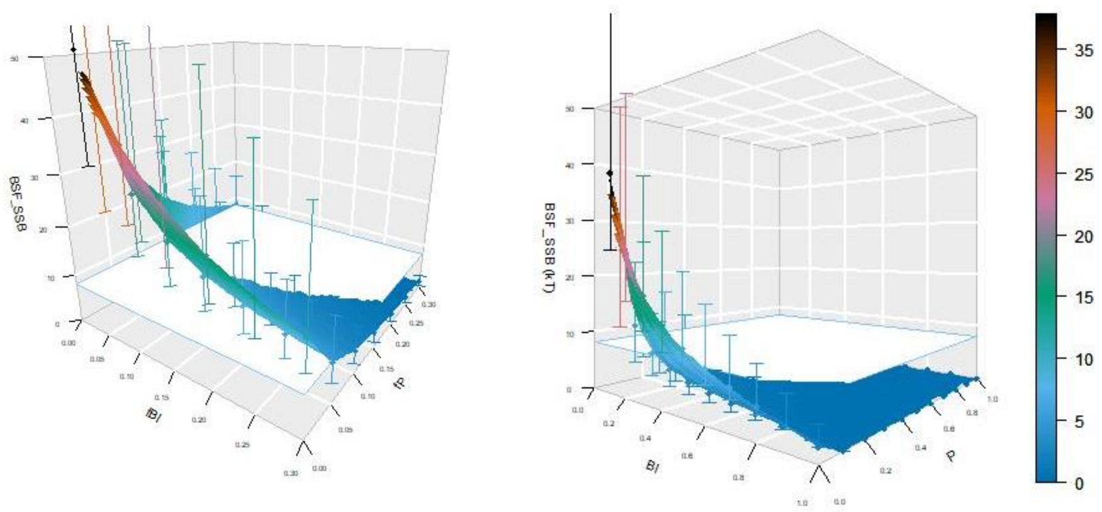


Figure 6. Black scabbardfish equilibrium SSB (Q95 and 95% credible interval) as function of the pair of BI and P fishing mortality rates (fBI and fP, respectively). Left panel represents fishing mortality rates varying in steps of 0.05; right panel represents fishing mortality rates varying in steps of 0.1. The white rectangle represents the B_{lim} threshold (B_{lim} = 8.07 kt).

The pairs of BI and P fishing mortality rates (fBI and fP, respectively) that correspond to the 95% quantile of SSB above the B_{lim} threshold, given B_{lim} = 8.07 kt, as shown in Figure 6, are presented in Table 2 (excluding the pairs where one of the fishing mortality rates is null).

Table 2. Black scabbardfish pairs of BI and P fishing mortality rates (fBI and fP, respectively) that correspond to the 95% quantile of SSB above the B_{lim} threshold (B_{lim} = 8.07 kt).

fBI	fP	Q95 BSF_SSB (kt)
0.05	0.05	25.550722
0.10	0.05	18.296066
0.05	0.10	15.577936
0.15	0.05	14.055516
0.20	0.05	11.264972
0.10	0.10	11.031017
0.05	0.15	10.670824
0.25	0.05	9.280955
0.15	0.10	8.414322

The reference points derived for black scabbardfish in NE Atlantic are summarised in Table 3.

Table 3. Black scabbardfish summary of estimated reference points.

Framework	Reference point	Value	Technical basis
Precautionary approach (PA)	B_{lim}	8.07 kt	B_{pa} / 1.4
	B_{pa}	11.3 kt	B_{loss} = SSB(2013) because fishing pressure has been low since HR has historically fluctuated between 0.06 and 0.13
	Pair F_{lim} (BI,P)	(fBI, fP) = (0.05, 0.15)	Pair F_{lim} (BI,P) corresponding to 50% long-term probability of SSB > B_{lim} and lowest dispersion of points around the median
	Pair F_{pa} (BI,P)	(see table 2)	Pair F_{pa} (BI,P) with an annual 5% probability of SSB < B_{lim}
MSY approach	MSY $B_{trigger}$	11.3 kt	Set as B_{pa}, hence corresponds to B_{loss}
	Pair F_{MSY} (BI,P)	(fBI, fP) = (0.15, 0.25)	Pair F_{MSY} (BI,P) that maximizes the yield surface of the medians

References

Farias, I., Morales-Nin, B., Lorance, P., Figueiredo, I. 2013. Black scabbardfish, *Aphanopus carbo*, in the northeast Atlantic: distribution and hypothetical migratory cycle. Aquatic Living Resources 26, 333–342. https://doi.org/10.1051/alr/2013061.

ICES. 2015. Report of the Benchmark Workshop on Deep-sea Stocks (WKDEEP), 3–7 February 2014, Copenhagen, Denmark. ICES CM 2014/ACOM:44. 119 pp.

ICES. 2020a. ICES Advice 2020 – bsf.27.nea. Black scabbardfish (*Aphanopus carbo*) in subareas 1, 2, 4–8, 10, and 14, and divisions 3.a, 9.a, and 12.b (Northeast Atlantic and Arctic Ocean). ICES Advice on fishing opportunities, catch, and effort Northeast Atlantic ecoregion. Published 10 June 2020. https://doi.org/10.17895/ices.advice.5816.

ICES. 2020b. Working Group on the Biology and Assessment of Deep-sea Fisheries Resources (WGDEEP). ICES Scientific Reports. 2:38. 928pp. http://doi.org/10.17895/ices.pub.6015.

ICES. 2021a. ICES fisheries management reference points for category 1 and 2 stocks; Technical Guidelines. In Report of the ICES Advisory Committee, 2021. ICES Advice 2021, Section 16.4.3.1. https://doi.org/10.17895/ices.advice.7891.

ICES. 2021b. Working Group on the Biology and Assessment of Deep-sea Fisheries Resources (WGDEEP). ICES Scientific Reports. 3:47. 944 pp. http://doi.org/10.17895/ices.pub.8108.

ICES. 2022. Workshop on ICES reference points (WKREF1). ICES Scientific Reports. 4:2. 70 pp. http://doi.org/10.17895/ices.pub.9749.

WD4 Using as Surplus Production Model (SPM) for the long-term forecasts to estimate BRP for mackerel

Henrik Sparholt 09/01/2022

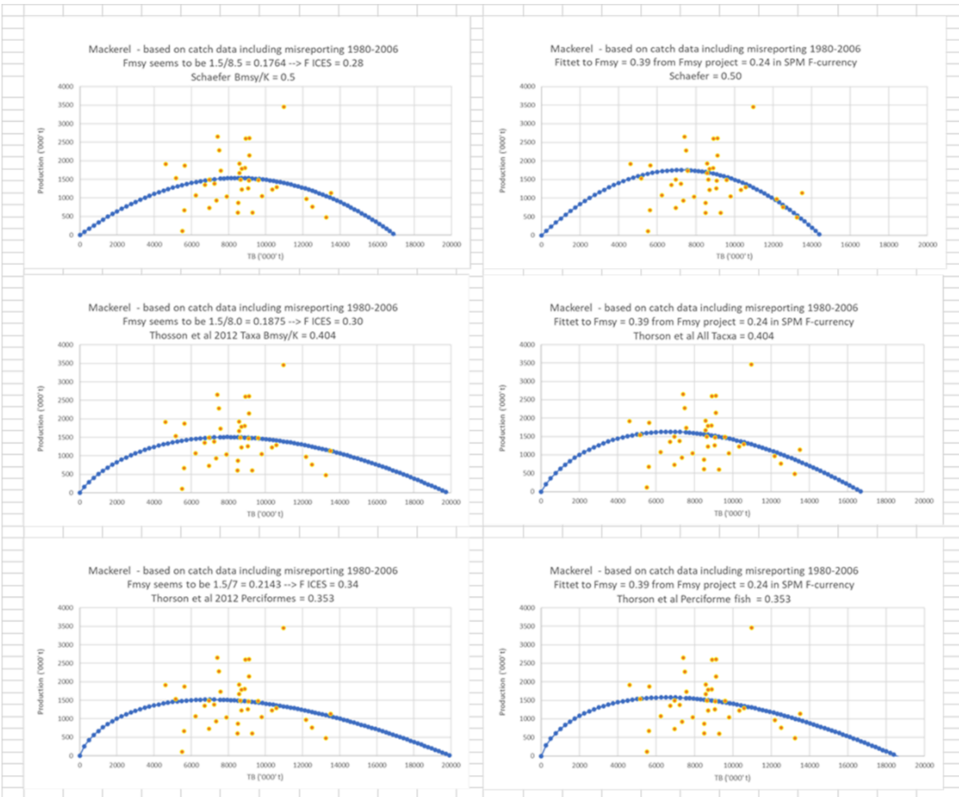


Figure 5.4. Alternative SPM models for NEA mackerel.

The first and most difficult stage in this solution is to find the best possible SPMs that are consistent with the available science for the stock in question. It involves checking for regime shifts and reviewing the relevant literature. We have included a case with Northeast Atlantic mackerel in Figure 3.1.1 and described in WK Doc for WKMSEMAC 2020 available on the SharePoint site for that group and for the present group (WKMSEMAC Doc HS1 and WKMSEMAC Doc HS2). For mackerel there was an issue of substantial misreporting (prior to 2006), that needed to be addressed. It can be noted that this misreporting issue is as much of a problem for the current ICES approach because it influences the S-R relationship.

Which SPM model to select could be decided using the normal AICc criteria. In this case it would be model #6 – (Table 5.1) which is the bottom right one in Figure 5.4. One could also use an ensamble approach giving weights to each model or use priors from Thorson *et al.* (2012) and Sparholt *et al.* (2020), to obtain the best SPM.

Table 5.1. The diagnostic and results of the fitted SPMs in Figure 3.1.1. Model #6 has the best AICc value.

SPM model	Number of para-meters estima-ted	Bmsy/K (curve shape parameter)	R^2	AICc	SSBmsy million t	MSY in million t	K (Carrying capacity) million t	MSY/ TBmsy
#0 Fmsy estimated Curve estimated	3	0.529	0.24	0.81	6.5	1.54	16.7	0.17
#1 Fmsy estimated – Schaefer	2	0.500	0.24	-2.50	6.4	1.53	17.5	0.17
#2 Fmsy estimated - Thorson et al. (2012) “all taxa”	2	0.404	0.24	-2.71	6.5	1.49	21.9	0.17
#3 Fmsy estimated - Thorson et al. (2012) “Perciformes”	2	0.353	0.25	-2.58	6.7	1.48	25.6	0.16
#4 Fmsy fixed –Schaefer	1	0.500	0.11	-2.38	4.9	1.68	14.0	0.24
#5 Fmsy fixed - Thorson et al. (2012) “all taxa”	1	0.404	0.20	-4.12	4.6	1.57	16.2	0.24
#6 Fmsy fixed –Thorson et al. (2012) “Perciformes”	1	0.353	0.22	-4.61	4.5	1.53	18.1	0.24

When the SPM is established then the forward projection could go like this (which is very much like an MSE calculation):

1) start with the observed TB (2021) from the assessment.
2) The real TB(2021) is obtained taking observation error into account.
3) Then the SP(2021) is obtained considering process error.
4) The real SSB(2021) is obtained by a linear link to TB influenced by F.
5) Then the observed SSB(2021) is obtained taking account of observation error.
6) Then intended F(2021) is obtained taking account of the HCR (linearly reduced when SSB < Btrigger).
7) The TAC(2021) is then obtained.
8) The realised yield(2021) is obtained taking implementation error into account.
9) The real TB for the following year is then obtained from the real TB the current year + real SP – realised yield.
10) The observed TB the following year is obtained from the real TB and observation error.
…repeat the sequence from stage 3) above for each year into the future in the simulations.

The results of these calculations could be shown as in Figure 5.4, where Yield, Risk to SSB, Inter-annual variation in catch, and SSB are shown as functions of Btrigger and F-target. For a given Btrigger, the F-target value giving the highest yield is Fmsy. Here it could be noted, that the Fmsy is not very sensitive to a quite wide range of Btrigger. The F.05 can be found as well from the right-hand top, panel.

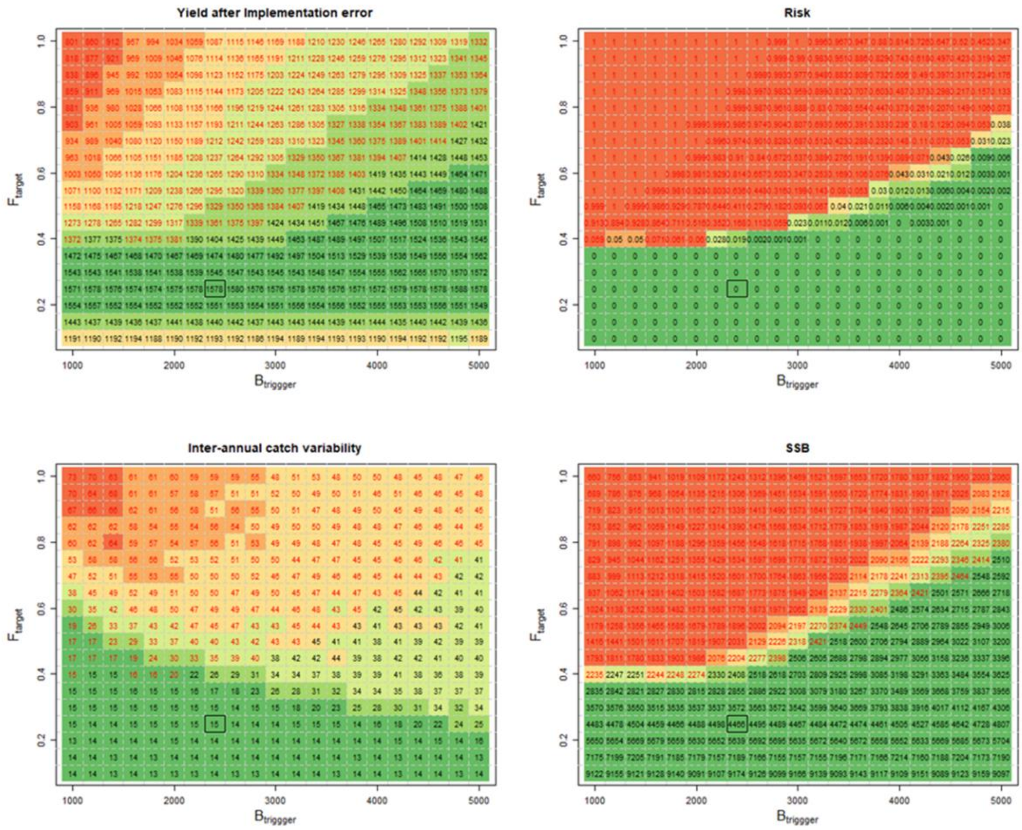


Figure 5.2. NEA mackerel. Plots of yield, Risk to SSB, Inter-annual variation in catch and SSB are shown as functions of Btrigger and F-target using the SPM model #6 from Figure 3.1.1.

Looking at the "big picture", the overall fishing pressure in the Northeast Atlantic has reduced substantially in the recent decade or two (Figure 5.5). At present it is only about half of that in the period of overfishing (1970-2000).

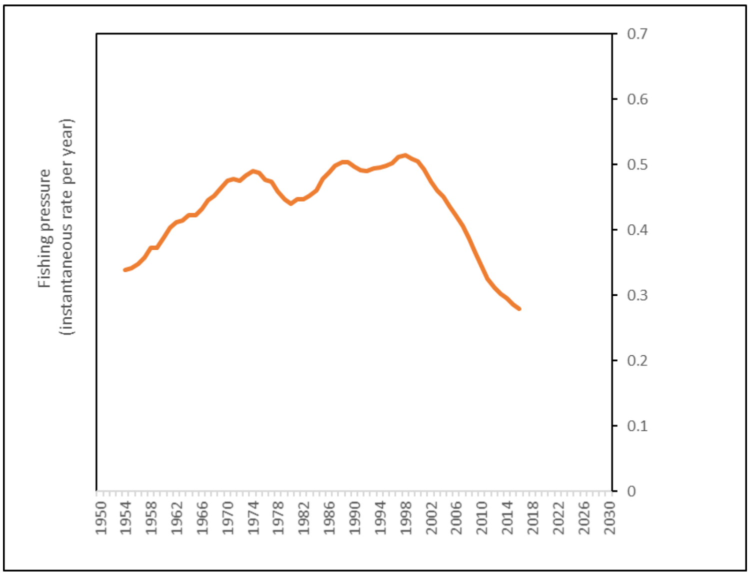


Figure 5.5. Mean fishing pressure in the Northeast Atlantic – mean of 53 ICES data rich stocks.

The total stock biomass of all stocks has increased in the ICES area by a factor of two the past 2 decades, especially the "3-big-pelagics" (mackerel, herring, and blue whiting) have increased (Figure 5.6). This means that BRP for the other species could be impaired in the recent decades due to predation of eggs and larvae and due to food competition with the "3-big-pelagics".

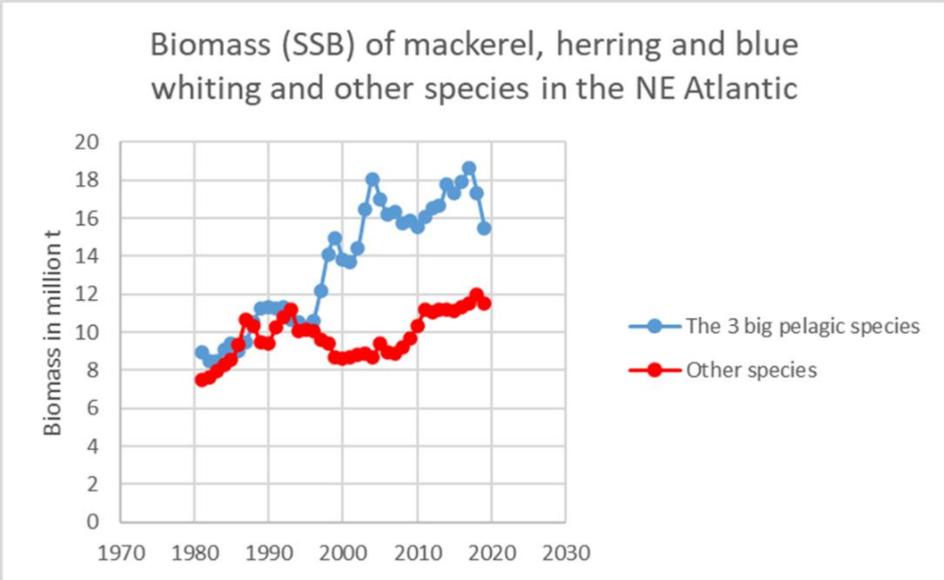


Figure 5.6 SSB of the ICES stocks in the Northeast Atlantic. From ICES summary tables 2020.

However, the catch has not increased (Figure 5.7). One could ask: Where is the "…long-term gain from the short-term pain…" ICES told managers about in the 1970-2000? Missing DD in 3 out of 4 parameters in ICES Fmsy calculations explains part of it.

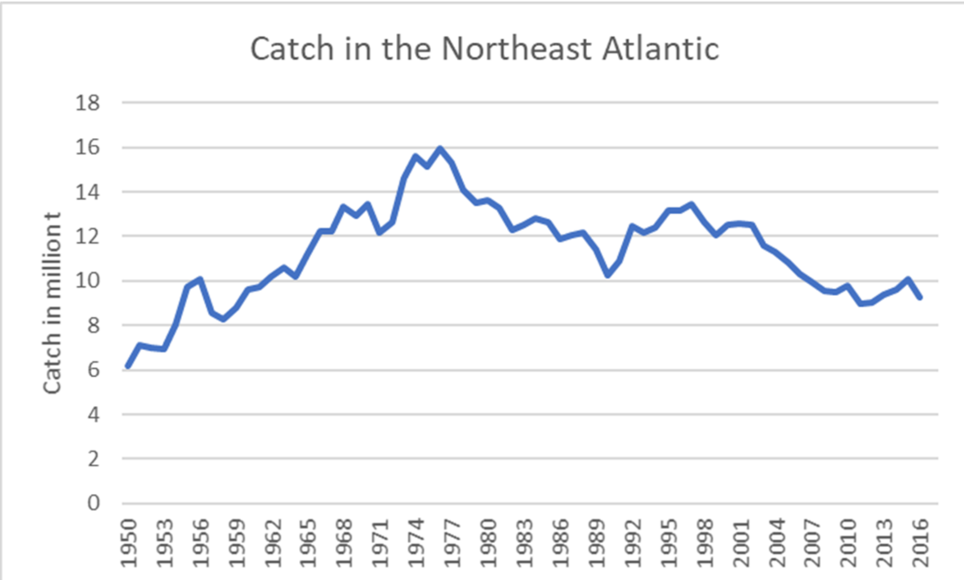


Figure 5.7. Catch in the Northeast Atlantic. From ICES summary tables 2020. From ICES database (http://www.ices.dk/marine-data/dataset-collections/Pages/Fish-catch-and-stock-assessment.aspx) except unreported catch (discards and IUU catch) which is from the "Sea Around Us"- database (http://www.seaaroundus.org/).