North Pacific Fisheries Commission

Review of Target and Limit Reference Points

Consultancy report

Laurence T. Kell

Imperial College, London

6 March 2019

2nd Floor Hakuyo Hall
Tokyo University of Marine Science and Technology
4-5-7 Konan, Minato-ku, Tokyo
108-8477 JAPAN

TEL +81-3-5479-8717
FAX +81-3-5479-8718
Email secretariat@npfc.int
Web www.npfc.int

Contents

1. Introduction **3**
- 1.1. Scientific Advice Framework . . . 3
- 1.2. Stocks . . . 4
  - 1.2.1. Pacific saury . . . 4
  - 1.2.2. Mackerel . . . 4
- 1.3. Population Dynamics . . . 5

2. Management Frameworks **6**
- 2.1. Reference Points . . . 7
- 2.2. Harvest Control Rules . . . 7

3. Management Strategy Evaluation **10**
- 3.1. Management Strategies . . . 11
- 3.2. Performance Metrics . . . 12
- 3.3. Best Practice . . . 12

4. Case Studies **14**
- 4.1. ICES . . . 14
  - 4.1.1. Precautionary Approach . . . 14
  - 4.1.2. MSY Advice Rule . . . 16
  - 4.1.3. Short-lived species . . . 16
- 4.2. Tuna RFMOs . . . 17
- 4.3. Mackerel . . . 19
  - 4.3.1. Blue mackerel: Eastern Australia . . . 19
  - 4.3.2. Blue mackerel: Western Australia . . . 20
  - 4.3.3. Atlantic mackerel: NE Atlantic . . . 21
  - 4.3.4. Jack mackerel: South Pacific . . . 21
- 4.4. Pacific Herring . . . 22
- 4.5. Pacific Sardine . . . 25
- 4.6. Capelin . . . 26

5. Stock Assessment **28**
- 5.1. Assessment Models . . . 28
- 5.2. Diagnostics . . . 29
  - 5.2.1. Residual Analysis . . . 29
  - 5.2.2. Hindcast . . . 30

6. Discussion **31**

7. Conclusions **34**

1. Introduction

The purpose of this document is to assist the NPFC Scientific Committee in setting management standards and developing harvest control rules for its target fisheries, by reviewing reference points used in pelagic species fisheries by other regional fishery management organisations (RFMOs).

Specific tasks are to

- Review target and limit reference points used in pelagic species fisheries by other general RFMOs and fishery management bodies, with focus on species which have similar life history to Pacific saury and chub mackerel;
- Make recommendations on possible options for NPFC with respect to target and limit reference points to be used for the priority species: Pacific saury and chub mackerel;

A number of case studies were chosen to illustrate the ways in which reference points are estimated, uncertainty considered, and advice provided by other RFMOs and management bodies. The strengths and weaknesses of the various approaches are discussed, and the main points of relevance to NPFC summarised.

1.1. Scientific Advice Framework

The legal requirements of fisheries management frameworks are specified in a variety of international instruments. Scientific advice frameworks, however, needs to take into account the stock, fisheries, data and management context.

The NPFC has already established a comprehensive Stock Assessment Protocol for Pacific saury and chub mackerel that requires the identification of the available data and knowledge, the main sources of uncertainty, the models and assumptions to be used, methods for diagnostics and validation, and how management will be formulated. This protocol provides a sound basis for the development of scientific management advice.

Stock Assessment Protocol

1. Identify the data that will be available to the stock assessment
2. Evaluate data quality and quantity and potential error sources (e.g., sampling errors, measurement errors, and associated statistical property (e.g., biased or random errors, statistical distribution) to ensure that the best available information is used in the assessment;
3. Select population models describing the dynamics of the stock and observational models linking population variables with the observed variables;
4. Develop base case scenarios and alternative scenarios for sensitivity analyses;
5. Compile input data and prior distributions for the model parameterization for the base case and alternative scenarios;
6. For each scenario, fit the model to the data, diagnostics of model convergence, plot and evaluate residual patterns, compare prior and posterior distributions for key model parameters, and evaluate biological implications of the estimated parameters;
7. Develop retrospective analysis to verify whether any possible systematic inconsistencies exist among model estimates of biomass and fishing mortality;
8. Identify final model configuration and model runs for each scenario;
9. For each scenario, estimate and plot exploitable stock biomass and fishing mortality (and their relevant credibility distributions) over time;
10. For each scenario, estimate biological reference points (e.g., MSY, Bmsy, Fmsy) and its associated uncertainty;
11. Identify target and limit reference points for stock biomass and fishing mortality;

12. Have the Kobe plot for each scenario;
13. Determine if the stock is âĂIJoverfishedâĂİ and âĂIJoverfishingâĂİ occurs for the base and sensitivity scenarios;
14. Finalize the base-case scenario;
15. Develop alternative ABCs for the projection (e.g., 5-year projection);
16. Conduct risk analysis for each level of ABC defined in Step (15) for the base-case scenario;
17. Develop decision tables with alternative state of nature;
18. Determine optimal ABCs based on decision tables developed in Step (17);
19. Provide scientific advice on stock status and appropriate catch level

1.2. Stocks

Pelagic fish inhabit the pelagic zone, the largest habitat on earth. Different species of fish are found throughout this zone and abundance and distributions vary regionally and vertically, depending on availability of light, nutrients, dissolved oxygen, temperature, salinity, and pressure. Stocks can demonstrate large fluctuations in abundance independently of fishing (Schwartzlose et al., 1999; Beare et al., 2004). Pelagic fish are also key components of marine food webs and can show wide spatial and temporal variability (Shephard et al., 2014).

Pelagic fish include forage fish such as anchovies, sardines, shad, and menhaden and the predatory fish that feed on them[1]. Pelagic fish can be categorised as coastal and oceanic fish, based on the depth of the water they inhabit. Coastal pelagic fish inhabit sunlit waters up to about 655 feet deep, typically over the continental shelf. While oceanic pelagic fish typically inhabit waters below the continental shelf, examples include larger fish such as swordfish, tuna, mackerel, and even sharks. There is no distinct boundary from coastal to ocean waters so some oceanic fish become partial residents of coastal waters, often during different stages of their lifecycle. Several species also perform large annual migrations across their range (Huse et al., 2010), however, true oceanic species spend their entire life in the open ocean.

1.2.1. Pacific saury

There is a single species of Pacific saury (*Cololabis saira*), which supports one of the most commercially important fisheries in the Western North Pacific. It has a wide distribution extending to the subartic and subtropical areas of the North Pacific Ocean from inshore waters of Japan and Kuril Islands eastward to Gulf of Alaska and southward to Mexico, and undergoes a seasonal migration, moving towards the northern waters of Japan in spring, staying in the Oyashio region in summer, and migrating southward in autumn (Syah et al., 2017; Kosaka, 2000).

The maximum lifespan is 2 years with age 1 reaching over 29cm in the fishing season between August and December (Suyama et al., 2006), and the length at maturity is 25 cm (Hatanaka et al., 1956). Oceanographic conditions are believed to influence spatial distribution and catches (Tian et al., 2003).

Currently the stock is assessed using a Bayesian state-space biomass dynamic model[2], which accounts for process and model error in addition to observation errors in the indices of abundance, and provided estimates of uncertainty in parameters and derived quantities such as Maximum Sustainable Yield (MSY) based reference points. As part of the assessment sensitivity analysis are run based on different assumptions about catchabilty (q) and intrinsic growth rate (r).

1.2.2. Mackerel

The scombrids (Family *Scombridae*) include many commercially important fish stocks including the tunas, bonitos, and Spanish mackerel (Juan-Jordá et al., 2013). As well as chub mackerel (*Scomber*

[1] http://firms.fao.org/firms/summaries/en
[2] https://www.npfc.int/meetings/3rd-meeting-technical-working-group-pacific-saury-stock-assessment

japonicus), which is found along the pacific coast of North and South America and in Asian waters, there are four other species of mackerel, namely Atlantic mackerel (*Scomber scombrus*), Atlantic chub mackerel (*Scomber colias*) and blue mackerel (*Scomber australasicus*) which is also known as Japanese mackerel or Pacific mackerel found in the Red Sea, Indian Ocean and the Pacific. In 2016 a new species of mackerel was discovered in the eastern Arabian Sea - Indian chub mackerel (*Scomber indicus*), this species was previously thought to be a variation of blue mackerel, but research showed that genetic differences meant that this was in fact a separate and distinct species.

Mackerel species are relatively long lived compared to Pacific saury and their age structure can be influenced by the periodic arrival of dominant year-classes[3], see **Figure 1**. Although biomass dynamic assessment models have been used, age based stock assessment models are commonly used; for example in the North East Altantic mackerel are assessed using the state-space assessment model SAM (Nielsen and Berg, 2014a), and methods such as Stock Synthesis have also been applied[4].

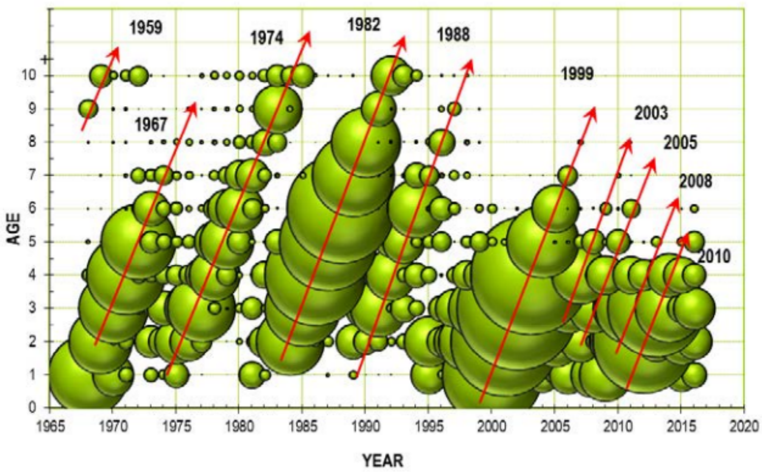

Figure 1: Catch-at-age in the commercial fishery of Northwestern Atlantic mackerel; red arrows indicate dominant year-classes and the size of the circles is proportional to fish abundance.

As of yet, however, there is no agreed stock assessment for chub mackerel in the North Pacific. Although a member countries of NPFC have developed stock assessments bsed on national data using a variety of methods. The Scientific Committee (SC) will complete the stock assessment of chub mackerel, however, as soon as possible in accordance with the terms of reference agreed at the Technical Working Group (TWG) CM meeting in December 2017.

1.3. Population Dynamics

Stock recruitment relationships are often required to be fitted in order to calculate reference points (Sissenwine and Shepherd, 1987). Many reference points assume stationary dynamics (Szuwalski and Hollowed, 2016), however, if stock dynamics are driven by environmental variability it is not clear always how best to estimate reference points. Furthermore Szuwalski et al. (2019) showed that that spurious, yet significant stock-recruit relationships can be generated if recruitment is environmentally driven and life histories are short, while processes such as natural mortality (Dickey-Collas et al., 2015; Kronlund and Shelton, 2014) may be confounded with recruitment and

MSE can be used to develop management strategies that are robust to uncertainty in the dynamics (see De Moor et al., 2011; Punt et al., 2016b; Siple et al., 2019; Haltuch et al., 2019). Studies have found

[3] http://waves-vagues.dfo-mpo.gc.ca/Library/40619576.pdf
[4] https://www.sprfmo.int/assets/SCW6-CJM-assessment/Report-of-Jack-mackerel-stock-assessment-workshop-SCW6.pdf

that modifying management strategies to include environmental factors does not necessarily improve the ability to achieve management goals much, if at all, and only if the manner in which these factors drive the system is well known (Punt et al., 2014). Until the skill of stock projection models improves, it seems more appropriate to consider the implications of plausible broad forecasts related to how biological parameters may change in the future as a way to assess the robustness of management strategies, rather than attempting specific predictions per se.

Haltuch et al. (2019) in a review of studies that incorporating environmental processes into recruitment forecasting, using both full-feedback MSE, and simulations to investigate harvest control rules suggested that the inclusion of environmental drivers into assessments and forecasting is most likely to be successful for species with short pre-recruit survival windows (e.g. sardine) and for those that have bottlenecks in their life history during which the environment can exert a well-defined pressure (e.g., anadromous fishes, those reliant on nursery areas). The effects of environment may be more complicated and variable for species with a longer pre-recruit survival window, reducing the ability to quantify environment-recruitment relationships.

Simulation testing can also be used to evaluate stock assessment methods that incorporate various levels of spatial complexity (for example Kell et al., 2009; Punt et al., 2018)

2. Management Frameworks

When managing fisheries decisions have to be made with incomplete knowledge, therefore a number of international agreements have requested the adoption and implementation of the Precautionary Approach (PA, Garcia, 1996). These include the United Nations Conference on Environment and Development (UNCED; UN, 1992), the United Nations Straddling Fish Stocks Agreement of 1995, also known as the UN Fish Stocks Agreement (UNFSA; UN, 1995) and the FAO Code of Conduct for Responsible Fisheries (FAO, 1995).

The PA requires that undesirable outcomes be anticipated and measures taken to reduce the probability of them occurring and to do this requires determining how well management measures achieve their objectives given uncertainty (Kirkwood and Smith, 1995). The United Nations Convention on the Law of the Sea (UNCLOS, UN, 1982), calls for the adoption of a maximum sustainable yield (MSY) approach to managing fisheries while the Johannesburg Declaration of the World Summit on Sustainable Development (WSSD; UN, 2002) called for the rebuilding of fisheries to levels that can produce MSY.

Due to UNCLOS and the PA many regional fisheries management Organisations (RFMOs) have implemented management frameworks based on reference points. A main objective of the use of reference points is to prevent overfishing, namely growth, recruitment, economic and target overfishing. Growth and recruitment overfishing are generally associated with limit reference points (LRPs), 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 is not necessarily considered serious unless a consistent bias becomes apparent. In contrast even a single violation of a limit reference point may indicate the need for immediate action.

In addition fish retailers and consumers are increasingly looking for assurances that the food they buy has been sustainably produced. For example to achieve Marine Stewardship Council (MSC) certification it has to be shown that stocks are above the point where recruitment is impaired (so that productivity is not compromised), are above or fluctuating around the MSY level and that harvest strategies are in place that will prevent overfishing. Certification as well as helping to ensure that fisheries are sustainable, brings benefits such as including help with marketing and securing a price premium. To achieve the MSC standard, however, requires a large amount of technical knowledge and fishery information, and failure to meet the minimum requirements or to provide the necessary information may result in conditions being placed on a fishery, which can increase the overall cost of achieving and maintaining certification. Although North various East Atlantic mackerel fisheries have been certified, recently due to low recruitment and high fishing pressure this certification has been suspended.

2.1. Reference Points

Annex II of the UNFSA[5] states that *âĂIJThe fishing mortality rate (F) which generates Maximum Sustainable Yield (MSY) should be regarded as a minimum standard for Limit Reference Points (LRPs)âĂIJ* and ensure that *âĂIJthe risk of exceeding LRPs is very lowâĂİ*. Many of the tuna RFMOs therefore refer to MSY and/or B_{MSY}, as an objective and there can be confusion as to whether F_{MSY} is a limit or a target.

Under UNFSA limit reference points set boundaries which are intended to constrain harvesting within safe biological limits. Where safe biological limits are interpreted as relating to highly undesirable states that are irreversible or slowly reversible, such as impaired recruitment (i.e. recruitment overfishing). Avoiding irreversible or slowly reversible impacts in the context of uncertainty is also the objective of applying the Precautionary Approach. A general target in UNFSA is therefore to maintain or restore stocks at levels capable of producing maximum sustainable yield while recognising uncertainty in understanding and variability of biological systems. The actual target is therefore recognised as being a management related issue, and the overall objective therefore is to maintain the highest long-term average catch with a low chance of being outside safe biological limits.

MSY is the largest average long-Âŋterm yield from application of a constant F (F_{MSY}) or from application of a variable F (harvest control rule where F varies as a function of stock size). In common practice MSY is estimated in this way taking realistic account of uncertainties/variability in productivity, stock status and fishery selectivity. However, in some cases, F_{MSY} may be determined assuming perfect knowledge and ignoring important sources of uncertainty. Where F_{MSY} is determined assuming perfect knowledge, the estimate of F_{MSY} should be used as a limit reference point as suggested in the UNFSA Annex II Guidelines. Consequently, the target F should be less than F_{MSY} so as to provide the precautionary buffer envisaged by the Guidelines. The use of F_{MSY} as a limit in most situations is expected to be very cautious because F_{MSY} is not usually associated with being beyond biologically safe limits, though a wide range of biomass outcomes for some stocks can be experienced at F_{MSY} because of variability in productivity (e.g. recruitment) and this should be examined on a case by case basis.

Figure 2 and **Table** 1 summarise the main reference points based on a yield per recruit and stock recruitment relationships. These were modelled using FLR[6] see the vignette for more details.

2.2. Harvest Control Rules

Reference points are commonly used within a harvest control rule (HCR) in order to achieve management objectives. Harvest control rules have become an important tool in modern fisheries management, and are increasingly adopted to provide continuity in management practices, to deal with uncertainty and ecosystem considerations, and to relieve management decisions from short-term political pressure (F. Kvamsdal et al., 2016).

A HCR is a formal decision rule which defines the agreed management action required when a given limit reference point is approached or breached. Once targets have been agreed limit reference points are

[5] The United Nations Agreement for the Implementation of the Provisions of the United Nations Convention on the Law of the Sea of 10 December 1982 relating to the Conservation and Management of Straddling Fish Stocks and Highly Migratory Fish Stocks

[6] www.flr-project.org

Table 1: Reference points

Acronym	Definition
MSY	Maximum Sustainable Yield
B_{MSY}	Biomass that will produce MSY
F_{MSY}	Fishing mortality (F_{MSY}) that will produce MSY
SPR_0	the spawner per recruit at virgin biomass.
SPR_{30}	corresponds to the point on the curve where SPR is 30% of SPR_0. In these cases the biomass, ssb and yield values are derived by multiplying the per recruit values by the average recruitment.
$F_{0.1}$	A proxy for F_{MSY} defined as the point on the yield per F curve where the slope equals 10% of that at the origin
F_{max}	A proxy for F_{MSY} defined as the maximum of the yield per F curve
$F_{30\%SPR}$	A proxy for F_{MSY} defined as the point on the SSB per recruit vs. F curve where the value of SSB per recruit is 30% of that where $F = 0$
$SSB_{F=0}$	Spawning stock biomass in the absence of fishing derived from a stock assessment.
B_{PA}	SSB reference point defined by ICES as part of the PA framework, it is intended to trigger management action before limit reference points are reached
B_{lim}	SSB limit reference point defined as part of the ICES framework defined as the level of SSB below which recruitment may be impaired
F_{lim}	Reference point corresponding to the level of fishing mortality that would drive the stock to extinction, estimated from where the SSB per recruit vs. F curve crosses the x-axis on the righthand side
F_{PA}	F corresponding to a SSB/R equal to the inverse of R/SSB at the Lowest Observed SSB - LOSS
F_{cap}	Maximum allowable fishing mortality
$B_{trigger}$	A level of SSB that if the stock falls triggers management action to rebuild the stock. Used as part of a harvest control rule
$B_{escapement}$	A buffer so that catch is only taken from biomass in excess of $B_{escapement}$
F_{crash}	A reference point corresponding to the level of fishing mortality that would drive the stock to extinction, estimated from where the SSB per recruit vs. F curve crosses the x-axis on the righthand side
MEY	represents the maximum economic yield, F_{MEY} corresponds to the level of exploitation that provides the maximum profit. Profit is obtained as the difference between the revenue (yield multiplied by prices) and the costs (cost per unit of F multiplied by the level of exploitation).

required to trigger management action when pre-specified indicators show that the fishery is reaching or breaching a specified value (Harley et al., 2009).

Uncertainty should be explicitly considered when formulating management advice, so that there is a low probability of exceeding safe biological limits and providing a high average long-term catches. For example in the case of ICES F_{MSY} is based on a stochastic projection with error in the resource dynamics (recruitment, M, maturity, growth, selectivity) as well as assessment/advice error. F_{MSY} is then calculated by finding the constant F which give the maximum yield. To ensure consistency between the precautionary and the MSY frameworks.

Figure 3 shows an example HCR (brown) plotted on a phase plot of harvest rate relative to F_{MSY} and stock biomass relative to B_{MSY}. The green quadrant defines a target area where the stock is above B_{MSY} and fishing mortality is below F_{MSY}. For a given value of SSB (x-axis) there is a target F, set by the brown line. The dark blue line is a trigger point ($Btrigger$) which if biomass falls bellow F is reduced, the vertical light blue line represents the B_{lim} a value which the stock should be above with high probability. If the stock is above the $B_{Trigger}$ then a constant F is used to set management advice, i.e. used to define the length of the fishing season, limit effort or to calculate a Total Allowable Catch (TAC). If F is set below F_{MSY} then it will help to maintain the stock above B_MSY.

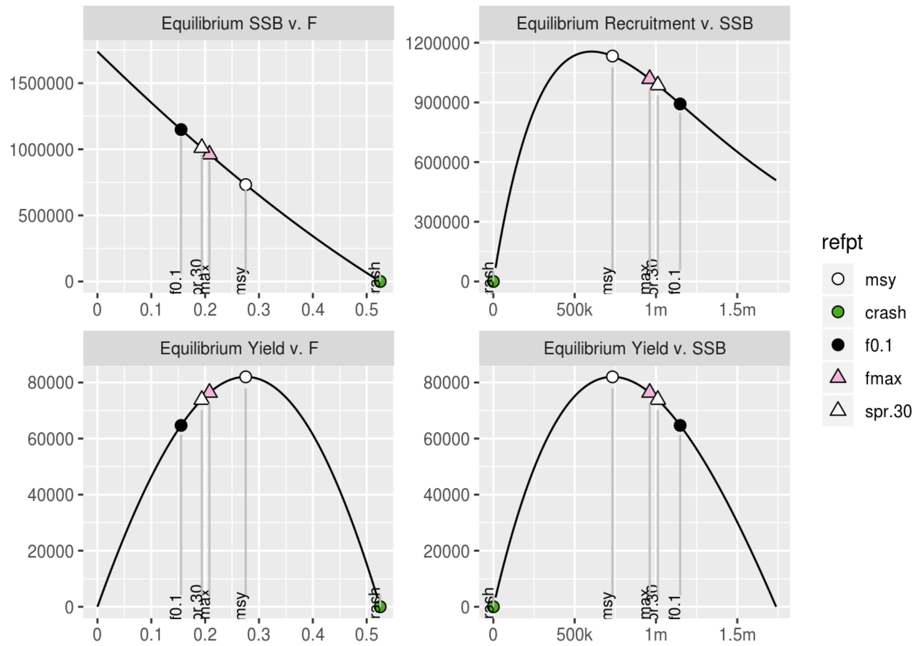

Figure 2: Reference points based on yield/spawner recruit analysis combined with a stock recruitment relationship.

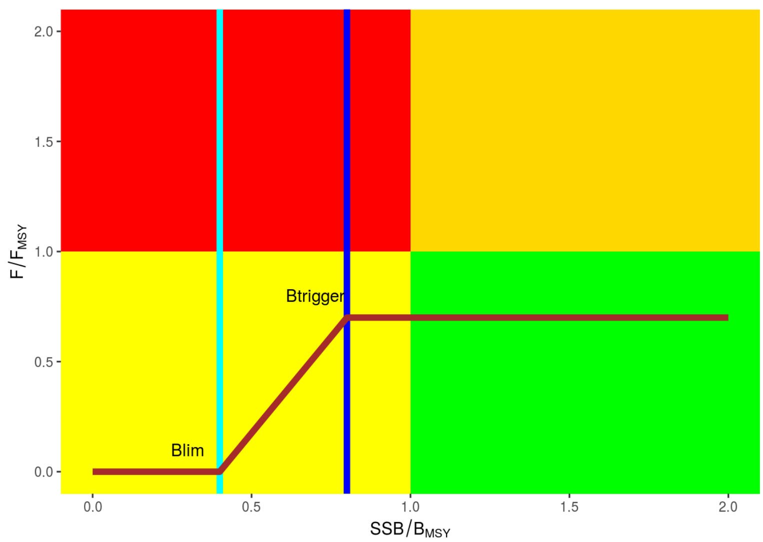

Figure 3: An example harvest control rule (brown) plotted on a phase plot of harvest rate relative to F_{MSY} and stock biomass relative to B_{MSY}.

3. Management Strategy Evaluation

To ensure management is robust requires showing that limits are avoided and targets achieved despite uncertainty, Management Strategy Evaluation (MSE) is therefore increasingly being used to simulation test alternative candidate reference points and HCRs prior to implementation (e.g. Edwards and Dankel, 2015; Carruthers et al., 2016). MSE therefore allows a fuller consideration of uncertainty as required by the PA and can be used to guide the scientific process by identifying where the reduction of scientific uncertainties will improve management and so help to ensure that expenditure is priortised to provide the best research. MSE can also help to provide stability since management objectives and the management actions required to meet them have to be pre-agreed

In MSE a simulation model, known as an Operating Model (OM), is used to represent the resource dynamics and the uncertainty about them (Kell et al., 2006). The OM is then used to simulate pseudo data in order to evaluate management procedures (MPs). A MP is the combination of monitoring data, analysis method and management measures as well as reference points and harvest control rules. Linking the OM and the MP requires an Observation Error Model (OEM) to generate fishery-dependent or fishery-independent resource monitoring data. The OEM reflects the uncertainties, between the actual dynamics of the resource and perceptions arising from observations and assumptions by modelling the differences between for example of an index of abundance and the actual value in the OM.

MSE therefore involves using simulation to compare the relative effectiveness for achieving management objectives of different combinations of data collection schemes, methods of analysis and subsequent processes leading to management actions. MSE can be used therefore to identify a *"best"* management strategy among a set of candidate strategies, or to determine how well an existing strategy performs. The ability of MSE to facilitate fisheries management achieving its aims depends on how well uncertainty is represented, and how effectively the results of simulations are summarised and presented to the decisionâĂŘmakers. Key challenges for effective use of MSE therefore include characterising objectives and uncertainty, assigning plausibility ranks to the trials considered, and working with decisionâĂŘmakers to interpret and implement the results of the MSE.

Management objectives reflect the social, economic, biological, ecosystem, and political (or other) goals. Management objectives typically conflict, and may include concepts such as maximising catches over time, minimising the chance of unintended stock depletion, and enhancing industry stability through low inter-annual variability in catches. In an MSE these objective are quantified in the form of Performance Statistics or Measures.

To conduct a MSE requires six steps (Punt and Donovan, 2007); namely

Identification of management objectives and mapping these to performance measures to quantify how well they are achieved

Selection of hypotheses about system dynamics for building **O**perating **M**odels (i.e. Simulation Models)

Building the simulation models, i.e. conditioning them on data and knowledge, and rejecting and weighting different hypotheses.

Identifying alternative management strategies, (i.e.the combination of pre-defined data, stock assessment methods, reference points and HCRs.

Running the simulations using the HCRs as feedback control procedures; and

Agreeing the Management Strategies that best meet management objectives.

The need to create a dialogue between managers, stakeholders and scientists means that these steps are iterative and because of the amount of work required and the need to build capacity require a multi-annual work plan.

MSE does not always need to rely on a very complex operating model; often a relatively simple operating model will provide a good idea of what management procedures are likely to perform well. To properly evaluate robustness, however, HCRs need to be tested under a wide range of models or even hypothetical situations.

MSE also be used to evaluate the Value-of-information, e.g. will there be an increase in revenue if there is an investment in reducing uncertainty

Conducting MSE can be a resource intensive process and so it may not be desirable to conduct MSE within an individual RFMO for more than one stock at a time.

3.1. Management Strategies

There are multiple examples of both empirical and model-Âŋbased Management Strategies.

A model-based management strategy is one which includes an explicit population dynamics model component and hence there are explicit âĂIJestimationâĂİ and âĂIJharvest control ruleâĂİ components to the strategy, for example the harvest control rule of ICES can be tested as full feedback control rule using Management Strategy Evaluation (see Kell et al., 2005a,b). Model-Âŋbased ones are attractive because they may be linked to the stock assessment results and generally have a greater capacity to *learn* about stock productivity. In some cases,however, the model used as part of the HCR is a much simpler model than the one used for actual assessments.

An empirical-based management strategy in contrast is based on data collected directly from the fishery. Monitoring data are not analysed in the context of a population dynamics model, but may be pre-processed (e.g. CPUE standardized).

Ultimately, the differences between the two types of strategies are largely academic as performance is often similar between them, but empirical management strategies can often be easier to explain to stakeholders and test using MSE

CCSBT developed an MP where The TAC is an average of candidate TACs obtained from two HCRs (Hillary et al., 2013).

The first HCR used a single index for the adult stock and then increased or decreased the current catch if that index was increasing or decreasing respectively, while the second compared the current value of an index to a reference period. The TAC is updated depending on the trend in an index (I)

$$TAC^1_{y+1} = TAC_y \times \begin{cases} 1 - k_1|\lambda|^\gamma & \text{for} \quad \lambda < 0 \\ \\ 1 + k_2\lambda & \text{for} \quad \lambda \geq 0 \end{cases} \tag{1}$$

where λ is the slope in the regression of $\ln I_y$ against year for the most recent n years. k_1 and k_2 are *gain* parameters and γ actions asymmetry so that decreases in the index do not result in the same relative change as as an increase.

The second HCR uses both an adult and juvenile indies i.e.

$$TAC^2_{y+1} = 0.5 \times \left(TAC_y + C^{\text{targ}}_y \Delta^R_y\right) \tag{2}$$

$$C^{\text{targ}}_y = \begin{cases} \delta \left[\frac{I_y}{I^*}\right]^{1-\varepsilon_b} & \text{for} \quad I_y \geq I^* \\ \delta \left[\frac{I_y}{I^*}\right]^{1+\varepsilon_b} & \text{for} \quad I_y < I^* \end{cases} \tag{3}$$

$$\Delta^R_y = \begin{cases} \left[\frac{\bar{R}}{\mathcal{R}}\right]^{1-\varepsilon_r} & \text{for} \quad \bar{R} \geq \mathcal{R} \\ \\ \left[\frac{\bar{R}}{\mathcal{R}}\right]^{1+\varepsilon_r} & \text{for} \quad \bar{R} < \mathcal{R} \end{cases} \tag{4}$$

where δ is the *target* catch; I^* the *target* adult index (e.g. a mean observed CPUE corresponding to a period where the stock was at a desired fraction of B_0 or M_{MSY}) and $\bar{R}$ is the average recent juvenile biomass i.e.

$$\bar{R} = \frac{1}{\tau_R} \sum_{i=y-\tau_R+1}^{y} R_i \tag{5}$$

$\mathcal{R}$ is a "limit" level derived from the mean recruitment over a reference period; while $\varepsilon[0,1]$ actions asymmetry so that increases in TAC do not occur at the same level as decreases.

Table 2: Derivative MP tunable parameters

Parameter	Symbol	Description	Default
Gain term	b	Sets change based on adult index in HCR 1	0.25
Gain term	r	Sets sets change based on recruit index HCR 1	0.75
Gain term	k_1	Sets decrease level when stock declines in HCR 2	1.5
Gain term	k_2	Sets increase level when stock increases in HCR 2	3.0
Exponent	γ	Additional decrease control in HCR 2	1

3.2. Performance Metrics

The results from a management strategy evaluation are compared using performance metrics. These should not be viewed as "the best" forecast of future stock status or likely catch levels. They cannot be, since MSE begins with admitting that a range of possible hypotheses must be considered due to uncertainty about the true underlying stock and fishery dynamics. The reason for this view are related to the fundamental goal of MSE: find MPs that are robust to uncertainties about stock dynamics. Typically the performance of each candidate MP is evaluated against a set of different plausible hypotheses about stock and fishery dynamics. The purpose of the performance measures is to allow MPs to be ranked relative to how well they meet prespecified objectives. In this sense, the performance measures represent the logical consequences of the specific scientific assumptions and data corresponding to each hypothesis. None can be said to be "the single best" forecast of future performance since they are contingent on the particular hypothesis used to derive the values.

The purpose of the simulations and evaluation of performance measures is to find at least one MP that performs acceptably well, regardless of which hypothesis about stock and fishery dynamics is most closely realized. Uncertainties in dynamics such as weight-at-age or natural mortality apply as much in future as now, and simply expand the range of outcomes for calculation of the performance measures.

3.3. Best Practice

Punt et al. (2016a) identified a set of "best practices" for conducting MSE. Ideally the âĂŸbest practicesâĂŹ should be followed as closely as possible, particularly when the intent is to use the MSE to develop a management strategy for a particular fishery. Conducting a MSE can still be useful, however, even if not all of the best practices are followed strictly. This is particularly the case when the aim of the MSE is to evaluate generic management strategies rather than to propose a management strategy for implementation to a specific stock.

Most critical that the primary aim of a MSE is to identify which uncertainties are most important in terms of achieving management objectives. A MSE consider all sources of influential uncertainty, even if they are not all represented in the operating models, considers all the management objectives, even if they cannot all be reflected in the operating models, and minimally allows for uncertainty in the information on which management advice is based.

Selection of objectives and performance metrics

- Involve decisionâĂŘmakers and stakeholders (e.g. using workshops) throughout the process to ensure the performance statistics capture the management objectives and are understandable.
- At a minimum, report statistics related to average catches, variation in catches and the impact on stock size.

Selection of uncertainties

- Consider a range of uncertainties, which is sufficiently broad that new information collected after the management strategy is implemented should generally reduce rather than increase this range.
- Include trials for each potential source of uncertainty (unless there is clear evidence that the source does not apply) and for the factors considered in Table 3.
- Consider the need for spatial structure, multiple stocks, predatorâĂŘprey interactions and environmental drivers on system dynamics; modelling the last by imposing trends on the parameters of the operating model is often sufficient to understand its implications.
- Include predation effects using minimum realistic models and examine the potential for technical interactions amongst major fished species, especially in multispecies fisheries.
- Divide the trials into âĂŸreferenceâĂŹ and âĂŸrobustnessâĂŹ sets.
- Use Bayesian posterior distributions to capture the parameter uncertainty for each trial if possible.

Identification of candidate management strategies

- This should be the primary responsibility of the stakeholders/decision-makers, but with guidance from the analysts given the limitations of the management strategy evaluation (MSE). Care needs to be taken that the management strategy can be implemented in practice.
- Evaluate the entire management strategy. In cases in which the management strategy is complex, this may be impossible computationally, in which case a simplification of the assessment method is needed - the nature of the simplification should be based on simulation analyses.

Simulation of the application of the management strategy

- Check that operating model and management strategy are consistent with reality; projections into the future should generate quantities, such as past assessment errors and levels of variability in biomass and recruitment, on the same scales as those estimated to have occurred in the past.
- Conduct tests of the software, for example using âĂŸperfectâĂŹ data before conducting actual analyses.
- Base recommendations for management actions in management strategies only on data which would (with near certainty) actually be available.
- Document any assumptions regarding parameters assumed known when applying the management strategy.

Presentation of results and selection of a management strategy

- Develop a process, so that the decisionâĂŘmakers understand the results of the MSE and the range of tradeâĂŘoffs which are available to them.
- Use effective graphical summaries which are developed collaboratively with the stakeholders.
- Identify whether there are âĂŸperformance standardsâĂŹ which must be satisfied to elimi-

nate some possible management strategies immediately and hence simplify the final decision process.

- Select a method for assigning a plausibility rank to each trial and take these ranks into account when making a final selection among candidate management strategies.

Other

- Include âĂŸExceptional CircumstancesâĂŹ provisions which specify the situations under which a management strategy's recommendations may be overâĂŘridden.
- Include a schedule for when formal reviews of the implemented management strategy will take place.

4. Case Studies

A number of case studies were selected to illustrate the ways in which reference points are estimated and used to provide advice.

4.1. *ICES*

ICES uses both a Precautionary Approach (PA) framework with reference points to trigger management action before limits are reached and an advice rule (AR) to achieve MSY. These can be based on either age or biomass based assessments. There are some extra considerations and differences, however for short-lived specie characterised by high natural mortality and for which year classes contribute significantly to the fishery for only one or a maximum of two years.

4.1.1. Precautionary Approach

When quantitative stock assessments are available, either age or biomass based, ICES uses precautionary approach (PA) reference points to trigger management action before limits are reached. The key reference point is the biomass limit B_{lim} as the other limit and precautionary reference points (B_{PA}, F_{lim}, and F_{PA}) are all derived from it.

B_{lim} is defined as a deterministic point below which the biomass (SSB) of a stock is considered to have reduced reproductive capacity or the recruitment dynamics are unknown. B_{lim} is commonly derived from a stock recruitment relationship. For stocks with a wide dynamic range of SSB, and evidence that recruitment is or has been impaired then the B_{lim} is set at the segmented regression change point.**Figure 4** shows a segmented regression fitted to a stock and recruitment dataset. The segmented regression divides the data into two parts; a left-hand set where the regression between stock size and recruitment is a line going through the origin and declines with stock size, and a right hand set where the regression is a horizontal line. The intercept of the two lines is defined as B_{lim}, as recruitment can be seen to be generally lower when SSB is below B_{lim}. F_{lim} is then defined as fishing mortality which would reduces SSB to B_{lim}.

The PA reference points are intended to prevent the limits from being reached. B_{PA} is therefore set at a level above which the stock is considered to have full reproductive capacity, and F_{PA} is the fishing mortality reference point below which exploitation is considered to be sustainable. The PA reference points are derived taking into accounting uncertainty by assuming an estimation error (δ) derived from a stock assessment. B_{PA} is then set at the 95_{th} percentile of B_{lim}, i.e. $B_{lim}e^{1.645\delta}$), to ensure that there is a less than a 5% chance of being below B_{lim}. Likewise F_{PA} is set at the 5_{th} percentile, $F_{lim}e^{-1.645\delta}$, and is intended ensure that exploitation is sustainable. In a few cases where the available information does not allow direct estimation of B_{lim}, B_{PA} is estimated instead and B_{lim} is then derived from B_{PA}. **Figure 5** shows the PA and limit reference points on the equilibrium curves.

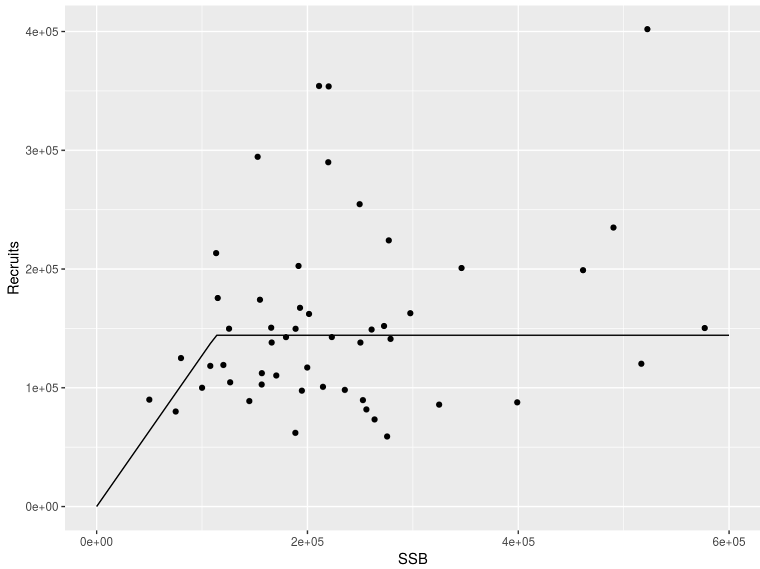

Figure 4: Segmented Regression fitted to stock recruitment pairs.

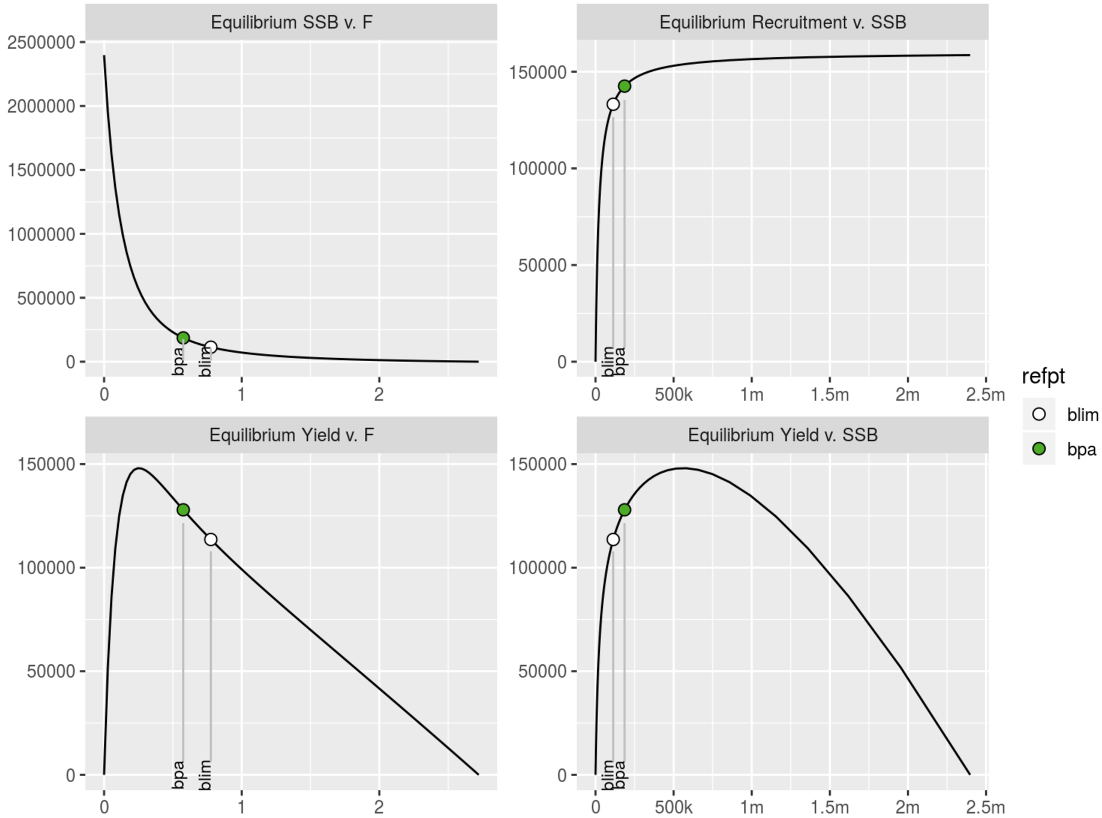

Figure 5: PA reference points

For stocks assessed using biomass dynamics models (i.e. where no age or length data are avalable) dynamics are modelled using a production function based a few parameters which models the combined effects of recruitment, growth, and natural mortality.

It such cases it is current practice to set reference points based on the deterministic equilibrium relationship between yield, F, and stock biomass using the production function, for example that of

(Pella and Tomlinson, 1969) i.e.

$$\frac{r}{p} \cdot \ B(1-(\frac{B}{K})^{p}) \tag{6}$$

Where r is the population growth rate at small populations size, (p) is the shape of the production function, B is the biomass and K is the carrying capacity. If $p = 1$ then the maximum productivity (MSY) is found halfway between 0 and K, as p increases MSY shifts to the right. F_{MSY} is then the harvest rate (catch/biomass) that maximizes production.

B_{lim} is the stock biomass corresponding to an equilibrium yield equal to half the maximum (i.e. $MSY/2$), and F_{lim} is the harvest rate that drives the stock to B_{lim}. So far no F_{PA} or B_{PA} reference points have been set by ICES for stocks assessed using a biomass dynamic assessment model. However, since biomass dynamic stock assessments can provide uncertainty estimates the probabilities of F exceeding F_{lim} or biomass being below B_lim in any assessment year can be directly calculated from the assessment.

4.1.2. MSY Advice Rule

To achieve MSY ICES has adopted an Advice Rule (AR) based on two reference points F_{MSY} and $MSY_{Btrigger}$, which can be used to set catches. F_{MSY} is the fishing mortality for a given fishing pattern and current environmental conditions that gives the long-term MSY. While $MSY_{Btrigger}$ is the lower bound of SSB fluctuation around B_{MSY} and is intended to trigger a cautious response so that in cases where SSB falls below $MSY_{Btrigger}$ fishing mortality is reduced to rebuild a stock to B_{MSY}. The reduction in fishing mortality is proportional to the ratio between the size of the spawning stock and $MSY_{Btrigger}$.

The advice rule therefore provides advice corresponding to a fishing mortality of:

- $F = F_{MSY}$ when the spawningâĂŞstock biomass is at or above $MSY_{Btrigger}$, and
- $F = F_{MSY} x SSB/MSY_{Btrigger}$ when the stock is below $MSY_{Btrigger}$

Advice is intended to be consistent with the precautionary approach and so should also conform to the overriding criterion of a greater than 95% probability that SSB remains at or above B_{lim}. For example if the stock is below B_{lim} advice is to bring the stock above B_{lim} in the short term, which may result in advice of zero catch.

4.1.3. Short-lived species

Short-lived species with a short lifespan typically exhibit high levels of natural mortality with large yearly variation largely caused by predation and environmental conditions, highly-variable recruitment, and a low age at first capture. Fishing mortality is generally lower than natural mortality and stock size is sensitive to recruitment because of the few age groups in the natural population. Therefore care must be given to ensure a sufficient spawning-stock size as the future of the stock is highly dependent on annual recruitment.

For short-lived stocks, as for long-lived ones, the ICES MSY approach is aimed at achieving a high probability (95%) of maintaining the stock biomass above the level required to produce MSY (i.e. $> B_{lim}$). To do this ICES uses two reference points, $MSY B_{escapement}$ and F_{cap}. Yearly catch advice corresponds to the estimated stock biomass in excess of $MSY B_{escapement}$, but constrained to allow a fishing mortality that is no higher than F_{cap}.

Additional rules include

1. If, under natural conditions of no fishing, the long-term annual probability of SSB being below $B_{lim} \leq 5\%$, then the same criteria as for long-lived stocks is used.

2. If, under natural conditions of no fishing, the long-term annual probability of SSB being below $B_{lim} > 5\%$, then the management plan/strategy is precautionary if the maximum probability that

SSB is below B_{lim} is $\leq 5\%$ (after the fishery) in any year when a fishery takes place. In all other years the fishery should be closed. Accepted plans with the above or more stringent criteria should not imply an increase of the long-term annual probability of SSB being below B lim by more than a factor of 2 compared to natural conditions of no fishing.

F_{cap} is defined to limit exploitation rates when biomass is high as a large stock is usually estimated with greater uncertainty, i.e. when the catch is taken, the uncertainty in the escapement biomass is greater. By capping the F, the escapement biomass is increased in proportion to stock size, maintaining a high probability of achieving the minimum amount of biomass left to spawn. The advised yearly catches correspond to the estimated stock biomass in excess of the MSY Bescapement, but constrained to allow a fishing mortality that is no higher than F_{cap}.

4.2. Tuna RFMOs

Tuna Regional Fisheries Management Organisations (tRFMOs) are intergovernmental organisations that are charged with data collection, scientific monitoring and the management of tuna and tuna-like species. There are five tRFMOs, namely the Commission for the Conservation of Southern Bluefin Tuna (CCSBT), the Inter-American Tropical Tuna Commission (IATTC), the International Commission for the Conservation of Atlantic Tunas (ICCAT), the Indian Ocean Tuna Commission (IOTC) and the Western and Central Pacific Fisheries Commission (WCPFC). The tuna RFMOs in contrast to ICES started with the objective of ensuring that stocks provide MSY, only more recently have they started to adopt precautionary limit reference points.

To provide consistent advice across the Tuna RFMOs have adopted a common management advice framework known as the Kobe Framework (De Bruyn et al., 2012) which has the main objective of keeping stocks above B_{MSY} and fishing below F_{MSY}. This requires assessment results to reported with respect to the probabilities of maintaining the stock above B_{MSY} and fishing mortality below F_{MSY}. Advice on stock status is therefore normally given in the form of a phase plot with a green quadrant corresponding to the target region (i.e. where the stock is neither overfished ($B \geq B_{MSY}$) nor subject to overfishing ($F \leq F_{MSY}$).

Figure 6 shows a Kobe phase plot, this combines two assessment methods, Stock Synthesis 3 (Methot and Wetzel, 2013, SS3) and a biomass dynamic model. The management objective is to keep the stock in the green quadrant with a high probability. Although both assessment methods estimate to only have a low probability of being subject to overfishing, one of the assessments shows that there is a greater than 50% that the stock is overfished.

A variety of assessment methods are used to provide advice; IATTC uses Stock Synthesis (SS Methot and Wetzel, 2013), WCPFC uses Multifan-CL (Hampton and Fournier, 2001), IOTC uses SS alongside a variety of other models, while ICCAT uses SS, Multifan-CL, virtual population analysis (VPA), and biomass dynamic models. Since all these methods can estimate stock status relative to MSY based reference the Kobe phase plot is a useful too for presenting management advice. Management advice is determined upon which quadrant a stock falls into. If the stock is in the green quadrant the objective is to maintain the stock in this state with high probability. While if the stock is in the red quadrant then management measures should be adopted immediately taking into account the biology of the stock and scientific advice to result in a high probability of ending overfishing in as short a period as possible.

Recently limit reference points have been developed (Preece et al.) to assess whether a stock is above the point where recruitment would be impaired. WCPFC uses $20\% SB_{F=0}$, the spawning stock biomass in the absence of fishing derived from a stock assessment. $20\% SB_{F=0}$ has the advantage of not depending on the selection pattern of the fleets like MSY based reference points and is relatively insensitive to the steepness of the stock recruitment relationship and as long as the relevant life-history

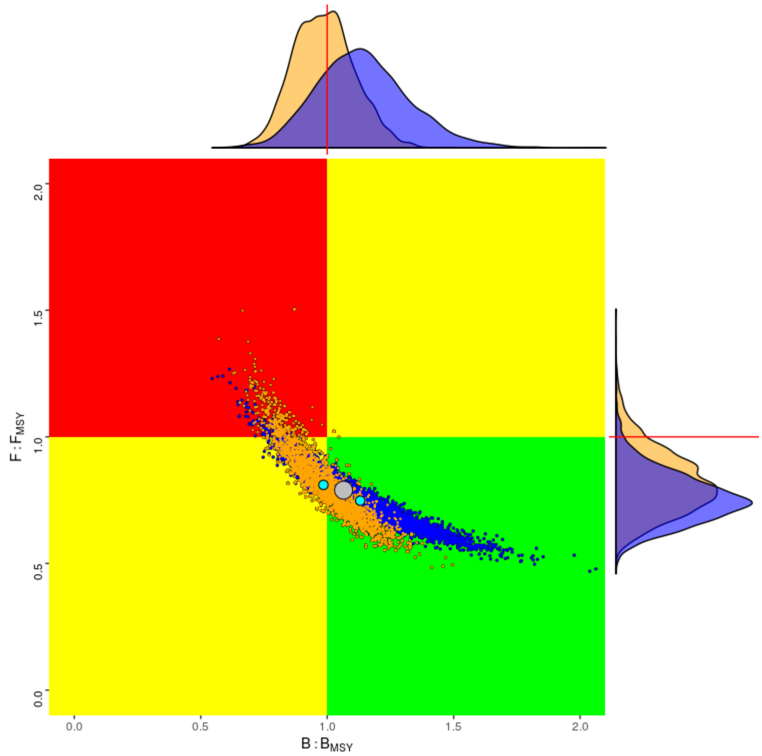

Figure 6: Kobe phase plot.

information is available (natural mortality, maturity, growth) offers a robust basis for comparing the exploited population relative to population subject to natural mortality only (Berger et al., 2013). The probabilities associated with the limit reference points can be displayed in a Majuro plot **Figure**

In the case of the tuna RFMOs have adopted the Kobe Framework Kell et al. (2016b). In the case of the Indian Ocean Tuna Commission (IOTC) the high level objectives are to maintain stocks in perpetuity with high probability, at levels not less than those capable of producing MSY as qualified by relevant environmental and economic factors including the special requirements of developing States in the IOTC area of competence[7].

Interim Target and Limit Reference Points (TRPs and LRPs) have been agreed for several stocks (i.e. albacore, swordfish, bigeye, skipjack, and yellowfin tuna). B_{TARGET} is set equal to the estimate of B_{MSY} and F_{TARGET} to F_{MSY} from the assessment, while the limits are set to multiples of B_{MSY} and F_{MSY}, e.g. for albacore, yellowfin and swordfish $B_{LIM} = 0.40B_{MSY}$, $F_{LIM} = 1.40F_{MSY}$, bigeye $0.50B_{MSY}$, $1.30F_{MSY}$, and skipjack $0.40B_{MSY}$, $1.50F_{MSY}$. These values reflect the life histories of thr species and the variability in productivity.

One of the key parameters in fisheries stock assessments is âĂIJsteepnessâĂİ, which is a measure of the productivity of the stock at low stock size (Simon et al., 2012). It is difficult to estimate steepness in stock assessments (Lee et al., 2012), however, as often there is insufficient data on recruitment at low stock size and recovery from depletion to enable steepness to be reliably estimated. Therefore WCPFC proposed a hierarchical approach to identifying limit reference points, with three levels based upon the biological knowledge and data available; namely i) Maximum Sustainable Yield (MSY); ii) spawning potential-per-recruit (SPR); and depletion based limit reference points. A simulation model was developed to evaluate the consistency and robustness of these reference points. From this work a three-level hierarchical approach to selecting and setting limit reference points for fishing mortality (F)

[7]http://iotc.org/cmm/resolution-1510-target-and-limit-reference-points-and-decision-framework

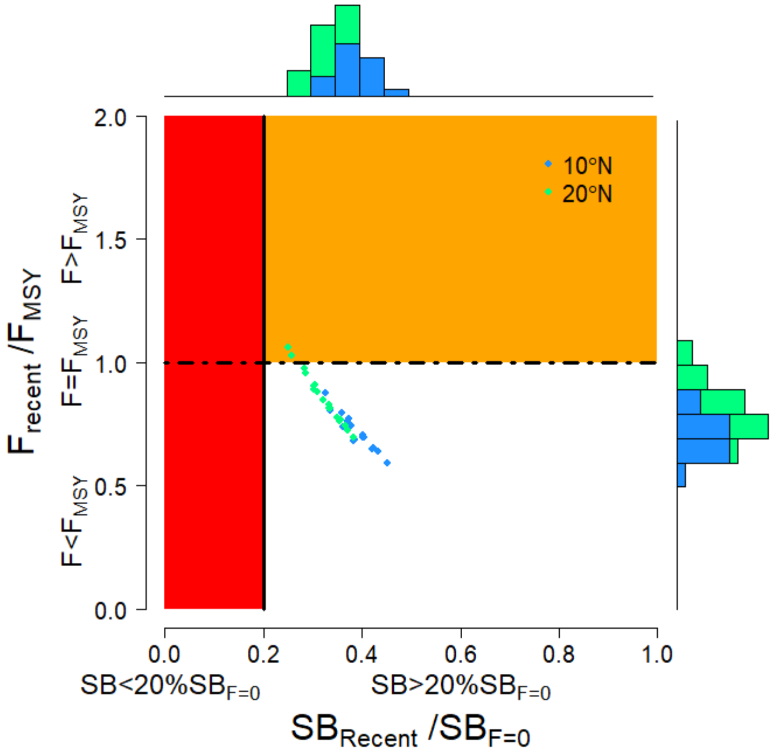

Figure 7: Majuro plot

and Spawning Stock Biomass (SSB) based on decreasing levels of available information was recommended. The first level uses F_{MSY} and B_{MSY} but only in the case where a reliable and precise estimate of steepness is available. The second level uses F_{SPR} and 20% of SSB_0 for cases in which uncertainty in steepness is high, but the key biological (natural mortality, maturity) and fishery (selectivity) variables are reasonably well estimated. The third level does not include an F-based limit reference point if the key biological and fishery variables are not well estimated, but simply uses a B_{limit} of 20% of SSB_0.

4.3. Mackerel

Worldwide there are 32 mackerel stocks exploited by 105 fisheries[8]. There is increasingly an interest amongst stakeholders in fisheries certification, for instance by the Marine Stewardship Council (MSC). Certification is a way of showing that a fishery meets international best practice for sustainable fishing. To date, however, only four mackerel stocks have certified fisheries; namely the Eastern and Western Australian stocks of blue mackerel, Atlantic mackerel in the North east Atlantic and Chilean jack mackerel in the South Eastern Pacific.

Here we present a summary of the assessments for the four certified stocks as an example of best practice. To achieve certification requires a limit reference point to indicate the stock level at which recruitment is impaired (PRI) which the stock should be above, and a biomass target that a stock should be fluctuating around a consistent with achieving MSY. A reason for this is the lack of stock assessments and reference points for use in management.

4.3.1. Blue mackerel: Eastern Australia

The Australian Small Pelagic Fishery (SPF) is not managed to achieve reference points for stock status; rather, harvest rates are applied at one of three precautionary tiers on the basis of expert judgment in

[8] https://www.fishsource.org

interpreting data from the most recent daily egg production method (DEPM) based on survey results.

The central tenet of the DEPM is that spawning biomass can be calculated by dividing the mean number of pelagic eggs produced per day throughout the spawning area (i.e. total daily egg production) by the mean number of eggs produced per unit mass of adult fish (i.e. mean daily fecundity).Total daily egg production is the product of mean daily egg production(P0) and total spawning area (A). Mean daily fecundity is calculated by dividing the product of mean sex ratio (by weight R),mean batch fecundity (number of oocytes in a batch, F) and mean spawning fraction (proportion of mature females spawning each night S) by mean female weight (W). Hence, spawning biomass(SB) is calculated according to the equation:

$$SSB = P0 \times A/(R \times F \times S/W) \tag{7}$$

The DEPM can be applied to fishes that spawn multiple batches of pelagic eggs during an extended spawning season(e.g. Lasker 1985). Data used to estimate the DEPM parameters are obtained during fishery-independent surveys during the spawning season.

The key assumptions are that: (1) the survey is conducted during the main (preferably the peak) spawning season; (2) the entire spawning area is sampled; (3) eggs are sampled without loss and identified without error; (4) levels of egg production and mortality are consistent across the spawning area; and (5) representative samples of spawning adults are collected during the survey period. The degree to which these assumptions are met affects the accuracy and precision of estimates of spawning biomass. Some assumptions, such as that the levels of daily egg production and mortality are consistent across the spawning area, are rarely, if ever, fully upheld

The SPF harvest strategy[9] is used to set recommended biological catches (RBCs) and total allowable catches (TACs). This includes a three-tier system that is applied separately to each stock. The tiered system is designed to allow greater levels of catch when higher-quality research information was available on stock status.

Under Tier 1, stocks with the highest quality of information on DEPM, have the largest potential RBC as a proportion of the estimated biomass. While for Tier 3, stocks with relatively poor-quality information, are set the smallest RBC. The SPF tier 1 decision rules used a maximum exploitation rate of 15% of estimated spawning biomass from a recent DEPM survey as the basis for setting RBCs. This is more conservative than the internationally recommended exploitation rate of 20-25% of current total biomass (Pikitch et al., 2012). The maximum duration for staying at tier 1 without a new DEPM survey was five years. Additional precaution was added to the RBC to account for biomass estimates in excess of five years. If a DEPM survey was not conducted within five years of the previous survey, the resource would be managed under tier 2 harvest control rules. Under tier 2, maximum RBCs are set based on a maximum exploitation rate of 7.5% of the estimated spawning biomass. For tier 3, the maximum RBC was set at 500 t for each stock, reflecting that a high level of precaution is warranted when information is lacking. Once the RBC was derived, an allowance for current state catches was deducted before setting the TACs.

4.3.2. Blue mackerel: Western Australia

As in Western Australia TACs are set based on the DEPM survey and the precautionary Tier approach, where RBCs and TACs are based on a DEPM survey and application of the tier 2 decision rule (using 7.5% of the 2005 spawning biomass estimate). Available age and length-frequency data were also examined and showed no trends of concern. MSE testing has been conducted for blue mackerel (west) (Giannini et al., 2010; Smith et al., 2015). These models suggest that the current harvest strategy is appropriate,

[9] AFMA 2014a, Southern and Eastern Scalefish and Shark Fishery management arrangements booklet 2014, Australian Fisheries Management Authority, Canberra

and its application would result in a low probability of the stock falling below 0.2B0 for more than 90% of the time, in line with the HSP[10].

4.3.3. Atlantic mackerel: NE Atlantic

An age-based stock assessment model (SAM Nielsen and Berg, 2014b) is used for the assessment of North East Atlantic mackerel, based on catch and tagging data and three survey indices. Partial discarding estimates are also included in the assessment but recent discarding is not thought to be significant

The stock is assessed by ICES using the PA and MSY framework relative to target, limit and precautionary reference points, i.e. F_{MSY}, F_{lim}, F_{pa}, $MSY_{Btrigger}$, B_{lim}, and B_{pa}. Despite being managed using this framework the stock is now assessed as being below B_{ps} due to several years of poor recruitment coupled with high F. It has just been announced by the Marine Stewardship Council (MSC) that all certification for all North East Atlantic mackerel fisheries will be suspended on March 2, 2019, since the stock has dropped below B_{pa}. ICES has recommended a significantly reduced catch of 318,403t, which represents a 68.2% cut in current catches to restore the stock to a sustainable level. Short-term projections suggest that catches in line with the ICES advice would recover the stock above the sustainable level by 2020-2021. Maintaining the current level of catches will result in the stock dropping the point where recruitment is impaired in 2020.

4.3.4. Jack mackerel: South Pacific

Jack mackerel is assessed and managed by the South Pacific Regional Fisheries Management Organization (SPRFMO). Conservation and management measures for the Jack mackerel fishery aim to allow the rebuilding of the stock. There are currently, however, no agreed target or limit reference points. The SCâĂŹs Multi-Annual Work Plan, however, does include work to evaluate alternative stock structure hypotheses and assessment models, review existing data, improve knowledge on growth estimations, recruitment under climatic drivers and Jack mackerel connectivity[11].

Benchmarks assessments are conducted every two years, with updates in odd years, i.e. adding new information without extensive model re-specifications and re-evaluating past yearâĂŹs recommendations. A statistical catch-at-age model i.e. the Joint Jack Mackerel Model (JJM) is used for assessment advice[12]. Input data included catch updates for each country; acoustic survey results from Chile (north and central-southern) and Peru; Catch per Unit Effort (CPUE) from Chile, Peru, China and EU; other indices from Chile and Russia. There are concerns, however, on the evidence of life history stage differences of the possible existence of different stocks. Therefore two scenarios are considered based on hypotheses about stock structure: namely 1) two stocks in Peruvian and Chilean waters that straddle the high seas; 2) a single shared stock. Hypothesis 2 is been used as the basis for the advice, as it provides a more precautionary biomass estimate. Similar trends are been estimated from the 1-stock and 2-stock models.

Medium and long-term projections are run under varying recruitment (average from 1970-2013 or from 2000-2013). Based on the rebuilding plan and given current stock status, catches could be potentially increased but considering the uncertainties in the assessment and under the one stock hypothesis, the SC adopted has adopted a precautionary approach and recommended catches to b kept at a lower level.

The indices of abundance used in the assessment are CPUE from the southern central Chile fleet and Chinese vessels. It has been recommended to use acoustic surveys which can estimate stock abundance and also have the potential to obtain oceanographic data.

In the most recent assessment (2018)[13] four assessment models were run i.e. JJM, SAM, SS3 and a

[10] http://data.daff.gov.au/data/warehouse/9aam/fsrXXd9abm$_{/}fsr16d9abm_{2}0160930/00_{F}ishStatus2016_{1}.1.0_{L}R.pdf$

[11] http://www.sprfmo.int/assets/SCW6-CJM-assessment/Report-of-Jack-mackerel-stock-assessment-workshop-SCW6.pdf

[12] http://www.sprfmo.int/jack-mackerel-sub-group/

[13] http://www.sprfmo.int/assets/SCW6-CJM-assessment/Report-of-Jack-mackerel-stock-assessment-workshop-SCW6.pdf

surplus production model.

Currently the stock is subject to a rebuilding plan where the value of F depends on where the stock is relative to $0.8B_{MSY}$ and B_{MSY} for SSB. This means that the catch advice for the stock can be highly variable between years when the SSB is estimated to be close to these boundaries.

The harvest strategy had been evaluated using MSE in 2014. The output of the 2013 jack mackerel stock assessment was used to develop the operating model, and based on two stakeholder consultations a number of alternative plans were developed and evaluated. A number of performance statistics were used, i.e. i) the rate of biomass growth during a certain time frame ii) expected catch and catch variability, iii) risks of biomass decline, and iv) expected time to reach X% of unfished SSB (a proxy representing 80% of B_{MSY})

In 2018 it was noted there are discontinuities and an alternative HCR was considered **Figure 8** that could consist of a continuous line for F throughout the entire SSB range, with a diagonal line linking the F values to be used when $SSB < 0.8B_{MSY}$ and when $SSB > B_{MSY}$. This could be used when the SSB is estimated to be between $0.8B_{MSY}$ and B_{MSY} with the intent to reduce the variability of the catch advice at these SSB boundaries. As part of the rebuilding plan the actual value of F_{low} will change from year to year, so the distance between F_{low} and F_{MSY} may be larger or smaller depending on the year. There is also a constraint on inter-annual variability in the TACs as a maximum change of 15% was established for biomass $> 0.8BMSY$

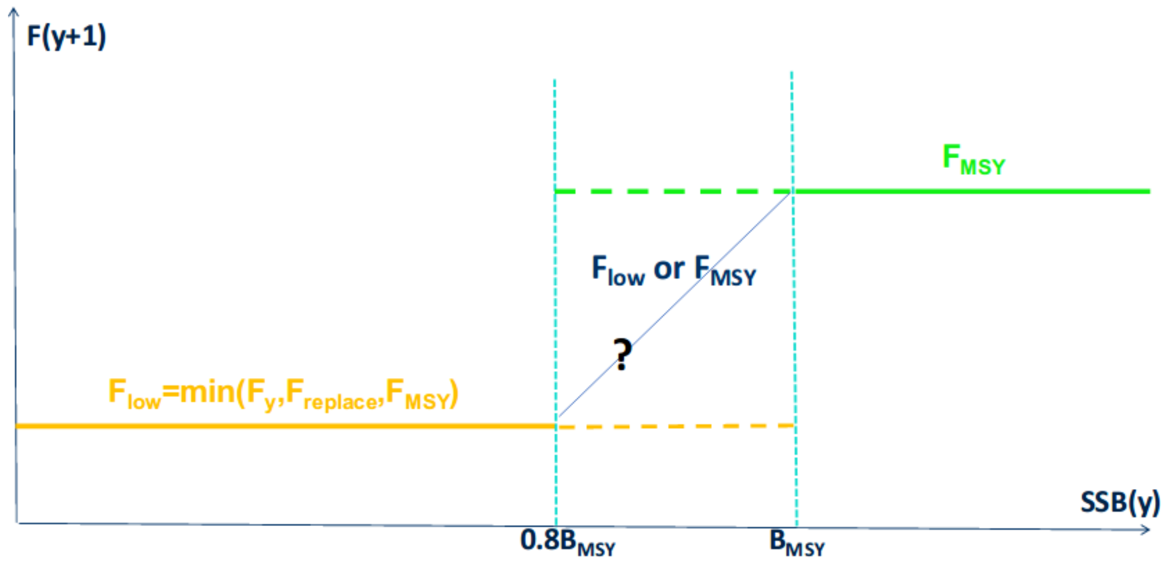

Figure 8: Jack mackerel rebuilding plan rule.

4.4. *Pacific Herring*

The Pacific herring (*Clupea pallasii*) stocks off the Canadian coast provides an important case study of how to integrate MSE with the traditional stock assessment approach. Rather than moving from a stock assessment to an MSE paradigm in one go it was decided to develop two parallel streams, based on traditional stock assessment and MSE, to allow the work to be done with existing resources and for lessons learnt to be applied as appropriate across all stocks. The main lessons learn were with respect to stakeholder engagement, stock assessment bias and assessment errors, the application of harvest control rules, and integration of science and management.

There are five major herring stocks off the Canadian coast, managed under a "Harvest Decision-Making Framework Incorporating the Precautionary Approach" policy (DFO, 2009). The framework is based on limit and target reference points and requires i) reference points that delineate Critical, Cautious and Healthy zones and a limit fishing mortality rate (**Figure 9**); ii) a harvest strategy and harvest control rules (Kronlund et al., 2014); iii) the need to take into account uncertainty and risk (Kronlund et al., 2014); and iv) a requirement to evaluate the performance of the management system against the objectives specified by the harvest strategy.

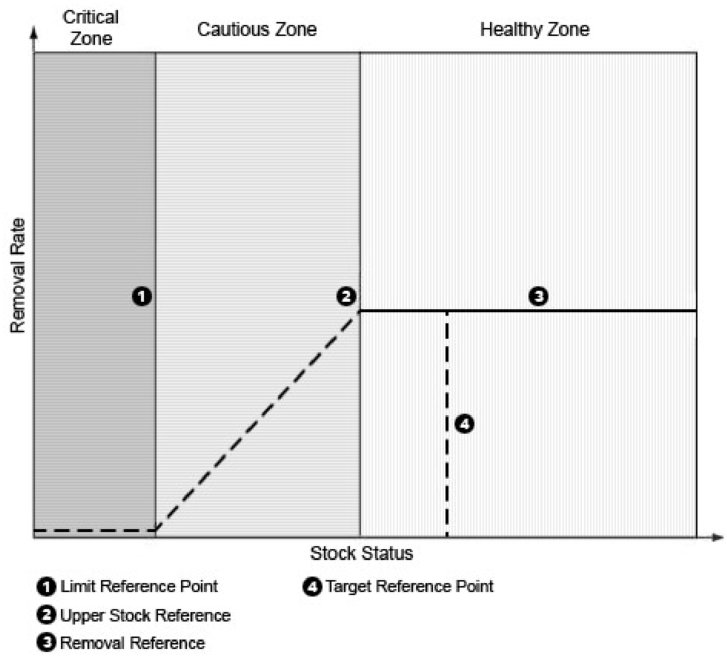

Figure 9: Critical, cautious and healthy zones, defined in terms of stock status (biomass) and removal rate.

However, despite extensive scientific and management effort there were no established reference points consistent with policy and extended closures had occurred due to low biomass estimates in three of the five major stocks. There are also conflicts about allocation, and concerns about ecosystem effects of harvesting forage fish. These factors coupled with high interest from the general public, as spawning and fisheries can occur in populous areas, led to criticism of the science and management and a call for renewal of the management system. Therefore in 2016 an MSE process started as part of the renewal of Pacific Herring management with the aim of supporting the development of management options and engagement of resource users.

The main objectives for the resource include to i) avoid the limit reference point of $0.30B_0$ with high probability (>75-95%); ii) maintain spawning stock biomass above the Upper Stock Reference (USR) of $0.60B_0$, with 50% probability; and iii) maintain spawning stock biomass at or above a target reference point biomass level of $0.75B_0$ with 75% probability. Where the probabilities are calculated over three herring generations (i.e. 15 years).

The OM was conditioned in two ways: i) time-varying M, and ii) constant but estimated M. Thus the OM had structural uncertainty.

Since previous work had indicated that M dominates other processes in the dynamics, M in the future was simulated in three alternative ways i.e. i) future natural mortality rates time-varying (i.e. random walk) around the long-term average (1951-2017) with pulse M at low populatin size; ii) around the recent average (2008-2017); and iii) constant at the long-term 1951-2017 estimate.

The stocks are assessed using a Bayesian statistical catch-at-age model, this is used to both provide advice and to condition the Operating Model. In the latter case a main uncertainty is the assumed natural mortality, and so three hypotheses were evaluated, i.e. related to time varying M

The Management Procedure uses the same stock assessment as used to provide stock assessment advice, with data sampled from the Operating Model. A hockey stick HCR was used with target and trigger reference points along with a catch cap to reduce the risk of assessment errors **Figure 10**.

As the same stock assessment model is used for all five stocks, lessons learnt when conducting the MSE for the two stocks examined could be transferred to the other stock areas, providing guidance for MSE work as analytical resources for the remaining three stocks become available. For example the assessment was shown to have large errors causing over-estimation of spawning biomass. Since the assessment method and HCR interact, the effects of such interactions cannot be predicted in advance of conducting the simulations. However, the simulation results allow those effects to be diagnosed and the MP tuned to guard against such bad behavior. The assessment errors meant that catch is set higher than intended and only simulation-evaluation can reveal this property, which is likely to apply to all five stocks. This led to investigation of the catch cap to provide a model-free means of reducing the impacts of assessment errors. It was found that the control (i.e., trigger) points are less important to conservation outcomes as the reduced harvest rate and cap mean trigger points reached less often, **Figure 10** shows an example of the HCR.

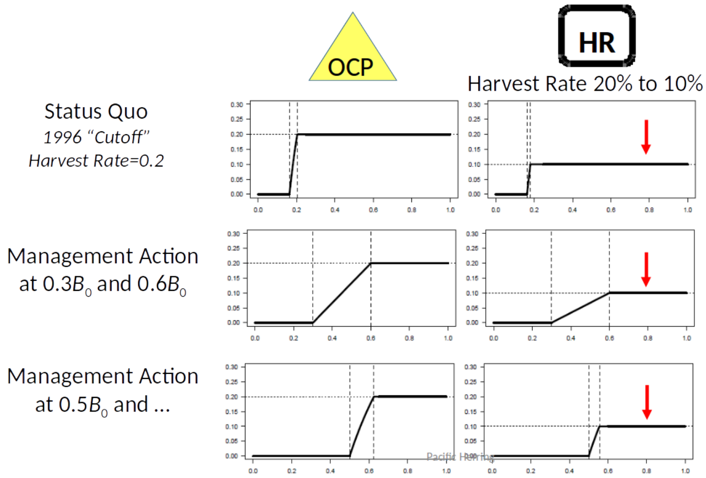

Figure 10: Example of the Pacific herring HCR

Integration of Science and Management was important and required much more time meeting with fishery managers than in previous years, for example to discuss results prior to scientific review meetings i.e., to ensure no surprises, joint-understanding of outcomes, and opportunity to improve communications. Co-developing communications (e.g., presentations and documents) with fishery managers for senior decision makers helped increase understanding and encouraged buy-in. It was also important for scientists and managers to meet with decision-makers to explain MSE approach, results, possible risks, limitations and to explain how MPs could be introduced into annual management. The meetings also allowed discussion of how MSE could be applied to other stocks in future. A result of such engagement was that clear direction was provided by senior decision-makers which helped clarify objectives and provide guidance for forthcoming development of the MSE processes for the other three major stocks.

The actual MPs developed will differ by stock, due to differences in objectives, exploitation history, population dynamics, and the magnitude of assessment errors. However process aspects of the MSE approach is the same regardless of area, as it requires identification of objectives, conditioning OMs, development of MPs, simulation testing of performance, and communication of results. If an acceptable MP can be identified, the consistent application of the selected MP over time is supported by the process steps that sought to engage resource users in identifying objectives and participate in iterative evaluation of results.

4.5. Pacific Sardine

The northern subpopulation of Pacific sardine ranges from northern Baja California, Mexico to British Columbia, Canada and extends up to 300nm offshore. Prior to and during summer months, large aggregations of the California Current sardine population migrate from the key spawning habitat off California to more northern waters. In common with many small pelagic fish species, sardine migration and population levels are heavily influenced by oceanic conditions that determine the survival and hence recruitment of juveniles into the adult stock. The biomass and catches of Pacific sardine increased rapidly during the 1930s until the mid-1940s, and then declined due to a combination of environmental conditions leading to poor recruitments and very high fishing mortality rates. The population eventually rebuilt during the 1980s, and by 1991 a directed fishery had been re-established. The stock is currently exploited by both the United States of America and Canadian. In Canadian waters the fishery is opportunistic and dependent on the migration of sardines into Canadian waters.

Fishery management was conducted by the State of California until 2000 when management authority was transferred to the Pacific Fishery Management Council (PFMC, Hill et al., 2011). Trends and status are currently assessed using the Stock Synthesis assessment model (SS Methot and Wetzel, 2013) and integrates data from research surveys and commercial catches (Hill et al., 2011). The assessment is conducted in the fall and annually estimates the pre-season age 1+ biomass.

Management between 1998 and 2012 was set using a harvest control rule (PFMC, 2011), i.e.

$$HG = (BIOMASS - CUTOFF) * FRACTION * DISTRIBUTION \tag{8}$$

which sets the total allowable catch (TAC) for each year, based on the current biomass (BIOMASS) less an escapement threshold (CUTOFF) below which fishing is prohibited. The harvest is determined by FRACTION which is a temperature-dependent exploitation fraction which ranges from 5 to 15%, and DISTRIBUTION the average proportion of the coastwide biomass in USA waters.

Following simulation testing using MSE the CUTOFF was set at 150,000t and a rule of the form adopted

$$TAC_y = B_{1+,y-1} - 150,000) \times d \times h \tag{9}$$

where: TAC_y in year y is based on last years biomass $B_{1+,y-1}$ which is the estimate of the biomass of Pacific sardine aged 1 and older obtained from an age-structured stock assessment model; 150,000mt, is the cut-off; d the a temperature-dependent exploitation fraction which ranges from 5 to 15%; and d is the average proportion of the coastwide biomass in USA waters, estimated at 0.87. In addition, there is a maximum allowable catch regardless of biomass such that $TAC_y \leq$ a maximum catch of 200,000 mt. The purpose of the cut-ff is to protect the stock when biomass is low. The purpose of d is to specify how much of the stock is available to the fishery when the biomass exceeds the cut-off. while d recognises that the stock ranges beyond USA waters and is subject to foreign fisheries.

h was determined on the basis of a 3-year running average of the temperature at Scripps Pier, La Jolla, USA. The overarching management plan for all coastal pelagic species (CPS) managed by the PFMC was modified in 2011 to be consistent with the 2006 reauthorisation of the Magnussen-Stevens Act. This involved formally introducing how the overfishing limit (OFL, the annual catch amount consistent with an estimate of the annual fishing mortality that corresponds to maximum sustainable yield) is calculated, as well as the acceptable biological catch (ABC, a harvest limit set below the OFL that incorporates a buffer against overfishing to take account of scientific uncertainty).

The inclusion of a 150,000t cutoff and the distribution factor reduces the effective harvest rate to less than h_{US} (e.g. < 0.15), and the degree of this reduction varies inversely with the magnitude of age 1+ biomass estimates $B_{1+,y-1}$. Due to the cutoff, the reduction will be greater at low population biomass estimates than at higher estimates.

The specifications of the harvest control rule adopted in 1998 were determined using simulations in which the population dynamics were represented by a production model where productivity was related to an environmental variable (, US). Results of assessments conducted after 1998 were analysed and suggested that the temperature at Scripps Pier no longer exhibited the same trends as most other measures of temperature for the offshore waters to the west of North America McClatchie et al. (2010). Rather, the relationship between recruitment, spawning biomass and temperature was strongest when temperature was based on sea surface temperature obtained from CalCOFI samples (PFMC).

The U.S. sardine HCR originated from a 1998 Coastal Pelagic Species Management Plan (Pacific Fishery Management Council 1998) and was one of thirteen candidate HCRs investigated with the objective of maintaining the sardine population at levels well above that which would occur with a single-species maximum sustainable yield (MSY). This HCR was selected because simulations indicated that it should maintain a relatively productive population, while providing forage for sardine predators. Also, the HCR was adopted based on its ability to meet a number of performance measures: high average biomass, high median biomass, high standard deviation of biomass, low percentage of years with biomass less than 400,000 t (a level at which, historically, the population appeared to be restricted to the area south of Point Conception), average catch near 147,000 t, high median catch, low standard deviation of catch, and a low percentage of years with no fishery.

The results from the February 2013 workshop formed the basis for developing a set of operating models for the northern subpopulation of Pacific sardine, as well as an initial set of candidate management strategies (PFMC). The process of selecting the operating models and the candidate management strategies was iterative, involving presentations by the analysts to the PFMC as well as its Scientific and Statistical Committee, Coastal Pelagic Species Advisory Panel and Coastal Pelagic Species Management Team. The PFMC took advice from these advisory bodies as well as from members of the public, including industry and environmental non-governmental organizations (ENGOs), and then directed the analysts. Hurtado-Ferro and Punt (2014) summarize the most recent MSE results, along with the specifications for the operating models and candidate management strategies.

Canada has also been applying[14] the same HCR to set the TAC for sardine in British Columbia (BC) waters where for step ii) the average seasonal migration rate of sardine into BC waters is used[15].

4.6. Capelin

Icelandic capelin has been acoustically surveyed since 1978, in 1979 it was recommended that the fishing season should open with a low, precautionary quota, which would then be revised upwards after an assessment survey of the mature stock within the fishing season (Vilhjálmsson, 1994). This plan was, however, never implemented and instead preliminary catch quotas were set without adequate knowledge of the condition of the stock. the resulting catches were probably too large; causing the spawning stock to collapse and a fishing moratorium was put in place in 1982 (Vilhjálmsson and Carscadden, 2002). After lifting the moratorium, a more cautious management plan was adopted (Vilhjálmsson and Carscadden,

[14] http://waves-vagues.dfo-mpo.gc.ca/Library/363736.pdf
[15] http://www.pcouncil.org/wp-content/uploads/2017/03/G5a_Stock_Assessment_Rpt_Full_ElectricOnly_Apr2017BB.pdf

2002), where preliminary catch quotas were based on acoustic surveys before fishing commenced in August, with a final TAC being set after an acoustic survey of maturing adults during the spawning migration in January.

The principal regulation at the time was to leave 400,000t to spawn under the assumption that the previous problems had to do with how to determine an appropriate preliminary quota and not the harvest control rule. Another moratorium was put in place in the early 1990s when the spawning stock collapsed again. When the fishery reopened, the basis for preliminary quotas was changed, with an October survey were used instead. Although the harvest control rule itself was kept. The objective was to improve the information on which the preliminary quota was set. In the early 2000s, the Icelandic capelin fishery was again at a low stock size as recruitment was lower than expected and preliminary quotas could not be set on the basis of juvenile indices from the autumn surveys. Recent research suggests that changes in the distribution, which were related to climate changes, affected the results from the autumn surveys (Pálsson et al., 2012; Carscadden et al., 2013).

Not until 2010, when the survey area was increased and the autumn surveys covered the full distribution of the stock, was it again possible to set a preliminary TAC. the consequences of not having a preliminary quota were that the fishery was not opened until after the adult stock acoustic surveys in late autumn or winter. However, the stock has been at lower levels in the 2000s than during the mid-1990s and the landings have been much smaller.the principal harvest control rule for the Icelandic capelin, to maintain a spawning stock of 400 thousand tonnes, has not changed since 1979. thus, harvest control rules have been in practical use before the general institutional framework came into place during the 1990s, as also seen in the Scientific Committee of the International Whaling Commission. the rule has been implemented via a two-step management plan with preliminary quotas that were adjusted when further information arrived. the government has followed the advice based on the HCR. Quotas are allocated through an Individual Transfereable Quota (ITQ) system with all of its general control and enforcement provisions. By and large, the management plan has been successful in sustaining a sufcient spawning stock, in particular if the low stock levels of the 2000s are attributed to climate changes that lie outside the scope of the current harvest control rule and management plan. As such, the current harvest control rule failed to realize its full potential in considering environmental changes. It is perhaps pertinent to mention here that the Barents Sea capelin has a history of repeated collapse and recovery, and these collapses are not necessarily tied to fishing but to changes in the ecosystem (Gjøsæter and Bogstad, 1998).

It thus seems that capelin is a type of fish that has an extraordinary sensitivity to environmental and ecological conditions, and that appropriate HCRs need to take this into account.the harvest control rule for the Icelandic capelin has been deterministic in the sense that point estimates of the spawning stock biomass have been taken at face value; that is, no provisions were made for empirical or structural (model) uncertainty. Recently, a new harvest control rule was introduced, again modelled after the harvest control rule for the Barents Sea capelin (Gjøsæter et al., 2009, 2015) where empirical uncertainty is taken into account. the new rule also reflects predation on capelin that takes place between the time of the acoustic measurements of the spawning migration and the spawning season. the final TAC will be set such that the probability of a spawning stock below a biomass limit of 150 thousand tonnes is less than 5%. the new HCR is considered to be in agreement with the precautionary approach. this change in formulation is expected to be more conservative in terms of harvest strategy than the earlier method, mostly because of higher the median of the probability distribution for the spawning stock level is expected to be close to 400,000t. Climatic considerations have until now not been taken into account in the harvest control rules despite evidence of impacts on capelin biology and distribution from climatic changes.

5. Stock Assessment

The provision of fisheries management advice requires the assessment of stock status relative to reference points, the prediction of the response of a stock to management, and checking that the predictions are consistent with reality (Kell et al., 2016a). There are a wide range of stock assessment models with a variety of data requirements and assumptions these include catch only, biomass dynamic, age-structured production models, virtual population analysis, statistical catch-at-age, and integrated analysis models (both length- and age-based).

There are currently two main trends in stock assessment, the using of generic tools such as stock synthesis and the development of models tailored to particular assessments. Which ever is the case it is important that robust use is made of statistical analysis to evaluate the performance of the methods, that they are tested via simulation where the assessment method is tested using data simulated from the same or from alternative models (Cadrin and Dickey-Collas, 2014) and models are validated. Validation examines if a model family should be modified or extended, and is complementary to model selection and hypothesis testing. Model selection searches for the most suitable model within a family, whilst hypothesis testing examines if the model structure can be reduced.

When different models are run often different datasets are used this means that AIC cannot be used to compare models. An alternative is predictive validation, where the model is used to predict (forecast) the systemâĂŹs behavior, and then comparisons are made between the systemâĂŹs behavior and the modelâĂŹs forecast to determine if they are the same, for example using a hindcast. To check that predictions are consistent with reality it is necessary to evaluate prediction skill (Walters and Punt, 1994; Patterson et al., 2001; Ralston et al., 2011); a statistical evaluation of the accuracy of a prediction relative to a reference model or dataset. Prediction skill can be used to compare alternative models or observations used for prediction to a reference set of estimates or data (Jin et al., 2008; Weigel et al., 2008; Balmaseda et al., 1995). If data are regarded as being representative of the dynamics of the stock then they can be used as a model-free validation measure (Hjorth, 1993), and the best performing scenarios (e.g., choice of models and data) can be identified by comparing predictions with observations. Stock biomass cannot actually be observed so if estimates of population abundance were compared in the hindcast this would be model-based validation.

5.1. Assessment Models

The current assessment model used for saury is a Bayesian state-space production model. A main feature of state-space models is that it is possible to separate observation noise from process noise. A feature that is important when developing reference points for use in a harvest control rule. Also most age based assessments consider process and error mainly in recruitment, but density dependence can occur in other processes such as growth and natural mortality (Bjoernstad et al., 2004) and patterns in time-series of estimates of recruitment could be an artefact due to variability in processes such as natural mortality (Dickey-Collas et al., 2015).

Recently the Bayesian state-space biomass dynamic framework has been extended to account for changes in selectivity and relative fishing mortality from multiple fisheries (JABBA-Select Winker et al., submitted). The steepness of the stock-recruitment relationship and the selectivity-at-age dependent mortality rates from an equilibrium age-structured model can be used to generate correlated multivariate normal priors on surplus-production shape and productivity parameters.

5.2. ***Diagnostics***

The provision of fisheries management advice requires fitting a model to data to assess the stock status, the prediction of the response of the stock to management, and checking that predictions are consistent with reality. The accuracy and precision of the prediction depend on the quality the model, the information in the data and the prediction horizon (i.e. how far ahead we wish to predict). Often small changes to the input data or assumptions can result in substantial differences to advice (Collette, 2017) which may undermine the scientific process.

As part of the stock assessment process, models have to be validated and the plausibility of alternative hypotheses to be considered. The stock assessment procedure adopted by the NPFC provides a good basis for doing this, and it is important to develop an objective set of diagnostics that can be run for the different types of stock assessment models that may be used, e.g. integrated, state space and Bayesian based on age, length or biomass. General questions that need to be answered include are there data conflicts? is the model misspecified? and does the model have prediction skill? Where prediction skill is a statistical evaluation of the accuracy of forecasts Huschke et al. (1959). Two important methdos are the use of residuals to identify problems due to data conflicts and possible model misspecification; and parameters and hincasting to evaluate prediction skill.

5.2.1. *Residual Analysis*

Plotting data over time is a simple method to examine trends, patterns, and variation in data and to look at the effect of changing model specifications and data conflicts. Inspection of residuals may be subject to individual interpretation and hence lead to disagreements. Residuals should be distributed in expected ways, and a runs test is an objective way of compares the probability of longest consecutive run (either positive or negative residuals) and the number of crossings with expected statistical behaviour.

A run chart is a simple line graph of a measure over time with the median shown as a horizontal line dividing the data points so that half of the points are above the median and half are below. Therefore if a series of catch per unit effort is an unbiased index of abundance then it should only show random variation, and the data points will be randomly distributed around the median. Random meaning that we cannot know if the next data point will fall above or below the median, but that the probability of each event is 50%, and that the data points are independent. Independence means that the position of one data point does not influence the position of the next data point, that is, data are not auto-correlated. If the process shifts, these desired conditions for residuals are no longer true and patterns of non-random variation may be detected by statistical tests. Non-random variation may present itself in several ways. If the process centre is shifting due to improvement or degradation we may observe unusually long runs of consecutive data points on the same side of the median or notice that the graph crosses the median unusually few times. The length of the longest run and the number of crossings in a random process are predictable within limits and depend on the total number of data points in the run chart (Anhoej, 2014, 2015)

A shift signal is present if any run of consecutive data points on the same side of the median is longer than the prediction limit, round(log2(n)+3), where n is the number of useful data points, that is, data points that do not fall on the median. Data points that fall on the median do not count, they do neither break nor contribute to the run (Schilling, 2012). A crossings signal is present if the number of times the graph crosses the median is smaller than the prediction limit, $qbinom(0.05, n-1, 0.5)$ (Chen, 2010). The shift and the crossings signals are based on a false positive signal rate around 5%. Therefore we have two statistical tests, based on residuals, that can be automated in order to detect problems in the estimation process for a number of select parameters in OMs.

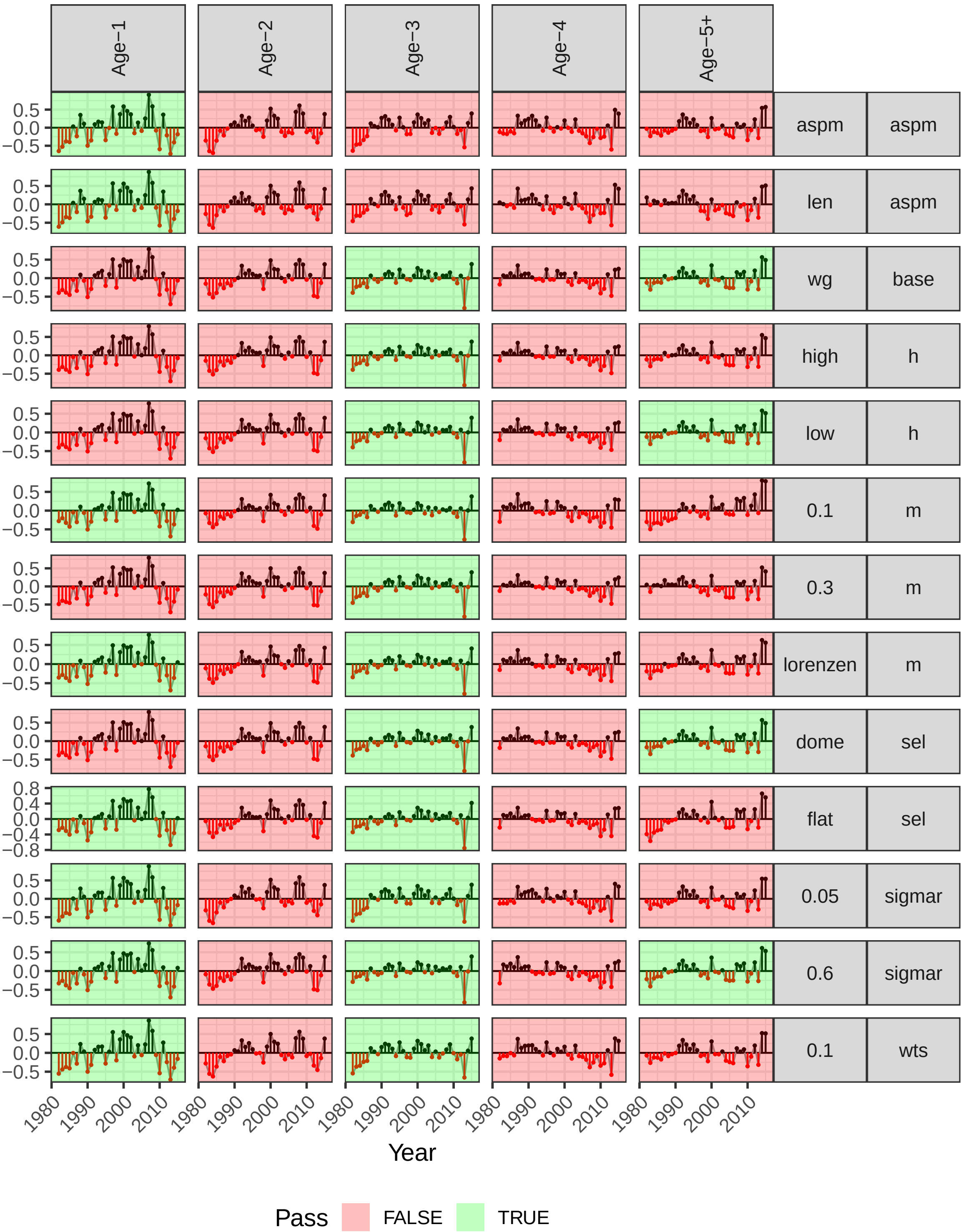

Figure 11: Example of a residual runs plot, red shaded areas indicate series that failed either the unusually long runs of consecutive data points on the same side of the median or notice that the graph crosses the median unusually few times

5.2.2. Hindcast

Prediction skill is a statistical evaluation of the accuracy of forecasts (Huschke et al., 1959), this can be evaluated use a hindcast, i.e. a forecast made retrospectively. The hindcast is a multi-step prediction where a model is fitted using a tail-cutting procedure, where data are deleted from year $t - n$ to t and then rerun using the data from year 1 to $t - n1$ to make predictions of what will happen in years tn to t.

When conducting projections to provide managers with advice, such as a total allowable catch (TAC), the short term is of primary importance as usually the immediate consequences of management advice is a major concern of some stakeholders.

Retrospective analysis is a calculation whereby a year of data is removed from the data set before the model is rerun. The removed year of data can then be thought of as âĂŸexternal dataâĂŹ and then predicted. More generally, to check that predictions are consistent with reality it is necessary to evaluate prediction skill Walters and Punt (1994); Patterson et al. (2001); Ralston et al. (2011); a statistical evaluation of the accuracy of a prediction relative to a reference model or dataset. Prediction skill can be used to compare alternative models or observations used for prediction to a reference set of estimates or data (Jin et al., 2008). If data are regarded as being representative of the dynamics of the stock then they can be used as a model-free validation measure (Hjorth, 1993), and the best performing scenarios (consisting of choices of models and data) can be identified by comparing predictions with external observations. Stock biomass cannot actually be observed so if estimates of population abundance were compared in the hindcast this would be model-based validation.

Various criteria are available for comparing the prediction skill of forecasts (Hyndman and Koehler, 2006). Although root-mean-square error (RMSE) is widely used to compare model predictions it is an inappropriate and misinterpreted measure of average error (Willmott and Matsuura, 2005). RMSE is inappropriate because it is a function of 3 characteristics of a set of errors, rather than of one (the average error), RMSE varies with the variability within the distribution of error magnitudes and with the square root of the number of errors (n1/2), as well as with the magnitude of the average-error magnitude (MAE). MAE is a more natural measure of average error, and (unlike RMSE) is unambiguous. Scaling the average errors using the Mean Absolute Scaled Error (MASE) allows forecast accuracy to be compared across series on different scales. A scaled error is less than one if it arises from a better forecast than the average one-step naÃŕve forecast, and it is greater than one if the forecast is worse than the average one-step naÃŕve forecast. Where a âĂIJnaÃŕveâĂİ forecast (or random walk) is where the forecast is equal to the last observation.

6. Discussion

To ensure management is robust to uncertainty requires that limits are avoided and targets achieved despite uncertainty, therefore additional trigger or threshold reference points are used (e.g. the PA reference points of ICES) to trigger management action before limits are reached. The PA reference points are intended to achieve a high probability (95%) of maintaining the stock biomass above the level required to produce MSY; therefore B_{PA} is set at the 95^{th} percentile of B_{lim}.

Reference points are commonly used as part of a harvest control rule (HCR) to achieve targets and to avoid limits. A HCR is a formal decision rule which defines the agreed management action required when a given limit reference point is approached or breached. Once targets have been agreed limit reference points are required to trigger management action when pre-specified indicators show that the fishery is reaching or breaching a specified value.

For example to achieve MSY ICES has adopted an Advice Rule (AR) based on two reference points F_{MSY} and $MSY_{Btrigger}$, which are used to set catches. F_{MSY} is the fishing mortality for a given fishing pattern and current environmental conditions that gives the long-term MSY. While $MSY_{Btrigger}$ is the lower bound of SSB fluctuation around B_{MSY} and is intended to trigger a cautious response so that in cases where SSB falls below $MSY_{Btrigger}$ fishing mortality is reduced to rebuild a stock to B_{MSY}. The reduction in fishing mortality is proportional to the ratio between the size of the spawning stock and $MSY_{Btrigger}$.

The legal requirements of fisheries management frameworks are specified in a variety of international

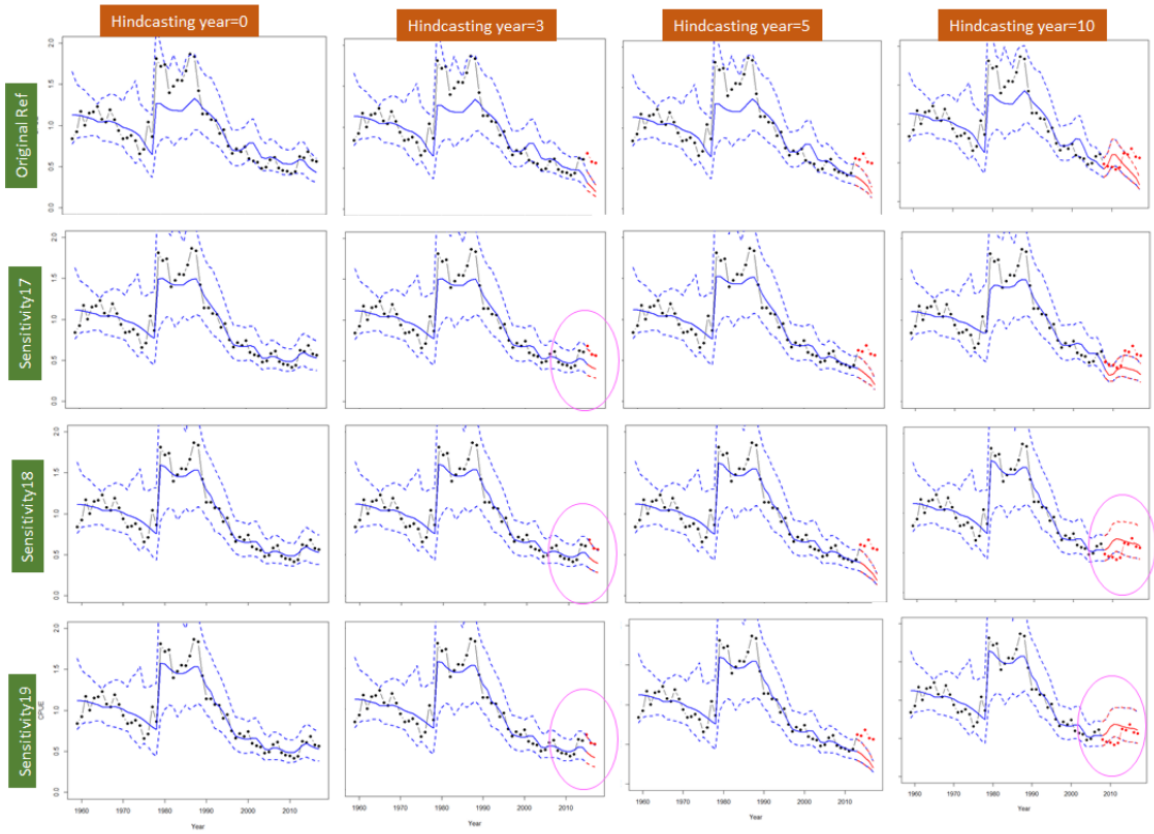

Figure 12: The predicted abundance indices in the hindcasting years for SS3 based 4 different sensitivity scenarios under 3 different hindcasting years (3, 5, and 10 years). Pink circles showed acceptable fitting.

instruments. Scientific advice frameworks, however, needs to take into account the particular characteristics of the stock, fisheries, data and the management context. It is therefore unlikely that frameworks developed elsewhere can simply be translated to the North Pacific. This reviewed showed that the life history characteristics and the dynamics of the stocks are important. Pelagic fish are key components of marine foodwebs and can show wide spatial and temporal variability and may fluctuations in abundance independently of fishing. Also when species are short lived fluctuations may occur independently of fishing, in such cases reference points derived from assumptions about surplus production derived from a stock assessment may not be robust.

Pacific saury, as a short-lived species, and can have large annual variation in recruitment and natural mortality in response to predation and environmental condition. This means that stock status may fluctuate widely even in the abscence of fishing. In the review of ICES and pacific stocks it was seen that it was insufficient to just develop limit and target reference points for use as in a harvest control rule, additional elements such as a cut-off that allowed a minimum spawning stock biomass each year, and a cap on F (F_{cap}) so that F was robust to assessment error were also required. It is also important to note that North East Atlantic mackerel although managed usin target and limit reference points stock became over fished due to recent poor recruitments. In some cases, e.g. (Pacific sardine) factors need to be included in the harvest control rule to adjust the permitted F depending on oceanographic conditions and stock distribution. Alternatively TACs can be set solely on a survey, for example like caplin where the TAC is based on an acoustic survey. Amother example of such an approach is the Australian Small Pelagic Fishery (SPF) which is not managed to achieve reference points for stock status; rather, harvest rates are applied at one of three precautionary tiers on the basis of expert judgment and a daily egg production method (DEPM). This also has the advantage that there is an economic benefit of improving data on the fishery.

Stock assessment methods used to provide advice can be data- and expertise-hungry, requiring annual updates of catch-at-age data and indices of abundance. Geromont and Butterworth (2014), however, showed that simple rules can achieve virtually equivalent catch and risk performance, with much less inter-annual TAC variability than traditional stock assessments. Harvest control rules, however, should be tested using management strategy evaluation as part of a management procedure that includes not just the reference points, but the stock assessment algorithm used to estimate them and the dataset to be used. This requires a focus on uncertainty about the resource dynamics, e.g. on hypotheses about stock recruitment, natural mortality and stock distribution and structure. The Stock Assessment Protocol of the NPFC provides an excellent basis for providing robust scientific advice.

Risk is an uncertainty that matters, and what matters are management objectives. Therefore to ensure that targets are met and limits avoided with high probability requires a consideration of uncertainty, i.e. about the data and knowledge of the resource dynamics.

When fitting assessment models there is often insufficient contrast in the data to estimate parameters for important population processes such as the steepness of the stock recruitment relationship or natural mortlaity. Therefore the data may appear equally likely given alternative model assumptions or parameter values (model and parameter uncertainty), while different data sets (CPUE, catch and length distributions) may show conflicting signals.

Therefore when conducting a stock assessment a variety of scenarios are commonly run to reflect scepticism about the capacity of the model to estimate key parameters. Scenarios are generally developed by evaluating the effect of fixing some parameters, assuming alternative functional form for processes, or by down weighting some datasets when fitting. In other cases the uncertainty from an agreed base case is estimated using methods such as Bootstrapping or Bayesian simulation.

This requires methods for proposing stock assessment scenarios and then accepting or weighting them, and to propagate the uncertainties through to management advice.

It is important that a stock assessment used for advice is shown to be robust to the key uncertainties, e.g. there is a lack of retrospective trends, good precision of key estimated variables and is statistically sound. When different models are run often different datasets are used this means that methods such AIC cannot be used to compare models. An alternative is predictive validation, where the model is used to predict (forecast) system behavior, and then comparisons are made between the systemâĂŹs behavior and the modelâĂŹs forecast to determine if they are the same, for example using a hindcast. To check that predictions are consistent with reality prediction skill can be evaluated.Prediction skill is a statistical evaluation of the accuracy of a prediction relative to a reference model or dataset. Prediction skill can be used to compare alternative models or observations used for prediction to a reference set of estimates or data. If data are regarded as being representative of the dynamics of the stock then they can be used as a model-free validation measure Hjorth (1993), and the best performing scenarios (e.g., choice of models and data) can be identified by comparing predictions with observations. Stock biomass cannot actually be observed so if estimates of population abundance were compared in the hindcast this would be model-based validation.

The ability of a HCR to achieve its aims depends on how well uncertainty is represented, and how effectively results are summarised and presented to decision-makers (Punt et al., 2016a). There has therefore been an increased use of MSE to simulation test harvest strategies, i.e. the combination of monitoring data, analysis method, harvest control rule and management measure. Harvest strategy and management procedure have the same elements but a management procedure is intended to be implemented as formally specified; it therefore has been simulation tested to demonstrate adequately robust performance in the face of plausible uncertainties about stock and fishery dynamics.

Whatever stock assessment method are used it is important to propagate the uncertainties through to management advice, i.e. stock status relative to reference points and how to achieve targets in the future. Therefore management strategy evaluation is increasingly being used to simulation test reference points and HCRs. A first step is to devlop an Operating Model of the resource dynamics, this is then used to test the HCRs as a feedback controller. The OM can be conditioned using a range of stock assessment scenarios reflecting a range of plausible hypotheses reflecting uncertainty about the resource dyamics.

Communication of results and uncertainty are important and can be done using interactive web based tools, e.g. shiny-App[16] developed to as a visualisation tool and MSE demonstrator. The app is in an initial stage of development and the intention is to allow stakeholder to evaluate the impact of uncertainty on management outcomes. The app currently conveys information on uncertainties that are relevant to the MSE process, information on the reliability of modelling, including a qualitative assessment of the modelling inputs, model validation, and the degree to which uncertainties were included.

MSE can still be used to simulation test elements of a HCR, for the example the robustness of a reference points to indicate stock status, without any intention of actually implementing a management procedure (e.g. Kell et al., 2005b,a). Further, there are examples where the MSE process (i.e. identification of potential strategies for testing, agreement on performance measures to compare strategies, predicting the performance measures for each potential strategy) has been successfully applied without modeling at all a wide range of expert judgments were used instead of a quantitative Operating Model to make these predictions and describe their uncertainty (Smith et al., 2018).

MSE can also be used to evaluate the benefits of reducing uncertainty and the value of new information (Mantyniemi et al., 2009), in such cases it will be important to consider a range of uncertainties, which is sufficiently broad that new information should reduce rather than increase any estimate of risk.

Conducting an MSE is also a lengthy process and requires a commitment to a multi-annual programme, it is therefore preferable to conduct an MSE initially for a single stock to develop expertise. Even if only elements of an MSE are conducted for a single stock there will be important benefits for other stocks, for example with respect to the treatment of uncertainty, validation of models, presentation of results and communication on objectives. For example a major uncertainty in stock assessment is whether dynamics are driven by a production function or variability in recruitment (Walters, 1987; Cury et al., 2014; Szuwalski and Hollowed, 2016; Vert-pre et al., 2013)

7. Conclusions

The purpose of this document is to assist the NPFC Scientific Committee in setting management standards and developing harvest control rules for its target fisheries. To do this reference points used in pelagic species fisheries by other regional fishery management organisations (RFMOs) were reviewed. The recommendations below are intended to help identify possible options for developing target and limit reference points for the priority species of Pacific saury and chub mackerel.

The maximum lifespan of saury is 2 years, in comparison mackerel species are relatively long lived, both however can demonstrate large variability independent of fishing. Therefore process error due to the impact of the environment on recruitment, growth and natural mortality is likely to be important. Currently saury is assessed using a Bayesian state-space biomass dynamic model, which accounts for process and model error. Currently there is no agreed assessment for chub mackerel but due to the importance of process error in the dynamics of pelagic species state space models should be considered, this will help in developing a common scientific advice framework and for lessons learnt from one stock to be transfered to the other.

[16] http://www.seaplusplus.co.uk/shiny-app-for-mse/

Stock Assessment continue the development of state-space models that distinguish between process and measurement error.

Target and Limit Reference Points

An hierarchical approach could be developed for limit and target reference points for setting advice, where tiers depend upon the biological knowledge, available data and dynamics of the stock; tiers could be based on

Maximum Sustainable Yield F_{MSY} and SSB_{MSY}: where age data are available and stock recruitment relationship and steepness or a production function can be estimated;

Spawning potential-per-recruit, SPR and $F_{0.1}$: where age based estimates are available but the stock recruitment relationship can not be estimated reliably .

Depletion where only indices of abudance a limit reference points limit such as a B_{lim} an be set equal to a value based on SSB_0

Simulation Testing

Management advice based on a biomass dynamic model may perform better than one based on an age based model if it has been designed to be robust to uncertainty, i.e. by developing a Management Strategy using Management Strategy Evaluation (MSE).

Even if MSE is not used to develop a formal management procedure, an operating model should be developed to evaluate the robustness of reference points across the tiers, by comparing their performance and the benefits of reducing uncertainty.

Initially any operating model can be relatively simple, e.g. an aged based model could be used to compare the robustness of the different types of reference points to non-stationarity in processes such as recruitment and natural mortality.

Harvest Control Rules

The use of reference points as part of a HCR should be evaluated. As well the limits and targets the HCR should include triggers to implement management before limits are reached, catch caps to reduce the risk due to assessment error, and features such as inter-annual bounds on catches to allow the fishing industry to plan its activities.

The purpose of MSE is to find a management Strategy that performs acceptably well, regardless of which hypothesis about stock and fishery dynamics is most closely realised. A range of uncertainties should be considered which is sufficiently broad that new information should reduce rather than increase any estimate of risk.

Model free harvest control rules (where no stock assessment is conducted) can also be developed based on empirical indicators (e.g. an acoustic survey), these should be tested first using MSE.

Stock Assessment Model Validation When conducting stock assessment often different model structures and frameworks have to be compared. This requires data conflicts and model misspecification to be identified and prediction skill evaluated. The stock assessment protocol provides an excellent basis for providing scientific advice, and two additional diagnostics could be developed; namely

Runs tests as an objective way or evaluating model fits to data and to identify data conflicts and potential model misspecification.

Hindcasting as an objective way to compare the prediction skill of different models and model structures

References

M. A. Balmaseda, M. K. Davey, and D. L. Anderson. Decadal and seasonal dependence of ENSO prediction skill. *J. Climate*, 8(11):2705–2715, 1995.

D. Beare, F. Burns, A. Greig, E. Jones, K. Peach, M. Kienzle, E. McKenzie, and D. Reid. Long-term increases in prevalence of north sea fishes having southern biogeographic affinities. *Marine Ecology Progress Series*, 284:269–278, 2004.

A. Berger, G. Pilling, C. Kirchner, and S. Harley. Proposed f-based limit reference points for bigeye, yellowfin and south pacific albacore tuna. Technical report, WCPFC-SC9-2013/MI-WP-03, 2013.

O. Bjoernstad, R. Nisbet, and J.-M. Fromentin. Trends and cohort resonant effects in age-structured populations. *Journal of Animal Ecology*, 73(6):1157–1167, 2004.

S. X. Cadrin and M. Dickey-Collas. Stock assessment methods for sustainable fisheries. *ICES Journal of Marine Science*, 72(1):1–6, 2014.

T. R. Carruthers, L. T. Kell, D. D. Butterworth, M. N. Maunder, H. F. Geromont, C. Walters, M. K. McAllister, R. Hillary, P. Levontin, T. Kitakado, et al. Performance review of simple management procedures. *ICES J. Mar. Sci.*, 73(2):464–482, 2016.

J. E. Carscadden, H. Gjøsæter, and H. Vilhjálmsson. A comparison of recent changes in distribution of capelin (mallotus villosus) in the barents sea, around iceland and in the northwest atlantic. *Progress in oceanography*, 114:64–83, 2013.

B. B. Collette. Bluefin tuna science remains vague. *Science*, 2017.

P. M. Cury, J.-M. Fromentin, S. Figuet, and S. Bonhommeau. Resolving hjort's dilemma how is recruitment related to spawning stock biomass in marine fish? *Oceanography*, 27(4):42–47, 2014.

P. De Bruyn, H. Murua, and M. Aranda. The precautionary approach to fisheries management: How this is taken into account by tuna regional fisheries management organisations (RFMOs). *Marine Policy*, 2012.

C. L. De Moor, D. S. Butterworth, and J. A. De Oliveira. Is the management procedure approach equipped to handle short-lived pelagic species with their boom and bust dynamics? the case of the south african fishery for sardine and anchovy. *ICES Journal of Marine Science*, 68(10):2075–2085, 2011.

DFO. A fishery decision-making approach incorporating the precautionary approach., 2009.

M. Dickey-Collas, N. Hintzen, R. Nash, P. Schön, and M. Payne. Quirky patterns in time-series of estimates of recruitment could be artefacts. *ICES J. Mar. Sci.*, 72(1):111–116, 2015.

C. Edwards and D. Dankel. *Fisheries management science: an introduction to simulation based methods.* Earthscan (Routledge), Oxford, UK, 2015.

S. F. Kvamsdal, A. Eide, N.-A. Ekerhovd, K. Enberg, A. Gudmundsdottir, A. H. Hoel, K. Mills, F. Mueter, L. Ravn-Jonsen, L. Sandal, J. Erik Stiansen, and N. Vestergaard. Harvest control rules in modern fisheries management. *Elementa: Science of the Anthropocene*, 4:000114, 07 2016. doi: 10.12952/journal.elementa.000114.

S. Garcia. The precautionary approach to fisheries and its implications for fishery research, technology and management: an updated review. *FAO Fisheries Technical Paper*, pages 1–76, 1996.

H. F. Geromont and D. S. Butterworth. Complex assessments or simple management procedures for efficient fisheries management: a comparative study. *ICES Journal of Marine Science*, 72(1):262–274, 2014.

F. Giannini, Hobsbawn, G. PI, Begg, and M. Chambers. Management strategy evaluation (mse) of the harvest strategy for the small pelagic fishery. 2010.

H. Gjøsæter and B. Bogstad. Effects of the presence of herring (clupea harengus) on the stock-recruitment relationship of barents sea capelin (mallotus villosus). *Fisheries Research*, 38(1):57–71, 1998.

H. Gjøsæter, B. Bogstad, and S. Tjelmeland. Ecosystem effects of the three capelin stock collapses in the barents sea. *Marine Biology Research*, 5(1):40–53, 2009.

H. Gjøsæter, B. Bogstad, S. Tjelmeland, and S. Subbey. A retrospective evaluation of the barents sea capelin management advice. *Marine Biology Research*, 11(2):135–143, 2015.

M. Haltuch, E. Brooks, J. Brodziak, J. Devine, K. Johnson, N. Klibansky, R. Nash, M. Payne, K. Shertzer, S. Subbey, et al. Unraveling the recruitment problem: A review of environmentally-informed forecasting and management strategy evaluation. *Fisheries Research*, 2019.

J. Hampton and D. A. Fournier. A spatially disaggregated, length-based, age-structured population model of yellowfin tuna (*Thunnus albacares*) in the western and central pacific ocean. *Marine and Freshwater Research*, 52(7):937–963, 2001.

S. Harley, S. Hoyle, J. Hampton, and P. Kleiber. Characteristics of potential reference points for use in wcpfc tuna stock assessments. *WCPFC-SC5-2009/ME-WP-02*, 2009.

M. Hatanaka et al. Biological studies on the population of the saury, cololabis saira (brevoort) part 1 reproduction and growth. *Tohoku journal of agricultural research*, 6(3):227–269, 1956.

K. T. Hill, P. R. Crone, N. C. Lo, D. A. Demer, J. P. Zwolinski, and B. J. Macewicz. *Assessment of the Pacific sardine resource in 2012 for US management in 2013*. Citeseer, 2011.

R. Hillary, A. Ann Preece, and C. Davies. MP estimation performance relative to current input cpue and aerial survey data. *CCSBT Extended Scientific Committee held in Canberra*, 1309(19), 2013.

J. U. Hjorth. *Computer intensive statistical methods: Validation, model selection, and bootstrap*. CRC Press, 1993.

R. E. Huschke et al. Glossary of meteorology. 1959.

G. Huse, A. Fernö, and J. C. Holst. Establishment of new wintering areas in herring co-occurs with peaks in the âĂŸfirst time/repeat spawnerâĂŹratio. *Marine Ecology Progress Series*, 409:189–198, 2010.

R. J. Hyndman and A. B. Koehler. Another look at measures of forecast accuracy. *International journal of forecasting*, 22(4):679–688, 2006.

E. K. Jin, J. L. Kinter III, B. Wang, C.-K. Park, I.-S. Kang, B. Kirtman, J.-S. Kug, A. Kumar, J.-J. Luo, J. Schemm, et al. Current status of enso prediction skill in coupled ocean–atmosphere models. *Climate Dynamics*, 31(6):647–664, 2008.

M. J. Juan-Jordá, I. Mosqueira, J. Freire, and N. K. Dulvy. Life in 3-d: life history strategies in tunas, mackerels and bonitos. *Reviews in Fish Biology and Fisheries*, 23(2):135–155, 2013.

L. Kell, M. Pastoors, R. Scott, M. Smith, F. Van Beek, C. O'Brien, and G. Pilling. Evaluation of multiple management objectives for northeast Atlantic flatfish stocks: sustainability vs. stability of yield. *ICES J. Mar. Sci.*, 62(6):1104–1117, 2005a.

L. Kell, G. Pilling, G. Kirkwood, M. Pastoors, B. Mesnil, K. Korsbrekke, P. Abaunza, R. Aps, A. Biseau, P. Kunzlik, et al. An evaluation of the implicit management procedure used for some ICES roundfish stocks. *ICES J. Mar. Sci.*, 62(4):750–759, 2005b.

L. Kell, J. A. De Oliveira, A. E. Punt, M. K. McAllister, and S. Kuikka. Operational management procedures: an introduction to the use of evaluation frameworks. *Developments in Aquaculture and Fisheries Science*, 36:379–407, 2006.

L. Kell, M. Dickey-Collas, N. Hintzen, R. Nash, G. Pilling, and B. Roel. Lumpers or splitters? evaluating recovery and management plans for metapopulations of herring. *ICES J. Mar. Sci.*, 66(8):1776–1783, 2009.

L. T. Kell, A. Kimoto, and T. Kitakado. Evaluation of the prediction skill of stock assessment using hindcasting. *Fisheries Research*, 183:119–127, 2016a.

L. T. Kell, P. Levontin, C. R. Davies, S. Harley, D. S. Kolody, M. N. Maunder, I. Mosqueira, G. M. Pilling, and R. Sharma. The quantification and presentation of risk. *Management Science in Fisheries: An Introduction to Simulation-based Methods*, page 348, 2016b.

G. Kirkwood and A. Smith. Assessing the precautionary nature of fishery management strategies. *Fisheries and Agriculture Organization. Precautionary approach to fisheries. Part*, 2, 1995.

S. Kosaka. Life history of the pacific saury cololabis saira in the northwest pacific and consideration on resource fluctuations based on it. *Bull. Tohoku Natl. Fish. Res. Inst.*, 63:1–96, 2000.

A. Kronlund, K. Holt, P. Shelton, , and J. Rice. Current approaches for the provision of scientific advice on the precautionary approach for canadian fish stocks: Harvest decision rules. *DFO Can. Sci. Advis. Sec. Res. Doc. 2013/080.*, page 29, 2014.

H. K. C. J. Kronlund, A.R. and P. Shelton. Current approaches for the provision of scientific advice on the precautionary approach for canadian fish stocks: Section 8 management strategy evaluation. *DFO Can. Sci. Advis. Sec. Res. Doc. 2013/080.*, page 29, 2014.

H.-H. Lee, M. N. Maunder, K. R. Piner, and R. D. Methot. Can steepness of the stock–recruitment relationship be estimated in fishery stock assessment models? *Fish. Res.*, 125:254–261, 2012.

S. Mantyniemi, S. Kuikka, M. Rahikainen, L. Kell, and V. Kaitala. The value of information in fisheries management: North sea herring as an example. *ICES J. Mar. Sci.*, 66(10):2278–2283, 2009.

S. McClatchie, R. Goericke, G. Auad, and K. Hill. Re-assessment of the stock–recruit and temperature–recruit relationships for pacific sardine (sardinops sagax). *Canadian Journal of Fisheries and Aquatic Sciences*, 67(11):1782–1790, 2010.

R. D. Methot and C. R. Wetzel. Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management. *Fish. Res.*, 142:86–99, 2013.

A. Nielsen and C. Berg. Estimation of time-varying selectivity in stock assessments using state-space model. *Fish. Res.*, xxx(xxx):xxx–xxx, 2014a.

A. Nielsen and C. W. Berg. Estimation of time-varying selectivity in stock assessments using state-space models. *Fisheries Research*, 158:96–101, 2014b.

Ó. K. Pálsson, A. Gislason, H. G. Guðfinnsson, B. Gunnarsson, S. R. Ólafsdóttir, H. Petursdottir, S. Sveinbjörnsson, K. Thorisson, and H. Valdimarsson. Ecosystem structure in the iceland sea and recent changes to the capelin (mallotus villosus) population. *ICES Journal of Marine Science*, 69(7): 1242–1254, 2012.

K. Patterson, R. Cook, C. Darby, S. Gavaris, L. Kell, P. Lewy, B. Mesnil, A. Punt, V. Restrepo, D. W. Skagen, et al. Estimating uncertainty in fish stock assessment and forecasting. *Fish and Fisheries*, 2 (2):125–157, 2001.

J. Pella and P. Tomlinson. *A generalized stock production model*. Inter-American Tropical Tuna Commission, 1969.

PFMC. Coastal pelagics species fishery management plan as amended through amendment 13. 2011.

P. F. M. C. (PFMC). Report of the pacific sardine harvest parameters workshop, 2013.

E. Pikitch, P. D. Boersma, I. Boyd, D. Conover, P. Cury, T. Essington, S. Heppell, E. Houde, M. Mangel, D. Pauly, et al. Little fish, big impact: managing a crucial link in ocean food webs. 2012.

A. Preece, R. Hillary, and C. Davies. Identification of candidate limit reference points for the key target species in the wcpfc mow1-ip/03 06 nov 2012 (wcpfc-sc7-2011/mi-wp-03).

A. Punt and G. Donovan. Developing management procedures that are robust to uncertainty: lessons from the International Whaling Commission. *ICES J. Mar. Sci.*, 64(4):603–612, 2007.

A. E. Punt, D. S. Butterworth, C. L. de Moor, J. A. De Oliveira, and M. Haddon. Management strategy evaluation: best practices. *Fish and Fisheries*, 17(2):303–334, 2016a.

A. E. Punt, A. D. MacCall, T. E. Essington, T. B. Francis, F. Hurtado-Ferro, K. F. Johnson, I. C. Kaplan, L. E. Koehn, P. S. Levin, and W. J. Sydeman. Exploring the implications of the harvest control rule for pacific sardine, accounting for predator dynamics: A mice model. *Ecological modelling*, 337:79–95, 2016b.

A. E. Punt, D. K. Okamoto, A. D. MacCall, A. O. Shelton, D. R. Armitage, J. S. Cleary, I. P. Davies, S. C. Dressel, T. B. Francis, P. S. Levin, et al. When are estimates of spawning stock biomass for small pelagic fishes improved by taking spatial structure into account? *Fisheries Research*, 206:65–78, 2018.

E. Punt, Andre, T. A'mar, N. A. Bond, D. S. Butterworth, C. L. de Moor, J. A. A. De Oliveira, M. A. Haltuch, A. B. Hollowed, and C. Szuwalski. Fisheries management under climate and environmental uncertainty: control rules and performance simulation. *ICES Journal of Marine Science: Journal du Conseil*, 71(8):2208–2220, 2014. doi: 10.1093/icesjms/fst057. URL http://icesjms.oxfordjournals.org/content/71/8/2208.abstract.

S. Ralston, A. E. Punt, O. S. Hamel, J. D. DeVore, and R. J. Conser. A meta-analytic approach to quantifying scientific uncertainty in stock assessments. *Fishery Bulletin*, 109(2):217–232, 2011.

R. Schwartzlose, J. Alheit, A. Bakun, T. Baumgartner, R. Cloete, R. Crawford, W. Fletcher, Y. Green-Ruiz, E. Hagen, T. Kawasaki, et al. Worldwide large-scale fluctuations of sardine and anchovy populations. *South African Journal of Marine Science*, 21(1):289–347, 1999.

S. Shephard, A. Rindorf, M. Dickey-Collas, N. T. Hintzen, K. Farnsworth, and D. G. Reid. Assessing the state of pelagic fish communities within an ecosystem approach and the european marine strategy framework directive. *ICES Journal of Marine Science*, 71(7):1572–1585, 2014. doi: 10.1093/icesjms/fsu005. URL http://dx.doi.org/10.1093/icesjms/fsu005.

M. Simon, J.-M. Fromentin, S. Bonhommeau, D. Gaertner, J. Brodziak, and M.-P. Etienne. Effects of stochasticity in early life history on steepness and population growth rate estimates: An illustration on Atlantic bluefin tuna. *PLoS one*, 7(10):e48583, 2012.

M. C. Siple, T. E. Essington, and É. E. Plagányi. Forage fish fisheries management requires a tailored approach to balance trade-offs. *Fish and Fisheries*, 20(1):110–124, 2019.

M. Sissenwine and J. Shepherd. An alternative perspective on recruitment overfishing and biological reference points. *Can. J. Fish. Aquat. Sci.*, 44(4):913–918, 1987.

A. Smith, K. C. Navaglio, E. F. Sainsbury, and D. Smith. Decline and partial recovery of a trawl fishery in south-eastern australia., 2018.

W. Smith, A, H. T, N. F, Klaer, E. Fulton, and A. Punt. Review and update of harvest strategy settings for the commonwealth small pelagic fishery: single species and ecosystem considerations. 2015.

S. Suyama, Y. Kurita, and Y. Ueno. Age structure of pacific saury cololabis saira based on observations of the hyaline zones in the otolith and length frequency distributions. *Fisheries Science*, 72(4):742–749, 2006.

A. F. Syah, S.-I. Saitoh, I. D. Alabia, and T. Hirawake. Detection of potential fishing zone for pacific saury (cololabis saira) using generalized additive model and remotely sensed data. In *IOP Conference Series: Earth and Environmental Science*, volume 54, page 012074. IOP Publishing, 2017.

C. S. Szuwalski and A. B. Hollowed. Climate change and non-stationary population processes in fisheries management. *ICES Journal of Marine Science*, 73(5):1297–1305, 2016.

C. S. Szuwalski, G. L. Britten, R. Licandeo, R. O. Amoroso, R. Hilborn, and C. Walters. Global forage fish recruitment dynamics: A comparison of methods, time-variation, and reverse causality. *Fisheries Research*, 214:56–64, 2019.

Y. Tian, T. Akamine, and M. Suda. Variations in the abundance of pacific saury (cololabis saira) from the northwestern pacific in relation to oceanic-climate changes. *Fisheries research*, 60(2-3):439–454, 2003.

P. F. M. C. (US). *Draft Amendment 8:(to the Northern Anchovy Fishery Management Plan) Incorporating a Name Change to the Coastal Pelagic Species Fishery Management Plan, Including Draft Supplemental Environmental Impact Statement: Public Review Document.* Pacific Fishery Management Council, 1998.

K. A. Vert-pre, R. O. Amoroso, O. P. Jensen, and R. Hilborn. Frequency and intensity of productivity regime shifts in marine fish stocks. *Proceedings of the National Academy of Sciences*, 110(5):1779–1784, 2013.

H. Vilhjálmsson. The icelandic capelin stock: Capelin, mallotus villosus (mü ller) in the icelandágreenlandájan mayen area. *Rit Fiskideildar*, 13, 1994.

H. Vilhjálmsson and J. E. Carscadden. Assessment surveys for capelin in the iceland–east greenland–jan mayen area, 1978–2001. *ICES Journal of Marine Science*, 59(5):1096–1104, 2002.

C. Walters and A. Punt. Placing odds on sustainable catch using virtual population analysis and survey data. *Can. J. Fish. Aquat. Sci.*, 51(4):946–958, 1994.

C. J. Walters. Nonstationarity of production relationships in exploited populations. *Can. J. Fish. Aquat. Sci.*, 44(S2):s156–s165, 1987.

A. Weigel, M. Liniger, and C. Appenzeller. Can multi-model combination really enhance the prediction skill of probabilistic ensemble forecasts? *Quarterly Journal of the Royal Meteorological Society*, 134 (630):241–260, 2008.

C. J. Willmott and K. Matsuura. Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance. *Climate research*, 30(1):79–82, 2005.

H. Winker, F. Carvalho, J. Thorson, L. Kell, D. Parker, M. Kapur, R. Sharma, A. Booth, and S. Kerwath. Jabba-select: Incorporating life history and fisheries' selectivity into surplus production models. *Fisheries Research*, submitted.