**Performance Review of Simple Management Procedures**

**Authors:**

Tom Carruthers, Laurence Kell, Doug Butterworth, Mark Maunder, Helena Geromont, Carl Walters, Murdoch McAllister, Richard Hillary, Toshihide Kitakado, Campbell Davies.

**Possible co-authors:**

Polina Levontin

**Abstract**

Using a management strategy evaluation approach, we compare a range of new and established management procedures (MPs) for setting catch-limits in fisheries. Performance is evaluated with respect to fish life-history type, level of stock depletion, and auto-correlation in recruitment strength and data quality levels. We also identify the core sensitivities of each management procedure with respect to simulated population dynamics and observation processes. We find that methods making use of even relatively imprecise and biased information regarding current absolute biomass or stock depletion offer the best overall performance and that this is consistent across life-history types, data qualities and stock depletion levels. In general methods are most sensitive to biases in reported catches, the selectivity to fishing of older age classes and relatively small temporal changes in somatic growth parameters. Our results indicate that in many cases tuning MPs to specific stock circumstances is important, though this may not be viable in data-poor assessment scenarios.

**Keywords**

Management strategy evaluation, management procedure, stock assessment, simulation, fisheries management, data-poor, data-limited

**1. Introduction**

Management advice for the majority of economically important fish stocks, and increasingly also for by-catch species, is based on stock assessment. Under the stock assessment paradigm, detailed fishery dependent and independent data are used fit to fit models of stock dynamics that are then used to assess historical stock status, derive reference points and predict the likely impact of alternative management options (e.g. total allowable catches and measures to change the relative mortality of age classes). Stock assessments are updated periodically to include new data and to assess whether management recommendations require revision according to changes in estimates of exploitation level, stock status and productivity. Typically the assumptions of the stock assessment are free to be adjusted by scientists each year (Hilborn, 2003).

Management Strategy Evaluation (MSE, Cochrane et al. 1998, Butterworth and Punt 1999, Butterworth 2007) is an alternative fisheries management paradigm, which focuses on the relative performance of alternative management procedures (MP, aka, harvest strategies) to meet specified management objectives. It differs from the assessment approach in that detailed fishery data are used to condition multiple Operating Models (OMs), simulation models that represent alternative plausible hypotheses about fishery and population dynamics, rather than selecting a “best assessment” and sensitivity tests as the basis for management advice. These simulations are then used to tune, evaluate and select an MP to be used as the basis for providing management recommendations. Commonly, the data, “assessment model” and decision rule that constitute the MP are functionally much simpler than a conventional stock assessment. For example, MPs often rely primarily on recent information regarding trends in abundance and catch data (e.g. Punt and Donovan 2007). Instead of using stock assessment as the primary source of management advice, the MP/HS approach makes routine management decisions using MPs while the operating model is updated to accommodate new data (e.g. Hillary et al. 2015). There is increasing evidence that simple MPs can perform as least as well as conventional stock assessments (Geromont and Butterworth 2014a).

MSE typically adds stability to the management decision process by first identifying realistic management objectives through stakeholder participation, followed by a thorough evaluation of trade-offs arising under alternative MPs, when accounting for a diverse range of plausible past and future scenarios and sources of uncertainty (e.g. Rockmann et al. 2012). MSE can also be used to guide the scientific process by identifying where the reduction of scientific uncertainty will improve, or not, performance in achieving management objectives and so help to ensure that expenditure is prioritised to provide the best research, monitoring and enforcement (Fromentin et al. 2014).

In most cases MPs have been developed in management settings where data would generally be deemed to be sufficiently informative to support conventional stock assessments (e.g. the Commission for the Conservation of Southern Bluefin Tuna, CCSBT 2011; South African hake, Rademeyer et al. 2008). Recently however there has been an increasing interest in quantitative methods to support management decision making in data-limited fisheries[[1]](#footnote-1). Approaches such as Depletion-Corrected Average Catch (DCAC, MacCall 2010), Depletion-Based Stock Reduction Analysis (DB-SRA, Dick and MacCall 2011) and fishing at a fixed fraction of natural mortality rate (Martell and Walters, 2002) are currently used in managing data-limited fisheries and have been subject to simulation testing (Carruthers et al. 2014). Many of these are closely related to stock assessments; they are based on comparable biological models and rely on many of the same assumptions (e.g. DB-SRA). Recent research has sought to develop and test new data-limited MPs that require fewer assumptions about underlying population dynamics and make management recommendations using only recent time-series data such as catches and catch per unit effort data (e.g. Geromont and Butterworth 2014b, Maunder 2014).

In this paper we compare the performance of a suite of candidate MPs that have been described in the primary and grey literature. We also include new approaches that operate on alternative fishery information. A number of the candidate MPs of this paper have been parameterized according to simulations specific to a particular management scenario. We refer to these as ‘tuned’ MPs. Examples include MPs applied in the management of Southern Bluefin Tuna (Hillary et al. 2015) and the index slope and target MPs described by Geromont and Butterworth (2014a). We also test ‘generic’ MPs that are intended to operate over a wider range of scenarios by attempting account for broad information about stock life-history or sustainable exploitation rate (for example fishing at a fixed fraction of the natural mortality rate). The generic MPs tested in this paper are new approaches that have higher data demands than the tuned MPs and rely on recent absolute abundance data.

We identify relevant performance metrics (i.e. summary statistics) and describe a reference set of OMs. The aims of this paper are to reveal the performance trade-offs amongst the MPs, identify the core sensitivities of the MPs to their inputs and the properties of the OMs, identify important interactions between MPs and life-history / data quality and confirm the potential performance advantages of simple MPs relative to more detailed stock assessments.

We evaluate the following hypotheses:

(1) The performance of the candidate MPs is specific to stock dynamics; in particular (a) longevity, (b) temporal variability in productivity and (c) stock status.

(2) Generic MPs that do not necessarily require tuning according to an OM fitted to available data and provide comparable or better performance to approaches that are currently applied in (a) data rich and (b) data-limited settings.

**2. Methods**

**2.1. Generic management procedures**

In this paper a number of new MPs are described that aim to use recent observations of absolute biomass *B*, and total annual catches *C*, to infer surplus production *S,* and therefore stock level relative to a productive stock size (Figure 1, panel a). Since these MPs require estimates of recent absolute biomass they are appropriate to data-rich situations where a conventional assessment has been used to quantify the catchability coefficient *q*, that scales a relative abundance index to predicted absolute biomass or alternatively data-moderate settings where a fishery independent survey is available that can provide an estimate of biomass in absolute terms that is not appreciably biased.

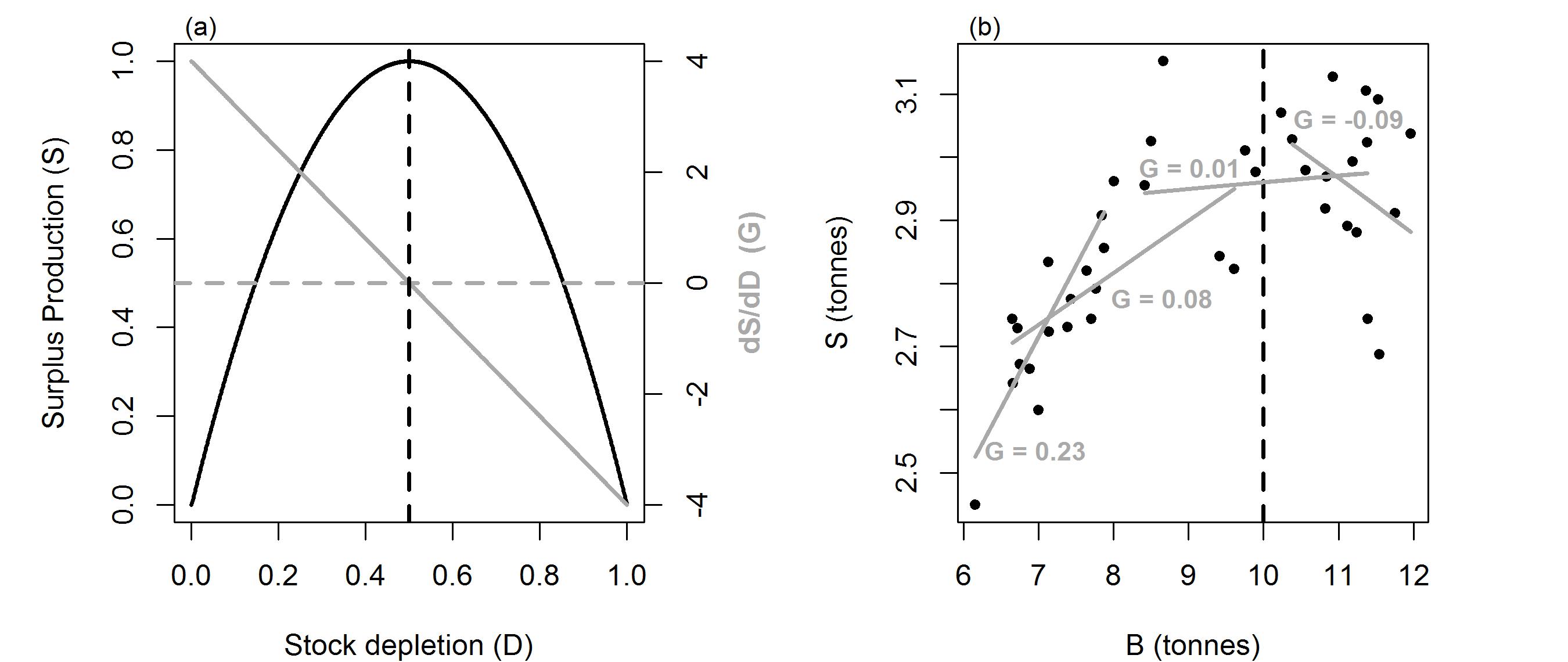


Figure 1. The theoretical model of gradient (*G*) in surplus production (*S*) with respect to stock depletion (*D,* biomass relative to unfished levels) according to the Schaefer production model (panel a). The vertical dashed line represents the simulated level of biomass at most productive stock size. Observations of catch and biomass (*B*) may be used to infer surplus production (*S*): *Sy*=*By*-*By-1*+*Cy-1* (panel b). In theory the gradient in surplus production with biomass can be used to modify management recommendations to move stock levels towards more productive stock sizes where *G* is close to zero (horizontal dashed line). Panel b illustrates estimates of *G* (grey lines) for four simulated time periods.

The generic MPs of this paper rely on the same calculation of surplus production *S*:

(1)

The gradient in surplus production with biomass *G* (d*S*/d*B*), may be used to move the stock towards a more productive stock size where *G* ≈ 0 (Figure 1). Negative *G* values imply that stock levels are above the most productive stock size; positive *G* values imply the stock is below productive stock sizes, while *G* values close to zero suggest that the stock is close to productive stock size. This concept does not rely on the assumption of a fixed position of productive stock size relative to unfished levels and may be able to adapt to temporal shifts in productivity, though the degree to which this is possible will depend heavily on the frequency and duration of productivity shifts and time lags caused by specific life-history characteristics such as age-at-maturity. All of the generic MPs have parameters that may be tuned to specific case-studies, such as the sensitivity of management updates to changes in *G* and the number of years of data used to calculate *G*. However, in this application we adopt rules with fixed parameter levels that are intended to operate over a range of population and fishing scenarios. All MPs tested in this paper are listed in Table 1. The equations of the generic MPs are presented in Table 2.

Table 1. Overview of the candidate management procedures and their data requirements.



Table 2. The equations of the generic management procedures. *Q*: a quota recommendation. *C:* a total annual catch observation. *B*: an absolute annual biomass estimate. *D*: an estimate of current stock depletion (biomass relative to unfished). *B0:* unfished biomass*,* *I*: an annual relative abundance index or catch rate observation. *M* and *FMSY/M* are imperfectly known simulated values of natural mortality rate and fishing rate at maximum sustainable yield relative to natural mortality rate.

|  |  |  |
| --- | --- | --- |
| **MP Name** | **Parameter values** | **Quota calculation** |
| Gcontrol | *gU*= 0.5  *gL*= 2 | *Gy* is the slope in surplus production *S* with biomass over the last ten years  where and are the biomass and catch predicted by a first degree Loess smoother with smoothing parameter alpha = 0.75, fitted to absolute abundance and catch observations over the last ten years |
| Rcontrol | As Gcontrol except:  where *r* is the demographically derived prior for intrinsic rate of increase (McAllister et al. 2001) and *D* is stock depletion (*By*/ *B0*) |
| Rcontrol2 | As Rcontrol except: |
| DynF |  | ,  As Gcontrol, *Gy* is the gradient in *S,* with biomass over the last 7 years. |
| Fadapt | ,    As Gcontrol, *G* is the slope in *S*, with biomass over the last 7 years. |
| SPslope |  | , , |
| SPmod |  |  |

**2.2. Tuned management procedures**

In many MSE applications MPs are tuned through simulation testing to achieve the pre-specified management objectives. The tuning is carried out over all OMS or a reference OM (CCSBT, 2011). In this simulation evaluation we choose not to re-tune these MPs to new simulated data. Rather we test the methods with their published parameter values and where possible choose more than one parameterization to span a range of biological precaution. In this way we aim to evaluate the structure of the MP rather than its exact parameterization, though we note that it is not always possible to treat these factors in such an independent manner. The tuned MPs of our analysis come from three sources: the MSE for Southern Bluefin Tuna (CCSBT, 2011) and two recent papers by Geromont and Butterworth (2014a, b) that seek to identify generic MPs for data-limited fisheries and stock-specific MPs for some data-rich fisheries, respectively. These tuned-MPs may be particularly appropriate for data-limited settings as they require only recent relative abundance data (or catch rate data) and recent annual catches (the equations for these methods can be found in Table 3).

In this paper we evaluate a simplified version of the second Southern Bluefin Tuna MP (SBT2) that modifies quotas to reach a predefined target catch level. The published rule derives target catches from an equation with parameters tuned to the specific SBT simulations, and involves both filtering of the two MP input indices via a two-stage relative abundance population model and both trend and target characteristics(CCSBT, 2011). In order to make the rule operate in this simulation framework we assumed that the target catch was MSY. To simulate imperfect knowledge in MSY as the target catch level we added bias to the true simulated MSY level. In this way we evaluated a more general version of the SBT2 MP.

Table 3. The equations of the tuned management procedures. *Q* is a quota recommendation, *C* is a total annual catch observation, *B* is an absolute annual biomass estimate, *I* is an annual relative abundance index or catch rate (CPUE) observation, *R* is an estimate of recruitment strength, *y\** refers to the first year in which the MP was implemented, *nt* is the number of length observations in a particular year *t*, *Lobs* is a length observation of a caught fish and *MSY* is Maximum Sustainable Yield (though this is not known perfectly).

|  |  |
| --- | --- |
| **MP Name** | **Quota calculation** |
| SBT1 | where *Gy* is the gradient of log CPUE over the last 10 years |
| SBT2 | , ,  , |
| GB\_CC | where *Cave* is the mean historical annual catch |
| GB\_slope | where *Cave*is the mean historical annual catch and *G* is the gradient of log CPUE over the last 5 years. |
| GB\_target | , ,  where *Cave*is the mean historical annual catches and *Itarg* is the abundance index corresponding to (true BMSY relative to unfished that is not known perfectly). |

Table 3 continued.

|  |  |  |
| --- | --- | --- |
| **MP Name** | **Parameter values** | **Quota calculation** |
| CC1 | *x* = 1 |  |
| CC4 | *x* = 0.7 |
| LstepCC1 | *x* = 1 | , , |
| LstepCC4 | *x* = 0.7 |
| Ltarget1 | *x* = 1  *v* =1.05 | ,  , , |
| Ltarget2 | *x* = 0.8  *v* = 1.15 |
| Islope1 | *x* = 0.8  *λ* = 0.4 | where *s* is the gradient of log CPUE over the last 5 years |
| Islope2 | *x* = 0.6  *λ* = 0.2 |
| Itarget1 | *x* = 1  *v* = 1.5 | ,  , ,  , *Itarg* is the abundance index corresponding to (true BMSY relative to unfished that is not known perfectly). |
| Itarget4 | *x* = 0.7  *v* = 2.5 |

**2.3. Reference methods**

In order to frame the performance of the generic and tuned MPs we include a series of reference MPs that represent conventional stock assessments or methods currently used in the management of data-limited stocks. We include a delay-difference assessment model (Deriso 1980, Schnute 1985) to represent a data-rich assessment. More complex stock assessment such as statistical catch-at-age models were too computationally intensive to be included in this MSE framework. Additionally it may be argued that detailed stock assessments involve many subjective decisions regarding data processing and model assumptions that cannot be properly replicated in an automated simulation evaluation. The purpose of including the delay-difference assessment approach is to evaluate the performance of approaches that rely on a long time-series of relative abundance data that are assumed to represent the exploitation history of the stock. We also test a variant of the delay-difference model that is combined with the ’40-10’ harvest control rule. Under this rule the stock is not fished when stock size is below 10% unfished biomass and fished at *FMSY* above 40% of unfished biomass. Between 10% and 40% unfished levels exploitation rate follows a linear increase from 0 to 100% *FMSY*.

The first of two data-limited methods included in the performance evaluation was DCAC (MacCall 2009) provides an estimate of “sustainable catch” based on an estimate of average annual catch four inputs: depletion, the ratio of *FMSY/M*, *M* and *BMSY/Bunfished*. DCAC aims to calculate a sustainable catch accounting for the removal of the “windfall harvest” of less productive biomass that may have occurred as the stock became depleted (the equations are included in the Appendix B.1.). DCAC is currently used by the Pacific Fishery Management Council to set catch-limits for data-limited stocks (PFMC, 2010).

The second data-limited method Fratio, simply fishes at a constant exploitation rate that is a fixed fraction of the natural mortality rate (Gulland 1971, Walters and Martell 2002). The North Pacific Fishery Management Council uses an Fratio method for managing stock complexes in situations where stock assessments are not available (NPFMC 2012, 2013). Under the Fratio method a catch limit is simply the product of the estimate of natural mortality rate *M*, the ratio of *FMSY/M* and a current estimate of absolute biomass. Since absolute biomass estimates are also a requirement of all of the generic MPs, the simpler Fratio MP provides a useful comparison.

**2.4. Operating model structure and simulation design**

The operating model is an age-structured, spatial population dynamics model of identical structure to that of Carruthers et al. (2014) (a full description of the operating model is given in Appendix A).

We constructed operating models using a factorial design encompassing 36 sets of operating model assumptions. The four factors were (1) life history with 3 levels, (2) temporal autocorrelation in recruitment with two levels, (3) starting stock depletion with two levels and (4) data quality with three levels (Table 4). For each of the 36 combinations we carried out 1000 simulations for each MP. Each simulation was then projected forward for 40 years adopting the quota recommendations of the management methods. We did not simulate implementation error and assumed that prescribed catches would be taken exactly up to a maximum instantaneous fishing mortality rate of 90%. The MPs were rerun and the quota updated every three projected years to approximate a typical assessment cycle.

Table 4. Overview of simulation design. Parameter ranges represent the lower and upper bounds of a uniform random variable. SD refers to the standard deviation. In the case of biases, the SD defines a range of potential biases that may be sampled among simulations. For example, given an SD of 20% three sampled biases could be 0.84, 1.05, 1.34 representing a consistent moderate negative bias for all observations of simulation 1, small positive bias for simulation 2 and a stronger positive bias for simulation 3. 

Three population life-history types of varying longevity were simulated based on the outputs of data-rich stock assessments for Pacific herring (DFO 2012), the eastern stock of Atlantic bluefin tuna (ICCAT 2012) and Pacific canary rockfish (Wallace and Cope 2011). The intention was to characterize broad life-history types rather than the status of particular stocks, so that the assessed level of stock depletion was not simulated.

Previous simulation evaluations have indicated that MPs can perform very differently for simulations depleted below *BMSY* (Carruthers et al. 2014), which is arguably the most important population level for evaluating performance, at least from a general policy perspective. We simulated two ranges of initial depletion (biomass relative to unfished): 2.5-15% and 15-35%, that approximately correspond to less than *BMSY*/2 and between *BMSY*/2 and *BMSY*, respectively.

We simulated autocorrelation in recruitment in order to evaluate the performance of the candidate MPs in situations where stock productivity varies over time (Appendix Eqn.A.1). This may not fully reflect step-changes in recruitment that have been observed in some fishery settings (Vert-pre et al. 2013). However, simulating recruitment autocorrelation does not invalidate the derivation and subsequent use of MSY reference points which are fundamental as inputs to several MPs and also in assessing performance of the MPs.

Bias and imprecision in the knowledge of the simulated system were generated for all of the inputs to the MPs (e.g.,natural mortality rate, current absolute biomass)(Appendix Table C.1 includes a summary of the observation error model). We simulated three different data quality levels corresponding to perfect information, data-rich and data-limited assessment settings. The perfect information simulations assume no error in knowledge of inputs to MPs, essentially removing the observation model and revealing performance with respect to operating model parameters only, such as age-at-maturity and recruitment compensation, and the specific levels of process error (recruitment variance in this case). Data-rich simulations assume that inputs to MPs are known imperfectly but may be subject to moderate bias and imprecision. For example among simulations, consistent bias in annual catches is sampled from a lognormal distribution with mean 1 and coefficient of variation *ϒC*, of 20% (Table 4). For example, a value of 0.82 could be drawn for the first simulation and be used as consistent downward bias for all observed catches. In addition to this bias we superimpose imprecision that is log-normal error in annual catch observations per simulation *σC*, allowing us to separate the effect on performance of bias and imprecision in inputs. Data-limited simulations include higher levels of bias and imprecision in the inputs to the MPs in an attempt to simulate the poorer quality of data that are typical in a data-limited setting.

Simulation testing was carried out in the R statistical environment (R Core Team 2014) using the R-package Data-Limited Methods toolkit (DLMtool 2014, v1.35). The package is freely available and includes all of the operating models and management procedures evaluated in this paper (computer code for reproducing our results are available online at https://github.com/tcarruth/Carruthers-et-al-2014-MP-MSE).

**2.5. Performance criteria**

Performance was summarized in regard to four performance metrics: long-term Expected Yield (EY), probability of achieving Pretty Good Yield (PGY), average annual Variability in Yield (VY) and the Probability of Severe Depletion (PSD) (the definition of the performance metrics is given in Table 5).

It is possible for an MP to obtain relatively high total yields in an unsustainable way by chronically depleting the stock (a mining strategy). We therefore focused on yields over the final five years of each projection, which are generally low in situations where the MP has mined the stock. The absolute yield in the last five years can vary widely among projections due to differences in simulated productivity and initial stock depletion. To standardise long-term yields on a simulation-by-simulation basis we used numerical optimization to find the fixed fishing mortality rate strategy (*Fopt*) that led to the maximum yield over the last five projected years given perfect information about future recruitment. For each simulation we then divided the absolute yields of the MPs by this ‘upper bound’ reference yield to obtain long-term relative yield Y (Table 5). We define two related quantities: EY and PGY. EY is simply mean Y across all simulations. Since EY may be high even when Y scores vary widely among simulations, we also calculate the metric PGY which is the fraction of Y values above 50%.

In most fisheries it is not desirable for catch limits to vary strongly between years. To address this we include an additional yield metric VY, which represents the average annual variability in yield. Finally we calculated PSD: the fraction of instances in which the biomass in the final year of the projection dropping below 25% of true simulated *BMSY* levels (typically between 5% and 10% of unfished biomass).

Table 5. Performance metrics of this simulation evaluation and their definition.

|  |  |  |
| --- | --- | --- |
| **Performance metric** |  | **Derivation** |
| Expected long-term yield | EY |  |
| Pretty good yield | PGY | , |
| Average annual variability in yield | VY |  |
| Probability of severe depletion | PSD | , |
| Where Y is the long term yield for a management procedure *j*, and simulation *i*, that is calculated:  where *ny* is the number of projected years and *C* are the true simulated catches of an MP *j*, or the *Fopt* strategy that is the fixed fishing mortality rate that maximizes catches over the final five years, *ni* is the number of simulations, *Bny* is the biomass in the final year of the simulations, and *BMSY* is the true simulated biomass at maximum sustainable yield. | | |

**2.6. Quantifying value of information**

Using multiple linear regression we aimed to relate long-term yield (Y*,* a metric of economic value) of the MPs to operating model variables in addition to imprecision and bias in the inputs to MPs. For each MP, two linear models were fitted in which Y was the dependent variable: (1) a linear model in which all operating model parameters were independent variables and (2) a linear model in which all observation error model parameters were independent variables. In the simulation, each independent variable (e.g. natural mortality rate *M*, observation error in catches *σC*) was sampled from a range of values that were considered to be credible *a priori* (Table 4 and Table App.A.1). In quantifying value of information we chose to standardize each independent variable to have a minimum value of 0 and a maximum value of 1. This standardization allows for the comparison of the slopes in long-term yield across independent variables over their respective credible ranges.

**3. Results**

The performance trade-offs among the MPs are illustrated in Figures 2 and 3. These plots feature the top ten performing methods defined as the shortest Euclidean distance from 100% EY, 0% PSD, 100% PGY and 0% VY. In doing so we scale performance linearly and weight metrics equally. This not intended to be a defensible representation of a utility function, but rather a means of simplifying the presentation of results to those MPs that were the most promising.

**3.1. Overall performance of generic and reference MPs**

Regardless of life-history, initial stock depletion level or quality of data, of the MPs that rely on estimates of absolute biomass or depletion, five consistently outperformed the others: the reference methods Fratio, DD, DD4010, and two adaptive fishing rate rules, DynF and Fadapt (Figures 2-3). An exception to this rule was the relatively high average annual variability in yields (VY) of the delay-difference assessment methods DD and DD4010 (Figure 3).

It is striking that for these simulations simply aiming to fish at a fixed fraction of natural mortality rate (Fratio) can be expected to perform as well or better than simple stock assessments relying on long time-series of relative abundance data or dynamic rules that attempt to modify the exploitation rate to suit changes in productivity (DD, Fadapt). This is particularly surprising given that we simulated reasonably large potential bias in current estimates of absolute biomass and FMSY: in the data limited observation model, observed absolute biomass estimates could easily be a factor of three different from the true values simulated. Dynamic rules DynF and Fadapt obtained comparable expected long term yield (EY) to the fixed fishing rate Fratio MP, but generally led to higher average annual variability in yield and a lower probability of obtaining pretty good yield (PGY) (Figure 3). Since DynF and Fadapt are based closely on the Fratio method, but attempt to incorporate dynamic information about surplus production, these findings imply that there may be difficulties in quantifying surplus production reliably using Equation 1. This is supported further by the generally poor performance of the Gcontrol, Rcontrol, Rcontrol2, SPslope and SPmod MPs that rely more heavily on inferred surplus production. In many cases these MPs failed to rank in the top ten and are absent from many of the panels of Figures 2 and 3. While SPslope could provide high expected yields, the chance of achieving pretty good yield was generally low.

**3.2. Overall performance of tuned MPs**

The CPUE index gradient methods Islope1 and Islope4 were standout performers in the class of tuned MPs that do not require estimates of current absolute biomass or stock depletion. These MPs could achieve expected yields almost as high as MPs using a direct estimate of current absolute biomass, for example in the data poor simulations (Figure 2, panels c, f, j) and the less depleted bluefin tuna simulations (Figure 2, panels d and e).

In general the SBT1, LStepCC1, Ltarget4, CC1, CC4, GB\_CC, and GB\_slope MPs performed consistently poorly in terms of expected yield and probability of severe depletion. These MPs are conspicuously absent from the top ten MPs plotted in Figures 2 and 3.

**3.4. Effect of life-history on performance**

The performance ranking of the MPs remained relatively similar across life-history types. The exceptions to this were the tuned MPs in the herring case study where the LStepCC4 and SBT2 rules provided reasonable overall performance (Figure 2), in some cases providing comparable expected yield and pretty good yield to the generic MPs that require additional absolute biomass data. The herring simulations also favoured GB\_target which could perform as well as the more data intensive assessments.

Interactions between life-history types and MP performance were relatively limited. Notable exceptions include the much higher inter-annual variability in yield of the delay-difference assessments DD and DD4010 for herring and bluefin and the higher ranking of Islope1 and Islope4 MPs for the herring stock. Unlike the rockfish and bluefin simulations DCAC performed well for the herring life history type but only in simulations that began above 15% unfished levels.

**3.5. Effect of auto-correlation in recruitment**

The performance of the MPs was virtually identical in simulations where high and low auto-correlation in recruitment was simulated (the trade-off plots for the high-recruitment autocorrelation simulations are available in the Appendix Figures App.D.1 and App.D.2). Typically the absolute difference in performance metrics was less than 5% and the rankings of the MPs were very similar.

**3.6. Effect of initial stock depletion**

The tuned MPs provided much better performance for simulations starting at intermediate depletion levels (15%-35% unfished). For simulations starting below these levels, tuned MPs generally had much higher probability of severe depletion and provided substantially lower relative long-term yields (Figures 2 and 3). Perhaps not surprisingly the generic MPs were more consistent across depleted and less depleted simulations. As has been identified before (Carruthers et al. 2014), DCAC performed reasonably well for stock sizes above heavily depleted, but below 15% of the unfished level it led to a high probability of severe depletion and much lower expected yields than potentially attainable. Consequently DCAC did not feature in the top ten methods for any of the simulations which started from a more depleted stock level.

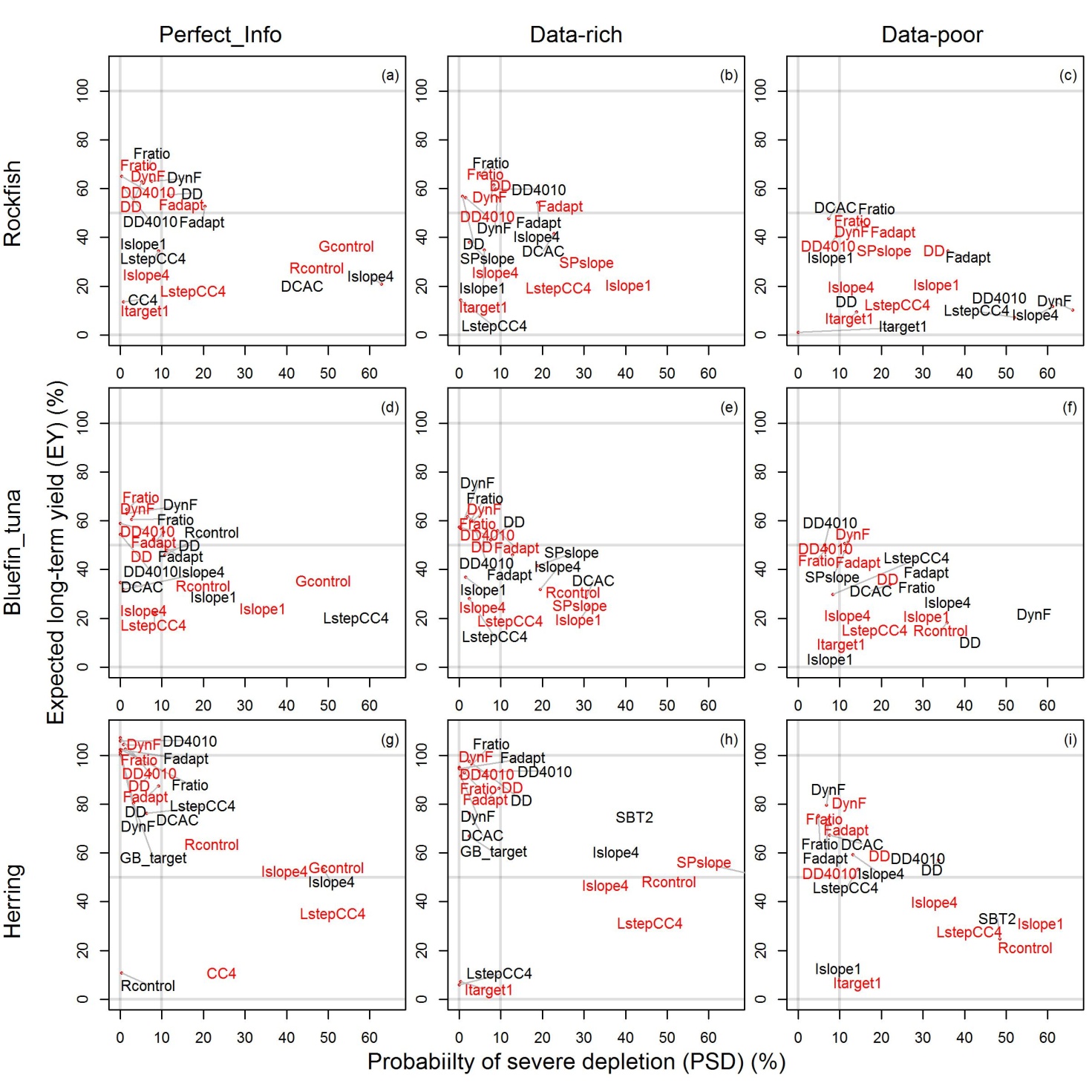


Figure 2. Performance for simulations with low temporal auto-correlation in recruitment Trade-off between probability of severe depletion (PSD, fraction of simulations ending below 10% *BMSY* ) and expected long-term yield (Y). The top-ten performing MPs for initial depletion below 15% of the unfished level are shown in red, while the top-ten for initial depletion between 15% and 35% are indicated in black..

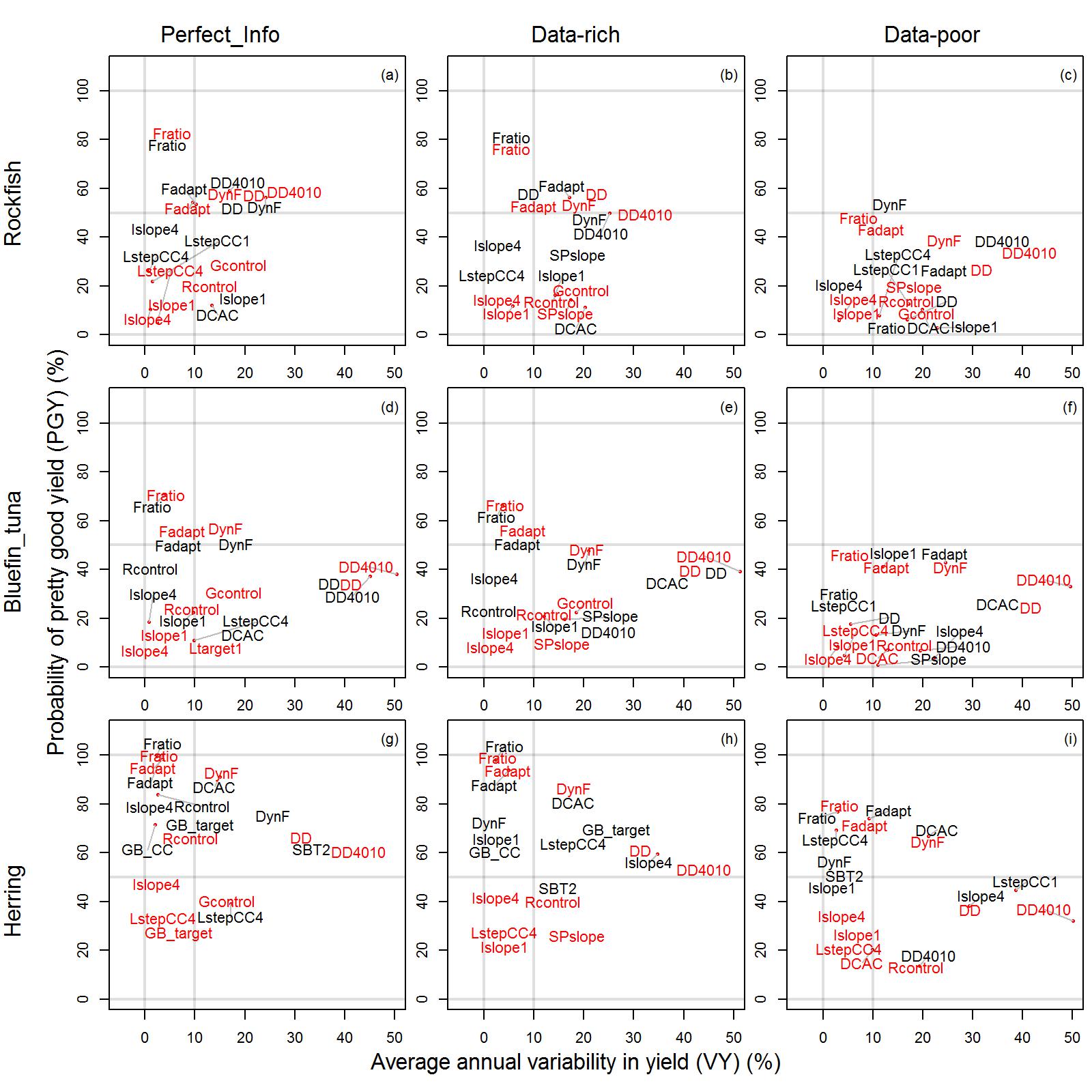
****

Figure 3. Performance for simulations with low temporal auto-correlation in recruitment. Trade-off between average annual variability in yield (VY) and pretty good yield (PGY, fraction of simulations obtaining over 50% relative yield). Plotted in red are the top-ten performing MPs given starting depletion below 15% unfished levels. Plotted in black are the top-ten performing MPs given starting depletion between 15% and 35% of unfished levels.

**3.7 Effect of data quality on performance**

Simulating poorer quality data (Table 4) did not lead to large performance differences for most of the MPs. In many cases moving from perfect information to the data-rich observation error model led to very similar results. The difference in performance among data-rich and data-limited observation error models was more pronounced but still marginal relative to the variability among MPs. This finding suggests that the performance picture as presented in Figures 2 and 3 may be largely driven by the theoretical inconsistency between MPs and the operating model. The interaction between data quality and MP performance was sometimes counter-intuitive. For example the expected long-term yield of DCAC actually improved with declining data quality whilst the probability of achieving pretty good yield using this MP declined. In terms of achieving pretty good yield, the LstepCC1 and LstepCC4 MPs appeared to offer much better performance relative to the other MPs for the data-poor observation model (Figure 3).

**3.8 Value of information and sensitivity analysis**

For most MPs the observation process that most strongly determined long-term yield was bias in observation of annual catches (Table 6). The sensitivity to bias in catch observations varied amongst MPs. For example the SBT2 MP was three times as sensitive to bias in catches as the SBT1 MP. The generic MPs were sensitive to other observation model variables that impact calculations of surplus production. In previous simulation evaluations, imprecision in catches and abundance indices has shown relatively little impact on the performance of data-limited assessment methods (Carruthers et al. 2014). However arising from the method used to calculate the surplus production (Eqn 1), generic MPs such as Rcontrol and SPslope are particularly sensitive to imprecision in catches and the relative abundance index respectively.

Despite simulating relatively trivial changes in the von-Bertalanffy growth parameters (on average less than   
+/- ¼ percent per year in maximum growth rate K and maximum length L-infinity)(Appendix Table A.1) these were amongst the strongest determinants of long-term yield (Y) for the MPs tested (Table 6). The starting level of stock depletion and related quantities such as the trajectory in fishing effort were also strong determinants of yield among MPs. The best performing MPs DD, DD4010, Fratio, Fadapt and DynF were most sensitivity to age selectivity dynamics (the relative selectivity of age classes to fishing). In particular the selectivity of the maximum age class (controlling the gradient in descending limb of the selectivity curve) was a key determinant. When this is low, exploitation of older age classes is lessened and productive biomass is essentially protected (cryptic biomass). It is therefore not surprising that in most cases there is a strong positive relationship with yield as most MPs implicitly assume full selection of older age classes.

Table 6. The sensitivity of long term relative yield (Y) in relation to operating model variables and bias and imprecision in the inputs to MPs. For each MP, two linear models were fitted in which Y was the dependent variable: (1) operating model parameters were independent variables and (2) observation error model parameters were independent variables. The table includes a row for each MP in which three variables of the operating model and up to two variables of the observation model are named that corresponded to p-values below 5%. For each named variable the slope in relative yield is also included indicating the sensitivity of Y with respect to the credible range of each independent variable. Slopes more extreme than =/- 25% are highlighted in bold. Slopes more extreme than +/-50% are shaded grey.



Table 6 continued.



**4. Discussion**

We hypothesized that the performance of the candidate MPs is specific to stock dynamics; in particular (1a) longevity, (1b) temporal variability in productivity and (1c) stock status. Our simulations indicate that for certain MPs, this may be rejected since they consistently outperformed other MPs regardless of the longevity of the simulated population, the degree of recruitment autocorrelation and the status of the stock. These were generic (e.g. DynF, Fadapt) or reference MPs (e.g. DD, Fratio) that require current data about absolute abundance or stock depletion. This result may simply support the view of many readers, that these data are particularly valuable and the related MPs should be considered quite separate from those that do not rely on these data.

The performance rankings of the MPs were similar for the moderately long-lived bluefin tuna simulations and the long-lived canary rockfish simulations, indicating that dynamics for shorter-lived species such as herring represent a distinct management problem (providing support for hypothesis 1a). The range in performance among the top performing MPs was wider for the data-poor simulations, particularly for the rockfish and bluefin tuna life-history types. This implies that MP selection may be more important for these cases.

We attempted to address hypothesis 1b by simulating autocorrelation in recruitment. Our simulations provided no evidence for a substantial effect on the absolute performance of the candidate MPs. This was perhaps surprising since, theoretically at least, the candidate MPs rely to a varying degree on assumptions of stationary stock productivity. For example, such assumptions are central to reference MPs such as the delay-difference assessments DD and DD\_4010. It is likely however that our simulations did not fully capture the abrupt and then lasting shifts in productivity that have been observed for many stocks (Vert-pre et al. 2013). Future testing should seek to re-evaluate performance of MPs subject to simulations that better reflect such regime shifts.

While reference MPs such as Fratio appear unaffected by stock status (hypothesis 1c), this cannot be said for tuned MPs that were generally more sensitive to particular depletion levels and life-history types. For example the SBT2 rule that otherwise performed relatively poorly, could perform well in the herring simulations at moderately depleted stock sizes (between 15% and 35% unfished). The trajectory in fishing effort was an important determinant of the performance of the tuned MPs. These results confirm the need to re-tune such MPs to status and exploitation history on a stock-by-stock basis.

The simulations generally support hypothesis 2 that non-tuned MPs can provide comparable or better performance to approaches that are currently applied in both (a) data rich and (b) data-limited settings. With respect to hypothesis 2a, this conclusion is largely driven by the relatively good performance of the data-moderate Fratio reference MP that is applied in setting catch limits for data-limited stock complexes in Alaska (e.g. sculpins NPFMC 2012). In the simulations Fratio could often out-perform the more data intensive delay-difference assessment over a range of data-qualities, depletion levels and life-history types. This suggests that if current absolute biomass can be estimated from a survey, or alternatively a relative abundance index can be scaled to absolute biomass by stock assessment, relatively good performance may be obtained from a simple and transparent Fratio approach.

In evaluating hypothesis 2b we compare the candidate MPs to DCAC that is currently used by the Pacific Fishery Management Council to set catch-limits for data-poor stocks (PFMC, 2010). As has been found previously (Carruthers et al. 2014) DCAC can lead to chronic overfishing at very low stock sizes (below 15% unfished levels in this analysis). In these circumstances Islope4 and LstepCC1 MPs often outperformed DCAC by a substantial margin. At more modest levels of stock depletion the same MPs provided comparable performance to DCAC, indicating that tuned MPs could be applied more widely, particularly as interim approaches as additional data become available. It should be noted that in some respects our simulations constitute an unfair evaluation of DCAC which was designed primarily as an interim approach to setting catch limits for relatively long-lived stocks (natural mortality rates less than 20%).

As with simulation testing in general, it is necessary to underline the role of the operating model and observation model specification in determining the relative performance of the candidate MPs. For example the data-poor simulations assume that the bias in absolute biomass estimates is sampled from a lognormal distribution with mean 1 and standard deviation 1.5 (Table 4). This results in 80% of the simulated bias between 1/5 and 5 times that of the true simulated biomass. It may be argued however that in many data-rich and data-limited settings bias in estimates of absolute biomass may be more extreme. Consequently our simulation evaluation may unfairly favour MPs such as Fratio that rely on current absolute biomass estimates. We counter this problem in two ways: we expose the sensitivity of long-term yield performance to the various inputs to identify those inputs that are critical in determining performance (Table 6).

An ongoing problem in the development, testing and adoption of MPs is that they are typically established using specific simulations that are often difficult to reproduce. In a new fisheries management setting it is therefore difficult to comparatively evaluate a wide range of candidate MPs and select an appropriate MP. In an attempt to address this issue we identified a reference set of simulations using software and code that are freely available. New MPs may be tested within the same framework, and results can be published that will be directly comparable to our results. An additional benefit is that our analysis may be modified to suit particular requirements. For example in a data-limited setting in which new data are to be collected, managers may seek MPs for use over a shorter interim period, and evaluate performance over for fewer projected years. Since the simulation data are reproducible exactly, readers can also frame the results using custom performance metrics that are tailored to their particular management framework.

In a data-rich setting it is clear that tuning an MP to simulations may have large benefits in terms of performance as demonstrated by the variable performance of MPs such as SBT2 across life-history types and depletion levels. Another major advantage of tuning is to standardize, to some degree, the broad management objective-level performance of a suite of candidate MPs. For example, the SBT MP as adopted (CCSBT, 2011; a more complex combination of SBT1 and SBT2 from this paper) was one of a suite of candidates – biomass dynamic, empirical, model-based – and all were tuned to meet the rebuilding objectives for the reference OM. For one, it ensures a base-level performance measure: any candidate MP can achieve the objectives for the most probable scenarios. Secondly, because all the candidate MPs have the same reference performance benchmark, “fair” comparisons on alternative robustness scenarios can actually be performed However, in data-poor settings it may not be clear how to specify a suitable operating model since, by definition, depletion is unknown. It may be necessary to simulate a wide-range of current stock depletion and provide suitable diagnostics of sensitivity in performance to initial depletion levels. Additionally, it might make more sense in data-poor scenarios to be broader in scope for both the nature of the tuning and the performance criteria. In data-poor settings in the real world, MSY is somewhat irrelevant even if it remains a policy objective, as we are never likely to know what it is and, hence, meaningfully measure performance in relation to it. Simpler criteria may be more appropriate, such as requiring MPs for (suspected to be depleted) stocks to be able to increase stock abundance (on average) without unduly decreasing catches; or focussing more on *status quo* scenarios where, for example, the simpler MP is a temporary management measure until better data are collected to allow future, more detailed MP or assessment work.

In this simulation evaluation all of the MPs provide management recommendations that are implemented perfectly up to a maximum harvest rate of 60%. Clearly scientific quota recommendations are rarely followed exactly due to a range of management considerations and fishery dynamics. For example, managers may deliberately apply a degree of quota inertia to prevent sudden declines in catch limits, it may not be profitable to fish to very low stock sizes, and a lack of enforcement may lead to quota overages. Considerations such as these may strongly affect the trade-offs presented in this paper, and it is important that future simulation evaluations attempt to tackle these issues. Some hypotheses can be proposed. Reduction of fishing effort as catch rates decline is likely to reduce the frequency of stock collapses leading to more comparable performance among MPs, reducing the importance of MP selection. Enforcement is likely to vary for alternative management regimes such as catch-limits, size-limits, gear restrictions, effort controls and spatial closures. Constructing credible enforcement models may be challenging (Coelho et al. 2013) but could strongly alter MP selection if, for example, there was a higher propensity for violation of catch-limits than gear restrictions.

Catch-limits have relatively high information requirements and are therefore most appropriate in data-rich settings. However many national fishery management organisations are now expected to provide catch-limits for all fisheries in a fishery management plan with the exception of some short-lived stocks (e.g. USA, Australia, New Zealand). For this reason we focused in this paper on MPs that provide catch-limits. However input controls such as gear restrictions and fishing effort can be expected to provide superior performance to output controls such as catch-limits in a wide range of fishery scenarios, and may be particularly appropriate in data-limited settings. Future simulation testing should be extended to include MPs that provide recommendations regarding input controls.

**Acknowledgements**

The development of DLMtool was supported by the Natural Resources Defense Council and the Kingfisher Foundation. Many thanks to Alec MacCall and Clay Porch for their help with the DLMtool initiative. This work was carried out under the ICCAT Atlantic-Wide Research Programme for Bluefin Tuna (GBYP), funded by the European Union, by several ICCAT CPCs, the ICCAT Secretariat and by other entities (see:http://www.iccat.int/GBYP/en/Budget.htm). The content of this paper does not necessarily reflect the point of view of ICCAT or of the other funders, who carry no responsibility for it; neither does it anticipate the Commission's future policy in this area.

**References**

Butterworth, D.S. 2007. Why a management procedure approach? Some positives and negatives. ICES J. Mar. Sci. 64: 613–617.

Butterworth, D.S., Punt, A.E., 1999. Experiences in the evaluation and implementation of management procedures. ICES J. Mar. Sci. 56, 985-998.

Carruthers, T.R., Punt, A.E., Walters, C.J., MacCall, A., McAllister, M.K., Dick, E.J., Cope, J. 2014. Evaluating methods for setting catch limits in data-limited fisheries. Fish. Res. 153: 48-68.

CCSBT. 2011. Report of the Sixteenth Meeting of the Scientific Committee. Bali, Indonesia. Commission for the Conservation of Southern Bluefin Tuna. Available at: <http://www.ccsbt.org/userfiles/file/docs_english/meetings/meeting_reports/ccsbt_18/report_of_SC16.pdf>

Cochrane, K L., Butterworth, D.S., De Oliveira, J.A.A., Roel, B.A., 1998. Management procedures in a fishery based on highly variable stocks and with conflicting objectives: experiences in the South African pelagic fishery. Rev. Fish. Biol. Fisher. 8, 177-214.

Coelho, M., Filipe, J.A., Ferreira, M.A.M. 2013. Modelling enforcement and compliance in fisheries: a survey. Int. J. Latest. Trends. Fin. Eco. Sc. 2(3): 464-469.

Deriso, R.B. 1980. Harvesting strategies and parameter estimation for an age-structured model. Can. J. Fish. Aqua. Sci. 37-268-282.

DFO. 2012. Stock assessment report on Pacific herring in British Columbia in 2012. Fisheries and Oceans Canada. Science Advisory Report 2012/062. Available at: <http://www.dfo-mpo.gc.ca/csas-sccs/Publications/SAR-AS/2012/2012_062-eng.pdf> [accessed 30/9/2014]

DLMtool v1.35. Online ref.

Dick, E.J., MacCall, A.D., 2011. Depletion-Based Stock Reduction Analysis: A catch-based method for determining sustainable yields for data-poor fish stocks. Fish. Res. 110, 331-341.

Fromentin, J-M., Bonhommeau, S., Arrizabalaga, H., Kell, L.T. 2014. The spectre of uncertainty in management of exploited fish stocks: the illustrative case of Atlantic bluefin tuna. Mar. Pol. 47:8-14.

Geromont, H.F., Butterworth, D.S. 2014a. Complex assessments or simple management procedures for efficient fisheries management: a comparative study. ICES J. Mar. Sci. doi:10.1093/icesjms/fsu017

Geromont, H.F., Butterworth, D.S. 2014b. Generic management procedures for data-poor fisheries; forecasting with few data. ICES J. Mar. Sci. doi:10.1093/icesjms/fst232

Gulland, J.A., 1971. The fish resources of the ocean. Fishing News Books, West Byfleet, UK.

Hillary et al. 2015

ICCAT. 2012. Report of the 2012 Atlantic bluefin tuna stock assessment session. International Commission for the Conservation of Atlantic Tunas. SCI-033/2012. Available at: <http://www.iccat.int/Documents/Meetings/Docs/2012_BFT_ASSESS.pdf>

MacCall, A.D., 2009. Depletion-corrected average catch: a simple formula for estimating sustainable yields in data-poor situations. ICES J. Mar. Sci. 66, 2267-2271.

Maunder, M.N. 2014. Management strategy evaluation (MSE) implementation in stock synthesis: application to pacific Bluefin tuna. Fifth meeting of the Scientific Advisory Committee. Inter-American Tropical Tuna Commission. SAC-05-10b. Available at: <http://www.iattc.org/Meetings/Meetings2014/MAYSAC/PDFs/SAC-05-10b-Management-Strategy-Evaluation.pdf> [accessed 3/10/2014]

McAllister, M.K., Pikitch, E.K., Babcock, E.A. 2001. Using demographic methods to construct Bayesian priors for the intrinsic rate of increase in the Schaefer model and implications for stock rebuilding. *Can. J. Fish. Aqua. Sci.* 58(9): 1871-1890.

NPFMC, 2012. Fishery management plan for groundfish of the Bering Sea and Aleutian Islands management area. North Pacific Fishery Management Council. Anchorage, Alaska. http://alaskafisheries.noaa.gov/npfmc/PDFdocuments/fmp/BSAI/BSAI.pdf

NPFMC,2013. Stock assessment and fishery evaluation reports for 2013. North Pacific Fishery Management Council. Anchorage, Alaska. <http://www.afsc.noaa.gov/refm/stocks/assessments.htm>.

PFMC, 2010. Pacific coast groundfish fishery management plan for the California, Oregon and Washington groundfish fishery, as amended through Amendments 23 and 16-5. Pacific Fishery Management Council. Portland, OR.

Punt, A.E., Donovan, G.P. 2007. Developing management procedures that are robust to uncertainty: lessons from the International Whaling Commission. ICES J. Mar. Sci. 64: 603-612.

Punt, A.E., Smith, D.C., Smith, A.D.M., 2011. Among-stock comparisons for improving stock assessments of data-poor stocks: the “Robin Hood” approach. ICES J. Mar. Sci. 68: 972-981.

Rademeyer, R.A., Butterworth, D.S. and Plagányi, É. E. 2008. A history of recent bases for management and the development of a species-combined Operational Management Procedure for the South African hake. Afr. J. Mar. 30: 291–310.

R Core Team. 2014. R: A language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Url: <http://www.R-project.org>.

Schnute, J., 1985. A general theory for analysis or catch and effort data. Can. J. Fish. Aquat. Sci. 42, 414-429.

Vert-pre, K.A., Amoroso, R.O., Jensen, O.P., Hilborn, R. 2013. Frequency and intensity of productivity regime shifts in marine fish stocks. Proc. Natl. Acad. Sci. USA. 110(5): 1779-1784.

Wallace, J.R., Cope, J.M., 2011. Status update of the U.S. canary rockfish resource in 2011. Report of the National Marine Fisheries Service. Available at: http:// www.pcouncil.org/ wp-content/ uploads/ Canary\_2011\_Assessment\_Update.pdf [accessed 3/10/2014]

Walters, C., Martell. S.J.D., 2002. Stock assessment needs for sustainable fisheries management. Bull. Mar. Sci. 70, 629‐638.

**Appendix A. Operating model**

## *A.1 Simulating stock dynamics*

A standard age-structured, spatial model identical to that of Carruthers et al. 2014 was used to simulate population and fishery dynamics. A range of parameters and variables are allowed to vary among simulations for a given stock (*e.g*., natural mortality rate *M*, gradient in recent fishing effort, targeting). All parameters that vary as random variables across simulations are denoted with a tilde (*e.g*., ). The probability distributions from which these parameters are sampled are detailed in Table App.A.1. Hence, each parameter or variable denoted with a tilde represents a sample from a distribution specific to each stock. This convention alleviates the need for a simulation and stock subscript for every parameter or variable described below. For example, the symbol  represents which is the sample of the parameter corresponding with the *i­­­*th simulation for stock *s*, drawn from a distribution function *f*(), from the stock specific parameters .

The numbers of individuals recruited to the first age group *Ny,a=1,r* in each year *y*, and area *r* is calculated using a Beverton-Holt stock-recruitment relationship with log-normal recruitment deviations:

App. A.1) 

where *h* is the steepness parameter, *R0* is the recruitment given unfished conditions, *SSBy,r* is spawning stock biomass in the previous year and *SSB0* is the spawning stock biomass under unfished conditions. The process error term *P*, is an autocorrelated random variable:

App.A.2) 

where, *v* controls the level of autocorrelation in recruitment deviations and *κ* is a normally distributed random variable of mean zero:

App.A.3) 

The spawning stock biomass, *SSB*, is given by:

App.A.4) 

where *ma* is the maturity-at-age *a,* and the maximum age *na* is specific to each stock. Maturity-at-age is assumed to follow a logistic relationship with age; the slope of the transition from immature to mature is determined by the precision parameter, where 50% of individuals are mature at:

App.A.5) 

Numbers at age are converted to biomass using the von Bertalanffy growth equation:

App.A.6) 

where *La* is the length of an individual of age *a*, the asymptotic length is *Linf*, and *K* is the slope at the theoretical age at zero length *t0*. *Linf* and *K* are assumed to be time-varying with mean percentage gradient *ΔLinf* and *ΔK*. Inter-annual variability in *Linf* and *K* is simulated from log-normal distributions with mean 1, and standard deviations *sdLinf* and *sdK*.

Weight at age *Wa*, is assumed to be related to length by:

App.A.7) 

For ages greater than 1, fishing mortality is assumed to occur before natural mortality and the numbers-at-age are calculated by:

App.A.8) 

where  is the rate of natural mortality. No “plus group” is modelled; instead the maximum age is set sufficiently high that survival to the maximum age is less than 1% under unfished conditions.

Movement is assumed to be constant over time and age of individuals, and to occur instantaneously at the end of each year. For example, for individuals of age *a,* moving from area *r* to area *k* for any year *y*:

App.A.9) 

where *ψ* is the probability of an individual moving from area *r*, to area *k* (Equation App.A.27).

**A.2** **Simulating fishery dynamics**

The selectivity at age *ωa,* was calculated using a double normal curve with age at maximum selectivity *msel*, an ascending limb standard deviation of *σsel1* and a descending limb standard deviation *σsel2*. These standard deviations were determined for each simulation by numerically solving for two user-specified quantities that are more intuitive: (1) the minimum age at 5% maximum selectivity , and (2) the selectivity of the oldest age class .

The ascending limb age selectivity *Aa* (before normalization to a maximum value of 1) is given by:

App.A.10) 

The descending limb selectivity *Da* is given by:

App.A.11) 

The selectivity at age is given by:

App.A12) 

Refuges from fishing are simulated here by a regional availability variable *R* that is 1 for at least one area*:*

App.A.13) 

where *R* is the regional availability of the stock to fishing, *pR* is the Bernoulli probability of failure (“failure to fish successfully” or “probability of a refuge”, Table App.A.1.) pre-specified for each stock.

Catch in numbers is calculated by:

App.A.14) 

where *F* is the fishing mortality rate.

Observed catch is calculated by multiplying simulated catch in numbers-at-age by weight-at-age and adding observation error:

App.A.15) 

The error term *ε*, is drawn from a standard normal distribution whose standard deviation *σobs* is sampled at random in each simulation:

App.A.16) 

Fishing mortality rate *F,* may increase relative to effort (*E*) over the historical period according to catchability *q* modified by a percentage increase in fishing efficiency each year :

App.A.17) 

Total effort is not related to biomass levels, and in historical and future projections can remain high even at very low biomass levels. The maximum fraction of the population that can be caught in any given year is restricted to a maximum of 80% to prevent the simulation of single year stock collapses from TAC recommendations that are occasionally very high.

Log-normal variability in effort is added to a general effort trend *V*:

App.A.18) 

The effort variability term *φy* is randomly sampled from a standard normal distribution that has a standard deviation, *σeff* drawn at random for each simulation:

App.A.19) 

A range of effort variability is sampled to assess how the degree of auto-correlation affected the performance of stock status classification methods. The general trend in effort is determined by a linear model of change in effort over time with slope *aE*, and intercept :

App.A.20) 

This functional form allows effort to increase, decrease or remain unchanged over time. This effort model is constrained by sampling positive values (effort is increasing at the start of the time series). The final annual change in effort , is specified by the user to control the sampling of increasing, constant and decreasing final effort trajectories:

App.A.21) 

For any simulated effort time series, the slope can then be calculated from the total number of years in the time series *ny*,and the sampled intercept:

App.A.22) 

Effort time series with negative values were discarded. All of the stocks had the same underlying variability in temporal effort dynamics.

In any given year, spatial fishing effort is assumed to be proportional to the distribution of the vulnerable biomass in the previous year, modified by a targeting parameter *λ*,that controls how strongly fishing effort will be distributed in relation to vulnerable biomass:

App.A.23) 

The values for *p* average 1 in any year so they can be used to distribute total effort *Ey* across areas in each year such that mean *F* among areas is the same as total annual *F*. Fishing is distributed evenly regardless of the vulnerable biomass in the previous year when the targeting parameter *λ* is zero. Spatial fishing will be distributed in favour of areas of high vulnerable biomass when *λ* is positive and distributed away from such areas when *λ* is negative. In order to simulate increases or decreases in targeting, the targeting parameter follows a linear change over time with intercept 0, and final targeting level  in the last historical year of the simulation *ny*:

App.A.24) 

Targeting is assumed to remain constant over projected years at the same level as the final year of the historical period.

## A.3 Initializing the population dynamics model and simulating movement

The initial biomass in each area is initialized according to an equilibrium assumption regarding age and spatial structure:

App.A.25) 

where *dr* is the initial spatial distribution proportion, and the *dr* sum to 1 over *r*. Note that the age structure is assumed to be the same across areas. The initial distribution vector of the stock over areas, *d=[d1,…,dn],* is the stationary distribution satisfying the condition:

App.A.26) 

where *d* is the positive eigenvector of the movement probability matrix *ψ*, corresponding to the first eigenvalue (this can also be determined numerically by repeatedly multiplying an initial distribution for *d* by *ψ*). Two user specified parameters are used to define the movement matrix *ψ:* the probability of remaining in area 1 between years (*ψ1,1* ) and the equilibrium unfished fraction of stock in area 1 (*d1*) are used to numerically solve for a matching set of *ψ* parameters.

## A.4 Parameterization of stock dynamics

Given the availability of full stock assessments with which to characterize their stock dynamics, we chose Pacific herring (DFO, 2012), Atlantic bluefin tuna (ICCAT, 2012), and canary rockfish (Wallace and Cope 2011) as case-studies that span a range of longevity. The values of input parameters and the sources of these inputs are detailed in Table App.A.1.

Table App.A.1. Summary of the variables/parameters that define each of the stock simulations, including values and/or the range over which they are sampled. Where two values are provided, variables are sampled from a uniform distribution with the lower and upper bounds listed..



**Appendix B: Reference methods**

### B.1 DCAC

In circumstances where the information available is insufficient to derive a catch-limit from stock assessment the NMFS advocates the use of Depletion Corrected Average Catch (DCAC, MacCall 2009). DCAC attempts to calculate average catch accounting for the removal of “windfall harvest” of less productive biomass that may have occurred as the stock became depleted. DCAC requires inputs for *M,* *FMSY/M (*or *c), BMSY/B0* (or *D*) and *Bcur/B0* (or *B*peak). A number of samples are drawn from the following distributions:

App.B.1a) *MDCAC* ~ dlnorm(*μ*=*M*, *SD*=0.5)

App.B.1b) *cDBSRA* ~ dlnorm(*μ*=*c*, *σ*=0.2)

App.B.1c) *DDBSRA* ~ dlnorm(*μ*=*D*, *σ*=0.2)

where, in keeping with MacCall’s (2009) approach, the SDs for *M* and c are set to 0.5 and 0.2, respectively.. MacCall (2009) states that “unlike the other parameters, the precision of [depletion *D*] is entirely dependent on the data and method used in its estimation, and there is no clear value of precision that can serve as a default”. Subsequently, Dick and MacCall (2011) assume a default distribution with a CV of 0.25. We adopt the same beta distribution for depletion to remain consistent with the assumptions made in simulating DB-SRA (detailed above in management scenario M1), i.e.:

App.B.2a) *DDBSRA* ~ dbeta(*μ*=*Dobs*, *CV* = 0.25) where *Dobs* < 0.5

App.B.2b) 1-*DDBSRA* ~ dbeta(*μ*=1-*Dobs*, *CV* = 0.25) where *Dobs* > 0.5

For each sample of these parameters, sustainable yield (*YS*) is calculated by:

App.B.3) 

where the *Cobs* are annual historical catches and *n* is the number of years of historical catches.

This stochastic approach produces numerous samples of the derived sustainable yield (YS) that may be used as a catch-limit.

### B.5 FMSY/M ratio ‘Fratio’

It has been suggested that ratios of *FMSY*/*M* (*c*)may be robust to broad life-history types and fisheries exploitation scenarios. Gulland (1971) proposed a simple method of setting maximum sustainable yield, in doing so assuming that *BMSY*/*B0*=0.5 and *FMSY/M* = 1. Subsequent publications have revised this *FMSY* recommendation downwards. The Fratio MP is simulated by generating imperfect knowledge regarding *M,* current absolute biomass and the ratio of *FMSY/M*.

### B.6 Delay-difference stock assessment (DD)

The performance of a delay-difference model (Deriso 1980, Schnute 1985) fitted to catch and effort data is evaluated to provide a reference for the performance of the other MPs. The delay-difference model requires additional auxiliary (independent) information regarding the form of the stock-recruit function, the fraction mature at age, somatic growth, *M*, and the selectivity-at-age curve. The delay-difference stock assessment method provides estimates of *B*curr and *F*MSY and therefore direct estimates of an appropriate catch limit.

The delay-difference model is fitted to annual total catch and effort data. The model is parameterized according to: maximum sustainable yield, *MSYDD* andharvest rate at maximum sustainable yield, *UmsyDD* . The catchability coefficient scaling effort to fishing mortality rate is also estimated. The growth parameters *α* and *ρ* of the Ford-Brody growth model (*Wa+1*=*α+ρWa*) are approximated from the known weight at age *W,* for each simulation:

App.B.4) ; 

where *W∞* is the maximum weight of an individual and *Vobs*is the observed age at 50% selectivity determined from the ascending limb of the selectivity curve *ω* (Eqn. App.A.12). Since bias in the age at 50% selectivity may strongly affect the delay-difference model, *Vobs* is simulated subject to imperfect knowledge (Table App.C.1 3). Survival rate at maximum sustainable yield is given by  so that the number of spawners per recruit, *SPR* is given by:

App.B.5) 

The Beverton-Holt parameter *αrec*, the maximum recruits per spawner as spawner abundance approaches zero, is calculated:

App.B.6) 

The derivative of yield with respect to harvest rate *ΔSPR,* evaluated at *UmsyDD*is given by:

App.B.7) 

where *S0* is unfished survival rate . The Beverton-Holt parameter *βrec* is calculated as:

App.B.8) 

Unfished recruitment *R0* is allocated to recruitments up to and including the age at recruitment to the fishery *Vobs* and is given by:

App.B.9) 

where unfished spawners per recruit *SPR0* is calculated using Equation App.B.5 when *S*msy is replaced by *S*0:

It follows that initial biomass *B1* is given by:  and initial numbers *N1* is given by. From this initialization, biomass dynamics are calculated by:

App.B.10)  ; 

where is the survival rate in year *y*, *N* represents stock numbers, *B* is the biomass, *Wk* is the weight of an individual at the age at 50% selectivity *k*, *M* is the natural mortality rate (assumed to be known exactly), *qDD* is the estimated catchability, *E*y is the observed fishing effort during year *y*, and *R*y represents the number of recruits during year *y*:

App.B.11) 

where catches *C,* are given by: .

The model is fitted to observed (simulated) catches by minimizing a global objective *O* that is calculated by the sum of the negative log likelihood of the catches:

App.B.12) 

where *σc* is the assumed standard deviation (in log space) of the observation error*.*

## Appendix C: Simulating imperfect information

Table App.C.1. Summary of the bias /error parameters and related distributions that control the accuracy and precision of knowledge of the simulated system that is subsequently used by the data-limited methods and harvest control rules. The log-normal distribution described in the table below (~dlnorm(*μ*,*σ*) is the exponent of the normal distribution with mean *μ* and standard deviation *σ*, parameters: . 

1. We define ‘data-rich’ as situations where data are available to conduct a conventional stock assessment (Punt et al. 2011). This includes simple stock assessment methods that typically require more than 15 years of relative abundance or fishing effort data in addition to catch data. We define all other data situations under the heading ‘data-limited’ which includes ‘data-moderate’ and ‘data-poor’. Data-moderate situations have some form of current information about stock levels that may be absolute abundance, relative abundance or stock depletion. Data-poor refers here to situations where only historical catches and some life-history information are available. [↑](#footnote-ref-1)