Impacts of spatial uncertainty on performance of age structure-based harvest strategies

Gavin Fay[a,b], André E. Punt[b], and Anthony D.M. Smith[a]

[a] CSIRO Marine and Atmospheric Research, Castray Esplanade, Hobart TAS 7001, Australia.
[b] School of Aquatic and Fishery Sciences, University of Washington, Seattle WA 98195-5020, USA.

Corresponding author: Gavin Fay
Phone:+61-3-6232-5321
Fax: +61-3-6232-5000
Email:gavin.fay_at_csiro.au

1. Abstract

Harvest Control Rules (HCRs), key components of fisheries management strategies, are used to calculate recommended catch levels given estimates of present stock status or levels of fishing mortality. The performance of HCRs when confronted with spatial variability, either from population dynamics, fishery operations, or in data collection, are poorly understood. Australia's Southern and Eastern scalefish and shark fishery (SESSF) uses a tier framework of HCRs, with the choice of which Tier rule to apply for a species reflecting the uncertainty in available information on stock status.

A Management Strategy Evaluation (MSE) approach is used to evaluate the performance of a 'data-poor' (Tier 3) HCR, which uses information from the age structure of the catch only, when applied to the fishery for blue eye trevalla (*Hyperoglyphe antarctica*): a long-lived, late-maturing species exhibiting spatial variability, potentially a result of structure in the population dynamics. Several versions of the Tier 3 HCR are tested, varying in the types of reference points used to determine management actions, and in the way spatial variability is accounted for when setting catch limits.

Results suggest effective implementation of the HCRs is challenging, and requires appropriate choice of reference points and estimators. Spatial disaggregation of data leads to uncertain estimates of current mortality. However, appropriate weighting of spatial estimates of stock status leads to improved conservation of the resource over 'pooled data' approaches. Variability in performance measures are dominated by uncertainties regarding whether the assumed value for the rate of natural mortality, *M* is correct or not, and the true value for the steepness of the stock-recruitment relationship. Indeed, simulated outcomes are sensitive to many uncertainties inherent to an information-poor, spatially-heterogeneous resource. Additional considerations besides the HCR should be taken to achieve a desired precautionary result in contrast to the situation for more data-rich scenarios.

2. Introduction

Harvest strategies (often termed Management Procedures) are well recognized as effective tools for conservation of natural resources and have been applied widely in fisheries management, principally in output control, data-rich fisheries (*e.g.* Butterworth et al. 1997, Butterworth and Punt 1999, Cooke 1999, Kell et al. 1999, 2005). Harvest strategies consist of the following components: data collection schemes, assessment methods, and harvest control rules (HCRs). The latter translate stock indicators from stock assessments into specifications for management actions (*e.g.* Restrepo and Powers 1999). A successful HCR should provide an appropriate response to deviations from management targets, be robust to key uncertainties, and emphasize precautionary action given uncertainty. The latter point is particularly important for so-called 'data-poor' situations, when the reliability of stock indicators is likely questionable. Simulation methods using a Management Strategy Evaluation (MSE) approach are well-developed, and offer powerful tools for comparing the performance of HCRs (*e.g.* De Oliveira et al. 2008, Butterworth and Punt 1999, Smith et al. 1999).

The blue eye trevalla (*Hyperoglyphe antarctica*) is a high-valued species in Australia's Southern and Eastern Scalefish and Shark fishery (SESSF). The fishery for this long-lived, late-maturing species is characterized by a large number of gear types operating in a range of areas, with uncertainty in stock structure, apparent spatial and seasonal variability in availability of different age classes, and low levels of sampling effort across the fishery (Smith and Wayte 2002, Fay 2007). Scientific advice for management in the SESSF takes the form of a Recommended Biological Catch (RBC) for each quota species (including blue eye trevalla) for the entire fishery to inform the setting of the Total Allowable Catch (TAC) (Smith et al. 2008). At present, the TAC for blue eye applies across the fishery, because there are few measures in place to allocate the TAC spatially[1] (a separate TAC is applied for one sector of the fishery, the trawl fishery in the Great Australian Bight (GAB)).

The SESSF adopted a formal harvest strategy framework (HSF) as a basis for setting RBCs in 2005 (Smith and Smith 2005, Smith et al. 2008). This framework is based on a tier system of HCRs, with the decision as to which tier a particular stock is placed in depending on the type of information available on which to base a stock status determination. The tier framework is intended to follow the precautionary approach, in that control rules should lead to lower RBCs, and result in maintaining the stock at higher levels of spawning biomass on average as information quality declines and progression through the tiers proceeds. The SESSF harvest strategies specify a biomass level B_{LIM} (currently 20% of unfished spawning biomass), below which targeted fishing should cease, and a target biomass B_{TARG}. The HCRs operate by specifying a maximum fishing mortality rate that defines overfishing (F_{LIM}), and a target fishing mortality rate that defines optimum utilization (F_{TARG}). Accounting for increasing uncertainty in stock status is reflected in the application of discounts to catch – the use of which is intended to achieve the same end as a decrease in the target fishing mortality rate as uncertainty about stock status increases.

[1] Catches by blue eye in the trawl fishery in the Great Australian Bight (GAB) are not included in the SESSF TAC, although the catch by other gears in this area are. This is not a major sector of the fishery, catches by trawl in the GAB have been at most on the order of 1-2% of the annual total catch for blue eye in the SESSF.

The “Tier 3” HCR has been applied to blue eye trevalla. This HCR is designed for stocks for which there exists no estimate of current biomass, but where an estimate of the current fishing mortality rate, F_{CUR}, is available, most frequently from the results of catch curve analysis applied to age composition data. The Tier 1 HCR is for the most information-rich case, and involves calculating RBCs from the results of fitting an integrated stock assessment model (*e.g.* Stock Synthesis, Methot 2007) to the available data. As the HSF was not tested before being implemented, it is not clear how well the tier framework of HCRs performs, and indeed whether scientific advice for management is more precautionary for species managed using the Tier 3 HCR, than would be the case had the species been data-rich and managed under Tier 1. Finally, it is not clear how best to cope with possibly conflicting information from multiple areas and gear types.

This paper uses MSE to assess the performance of the Tier 3 HCRs for blue eye trevalla given key uncertainties. Implementation of MSE typically involves assessing the consequences of a range of management options, and transparently deals with trade-offs among performance criteria given a specified set of management objectives. The performance of HCRs is assessed based on how well they meet management targets and objectives, including risk specifications. The performance of several variants of the Tier 3 HCR that use different specifications for the various reference points and/or utilize different estimation methods are compared. These alternatives increase correspondence with the Tier 1 HCR, and include calculation of biomass estimates and assumptions regarding the stock-recruitment relationship, negating the need for the RBC to rely directly on previous year’s catch levels. HCR performance is considered both when there is no spatial structuring of the population or fishery, and when there exist two regions in which the fishery operates, with uncertainty related to exploitation pattern and selectivity by region, and also given different assumptions regarding the spatial structure and degree of mixing of the fished stock between regions.

Emphasis is placed on presenting key results and demonstrating HCR behaviour given different approaches regarding how to improve the performance and precautionary nature of the tier framework. Comparisons with data-rich scenarios are presented for some cases. While the MSE is restricted to a case study of a single species and fishery, the nature of the studied resource is relevant to other fished populations, and the discussion outlines general points that may be taken into account when applying these methods to other systems, particularly when faced with issues related to spatial uncertainty, either with respect to the resource or the fishery.

3. Methods

3.1 Simulation protocol

Performance of the HSF for blue eye is evaluated using a simulation modelling framework that incorporates feedback between the HCR and the population dynamics. Attention is focused in this section on describing the HCRs and the various modifications made to their implementation, rather than describing the technical details of the operating model, which are provided in full in Appendix 1. The general approach on which the operating model is based consists of tuning a spatial age-structured model to represent a set of hypotheses for the dynamics of the blue eye trevalla population and fishery. The values for the parameters of the operating model are not based on the results of a stock assessment, as no model for the population dynamics exists at present for blue eye in the SESSF. Rather, values for parameters were either sourced from the literature, or derived via preliminary estimation and trial and error analyses in order to mimic the general characteristics of the available data for blue eye.

The operating model is projected over a historical period given the known catch history for blue eye, and age-composition data are generated given the known 'true' population. The chosen HCR is then used to determine the RBC for the following year(s), given an estimation method (catch curve analysis) and the selected parameters governing the HCR. The RBC is then allocated to fleet and region within the operating model, the population size is updated given this new catch, and additional data are generated. This assessment / population update cycle is repeated for 20 years, with annual assessment and updating of the RBC. A ***scenario*** is defined as the combination of a set of parameter values for the operating model, a data collection scheme, and a specific version of the Tier 3 HCR. One hundred simulations were conducted for each scenario, each differing due to process error in the population dynamics, observation error when generating the age-composition data, and error associated with implementation of the estimation method and application of the HCR[2]. A series of summary statistics is calculated at the end of the projection period. The summary statistics for each simulation are further analysed to derive a set of performance measures, which are used to compare results among scenarios.

3.2 Operating model

The operating model consists of an age-structured population dynamics model that can be parameterised to include spatial regions (with movement of fish among regions), and multiple fleets, to capture key dynamics for blue eye trevalla. Full technical specifications for the operating model are detailed in Appendix 1. Analyses detailed in this chapter consider two

[2] The results obtained from 100 simulations were almost identical to those obtained when 1,000 simulations were used to characterise a scenario for a subset of the scenarios described.

versions of the operating model: (a) a single population occupying a single region and exploited by a single fishing fleet, and (b) a population occupying two regions with movement between regions, and exploited by one or two fishing fleets (with different selectivity patterns). Several parameterizations of each version of the operating model are considered to investigate the implications of key uncertainties. The parameterization of the operating models, along with the values for biological parameters for blue eye considered for the various scenarios, are given in Tables 1 and 2, and Figure 1.

Scenarios using the two-region version of the model are designed to mimic general assumptions regarding the nature of the blue eye trevalla fishery, rather than the actual spatial structure. Two ‘continental slope’ regions with differing exploitation histories are assumed, with levels of stock mixing between the two regions, ranging from full mixing, in that the impacts of spatial variability are minimal, to almost no mixing, indicating a high degree of spatial structuring in the blue eye population. Spatial differences in population responses to exploitation are more likely to be observed under the latter scenario. The regional catch histories used to drive the population dynamics (Figure 1f) are taken from the landings data from the relevant zones of the SESSF, with the geographic split in these data being catches taken east and west of Tasmania.

3.3 Harvest strategies

The harvest strategies consist of a data collection scheme, a method to estimate the current fishing mortality rate F_{CUR}, and an HCR. Scenarios are limited to instances where the harvest strategy is applied every year of the projection period, consistent with the current practice of annual setting of TACs within the SESSF. The two forms for the Tier 3 HCR shown in Figure 2 are tested, with two methods for estimating F_{CUR}. Variations of the HCR that utilise different reference points and have differing data requirements are implemented as outlined below.

3.3.1 Data and estimation methods

Data available for the Tier 3 analyses were limited to fishery-dependent age-composition data (*i.e.* no index of abundance or fishery-independent data), with an annual multinomial sample size of 100 allocated by fleet and region in the same proportions as the annual catch (*i.e*. the total sample size was 100). While this sample size is a good deal less than the number of otoliths that have been aged annually in recent years for blue eye (on the order of 500 per year), it is perhaps unreasonable to think that these data truly constitute a random sample of size equal to the number of aged otoliths (e.g. Candy 2008, Miller and Skalski 2006). A random sample of 100 represents a reasonably good sample size that might be hoped for from SESSF species. Evaluation of HCR performance when sample size is reduced or not randomly determined from the catch is not considered in this chapter. Four years of age-composition data were assumed to be available to the estimators. Two catch curve estimation methods were employed: (a) the Chapman and Robson (1960) catch curve estimator (CR), and (b) a multi-year equilibrium *F* age-structure based-estimator (MYEF). The estimators aim to estimate total mortality, *Z*, with estimation of *F* then achieved given an assumed value for the rate of natural mortality, *M* (denoted “assumed *M*” in Tables 1 and 2). For the CR method, catch curves were applied to the annual age-composition data, with F_{CUR} calculated as an inverse-variance weighted average of the annual estimates. In contrast, MYEF integrates over all years, therefore averaging over years is not required to obtain an estimate of $F_{CUR.}$ for this estimation method.

a) Chapman and Robson catch curve estimator (CR)

The CR estimator assumes that the population is in equilibrium, and that recruitment is constant over time. The estimate of *Z*, from a sample of the age composition for a given year is:

$$Z_y = \ln\left(\frac{1+\bar{a}_y - 1/n_y}{\bar{a}_y}\right) \tag{1}$$

where $\bar{a}_y$ is the mean age (above the recruitment age) of the sample and n_y is the sample size for year *y*. A single estimate of *Z* is required to calculate the RBC, and so weighted averages of estimates of Z_y from the most recent four years of age-composition data were calculated, with weighting inverse to the variance estimate for each year:

$$Var(Z_y) \approx \frac{\left(1-e^{-Z_y}\right)^2}{n_y e^{-Z_y}} \tag{2}$$

Catch curve estimators are known to be sensitive to the age-range of the data used (Chapman and Robson 1960, Dunn et al. 2002). For the analyses presented here, the recruitment age was set at that for which the numbers–at-age were greatest, with the maximum age being determined from the sample. CR assumes uniform selectivity for ages above the recruitment age, likely biasing estimates of vulnerable biomass when selectivity is actually dome-shaped.

(b) Multi-year equilibrium F age-structure based-estimator (MYEF)

Estimation of F_{CUR} using MYEF involves fitting an equilibrium-based age-structured population dynamics model to the available age-composition data, with the population model being of the form:

$$N_a = \begin{cases} 1 & \text{if } a = 0 \\ N_{a-1} e^{-(s_{a-1}F+M)} & \text{if } 0 < a < 100 \\ \dfrac{N_{a-1} e^{-(s_{a-1}F+M)}}{\left(1 - e^{-(s_a F + M)}\right)} & \text{if } a = 100 \end{cases} \tag{3}$$

where the N_a 's are the numbers-at-age, s_a is the (estimated) selectivity at age (assumed to be asymptotic and to follow a logistic curve), F is the estimated rate of fishing mortality, and M is the assumed rate of natural mortality. The values for F and the parameters which define s_a are determined by maximizing the following log-likelihood function:

$$\ln L = \sum_y n_y \sum_a O_{y,a} \ln\left(\frac{\tilde{N}_a}{O_{y,a}}\right) \tag{4}$$

where $O_{y,a}$ is the observed proportions in the sample by age in year y, n_y is the sample size for year y, and $\tilde{N}_a$ are the predicted proportions of catch-at-age:

$$\begin{aligned} \tilde{N}_a &= \bar{N}_a / \sum_{a'} \bar{N}_{a'} \\ \bar{N}_a &= \frac{N_a s_a F}{\left(s_a F + M\right)} \left(1 - \exp^{-(s_a F + M)}\right) \end{aligned} \tag{5}$$

Maximisation of (4) was achieved using AD Model Builder (ADMB Project 2009). Differences between MYEF and CR are that MYEF accounts for selectivity, data from all ages are used, and the likelihood is multinomial. Unlike the CR estimator, no averaging of annual mortality estimates is necessary to calculate the RBC under MYEF because F is calculated using all the available data simultaneously.

The scenarios outlined in Tables 1 and 2 include uncertainties related to applying the estimation methods. Importantly, the impact of assuming the incorrect value for M when conducting the estimation is examined.

3.3.2 Harvest Control Rules

Concern about the performance of the Tier 3 HCR has been noted following implementation (Klaer et al. 2009). There is concern that the original nature of the calculation of RBCs for Tier 3 (applying an appropriate multiplier to recent average catch levels, Figure 2b) could produce a ratchet effect of continually increasing or decreasing catches, even though information suggests that the target level has been reached. A revised harvest control rule (Klaer et al. 2009), which shows consistency with the more data-rich tier rules in terms of reference points, was applied in

2008 (Figure 2c). Unlike the ‘old’ rule, this ‘new’ rule does not have a cap on annual catch increases.

Each of the scenarios outlined in Table 1 were projected using three variants of the Tier 3 HCR (Figure 2), which differed either by adopting the ‘old’ or ‘new’ Tier 3 rule, and in the choice for the target and limit reference points:

1. The shape of the HCR follows the ‘old’ rule (Figure 2b), with $F_{\mathrm{TARG}} = M$,
2. The shape of the HCR follows the ‘new’ rule (Figure 2c), with $F_{\mathrm{TARG}} = 0.5M$ and $F_{\mathrm{LIM}} = M$, and
3. As for 2), but with the reference points adopting a Tier 1-like approach with $F_{\mathrm{TARG}} = F_{40}$, and $F_{\mathrm{LIM}} = F_{20.}$ (F_{40} and F_{20} are the fishing mortality rates which will result in [under equilibrium age structure] spawning biomasses of 40% and 20% of unfished spawning biomass [corresponding to the B_{MSY} proxy and B_{LIM} under the SESSF Tier 1 HCR]).

The values for M used in the HCRs (and that used to calculate F) are the ‘assumed M’ values as detailed in Tables 1 and 2.

Empirical investigation suggests that the assumption of $F_{MSY} \approx M$ is too high for blue eye trevalla (Figure 3). Walters and Martell (2004) suggest $F_{msy} = cM$, with values for c including 0.8 in general, but 0.6 or less for commonly fished species (Walters and Martell 2004). For U.S. west coast groundfish species, the average is c = 0.62 (MacCall 2007), and so 0.5M was chosen for the analyses here to adopt a conservative estimate.. Calculation of F_{40} and F_{20} depends on the values for the parameters of the stock-recruitment relationship (assumed to follow a Beverton-Holt form), and requires estimates of the steepness parameter h (Mace and Doonan 1988), information on growth and fecundity, and selectivity in addition to an estimate of M. The values for these reference points used in the HCR were calculated based on the estimates of selectivity from the estimators, an assumed value for h of 0.75, and the ‘correct’ values for growth and fecundity (Figure 1a-c). In contrast, versions 1) and 2) of the Tier 3 HCR rely only on an estimate of M to calculate the RBC given F_{CUR}. However, version 3) more appropriately accounts for biology when determining the likely response to fishing.

Calculation of the RBC under version 1) is achieved by applying the appropriate multiplier from Figure 2b to C_{cur}, defined as the average catch over the four years prior to the year for which an RBC is needed. Under versions 2) and 3), the RBC is calculated by first obtaining F_{RBC} given Figure 2c, and applying the formula:

$$RBC = C_{cur} YPR(F_{RBC}) / YPR(F_{cur}) \quad (6)$$

where $YPR(F)$ is the yield-per-recruit obtained given a fishing mortality F. Note that equation (6) allows for greater increases in catch than does the old Tier 3 HCR (maximum increase of 20% above the recent average, Figure 2b), if F_{CUR} is estimated to be below the target level. Irrespective of this, the maximum allowable change in the catch (RBC) from one year to the next was restricted to 50%[3].

Comparison of Tier 3 performance with that expected under Tier 1 is achieved by calculating the projected spawning stock biomass trajectories for a subset of the scenarios in Table 1 under the Tier 1 HCR. This involved generating additional data (cpue (20 yrs, CV=0.3), and length composition (10 years, n_y=100)), and applying Stock Synthesis 2 (Methot 2007) to this data set each year of the projection period. Results for Tier 1 HCR are simply shown for comparison purposes because the focus of this chapter is the Tier 3 HCRs.

3.3.3 Accounting for fleet/spatial structure

Uncertainty in spatial structure through the scenarios in Table 2 presents additional challenges when implementing the Tier 3 HCRs, because of the need for decisions regarding what combinations of fleet and region are to provide the parameters used when calculating the RBC, and how to choose among potentially differing estimates of the fishing mortality rate. The scenarios in Table 2 were crossed with the following options to investigate how performance given spatial structure changes with assumptions regarding how the data are analysed:

1. spatial complexity is ignored, and a single analysis (CR, MYEF) is conducted using the data pooled across regions (added together as samples are allocated by region relative to the size of the catch) – this is how the Tier 3 control rule is applied in the SESSF at present,
2. the data from the two regions are analysed separately to obtain two estimates of current fishing mortality / stock status; these estimates are then weighted by the inverse of their variance estimates to obtain the RBC.
3. separate analyses as in 2, but the maximum estimated F is used to calculate the RBC.

The variance estimates for F_{CUR} are (primarily) driven by sample size, and so option 2 effectively weights the regional estimates by the current catch allocation. Option 3 is potentially more conservative because it bases the RBC on the parameters for the region with the highest estimated fishery mortality rate. However, this option can be expected to be more prone to inaccurate estimates of F that might result from low sample sizes.

[3] Within the HSF of the SESSF, a rule exists where the TAC for quota species cannot change by more than 50% from one year to the next.

3.4 Performance measures

Performance of the HCRs is evaluated using a set of performance measures:

1. The median (over simulations) spawning stock status at the end of the projection period (final spawning biomass as a fraction of unfished spawning biomass, B_0), [median final depl.].
2. The inter-quartile range of the spawning stock status (relative to B_0) at the end of the projection period [IQR final depl.].
3. The probability of the spawning biomass being below the Tier 1 limit reference point (B_{20}) at the end of the projection period [P(final B < B_{LIM})].
4. The probability of the spawning biomass going below the Tier 1 limit reference point (B_{20}) at some point during the projection period [P(B < B_{LIM} anytime)].
5. The median of the average annual catch during the projection period [median (avg. TAC)].
6. The median (over simulations) of the CV of the annual catches during the projection period [median(CV TAC)].
7. The mean (over simulations) number of years for which the RBC is less than 4t[4] [mean(#yrs collapse)].

Performance measures 1-4 relate to the effect of implementing the HCR on spawning biomass, while measures 5-7 provide information regarding the catch performance of the HCR.

4. Results

The results of the simulations are displayed in the form of boxplots of the performance measures across scenarios to compare among the HCRs and methods for obtaining *F* estimates. Simple linear models are used to evaluate the contribution of the different scenario specifications to the values for the performance measures. The scenario characteristics as defined in Tables 1 and 2, the catch curve estimation type, the choice of HCR, and (for the spatial analyses) the method for obtaining a single *F* estimate, were included as factors in the linear predictors of these models, fitted separately for the seven performance measures. Interaction terms involving some of the factors were also considered.

4.1 Non-spatial analyses

The performance of the three versions of the Tier 3 HCRs, given estimation using CR and MYEF, are compared in Figure 4. The old Tier 3 HCR leads to levels of spawning biomass that are well below the Tier 1 target and limit biomass reference points (40 and 20% of unfished

[4] 4t is 1% of the total catch prior to implementation of the harvest strategy, this performance measure is intended to reflect the frequency of fishery collapse.

spawning biomass) at the end of the projection period, (Figure 4a) with high probabilities of dropping below B_{20} during the projection period (Figure 4d). The performance of the new HCR varies considerably among scenarios and the projections under this HCR lead to more variable catches (Figure 4f). However, the results for this HCR are generally more optimistic regarding stock status (Figure 4a, 4c), although many scenarios still remain below the target biomass at the end of the projection. Comparison of performance for three scenarios suggests that, for these scenarios at least, the Tier 3 HCRs are not precautionary compared to the Tier 1 HCR, because the Tier 1 HCR leads to higher relative biomass, a lower probability of dropping below the limit reference biomass, and lower, less variable annual catches (Figure 5).

The changes in performance with respect to final stock status, the risk of going below the limit reference point, and the magnitude in catch levels are largely independent of the estimation method (CR versus MYEF), but are a function of the HCR (Table 3, 'new HCR' rows). Adopting $F_{\mathrm{TARG}}=F_{40}$ and $F_{\mathrm{LIM}}=F_{20}$ results in an increase in the median final relative spawning biomass (Table 3, new HCR, Figure 4a). The differences between applying the old and new HCRs, for the two estimation methods can be seen clearly in Figure 4, which shows the distribution of the values for the performance measures across all of the scenarios in Table 1.

Performance of the new Tier 3 HCR is also determined by the choice of reference points. The HCR based on the spawner-recruitment reference points ($F_{\mathrm{TARG}}= F_{40}$ and $F_{\mathrm{LIM}}= F_{20}$) tend to lead to higher values for relative spawning biomass, lower probabilities of dropping below the limit reference point, and lower median annual catches than the $F_{\mathrm{TARG}}= 0.5M$ version of the new Tier 3 HCR (Figure 4a, c-e). However, performance is variable among scenarios, and the probability of dropping below the limit biomass remains very high for a number of the scenarios (Figure 4c-d).

Variability in the values obtained for the performance measures was caused not only by the choice of HCR. Scenarios where the true value for steepness was low ("*h=0.3*") resulted in lower final biomass, an increased probability of dropping below B_{LIM}, and increased variability in the annual catches (Table 3). Likewise, more productive stocks ("*h=1.0*") resulted in higher final biomass levels and a lower probability of being below the limit reference point at the end of the projection period. Scenarios in which the initial (prior to implementation of the HCR) relative stock size was low ("InitDepl =0.2") resulted in lower levels of catch (albeit more variable). Under the old Tier 3 HCR, a higher initial stock size ("InitialDepl = 0.75") led to higher catches, an increase in the final relative stock size and a lower probability of being below the limit reference point at the end of the projection period. However, an interaction between the initial stock size and the choice of HCR meant that under the new HCR, scenarios starting at high relative biomass resulted in large, unsustainable catches being taken, and a general poor performance of the HCR in terms of maintaining stock status, and near ubiquitous probability of ending the projection below the limit biomass (Table 3, “new HCR : Init Depl interactions”). This change in behavior between the old and new Tier 3 HCR for the high initial stock size scenarios was largely a result of the difference in maximum allowable increases in catch (old Tier 3 has a cap of 20% increase versus 50% for the new HCR).

In terms of magnitude, aside from the initial stock size, the factor with the largest impact on the biomass-related performance measures was whether the assumed value for M was correct or not. Assumed values for M less than the true value resulted in more optimistic outcomes in terms of stock status, with higher final biomasses, and lower probabilities of dropping below the limit reference point (Table 3, ‘assumed M < true M). Average catches were also lower. Conversely, assuming a value for M greater than the true value resulted in an under-estimation of F, and consequently, outcomes with lower final relative biomass and a higher risk of dropping below the limit (Table 3, ‘assumed M > true M ’).

4.2 Spatial analyses

The results for the ‘spatial’ two-region scenarios for the F_{TARG}=F_{40} version of the new Tier 3 HCR are shown in Table 4, and include results for the three options related to how to deal with the spatial data (see Section 3.3). Results for this implementation of the HCR (new form, F_{TARG}=F_{40}) are shown because this option appeared to best satisfy management objectives in the non-spatial analyses described above. A decrease in the connectivity of the regions results in a decrease in the probability of dropping below B_{20}, and increases in final spawning depletion for scenarios when the stock is initially at the biomass target (Table 4, F option 1). This is presumably because the decrease in movement between regions increases the signal in the data, as the initially exploited region must be driven to low levels before implementation of the harvest strategies. However, this sensitivity to the degree of connectivity is lost when the initial spawning depletion is either at high or low levels (Table 4, scenarios 1-7). The initial status of the stock therefore appears to be at least as important (if not more so) in determining the values for the performance measures as the connectivity among the regions. Indeed, the results of linear modeling to predict the values of performance measures suggests that the effect of the degree of mixing between regions is not important (Table 5).

There was no impact of moving from the intermediate (in which the average mixing rate is 20%) to the limited (5%) level of mixing (the magnitudes of the coefficients for the intermediate and limited mixing scenarios given a particular performance measure were almost the same). The age structure of fish mixing between regions appeared less important in driving performance.

Whether selectivity was dome-shaped, or modelled differently by region was a major determinant of performance, with the amount of dome-shaped selectivity (in 1 region or 2) leading to higher final relative spawning biomasses and lower probabilities of going below the B_{20} limit (Table 4 and Table 5, scenarios 8-9, ‘Different selectivities by region’ and ‘Selectivity dome-shaped in both regions’). Lower catches resulted from selectivity being dome-shaped in both regions (Table 5, ‘Selectivity dome-shaped in both regions’).

Analysing the data by region and then choosing the maximum estimated F to set the RBC (Table 4, F option 3) unsurprisingly led to the most optimistic results regarding spawning stock biomass, and the probability of going below the limit reference point (Figure 6). This choice also resulted in tighter intervals for the biomass. The relative performance of the different scenarios is very similar when data from both regions are analysed together and when the regional estimates are weighted by their variance (Table 4, F options 1 and 2), although the latter case appears more variable. An exception is when movement between regions is limited to pre-recruits (Table 4, scenario 8). In this instance, aggregating the data and conducting a single analysis appears to be a much more conservative way to determine RBCs, because the relative biomass is well below B_{20} when regional estimates of F are weighted by the inverses of their variances. As with the non-spatial analyses, a large proportion of the variation in the performance measures among scenarios can be attributed to the relative stock size prior to implementing the harvest strategy, rather than the specifications for the particular HCR implemented. The increase in final biomass and decrease in the risk of dropping below B_{20} associated with choosing the maximum regional F estimate are naturally associated with lower catches, however do not result in a decrease in variability in catches, nor a decrease in the relative frequency of fishery collapse (Figure 6f-g). This option also appears to mitigate the change in performance associated with the spatial connectivity among regions, as the values for the performance measures do not change with decreasing connectivity as for the case when the regional data are aggregated prior to analysis (Table 4, compare scenarios 1,2, and 5 between F options 1 and 3). This suggests the degree to which sampling error has on the performance of the HCR, as for scenario 1, the true exploitation rate is the same in both regions yet F option 3 results in higher final spawning stock biomass.

Although the results suggest that reasonable performance can be achieved using the new Tier 3 HCR given an appropriate choice of reference points and decision rule for dealing with space (at least compared to the original Tier 3 HCR), Figure 7 suggests that performance of these HCRs is not particularly satisfactory, because higher relative spawning biomass may be a result of closing the fishery for a number of years following a series of successive increases (or decreases) in the RBC. This is also reflected in the values for the mean number of years in which the fishery collapses (Figure 4g, Figure 5g, Figure 6g). The trajectories in Figure 7 suggest, as inferred above, that for a species like blue eye, the catch curve is fairly unresponsive in detecting changes in F. This can be expected for a long-lived species, where there would presumably be considerable inertia in the age structure. As such, the estimates of F obtained may not be reflecting the current fishing mortality rate.

5. Discussion

Management based on rapid stock assessment is attractive for fisheries where there are limited data, and methods for such assessment, including catch curve analysis, are well-established (albeit also with well-known shortcomings related to unrealistic assumptions). The MSE testing of the Tier 3 HCRs presented here suggests that it is indeed possible to formulate HCRs based on the results of catch curves that address management objectives (i.e. maintain spawning stock biomass at or above target levels), despite some of these shortcomings. However, it is also clear that implementing the Tier 3 HCRs can result in undesirable behaviour, and that outcomes can be sensitive to many of the known shortcomings of the associated estimation methods.

Assessing performance of the HCRs through their ability to conserve stock biomass may not be an appropriate choice – the spawning biomass trajectories in Figure 7 suggest that satisfactory outcomes for a scenario (for example, a low probability of being overfished) can be achieved with undesirable system properties (such as complete closure of the fishery following a ratcheting increase in catch), even for the “new” Tier 3 HCR. Klaer et al. (2009) and Smith et al. (2008) address issues related to the unresponsiveness and ratcheting behaviour of the Tier 3 HCR. These undesirable properties are likely to be more pronounced for longer-lived species because the catch curve does not relate to current conditions. Unresponsiveness in the Tier 3 HCR is also a consequence of restrictions on the magnitude of permitted changes in management actions (the RBC is only allowed to change by 50% in a given year even if the estimate of F changes dramatically). The results suggest that such a behaviour appears to result in higher final biomass levels for stocks that are at low relative stock size prior to implementation of the HCRs compared with those achieved for stocks that are at or above management targets Table 3 and Table 5). Increasing the time period over which the catch is averaged will mitigate the ratcheting effect of RBCs (concurrent increases or decreases), however doing so effectively downweights the influence of previous management actions.

Differences in the performance of the Tier 3 HCRs appeared to be related to both the form of the HCR, and the values chosen for the reference points (e.g. Figure 4). Tier 3 HCRs that used the spawner-recruit-based reference points resulted in the best perceived performance (Figure 4). However, performance of the rules using a target of $0.5M$ was generally only marginally different than those using spawn-recruit-based reference points, even though the data requirements were markedly less (Table 3 and Figure 4). As estimates of *M* already tend to be uncertain (with results being very sensitive to assuming the wrong value), including additional uncertainty associated with estimating the compensation of the spawner-recruit curve (steepness) is perhaps unnecessary. However, Figure 3 clearly shows that 0.5*M* is not necessarily an appropriate target rate of fishing mortality (when compared with Tier 1 reference points) for all instances (*e.g.* when steepness is low). Note that even the 'poorly' performing HCRs require an estimate of *M*, typically derived from longevity and growth information (*e.g.* Hoenig 1983, Jensen 1996, Pauly 1980). While such information generally tends to be available, the nature of a 'data-poor' fishery may mean that these estimates are uncertain. The CV of *M* estimates from the Pauly and Hoenig methods are 0.53 and 0.61 respectively (MacCall 2009), and therefore it might be unreasonable to assume greater certainty in *M* than this for a data-poor stock (MacCall 2009 recommends a CV of at least 0.5).

The results clearly demonstrate the need for careful application of 'common sense' when applying methods such as the Tier 3 HCRs. For example, the implications of dome-shaped selectivity are that mortality is over-estimated, leading to specification of lower catches, but it would be somewhat foolish to use this conservation of stock biomass as a reason for implementation of the Tier 3 HCRs when selectivity is known to be dome-shaped. Having an accurate estimate for *M* appears to be very important for HCR performance, with scenarios where the chosen value for *M* is higher than the true value resulting in high probabilities of dropping below the limit reference point. Similarly, scenarios for which the assumed value for *M* is lower than the true value are among the most conservative in terms of biomass relative to the unfished state at the end of the projection. These results are unsurprising, as the estimate of *F* is clearly negatively correlated with *M*. The impact of selectivity being dome-shaped is similar to that of under-estimating *M*, in that *F* is over-estimated (because the estimators assume selectivity to be asymptotic), resulting in lower RBCs and higher spawning stock biomasses.

Although the analyses examined the impact of collecting data from multiple regions, and where the regional allocation of catches was changing, the data were generated in proportion to the catch, with no over-dispersion or bias in the sample other than the stochasticity imposed on the data through sample size and multinomial sampling. The low sampling effort present in the actual blue eye data set, coupled with seasonal differences in availability, means that the age and length data are not representative of the fishery as a whole. Indeed, the sample size of the age data for blue eye is such that pooling age data across years to obtain an age-length-key and then applying this to the year/region-specific length composition data is the most likely means of estimating the age-composition of the catch (*e.g.* Klaer 2008). While the analyses investigating the impacts of region-specific selectivity go some way to addressing these questions, it is likely that incorporation of bias and non-representative sampling into the MSE framework will further degrade HCR performance. It is also not clear how the way in which future catches are allocated by region/fleet impacts the results of this paper. The lack of difference between scenarios that varied in the degree of spatial connectivity can be attributed somewhat to the allowance of a shortfall in the catch in one region (as a result of insufficient available biomass) to be taken in the other region if required.

Most fisheries and also fished populations exhibit spatial structure, creating spatial heterogeneity in realized exploitation rates and biomass trends, with the extent of heterogeneity depending on the level of mixing in the stock. However, most management agencies lack the ability (or rather, the infrastructure) to specify the TAC at the level of this spatial structuring. HCRs that show robustness to spatial differences are therefore desirable. Disaggregating the data by fleet and region, analyzing these data separately, and then choosing the maximum estimate of F to determine the RBC appears to produce the most conservative results irrespective of the true nature of stock connectivity and fishery behaviour, and also resulted in the most consistent values for the performance measures among scenarios that varied in spatial structure. However, application of this version of the HCR leads to perhaps unnecessarily low catches when the connectivity of the stock between regions is high, and reflects the impact that sampling error can be expected to have on HCR performance in such scenarios. The maximum F option would also be inappropriate if the maximum F estimate came from a sector of the fishery which was a minor component of the catch, as it would be more likely that such an F estimate would be both uncertain and not representative of the overall exploitation rate. Weighting fleet and regional estimates of F by their variance accounts for this if the data are collected proportionally with the catch, as was implemented here. If not, then additional rules to determine how to proceed will be needed. For example, weighting the F estimates by catch rather than variance. Such methods however will not accommodate the effects of dome-shaped selectivity have on over-estimation of F. Spatial disaggregation of data that already has low sample size will result in more variable estimates of mortality than might be expected given population dynamics, particularly when constructing annual catch curves.

The use of an MSE approach enables the examination of control rules used to set catch limits, by evaluating performance given the known true state of the system. Such a framework can be used to identify strategies that perform poorly in the fairly well-ordered structure of the simulation. Perhaps more importantly, the relative performance of different strategies can be compared. The adoption of a precautionary approach to management of exploited marine resources is increasingly common, and it is clear that testing of harvest control rules is necessary to understand whether these rules can be expected to perform as intended. The analyses described focus on parameterizations of the operating model which mimic blue eye trevalla, but the system can be extended to examine the performance of the tier framework given different life histories. Indeed, a natural avenue for further extension of these analyses would be to examine whether the relative performance of the various Tier 3 control rules depends on the life-history of the species of interest, and whether the various HCRs need to be modified with life history.

While improved performance and conservation of stock biomass is achieved under the new Tier 3 HCR over that of the old form, the variability around the stock biomass, and in catches under this HCR are greater than that expected for a more data-rich scenario (e.g. integrated assessment using Stock Synthesis). This is to be expected; data-poor methods should estimate quantities of interest with greater uncertainty than those for which more data are available. While the Tier 3 HCR based on reference points such as F_{40} and F_{20} is more equivalent to the Tier 1 HCR, care should be taken regarding the ability to estimate F sufficiently well to be able to apply this rule successfully. Application of these reference points under the Tier 3 HCR requires an estimate of the value for steepness, which cannot be obtained during the analysis and was assumed to be 0.75 regardless of the true value. Consequently, performance of the HCRs was poor when the true value for steepness was lower than this. However, the approach taken here is not much different than that employed for data-rich scenarios in the SESSF, as estimation of steepness within stock assessments for this fishery is restricted to the assessments for tiger flathead (*Neoplatycephalus richardsoni*, e.g. Klaer 2010) and eastern gemfish (*Rexea solandri*, Little and Rowling 2009). In general, although the new Tier 3 HCR only relies on one source of current data from the fishery, application of the HCR requires estimates be available for the majority of biological parameters that would be included in a more formal stock assessment based on a statistical catch-at-age model. The performance of such estimators (e.g. Stock Synthesis) given paucity of information, the robustness of such models to mis-specifications such as those investigated here, and comparison of performance with the Tier 3 HCRS warrants interest.

The desired F to be estimated is the current rate of fishing mortality, whereas the annual catch curve integrates over the fishing mortality rates experienced by the stock for the length of the age structure, which may either not correspond well with recent trends in F, or, if data are noisy, may impede estimation of F. Poor ability to estimate F may mean a lack of ability to truly discriminate between the reference points involved in the HCR. This may be particularly important for long-lived, late-maturing species where F_{40} and F_{20} are similar. Successful implementation of a harvest control rule relies on being able to distinguish between values for stock indicators that result in changes in management action. Approximate confidence intervals for the current rate of fishing mortality on blue eye trevalla based on application of the MYEF estimator to data from the auto-longline fishery are wider than the range of F's over which changes in management actions are indicated given the HCR (Fay 2009).

Precaution with respect to Tier 1 is not explicitly built into the Tier 3 HCR at present, particularly as the quantities for F_{TARG} and F_{LIM} are the same as for Tier 1 (even though their estimates may be different). Conservation of stock biomass under the Tier 3 HCR arises from the behaviour of the rules. Additional measures to modify the Tier 3 HCR such that there is equivalency of risk with the Tier 1 HCR could involve the choice of alternative reference points (*e.g.* $F_{TARG} = F_{50}$), the application of a discount to the RBC for being at a less data-rich tier level (Smith et al. 2008), or perhaps application of current HCRs with a more conservative value for F_{CUR}, based on some percentile of the confidence interval of the estimate. Further simulation testing to address the efficacy of such approaches is clearly warranted, and is a suitable candidate for future work.

6. Acknowledgements

Funding for GF was provided by CSIRO Marine and Atmospheric Research as part of the SESSF stock assessment project. The members of the SESSF stock assessment team, J. Day, M. Haddon, N. Klaer, R. Little, R. Thomson, G. Tuck, J. Upston, and S. Wayte, are thanked for extensive discussions and input.

7. References

ADMB Project. 2009. AD Model Builder: automatic differentiation model builder. Developed by David Fournier and freely available from admb-project.org.

Butterworth, D. S., K. L. Cochrane, and J. A. A. De Oliveira. 1997. Management Procedures: a better way to manage fisheries? The South African experience. Pages 83-90 in E. L. Pikitch, D. D. Huppert, and M. P. Sissenwine (Eds). Global trends: Fisheries Management. American Fisheries Society, Symposium 20, Bethesda, Maryland.

Butterworth, D. S. and A. E. Punt. 1999. Experiences in the evaluation and implementation of management procedures. ICES Journal of Marine Science 56: 985-998.

Candy, S. G. 2008. Estimation of effective sample size for catch-at-age and catch-at-length data using simulated data from the Dirichlet-multinomial distribution. CCAMLR Science 15: 115-138.

Chapman, D. G. and D. S. Robson. 1960. The analysis of a catch curve. *Biometrics* 16: 354-368.

Cooke, J. G. 1999. Improvement of fishery-management advice through simulation testing of harvest algorithms. ICES Journal of Marine Science 56: 797-810.

Dunn, A., R. I. C. C. Franics, and I. J. Doonan. 2002. Comparison of the Chapman-Robson and regression estimators of *Z* from catch-curve data when non-sampling stochastic error is present. Fisheries Research 59: 149-159.

De Oliveira, J. A. A., L. T. Kell, A. E. Punt, B. A. Roel, and D. S. Butterworth. 2008. Managing without best predictions: the Management Strategy Evaluation framework. Pages 104-134 in Advances in Fisheries Science; 50 years on from Beverton and Holt. Blackwell Publishing. Oxford.

Fay, G. 2007. Tier 3 calculations for blue eye trevalla (*Hyperoglyphe antarctica*) using data up to and including 200 Pages 524-531 in Tuck, G. N. (ed.) 2007. Stock Assessment for the Southern and Eastern Scalefish and Shark Fishery 2006-2007. Volume 1: 2006. Australian Fisheries Management Authority and CSIRO Marine and Atmospheric Research, Hobart. 577 p.

Fay, G. 2009. Tier 3 calculations for blue eye trevalla (*Hyperoglyphe antarctica*) using data up to and including 2007. Pages 211-225 in Tuck, G. N. (ed.) 2009. Stock Assessment for the Southern and Eastern Scalefish and Shark Fishery 2008. Part 2. Australian Fisheries Management Authority and CSIRO Marine and Atmospheric Research, Hobart. 331 p.

Hoenig, J. M. 1983. Empirical use of longevity data to estimate mortality rates. Fishery Bulletin 82: 898-903.

Jensen, A. L. 1996. Beverton and Holt life history invariants result from optimal trade-off of reproduction and survival. Canadian Journal of Fisheries and Aquatic Sciences 53: 820-822.

Kell, L. T., C. M. O’Brien, M. T. Smith, T. K. Stokes, and B. D. Rackham. 1999. An evaluation of management procedures for implementing a precautionary approach in the ICES context for North sea plaice (*Pleuronectes platessa* L.). ICES Journal of Marine Science 56: 834-84

Kell, L. T., M. A. Pastoors, R. D. Scott, M. T. Smith, F. A. Van Beek, C. M. O'Brien, and G. M. Pilling. 200 Evaluation of multiple management objectives for Northeast Atlantic flatfish stocks: sustainability vs. stability of yield. ICES Journal of Marine Science 62: 1104-1117.

Klaer, N. L. 2008. Yield and total mortality values and Tier 3 estimates for selected shelf and slope species in the SESSF 2008. In Tuck, G.N. (ed.) 2009. Stock Assessment for the Southern and Eastern Scalefish and Shark Fishery 2008. Australian Fisheries Management Authority and CSIRO Marine and Atmospheric Research, Hobart. 645 p.

Klaer, N. L. 2010. Tiger flathead (*Neoplatycephalus richardsoni*) stock assessment based on data up to 2008. Pages 164-189 in Tuck G, N. (ed.) 2010. Stock Assessment for the

Southern and Eastern Scalefish and Shark Fishery 2009. Part 1. Australian Fisheries Management Authority and CSIRO Marine and Atmposheric Research, Hobart. 334 p.

Klaer, N. L., S. E. Wayte, A. E. Punt, L. R. Little, A. D. M. Smith, R. B. Thomson, and G. N. Tuck. 2009. Simulation testing of alternative Tier 3 assessment methods and control rules for the SESSF. Pages 42-70 in Wayte, S.E. (ed.) 2009. Evaluation of New Harvest Strategies for SESSF Species. CSIRO Marine and Atmospheric Research, Hobart and Australian Fisheries Management Authority, Canberra. 137 p.

Little, L. R., and K. Rowling. 2009. Eastern Gemfish (*Rexea solandri*) stock assessment based on 2008 survey data. Pages 167-208 in Tuck G, N. (ed.) 2009. Stock Assessment for the Southern and Eastern Scalefish and Shark Fishery 2008. Part 1. Australian Fisheries Management Authority and CSIRO Marine and Atmposheric Research, Hobart. 344 p.

Mace, P. M., and I. J. Doonan. 1988. A generalized bioeconomic simulation model for fish population dynamics. New Zealand Fisheries Assessment Research Document 88/4, MAFFish Fisheries Research Centre, Wellington.

MacCall, A. 2007. Appendix D - Depletion-adjusted average catch. Pages 27-30 in: Rosenberg, A., D. Agnew, E. Babcock, A. Cooper, C. Mogensen, R. O'Boyle, J. Powers, G. Stefansson, and J. Swasey, 2007. "Setting Annual Catch Limits for U.S. Fisheries: An Expert Working Group Report". Lenfest Ocean Program, Washington DC. 36 p.

MacCall, A. 2009. Depletion-corrected average catch: a simple formula for estimating sustainable yields in data-poor situations. ICES Journal of Marine Science, 66: 2267–2271.

Methot, R. D. 2007. User Manual for the Integrated Analysis Program Stock Synthesis 2 (SS2). Model Version 2.00b. 84 p.

Miller, T.J. and J.R. Skalski. 2006. Integrating design- and model-based inference to estimate length and age composition in North Pacific longline catches. Canadian Journal of Fisheries and. Aquatic Sciences 63: 1092–1114.

Pauly, D. 1980. On the interrelationships between natural mortality, growth parameters, and mean environmental temperature in 175 fish stocks. J. Cons. int. Explor. Mer 39: 175-192.

Restrepo, V. R., and J. E. Powers. 1999. Precautionary control rules in US fisheries management: specification and performance. ICES Journal of Marine Science 56: 846-852.

Smith, A. D. M., K. J. Sainsbury, and R. A. Stevens. 1999. Implementing effective fisheries management systems – management strategy evaluation and the Australian partnership approach. ICES Journal of Marine Science 56: 967-979.

Smith, A. D. M., and D. C. Smith. 200 A harvest strategy framework for the SESSF. Report to the Australian Fisheries Management Authority, Canberra, June 200

Smith, A. D. M., D. C. Smith, G. N. Tuck, N. Klaer, A. E. Punt, I. Knuckey, J. Prince, A. Morison, R. Kloser, M. Haddon, S. Wayte, J. Day, G. Fay, F. Pribac, M. Fuller, B. Taylor, and L. R. Little. 2008. Experience in implementing harvest strategies in Australia's south-eastern fisheries. Fisheries Research 94: 373-379

Smith, A. D. M. and S. E. Wayte (eds). 2002. The South East Fishery 2001. Fishery Assessment Report compiled by the South East Fishery Assessment Group. Australian Fisheries Management Authority, Canberra.

Walters, C. J., and S. J. D. Martell. 2004. Fisheries Ecology and Management. Princeton University Press. Princeton, USA.

Table 1 : Parameterisation of the operating model for the non-spatial scenarios. h is the steepness parameter of the spawner-recruit relationship, B_{curr}/B_0 is the spawning biomass relative to unfished prior to implementation of the HCRs, and n_A is the annual sample size for the age composition data.

#	Scenario	Type of selectivity curve	true M	h	B_{curr}/B_0	n_A	assumed M
1	base-case	asymptotic	0.08	0.75	0.40	100	0.08
2		asymptotic	0.08	0.75	0.20	100	0.08
3		asymptotic	0.08	0.75	0.75	100	0.08
4		asymptotic	0.12	0.75	0.40	100	0.12
5		asymptotic	0.12	0.75	0.20	100	0.12
6		asymptotic	0.12	0.75	0.75	100	0.12
7		asymptotic	0.18	0.75	0.40	100	0.18
8		asymptotic	0.18	0.75	0.20	100	0.18
9		asymptotic	0.18	0.75	0.75	100	0.18
10	low steepness	asymptotic	0.08	0.30	0.40	100	0.08
11		asymptotic	0.08	0.30	0.20	100	0.08
12		asymptotic	0.08	0.30	0.75	100	0.08
13	high steepness	asymptotic	0.08	1.00	0.40	100	0.08
14		asymptotic	0.08	1.00	0.20	100	0.08
15		asymptotic	0.08	1.00	0.75	100	0.08
16	dome-shaped selectivity	dome-shaped	0.08	0.75	0.40	100	0.08
17		dome-shaped	0.08	0.75	0.20	100	0.08
18		dome-shaped	0.08	0.75	0.75	100	0.08
19	assume wrong M	asymptotic	0.08	0.75	0.40	100	0.05
20		asymptotic	0.08	0.75	0.40	100	0.12
21		asymptotic	0.12	0.75	0.40	100	0.08
22		asymptotic	0.12	0.75	0.40	100	0.18

Table 2 : Parameterisation of the operating model for the spatial scenarios. 'Full' connectivity between regions implies single stock dynamics, 'intermediate' has 20% annual movement rate from one region to the other, while the 'limited' scenario only has a 5% annual movement rate. The movement patterns are as shown in Figure 1.

#	Scenario	Type of selectivity curve region 1	Type of selectivity curve region 2	true M	h	B_{curr}/B_0	n_A	assumed M	connectivity	movement pattern
1	base-case, full mixing	asymptotic	asymptotic	0.08	0.75	0.40	100	0.08	full	constant
2	intermediate mixing	asymptotic	asymptotic	0.08	0.75	0.40	100	0.08	intermediate	constant
3		asymptotic	asymptotic	0.08	0.75	0.20	100	0.08	intermediate	constant
4		asymptotic	asymptotic	0.08	0.75	0.75	100	0.08	intermediate	constant
5	limited connectivity	asymptotic	asymptotic	0.08	0.75	0.40	100	0.08	limited	constant
6		asymptotic	asymptotic	0.08	0.75	0.20	100	0.08	limited	constant
7		asymptotic	asymptotic	0.08	0.75	0.75	100	0.08	limited	constant
8	movement declines with age	asymptotic	asymptotic	0.08	0.75	0.40	100	0.08	intermediate	pre-recruit
9	movement increases with age	asymptotic	asymptotic	0.08	0.75	0.40	100	0.08	intermediate	adult
10	dome-shaped selectivity	dome-shaped	dome-shaped	0.08	0.75	0.40	100	0.08	intermediate	constant
11	differing selectivities	asymptotic	dome-shaped	0.08	0.75	0.40	100	0.08	intermediate	constant

Table 3 : Coefficients estimated from the linear models for the non-spatial analyses, by performance measure. Numbers shown represent terms that were assessed to be significant at the α = 0.05 level in a full model that included all terms listed.

	Performance measure						
Linear predictor term	med(B_{final}/B_0)	IQR(B_{final}/B_0)	P($B_{final}<B_{lim}$)	P($B_{proj}<B_{lim}$)	med(avg TAC)	med(CV TAC)	# yrs collapse
base intercept		0.17	0.79	0.87	721	0.16	
(CR, old HCR, InitDepl=0.4, *h* =0.75, asymptotic Sel)							
MYEF							
new HCR, Ftarg=0.5*M* , Flim=*M*	0.14	0.10	-0.17	-0.13	-209	0.96	5.9
new HCR, Ftarg=F40, Flim=F20, adjust for h	0.17	0.15	-0.22	-0.15	-249	0.95	5.4
assumed *M* > true *M*	-0.14	-0.26	0.35	0.22		0.54	4.6
assumed *M* < true *M*	0.56		-0.64	-0.75	-468	0.85	5.7
Initial depletion = 0.75	0.39	0.24	-0.59	-0.64	476	0.20	
Initial depletion = 0.2				0.12	-394	0.62	4.8
h = 1.0	0.15		-0.17	-0.08			
h = 0.3	-0.14	-0.15	0.32	0.11		0.39	3.6
true *M* = 0.12							
true *M* = 0.18					150		-1.3
new HCR, Ftarg=0.5*M* : Init Depl = 0.75 interaction	-0.58	-0.51	0.90	0.84	1279		-2.4
new HCR, Ftarg=F40 : Init Depl = 0.75 interaction	-0.60	-0.52	0.94	0.87	1108		
new HCR, Ftarg=0.5*M* : Init Depl = 0.2 interaction	0.19		-0.32			-0.30	-2.8
new HCR, Ftarg=F40 : Init Depl = 0.2 interaction	0.23		-0.34				
dome-shaped selectivity : old HCR interaction							
dome-shaped selectivity : new HCR 0.5*M* interaction	0.12		-0.20	-0.14	-216		
dome-shaped selectivity : new HCR F40 interaction							

Table 4 : Values for the performance measures for the spatial analyses given estimation under MYEF, application of the new Tier 3 HCR, with $F_{\mathrm{TARG}} = F_{40}$ and $F_{\mathrm{LIM}} = F_{20}$. '$F$ option' is the method used to obtain a single F_{CUR} estimate given the fleet and regional data (Section *3.3.3*): 1) aggregate fleet and regional data, 2) obtain fleet and region-specific F estimates, weight by inverse variance, 3) separate analyses as in 2, but use the maximum estimated F to calculate the RBC. Scenario numbers as in Table 2.

		Performance measure						
F option	Scenario #	med(B_{final}/B_0)	IQR(B_{final}/B_0)	P(B_{final}<B_{lim})	P(B_{proj}<B_{lim})	med(avg TAC)	med(CV TAC)	# yrs collapse
1)	1	0.12	0.36	0.61	0.77	491	0.99	3.9
	2	0.24	0.35	0.42	0.69	479	0.95	2.6
	3	0.46	0.24	0.09	0.90	114	1.58	7.1
	4	0.00	0.00	0.97	0.98	1,944	1.48	5.3
	5	0.21	0.34	0.50	0.74	478	1.03	3.8
	6	0.47	0.22	0.07	0.86	90	1.56	7.2
	7	0.00	0.00	0.94	0.97	2,122	1.41	4.6
	8	0.21	0.42	0.47	0.58	458	1.02	4.1
	9	0.14	0.34	0.61	0.73	511	0.92	3.2
	10	0.47	0.44	0.26	0.33	288	1.00	2.6
	11	0.33	0.39	0.32	0.54	404	0.96	2.6
2)	1	0.15	0.42	0.55	0.72	436	1.04	4.1
	2	0.15	0.33	0.56	0.75	481	0.97	3.3
	3	0.44	0.22	0.09	0.87	109	1.53	6.8
	4	0.00	0.00	0.99	0.99	2,006	1.46	5.1
	5	0.19	0.42	0.56	0.71	479	1.05	3.7
	6	0.45	0.25	0.14	0.88	98	1.53	6.7
	7	0.00	0.00	0.97	0.97	2,161	1.47	5.2
	8	0.00	0.07	0.83	0.91	494	1.39	7.0
	9	0.07	0.36	0.60	0.73	508	0.92	3.3
	10	0.37	0.41	0.29	0.45	344	0.92	2.2
	11	0.22	0.40	0.46	0.60	443	0.94	2.9
3)	1	0.38	0.28	0.24	0.47	353	0.95	2.6
	2	0.35	0.36	0.26	0.48	367	0.98	2.9
	3	0.53	0.16	0.05	0.87	71	1.72	8.9
	4	0.00	0.11	0.83	0.84	1,899	1.27	3.5
	5	0.39	0.41	0.29	0.48	310	1.05	3.0
	6	0.52	0.18	0.02	0.83	46	1.86	10.2
	7	0.00	0.06	0.81	0.82	1,834	1.27	3.9
	8	0.62	0.23	0.01	0.07	135	1.12	3.5
	9	0.45	0.34	0.21	0.39	322	0.90	1.8
	10	0.59	0.27	0.12	0.22	205	1.02	2.9
	11	0.49	0.35	0.14	0.30	282	1.10	3.3

Table 5 : Coefficients estimated from the linear models for the spatial analyses. Numbers shown represent terms that were assessed to be significant (p<0.05) in a full model that included all terms listed. Base intercept values were: *MYEF*, new HCR with F_{TARG}=*F40* and F_{LIM}=*F20*, obtain a single *F* estimate with all data, full mixing between regions, movement constant with age, initial depletion = 0.4, and h=0.7

	Performance measure						
Linear predictor term	med(B_{final}/B_0)	IQR(B_{final}/B_0)	P($B_{final}<B_{lim}$)	P($B_{proj}<B_{lim}$)	med(avg TAC)	med(CV TAC)	# yrs collapse
base intercept	0.19	0.37	0.51	0.71	468	0.97	3.5
wt regional *F* estimates by variance							
choose highest regional *F*	0.15		-0.21	-0.21	-141		
Initial depletion = 0.75	-0.25	-0.34	0.49	0.29	1562	0.39	1.4
Initial depletion = 0.2	0.22	-0.16	-0.36	0.23	-344	0.63	4.6
Different selectivities by region							
Selectivity dome-shaped in both regions	0.23		-0.21	-0.31	-145		
intermediate mixing							
limited mixing							
juveniles move only		-0.12				0.19	1.8
adults move only							

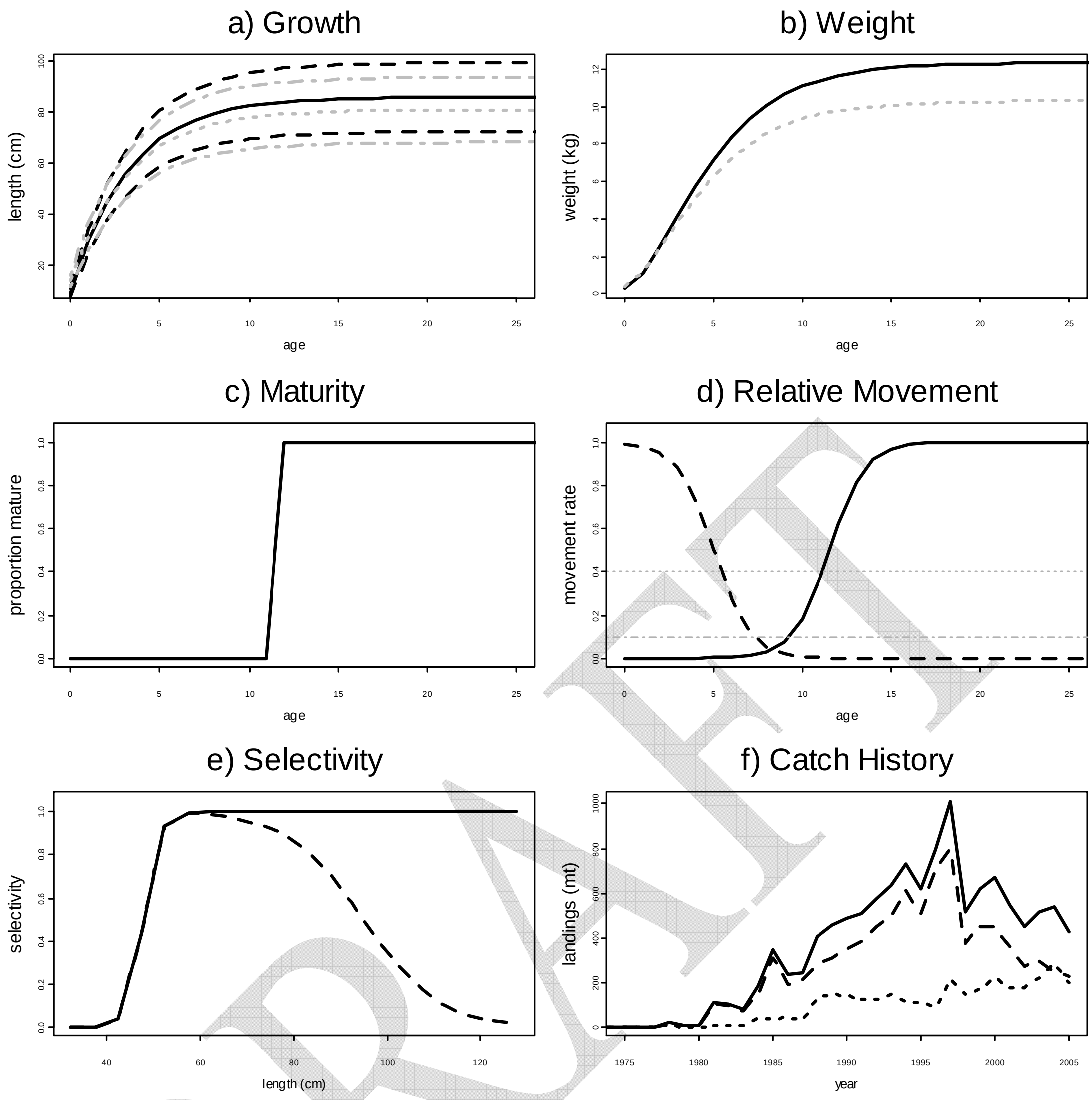

Figure 1 : Biological and fishery-related parameters. Top row of panels: values for females shown in black lines, males in grey. Solid lines in Growth panel represent mean lengths-at-age, dashed lines correspond to the 95% intervals of the distribution for length–at-age. Relative Movement panel shows pattern of relative movement rate for (solid) adult-only movement, and (dashed) pre-recruit movement. Gray dotted and dot-dashed lines indicate rates of movement (relative to “full” mixing scenario) for the “intermediate” and “limited” movement scenarios. Selectivity panel shows both asymptotic (solid line) and dome-shaped (dashed line) patterns with length. Catch history panel indicates both total catches (solid line) and regional catch histories used in the spatial analyses, with dashed line indicating catches from region 1, and dotted line indicating the catch from region 2.

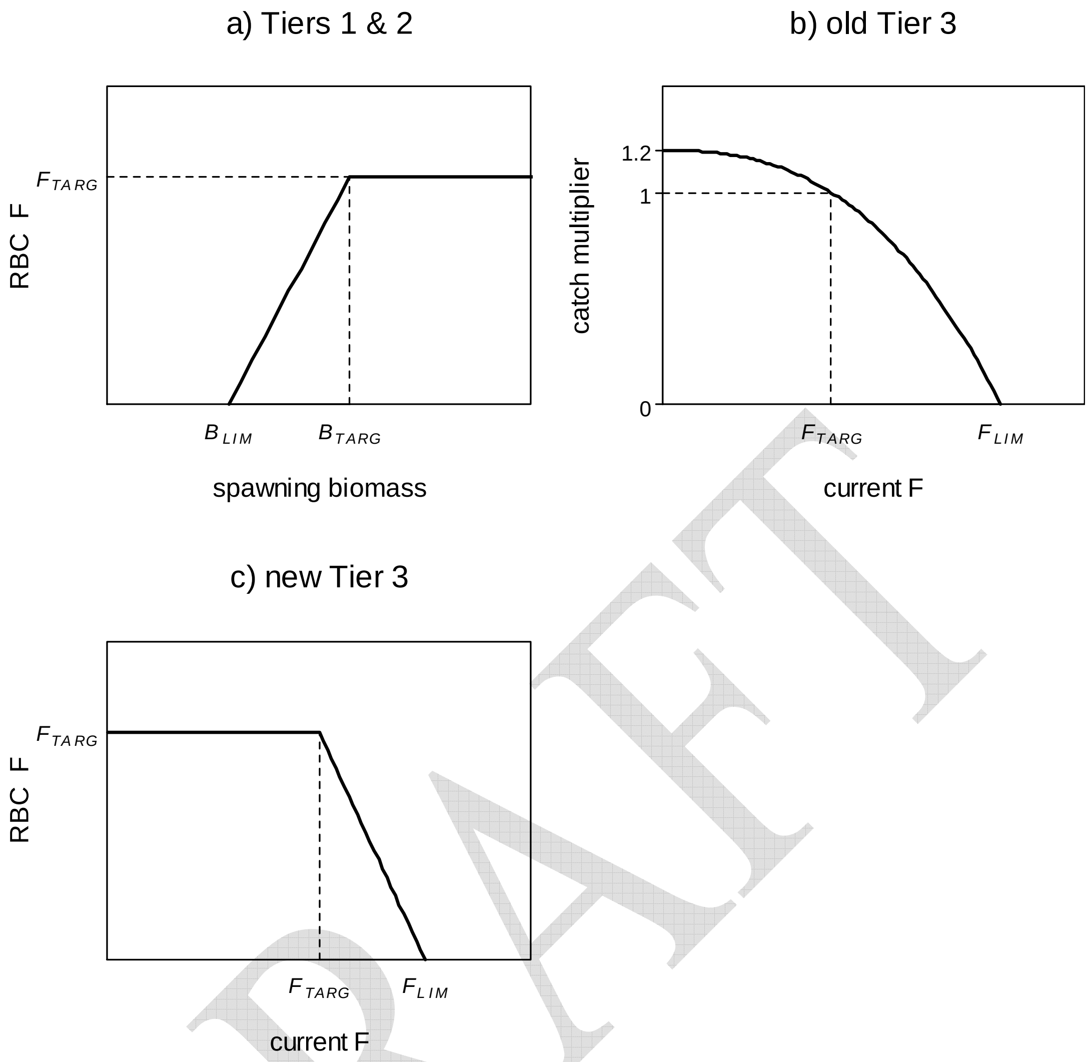

Figure 2 : Forms for the Harvest Control Rules (HCRs) for Tiers 1 and 2 (top-left panel), old Tier 3 (top-right panel), and new Tier 3 (bottom-left panel). The estimated value for the stock indicator on the x axis is used to derive either the RBC rate of fishing mortality (Tier 1 and new Tier 3), or the multiplier to the current catch (old Tier 3).

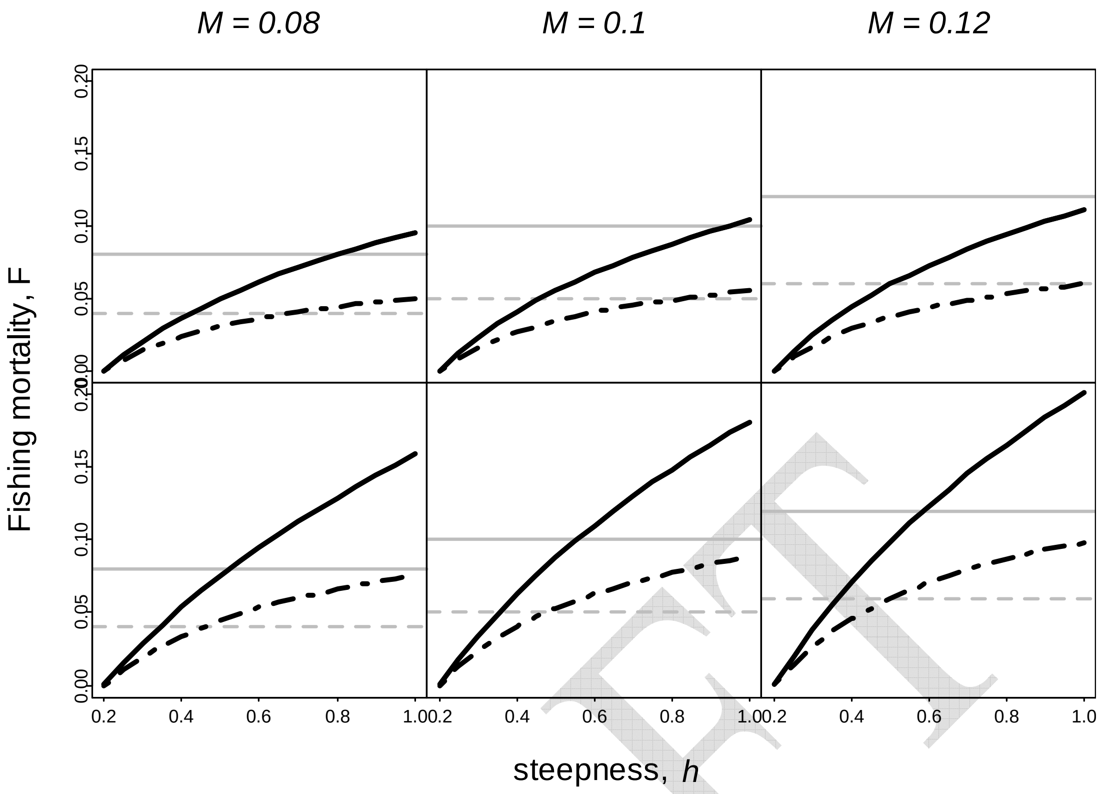

Figure 3 : Relationship between F_{40} (solid black line) and F_{20} (dot-dashed black line) and M, and h. The solid and dashed grey lines are M and $0.5M$ respectively. Top row of panels corresponds to an age-at-maturity of 12 yrs, as used in the analyses presented here. The bottom row of panels shows the change for an age-at-maturity of 6 yrs.

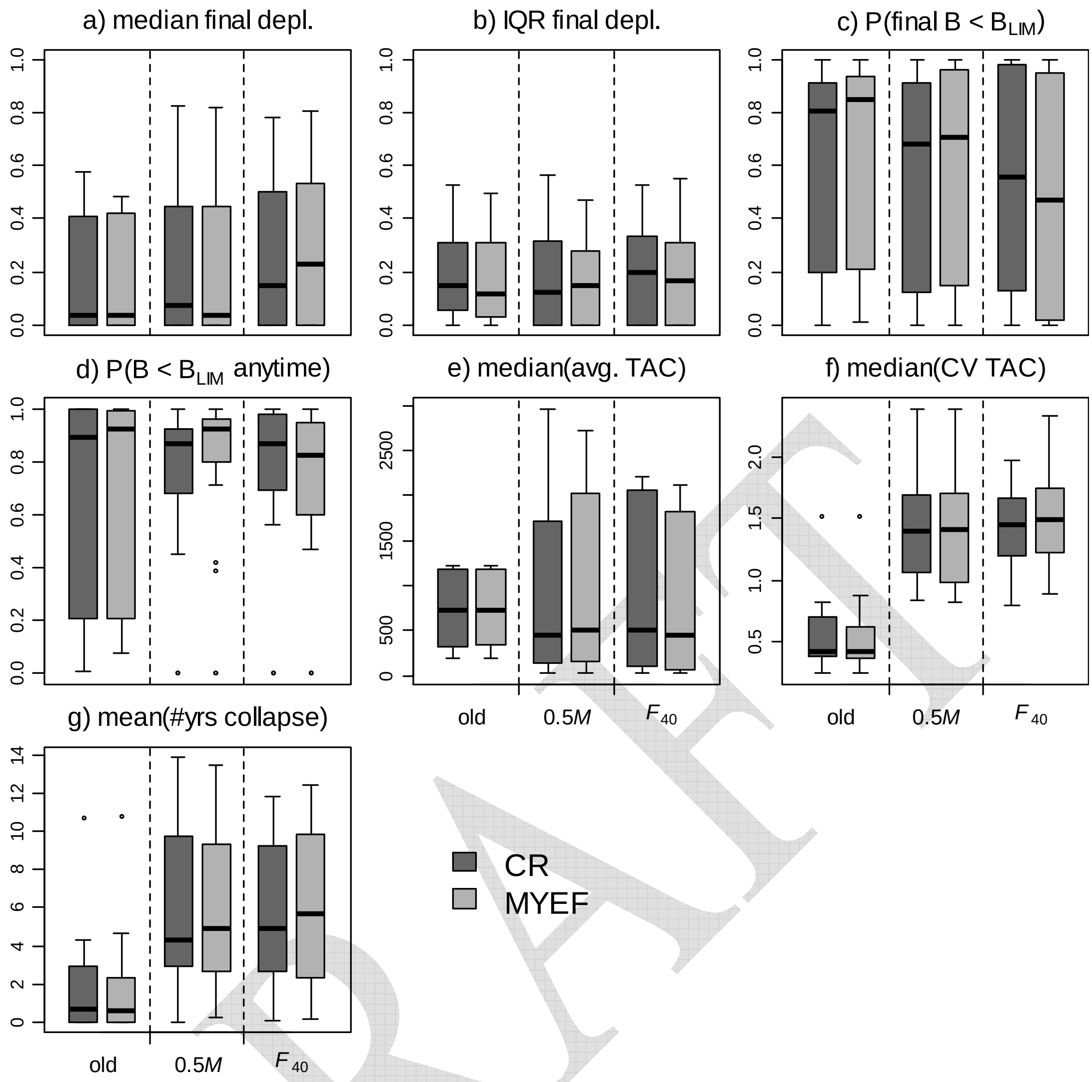

Figure 4 : Distribution of the performance measures across scenarios for the non-spatial analyses, for the two estimation methods, CR and MYEF, given application of the three HCRs (old = old Tier 3; *0.5M* = new Tier 3 HCR with F_{TARG}= *0.5M*. and F_{LIM}= *M*; F40 = new Tier 3 HCR with F_{TARG}= F_{40}. and F_{LIM}= F_{20}).

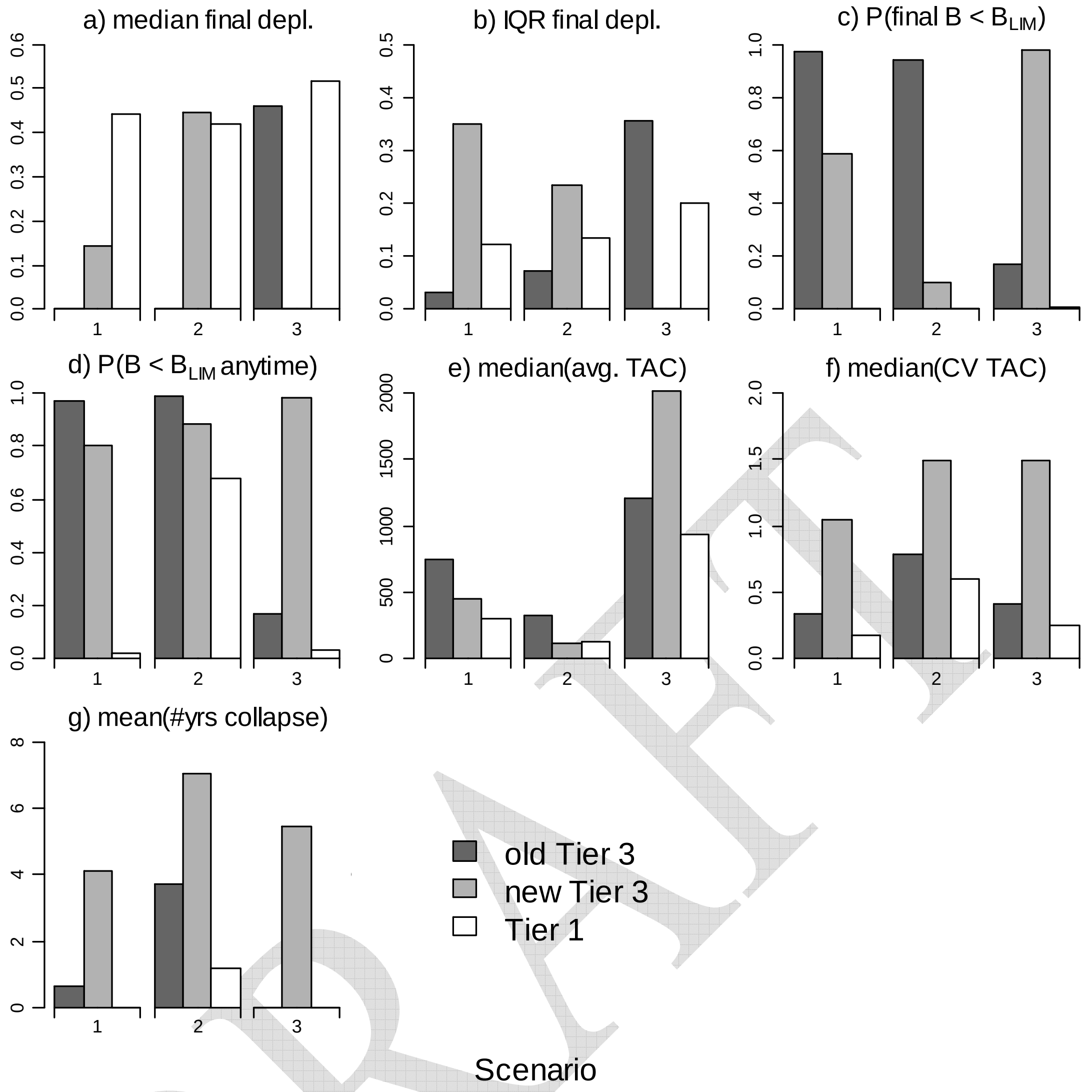

Figure 5 : Comparing performance of the old Tier 3 HCR, the new Tier 3 MYEF $F_{TARG} = F_{40}$, and the Tier 1 HCR, for three of the non-spatial scenarios. Numbers correspond to scenarios listed in Table 1.

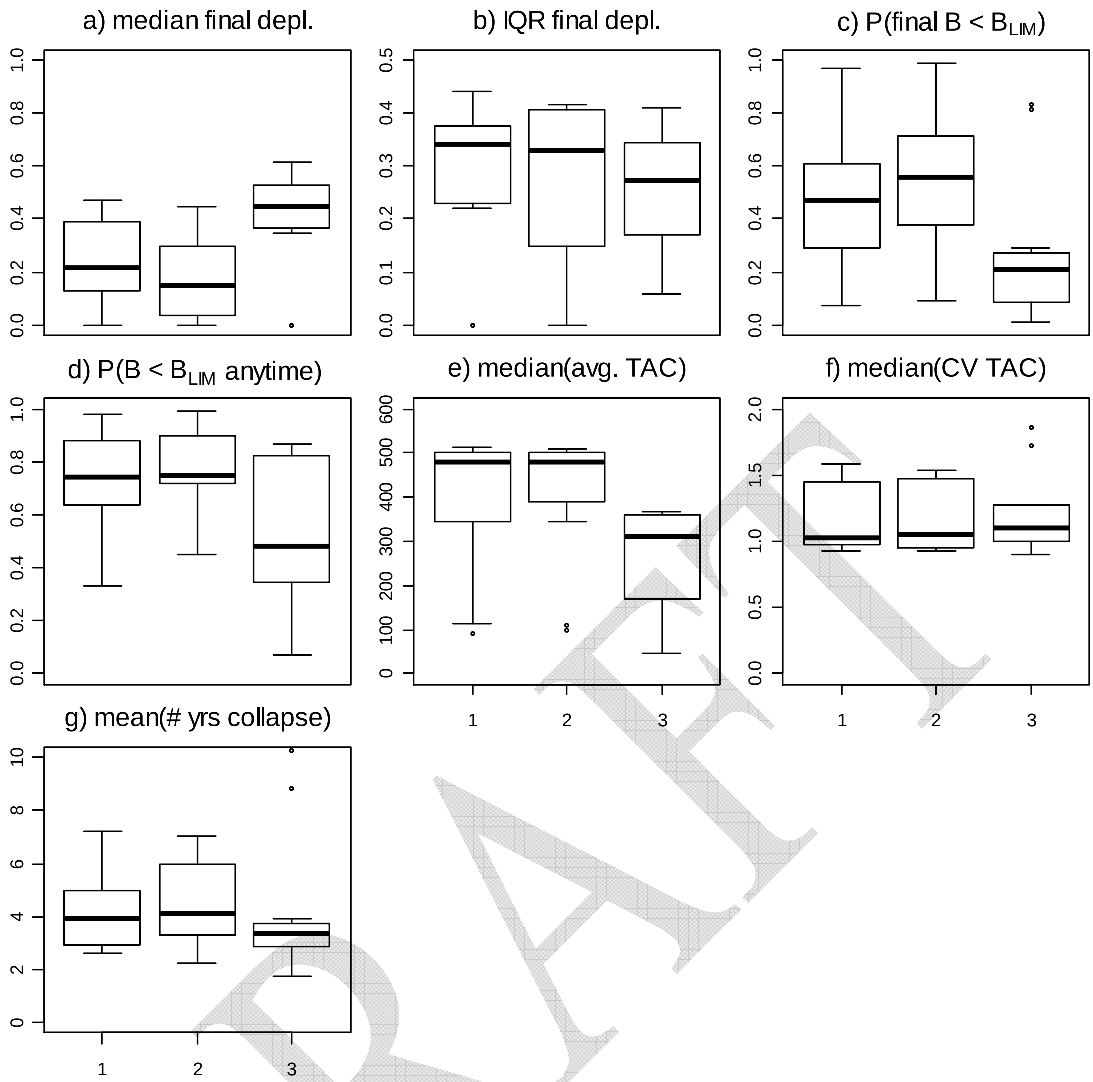

Figure 6 : Distribution of the performance measures across scenarios for the spatial analyses, for the different ways of choosing the annual F estimate (1 = aggregate data, 2 = analyse by region, weight estimates by variance, 3 = analyse by region and choose the maximum estimated F). Estimation is using MYEF with the new Tier 3 HCR with $F_{TARG}= F_{40}$. and $F_{LIM}= F_{20}$.

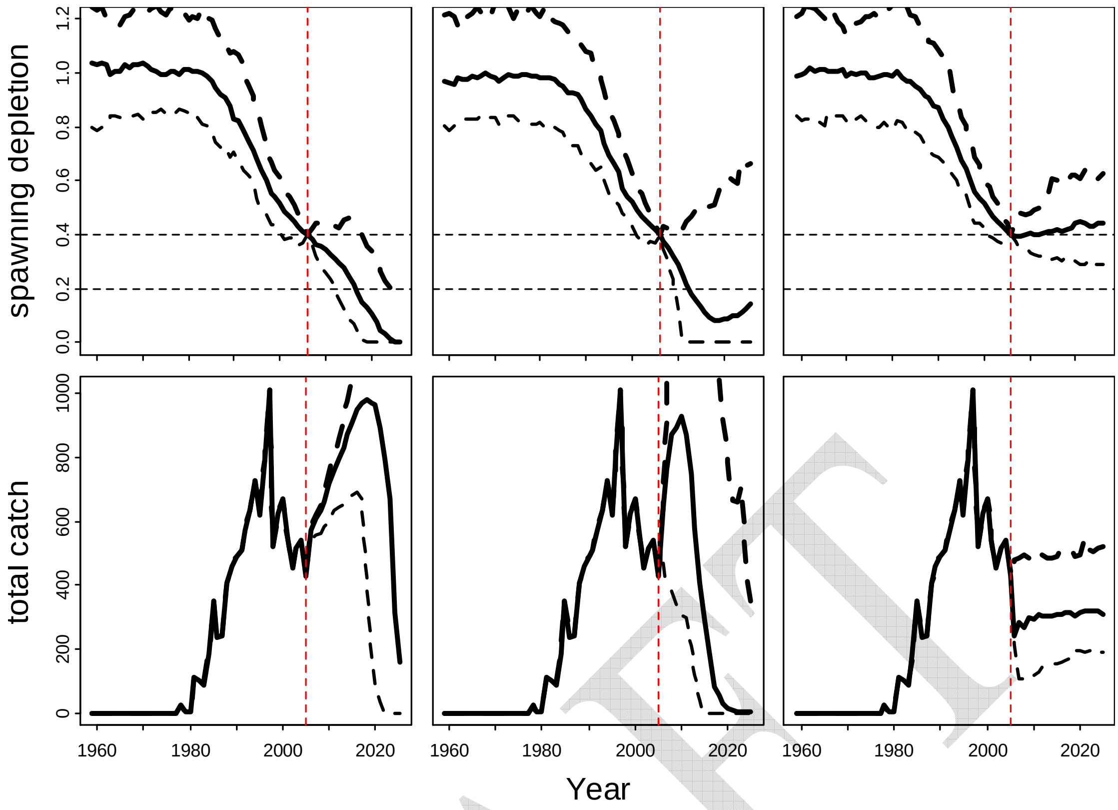

Figure 7 : Relative spawning biomass and catch trajectories for scenario 1 of the non-spatial analyses for (left) old Tier 3 HCR, (center) new Tier 3 HCR with MYEF and F_{TARG}=F_{40}, and (right) Tier 1. Plotted are the median (solid line) and central 95% interval (dashed lines) from the simulations. Red vertical line indicates the start of implementation of the harvest strategy, horizontal black dotted lines indicate the target (B_{40}) and limit (B_{20}) reference points.

Appendix 1. Operating Model Specifications

The operating model consists of an age-structured population dynamics model, a data-generation module, and a component to allow future projections of the population model given input from estimation methods and HCRs. The operating model can be appropriately dimensioned and parameterised to account for several spatial regions and multiple fleets in order to capture key dynamics for blue eye trevalla. The specifications are a simplified version of those used to evaluate management strategies for a variety of species within the SESSF (Fay et al. 2009).

A1.1 Population dynamics

The operating model includes one or more regions. Population dynamics operate at the level of the fish stock, with a single stock occupying one or more regions. Fishing fleets operate in one or more regions.

A1.1.1 Abundance dynamics

The number of animals of sex *s* and age *a* in region *r* at the start of year *t*, $N^r_{s,a,t}$ is given by:

$$N^r_{s,a,t} = \begin{cases} \tilde{N}^r_{s,a-1,t-1} & \text{if } 1 \le a < x \\ \tilde{N}^r_{s,x-1,t-1} + \tilde{N}^r_{s,x,t-1} & \text{otherwise} \end{cases} \tag{7}$$

where $\tilde{N}^r_{s,a,t}$ is the number of animals of sex *s* and age *a* in region *r* following mortality (all sources) and movement during year *t*:

$$\tilde{N}^r_{s,a,t} = \bar{N}^r_{s,a,t} + \sum_{r' \neq r} \bar{N}^{r'}_{s,a,t} X^{r',r}_{s,a} - \sum_{r' \neq r} \bar{N}^r_{s,a,t} X^{r,r'}_{s,a} \tag{8}$$

$$\bar{N}^r_{s,a,t} = N^r_{s,a,t} e^{-M} (1 - u^r_{s,a,t}) \tag{9}$$

$X^{r',r}_{s,a}$ is the proportion of animals of sex *s* and age *a* moving from region *r*’ to region *r*,

M is the rate of natural mortality,

$u^r_{s,a,t}$ the exploitation rate (due to all fleets) on animals of sex *s* and age *a* in region *r* during year *t*:

$$u^r_{s,a,t} = \sum_f \tilde{u}^{f,r}_t s^f_{s,a} \tag{10}$$

where:

$$\tilde{u}^{f,r}_t = \frac{C^{f,r}_t}{\sum_s \sum_L w_{L,s} S^f_L \sum_a \Phi_{L,s,a} N^r_{s,a,t} e^{-0.5M}} \tag{11}$$

$C^{f,r}_t$ is the retained catch by fleet *f* in region *r* during year *t*,

$S^f_{L,t}$ is the selectivity of fleet *f* on animals in length bin *L* during year *t*,

$\Phi_{L,s,a,t}$ is the proportion of fish of sex s and age a in length bin L during year t,

$s^f_{s,a,t}$ is the selectivity of fleet f on animals of sex s and age a during year t,

$w_{L,s}$ is the mean weight of a fish of sex s in length bin L, and

x is the maximum age (treated as a plus-group).

A1.1.2 Selectivity

The sex- and age-specific selectivity pattern for fleet f is calculated from the inputted length-specific selectivity pattern:

$$s^f_{s,a} = \sum_L S^f_L \Phi_{L,s,a} \tag{12}$$

$$\Phi_{L,s,a} = \begin{cases} \tilde{\Phi}\left(\dfrac{l^{lo}_L - \bar{l}_{s,a}}{\sigma_{l_{s,a}}}\right) & \text{if } L=1 \\ \tilde{\Phi}\left(\dfrac{l^{lo}_{L+1} - \bar{l}_{s,a}}{\sigma_{l_{s,a}}}\right) - \tilde{\Phi}\left(\dfrac{l^{lo}_L - \bar{l}_{s,a}}{\sigma_{l_{s,a}}}\right) & \text{if } 1<L<N_L \\ 1-\tilde{\Phi}\left(\dfrac{l^{lo}_L - \bar{l}_{s,a}}{\sigma_{l_{s,a}}}\right) & \text{if } L=N_L \end{cases} \tag{13}$$

where $L = l^{lo}_L + 0.5\left[l^{hi}_L - l^{lo}_L\right]$,

l^{hi}_L and l^{lo}_L are upper and lower limits of length bin L,

$\tilde{\Phi}$ is the standard normal cumulative density function,

$\bar{l}_{s,a,t}$ is the mean length of a fish of sex s and age a in the middle of year t,

$\sigma_{l_{s,a,t}}$ is the input standard deviation of the length of a fish of sex s and age a.

A1.1.3 Growth

The mean length-at-age by sex in year t is calculated by:

$$\bar{l}_{s,a,t} = L_{\infty,s}\left(1-\exp\left[k_s\,(a-t_{0,s})\right]\right) \tag{14}$$

where $L_{\infty,s}$, k_s, and $t_{0,s}$ are the input growth parameters for animals of sex s.

Weight-at-length is governed by a length-power relationship:

$$w_{L,s} = \alpha_s\,(L)^{\beta_s} \tag{15}$$

where α_s and β_s are the input parameters of the weight-length relationship for sex *s*.

A1.1.4 Recruitment

The annual recruitments (by region) are log-normally distributed about an underlying Beverton-Holt stock-recruitment relationship (SRR):

$$N^r_{s,0,t} = 0.5 R^r_t \, e^{\varepsilon^r_t - 0.5\sigma^2_R} \tag{16}$$

$$R^r_t = \lambda^r_t \left(\frac{4hR_0 SB_t}{SB_0 (1-h) + SB_t \, (5h-1)} \right) \tag{17}$$

where ε^r_t is the recruitment residual for region *r* and year *t*, which can be correlated among regions:

$$\boldsymbol{\varepsilon}_\mathbf{t} \sim MVN(0, \Sigma_t) \tag{18}$$

$$\Sigma_t = \sigma^2_R \begin{pmatrix} 1 & \rho^{r_i r_j} \\ \rho^{r_i r_j} & 1 \end{pmatrix} \tag{19}$$

R_0 is the number of age-0 animals at pre-exploitation equilibrium,

h is the steepness of the stock-recruitment relationship,

SB_0 is the spawning biomass at pre-exploitation equilibrium (when recruitment equals R_0),

σ_R is the standard deviation of the recruitment residuals,

$\rho^{r_i r_j}$ is the correlation between the recruitment residuals for regions r_i and r_j, set to 1 (perfect correlation among regions) for this paper, and

λ^r_t is the expected fraction of the number of age-0 animals assigned to region *r* during year *t*:

$$\lambda^r_t = \tilde{S}B^r_t / SB_t \tag{20}$$

The total spawning biomass during year *t* is given by:

$$SB_t = \sum_r \tilde{S}B^r_t = \sum_r \sum_{a=1}^{x} N^r_{\text{fem},a,t} \tilde{w}_{\text{fem},a} \, f_a \tag{21}$$

where f_a is the fraction of females of age *a* that are mature, and

$\tilde{w}_{\text{fem},a}$ is the weight at age of a female of age a at the start of the year:

$$\tilde{w}_{\text{fem},a} = \sum_L \Phi_{L,\text{fem},a} w_{L,\text{fem}} \tag{22}$$

A1.1.5 Movement

The probabilities of moving among regions are determined by:

$$X_{s,a}^{r',r} = \frac{\bar{X}_{s,a}^{r',r}}{\sum_{r'} \bar{X}_{s,a}^{r',r}} \tag{23}$$

where $\bar{X}_{s,a}^{r',r}$ is the average probability of an animal of sex s and age a moving from region r' to region r:

$$\bar{X}_{s,a}^{r',r} = \begin{cases} T_s^{r',r} m_{s,a} & \text{if } r' \neq r \\ 1 - \sum_{r' \neq r} T_s^{r',r} m_{s,a} & \text{otherwise} \end{cases} \tag{24}$$

$T_s^{r',r}$ is the maximum average probability of moving from region r' to region r, with $T_s^{r'+} = 1$, and

$m_{s,a}$ is the relative age-specific movement rate for an animal of sex s.

A1.1.6 Initial Conditions

The initial (t=1) numbers at age for each sex by region are determined by solving the following set of linear equations:

$$\mathbf{N}_{s,1} = \left(\mathbf{I} - \mathbf{G}_s\right)^{-1} \tilde{\mathbf{R}}_{s,1} \tag{25}$$

where $\mathbf{N}_{s,1}$ is an (x+1) x *Nreg* (number of regions) vector containing the initial age-structure for animals of sex s,

$\tilde{\mathbf{R}}_{s,1}$ is the corresponding vector of recruits with elements:

$$\tilde{R}_{s,a,1}^r = \begin{cases} 0.5 \lambda_0^r R_0 & \text{if } a = 0 \\ 0 & \text{if } 1 \leq a \leq x \end{cases} \tag{26}$$

where λ_0^r is the fraction of recruits allocated to region r in equilibrium, the value for which is solved for in order to satisfy equation (20), and

$\mathbf{G}_s$ is a square transition matrix with the same dimension as $\mathbf{N}_{s,1}$, describing the mortality and movement pattern, the elements of which are obtained from the equations for the population update:

$$G^{r_p,r_q}_{s,a_p,a_q} = \begin{cases} X^{r_q,r_p}_{s,a_p} e^{-M} & \text{if } a_p = a_q \text{ -1} \\ X^{r_q,r_p}_{s,a_p} e^{-M} & \text{if } a_p = a_q = x \\ 0 & \text{otherwise} \end{cases} \tag{27}$$

a_p is the age associated with row *p*,

a_q is the age associated with column *q*,

r_p is the region associated with row *p*, and

r_q is the region associated with column *q*.

A1.2 Generating Age-composition Data

The observed catch-at-age proportions by region, sex and fleet are a multinomial sample of size $n^{f,r}_{s,t}$ from the true catch-at-age proportions. The proportion of the catch that is of age *a* during year *t* for fleet *f* and sex *s* in region *r* is:

$$p^{f,r}_{a,s,t} = \frac{\tilde{C}^{f,r}_{s,a,t}}{\sum_a \tilde{C}^{f,r}_{s,a,t}} \tag{28}$$

where:

$$\tilde{C}^{f,r}_{s,a,t} = I^{f,r} \sum_L S^f_L \Phi_{L,s,a} N^r_{s,a,t} e^{-0.5M} \tag{29}$$

$I^{f,r}$ is an indicator equal to 1 if fleet *f* operates in region *r*, and zero otherwise.

The individual $n^{f,r}_{s,t}$'s by fleet and region are derived from a multinomial sample of the total annual age–composition with relative probability given by $C^{f,r}_t$ and sample size 100. As such, the ageing samples are (on average) proportionally allocated by fleet and region with respect to the annual catch.

A1.3 Allocation of TAC by fleet and region during projection period

For each year of the projection period, the catches for each fleet and region are calculated using a multinomial allocation of the total TAC for that year. The expected proportions of the catch for each fleet/region are:

$$p_{C,t}^{f,r}=\frac{C_t^{f,r}}{\sum_{f'}\sum_{r'}C_t^{f',r'}}=\frac{\tilde{p}_{C,t}^{f,r}}{\sum_{f'}\sum_{r'}\tilde{p}_{C,t}^{f',r'}}$$

$$\tilde{p}_{C,t}^{f,r}=\frac{\xi^{f,r}+\psi^{f,r}\left(B_t^{f,r}\right)^{\varsigma^{f,r}}}{\sum_{f'}\sum_{r'}\left[\xi^{f',r'}+\psi^{f',r'}\left(B_t^{f',r'}\right)^{\varsigma^{f',r'}}\right]} \qquad (30)$$

where $\xi^{f,r}$, $\psi^{f,r}$, and $\varsigma^{f,r}$ are the parameters of the relationship between biomass distribution and catch allocation, and

$B_t^{f,r}$ is the vulnerable biomass in region *r* for fleet *f* in the middle of year *t*:

$$B_t^{f,r}=I^{f,r}\sum_{s}\sum_{L}w_{L,s}S_L^f\sum_{a}\Phi_{L,s,a}N_{s,a,t}^{r}e^{-0.5M} \qquad (31)$$

The values of the parameters $\xi^{f,r}$, $\psi^{f,r}$, and $\varsigma^{f,r}$ are determined by fitting the multinomial model in equation (30) to the (known) historical catch proportions by fleet and region.

A1.4 References

Fay, G., A. E. Punt, and A. D. M. Smith. 2009. Operating model specifications. Pages 125-133 in Wayte, S.E. (ed.) 2009. Evaluation of New Harvest Strategies for SESSF Species. CSIRO Marine and Atmospheric Research, Hobart and Australian Fisheries Management Authority, Canberra. 137 p.