TWO CLASSES OF MULTI-MODEL CANDIDATE MANAGEMENT PROCEDURES FOR ATLANTIC BLUEFIN TUNA

Sean P. Cox[1,2,a], Samuel D. N. Johnson[1,2,b], Steven P. Rossi[1,2,c]

SUMMARY

Two classes of multi-model candidate management procedures for Atlantic bluefin tuna were developed and tested. Procedures were based on spawning biomass estimation methods scaled to five operating models selected via cluster analysis from the reference OM grid. For the empirical class, OM catchability and a constant stock mixing distribution were used to estimate area biomass from the larval indices. For model-based MPs, five delay difference assessment models were scaled to each of the five operating models, matching stock recruit steepness and biomass for the recent historical period from 1965 - 2016. At each time step, estimates of current (empirical) or projected (model-based) biomass were generated from approved management indices and used in harvest control rules to generate area-specific TACs, and the five TACs were averaged to produce harvest advice for the East and West area. Multi-model CMPs scaled to the five OMs performed well across the full range of 96 operating models with minimal tuning; however, some CMPs were overly conservative and would benefit from refinement to reduce overfishing when stock biomass is overestimated.

RÉSUMÉ

Deux catégories de procédures de gestion multi-modèles potentielles pour le thon rouge de l'Atlantique ont été développées et testées. Les procédures étaient basées sur des méthodes d'estimation de la biomasse du stock reproducteur échelonnées à cinq modèles opérationnels sélectionnés par analyse de grappes à partir de la grille de référence du modèle opérationnel (OM). Pour la classe empirique, la capturabilité du OM et une distribution constante du mélange des stocks ont été utilisées pour estimer la biomasse de la zone à partir des indices larvaires. Pour les procédures de gestion basées sur des modèles, cinq modèles d'évaluation à différences retardées ont été échelonnés à chacun des cinq modèles opérationnels, en faisant correspondre la pente à l'origine de la relation stock-recrutement (steepness) et la biomasse pour la période historique récente de 1965 à 2016. À chaque étape, des estimations de la biomasse actuelle (empirique) ou projetée (basée sur le modèle) ont été générées à partir d'indices de gestion approuvés et utilisées dans les règles de contrôle de l'exploitation pour générer des TAC spécifiques à la zone. La moyenne des cinq TAC a été calculée pour produire un avis de capture pour la zone Est et Ouest. Les CMP multi-modèles échelonnées aux cinq OM ont donné de bons résultats dans toute la gamme des 96 modèles opérationnels avec un calibrage minimal ; toutefois, certaines CMP étaient trop conservatrices et gagneraient à être affinés pour réduire la surpêche lorsque la biomasse du stock est surestimée.

RESUMEN

Se desarrollaron y probaron dos clases de procedimientos de ordenación candidatos con varios modelos para el atún rojo del Atlántico. Los procedimientos se basaban en los métodos de estimación de la biomasa reproductora escalada a cinco modelos operativos seleccionados de la matriz de referencia de OM mediante un análisis de conglomerados. Para la clase empírica, se utilizaron la capturabilidad de los OM y una distribución constante de la mezcla del stock para estimar la biomasa del área a partir de los índices larvarios. Para los MP basados en modelos, cinco modelos de evaluación de diferencia retardada se escalaron a cada uno de los cinco modelos operativos, haciendo corresponder la inclinación stock recluta y la biomasa para el reciente periodo histórico desde 1965 a 2016. En cada fase temporal, las estimaciones de la

[1] School of Resource and Environmental Management, Simon Fraser University, 8888 University Drive, Burnaby, BC, Canada
[2] Landmark Fisheries Research, 301-220 Brew Street, Port Moody, BC, Canada (a: spcox@sfu.ca; b: samuelj@sfu.ca; c: spr1@sfu.ca)

biomasa actual (empírica) o proyectada (basada en el modelo) se generaron a partir de índices de ordenación aprobados y utilizados en normas de control de la captura para generar TAC específicos de áreas, y los cinco TAC se promediaron para formular un asesoramiento sobre captura para las zonas del este y del oeste. Los CMP de varios modelos escalados a los cinco OM funcionaron bien en toda la gama de 96 modelos operativos con una calibración mínima, sin embargo, algunos CMP eran excesivamente conservadores y requerirían un refinamiento para reducir la sobrepesca cuando la biomasa del stock está sobrestimada.

KEYWORDS

Atlantic bluefin tuna, Multi-model CMPs, Oms

1. Introduction

Two classes of multi-model CMPs were developed and tested in the Atlantic Bluefin tuna MSE framework. The first class is based on an empirical index-based method of estimating spawning stock biomass, and the second uses a multi-stock state-space delay difference stock assessment model. Both classes are made up of five biomass estimation and harvest control rules that are scaled to one of five reference grid operating models. Catch advice from each sub-procedure is then combined to produce area-specific TACs.

2. Methods

Our candidate management procedures combined estimates of biomass and associated catch advice from sub-procedures tuned to a subset of the reference grid of multi-stock operating models. This multi-model formulation allowed us to capture the biological uncertainty represented by the reference grid of operating models.

We defined two classes of MP: empirical and model-based. The empirical MPs scale larval indices by OM estimates of catchability, while the model-based MPs fit state space delay difference assessment model to management indices and pre-2015 spawning biomass.

2.1 Scaling subset of operating models

We scaled our estimation procedures on a subset of OMs which were chosen to be representative of the main OM uncertainty factors (1: steepness, 2: maturity/M, 3: stock mixing, 4: mean SSB, 5: length composition weight) and the SSB range in the OMs. To choose representative OMs, we ran a clustering algorithm on the SSB time-series to identify *k* cluster centers or "medoids", then checked the factor levels associated with each medoid. We considered a set of OMs as adequately representative if (i) SSB trajectories in the subset spanned approximately the same range as the SSB in the OMs, and (ii) each uncertainty factor level was represented at least once. OMs with changing recruitment steepness in the projection years were excluded from the clustering analysis, leaving a total of 64 OMs to be clustered.

We used the Time-series Anytime Density Peaks (TADPole) algorithm to cluster SSB time-series (Begum et al. 2016). To simultaneously cluster the SSB in each area, we first created a single SSB "time-series" for each OM by adding the west SSB time-series onto the end of the east SSB time-series. We then ran the clustering algorithm on the resulting 32 combined time-series for *k*=4 to *k*=8. The weight associated with each cluster was calculated as the proportion of individual time-series in the cluster.

2.2 Empirical estimation procedure

The empirical estimation procedure used a simple moving average of the two spawning stock larvel survey indices. For each time step t, and stock $s \in \{GOM, MED\}$ the current average larvel index is calculated as

$$\bar{I}_{s,t} = \frac{\sum_{t\prime=t-3}^{t} I_{s,t\prime}}{4},$$

where $I_{s,t}$ is the larval index for stock s at time t. We then scaled the smoothed survey indices to multiple biomass estimates, one for every operating model in the tuning subset, by using the appropriate larval survey catchability parameter, e.g

$$\hat{B}_{s,t,j} = \frac{\bar{I}_{s,t}}{q_{s,j}},$$

where $q_{s,j}$ is the stock larval index catchability parameter from the tuning operating model $j \in \{37, 14, 53, 31, 89\}$.

Spawning stock biomass was then translated to area biomass using an assumed constant distribution of biomass for stock mixing

$$P = (p_{s,a}) = \begin{pmatrix} .898 & .102 \\ .1 & .9 \end{pmatrix},$$

found by averaging the proportion of stock biomass in each area over the historical period in the tuning set of OMs. The rows of matrix P are indexed by spawning stocks $\{MED, GOM\}$, and the columns are indexed by management area $\{E, W\}$. This generates an area specific biomass

$$B_a = \sum_s p_{s,a} B_s,$$

where $s \in \{MED, GOM\}$, $a \in \{E, W\}$.

2.3 Delay difference model

For the model-based class of procedures, the simple index scaling method of the empirical procedure is replaced by a state-space multi-stock Deriso-Schnute delay difference assessment model (Deriso 1980; Schnute 1985).

The two spawning stocks are assumed to be distinct spawning stocks for the larval surveys, but are mixed using the same mixing distribution *P* as the empirical CMP for the other management indices and the observed catches.

2.3.1 Equilibria

Because there is no mixing of spawners, each stock has the simple Delay Difference equilibrium states, which can be expressed as a function of long-term fishing mortality $f = F_s$. The equations for equilibria in **Table 1** were used to initialise the model at a fished equilibrium in 1965, as well as estimate biological reference points for use in the harvest control rules.

2.3.2 Biomass time series reconstruction

Each assessment model was initialised in a fished state in 1965. To do so, we estimated an initial fishing mortality rate $F(s, 0)$, and assumed that the stock was at the fished equilibrium defined in **Table 1.** Other free parameters were unfished biomass $B_{s,0}$, catchability q_g and observation error uncertainty τ_g for biomass and abundance indices, and recruitment process error deviations $\omega_{s,t}$ for $t \in 1966, \ldots, 2014$ (**Table 2**). Stock recruit steepness h_s and natural mortality M_s were fixed to the estimated values from the associated operating model (Note: west *h* shifts from 0.6 to 0.9 in 1975 in some OMs. In these scenarios, we set west *h* to 0.9).

The proportion $p_{s,a}$ of stock $s \in \{MED, GOM\}$ in area $a \in \{E, W\}$ was assumed to be constant, allowing the fluctuations in stock biomass to account for the mixing dynamics. Both stocks were also assumed to be homogeneously mixed in each area, so that total stock specific catch was the sum of the area specific catches, split according to the proportion to the stock biomass in each area.

Instantaneous fishing mortality was solved via a numerical Newton-Rhapson method within the delay difference model for all time steps $t \geq 1$, which were used in the survival calculation to progress numbers and biomass to the next time step.

Recruitment process errors were modeled as a simple random walk, rather than independent deviations from the stock recruit curve, chosen because it was more able to capture dynamic recruitment regimes. Furthermore, random walks would also make projected recruitments, which are an important component of projected biomass in a delay difference formulation, less likely to be average, and more like the estimated recruitment in the last year of the assessment period.

Because this is a model-based MP and the time-series of approved management indices are rather short, two additional indices were included. These indices were the M3 model yearly spawning stock biomass estimates for the East and West stocks. These biomass time series were assumed to be observed with catchability $q = 1$ and an observation error CV of 5%, allowing the delay difference AMs to estimate catchability for the approved indices and scale them appropriately to the spawning stock biomass from the associated operating model. This scaling was required for extending the fit to the approved indices in the projections, as well as making sensible estimates of biological parameters despite the complexity mismatch between the operating model and the assessment models.

Finally, a log-normal prior distribution was applied to unfished biomass $B_{0,s}$ to prevent the unfished biomass from being estimated too close to the initial biomass for the assessment period, and to avoid producing optimistic estimates of initial and current biomass depletion. For each AM, the prior mean was defined as the estimate from the associated operating model with a log standard deviation of 0.05.

2.3.3 Reference Points

Reference points for each stock were estimated from the equilibrium values in (**Table 1**). The optimal fishing mortality rate F_{MSY} was found by numerically solving for the stationary point of the yield curve, and was in turn used to estimate the optimal equilibrium biomass B_{MSY} associated with that mortality rate.

Area based reference points for use in harvest control rules were defined as the stock-area biomass weighted averages of the stock-specific reference points, i.e.

$$B_{MSY,a} = \sum_s \frac{B_{s,a}}{B_a} B_{MSY,s},$$

$$F_{MSY,a} = \sum_s \frac{B_{s,a}}{B_a} F_{MSY,s},$$

where $B_{s,a} = p_{s,a} B_s$.

2.4 Harvest control rules

For each of the five sub-procedures, a ramped harvest control rule was defined with upper and lower control points and a maximum harvest rate (**Figure 1**). The control points and maximum harvest rates for these rules were either based on the associated operating model's biological reference points and a precautionary TAC cap to limit removals in cases of bias in projected biomass or reference point estimates.

Our ramped harvest control rule was based on the rules used for albacore tuna. These rules require an upper control point (UCP), a lower control point (LCP) and a maximum target harvest rate U_{max}. The general form of the rules (ignoring AM, area, and stock indices) was

$$U_{targ} = \begin{cases} 0.1 \cdot U_{max} & B \le LCP \\ U_{max} * (0.1 + 0.9 \dfrac{B - LCP}{UCP - LCP}) & LCP \le B \le UCP \\ U_{max} & UCP \le B \end{cases}$$

For all HCRs, we set $LCP = .4UCP$, and tested two options for the upper control point: $UCP = B_{MSY}$ or $UCP = .4B_0$. We tested two options for the maximum target harvest rate: (i) setting U_{max}=F_{MSY}, or (ii) setting U_{max} equal to some multiple of *M*. For the model-based MPs, we tested $U_{max} = \frac{2}{3}M$. Initial trials indicated that harvest rates under $U_{max} = \frac{2}{3}M$ for the empirical MPs were too high, so we tested a range of more conservative values for U_{max} (see 2.6). Under the empirical management procedures, control points were taken from the tuning grid of operating models, while under the model-based MPs, control points were taken as the delay difference model equilibria.

We also tested a procedure wherein TACs were capped at MSY to provide precautionary harvest rates in the presence of potentially large biomass estimation errors and biases in reference point calculations, both of which are caused by the complexity mismatch between the AMs and the OMs. Stock- and area-specific MSY estimates were based on assessment model equilibria.

Both home-stock- and area-based harvest control rules were applied for East and West area TACs. For example, in the East area, HCRs were applied based on East (MED) spawning stock biomass compared to the East stock control points and harvest rates, and East area biomass compared to mixed stock control points and harvest rates. Mixed area based control points and harvest rates were averaged over the two stocks present in the area, weighted by the proportion of stock specific biomass in that area (e.g. see area-based reference points calcs in previous section), and the lower of the two TACs was chosen. From this, the West TAC would almost always be managed according to the Gulf of Mexico spawning stock harvest control rules, and the East TAC would almost always be managed according to the mixed East Area harvest control rule, as this includes the weaker stock.

2.5 Providing catch advice by area

To provide a single TAC at each time step, the five TACs advice from the sub-procedures were averaged with the AIC based weights for the model-based CMPs, and weights from the cluster analysis for the empirical CMPs. Averaged TACs were smoothed with respect to the previous management interval's TAC allowing a maximum increase of 20%, and a maximum decrease of 50%, compared to the previous management interval.

2.6 Simulation experiments and performance metrics

For model-based CMPs, procedures with and without TAC caps were tested, as well as CMPs where U_{max} was linked to either F_{MSY} or M (**Table 3**).

For the empirical CMPs, we varied our treatment of both U_{max} and the UCP (Table 4). For procedures that set U_{max} to M, we defined $U_{max} = M/\lambda$ and ran an grid search from $\lambda = 1$ to $\lambda = 4$ in increments of 0.5 to find a harvest level that produces median Br30 close to 1. We capped the TAC in all empirical procedures at MSY.

All MPs were evaluated over the 96 reference grid deterministic operating models, labelled OM_1d through OM_96d.

3 Results

3.1 Operating Model Cluster Analysis

Each of the cluster sets with $k \geq 5$ covered the historical OM SSB range reasonably well (**Figure 2**) and also included at least one level from each factor (Table 1). The differences between $k = 5$ and $k = 6$ were minimal, so in the interest of parsimony we selected $k = 5$ for CMP scaling. The OMs chosen under $k = 5$ and their cluster weights were OM 37 (0.33), OM 14 (0.25) , OM 53 (0.20), OM 31 (0.17) and OM 89 (0.05).

The five cluster centres were well separated across the full range of biomasses, and the clusters themselves were not too spread out. The TADpole algorithm uses a non-standard loss function when optimizing clusters, and did not produce a measure of within cluster and between cluster variance, which is more interpretable for understanding cluster performance. We approximated between cluster variance by calculating pair-wise root-mean-squared-error (rMSE) between cluster medoids, and within cluster variance by calculating the average of the pair-wise rMSEs for all OMs in a cluster. Average within cluster rMSE ranged within 42.91 kt – 62.34 kt, with a mean rMSE of 52.03 kt, and between cluster rMSE ranged within 111.95 kt – 556.87 kt, with a mean rMSE of 297.21 kt.

3.2 Fits of Delay Difference AMs to historical data

Fitting the delay difference assessment model to operating model biomass faithfully reproduced the historical operating model biomass series (**Figures 3 & 4**). In contrast, assessment model estimates of biological parameters and equilibria, and operating model estimates of the same quantities, were less consistently similar (**Figure 5**). While the log-normal prior on unfished biomass worked well to keep $B_{0,s}$ estimates close to the operating model values, the model equilibria were often biased. For example, in AMs 31 and 37, the DD model F_{MSY} values for the east spawning stock were around 0.30, in contrast with the operating model values, which were near 0.10.

Similarly, in the same assessment models, B_{MSY} values were much higher than the corresponding operating model values. These biases have implications for the performance of MPs that rely on the AM reference point estimates to set TACs, which we describe below.

3.3 Performance in projections

3.3.1 Delay Difference Model-based CMPs

CMP performance varied for the two stocks (**Figure 6**). Uncapped procedures generally produced much higher catches from the east stock (though median catch in projection years 21-30 was slightly higher under capped procedures), while procedures that linked U_{max} to F_{MSY} tended to produce higher catches from the west stock. In both areas, however, stocks were overharvested (Br30 < 1) and, in some cases, crashed, by the uncapped procedure that linked U_{max} to F_{MSY}. In contrast, CMPs that capped the TAC at MSY tended to be too conservative (i.e., most Br30 values were above 1), particularly for the west stock.

The uncapped CMP that linked U_{max} to *M* (DD-NoCapFM) had the best overall performance; Br30 values were close to 1 for the east stock for nearly all OMs, while median Br30 in the west was also close to 1. Additionally, neither stock crashed for any OM under DD-NoCapFM (lowest depletion ranged from 0.09-0.57 with median 0.22).

Recruitment appeared to be the main OM factor affecting MP performance (**Figure 7**). MPs generally performed well (i.e., Br30 close to 1) on OMs with low, constant steepness for both stocks (factor level 2). In contrast, MPs crashed most often under OMs in which steepness declines after 10 projection years (factor level 3). To further explore this result, we used a meta-modelling approach to estimate the effect of each factor level on Br30 values (**Figure 12**), fitting a generalized linear model with OM factors as the observations, and Br30 values as the response. Regression coefficients for recruitment factor level 3 were negative for all MPs, indicating that MPs performed relatively worse for OMs with shifting recruitment in the projections.

The poor performance of the CMPs that link U_{max} to F_{MSY} is caused partly by the complexity mismatch between the OM and AM, biasing the delay difference reference points, with F_{MSY} positively biased and B_{MSY} negatively biased. Because of this, although some AM estimates of current and projected biomasses were often very close to the OM biomass, the biomass relative to B_{MSY} was above the upper control point, so the positively biased maximum target harvest rate was applied, leading to overfishing.

To further explore the influence of assessment model bias on CMP performance, we calculated the relative error in biomass estimates between the OM and the top AIC-ranked assessment model at each assessment interval (**Figure 9**). Significant errors in biomass estimates exist, even for OMs which were used to tune the assessment models – only OM 31 in the west area was close to unbiased on average. Larger bias occurs because AIC does not always select the "correct" AM (i.e., when running a CMP on OM 89, the AM that was tuned to OM 37 often has the lower AIC score than the AM that was tuned to OM 89). As expected, there was a negative relationship between relative error in biomass estimates and Br30, with CMPs that under-estimate spawning biomass producing higher Br30 values, on average (**Figure 10**).

3.3.2 Empirical CMPs

The empirical CMPs were prone to overharvest both stocks; only when a very modest harvest rate was used (*F*=*M*/4) did CMPs avoid crashing the stocks (**Figure 11**). However, the empirical CMPs tended to produce median Br30 values that were close to or above 1. Catches in the east often hit the TAC cap (particularly in the first 20 projection years) while catches in the west did not. We initially tested uncapped empirical procedures, however, these CMPs were too severe in the east and made little difference in the west.

As with the model-based procedures, recruitment appeared to be the most important uncertainty axis (**Figures 12 & 13**). Unlike the model-based procedures, however, the OMs associated with the lowest empirical Br30 values were those with recruitment level 2.

4. Discussion

In this paper, we tested empirical and model-based candidate management procedures (CMPs) for Atlantic Bluefin Tuna. These CMPs were based on multi-model inference, where TACs from 5 sub-procedures, each scaled to a specific reference grid OM, were combined as a weighted average to produce TAC advice, with weights either taken from an initial cluster analysis of historic operating model biomasses (empirical) or based on Akaike's Information Criterion (model-based).

Multi-model CMPs are a viable option for managing Atlantic Bluefin Tuna fisheries across a wide range of operating model hypotheses. Clustering OMs by biomass to choose a small representative subset for sub-procedure scaling produced CMPs with reasonable performance according to the Br30 metric. Indeed, with only a small amount of tuning, the median Br30 values across the full operating model grid were at or above Br30 = 1.0 for all capped model-based and empirical CMPs. All uncapped CMPs were, as expected, more aggressive, which crashed stocks more often.

There was no evidence that scaling model-based sub-procedures by fitting to historical OM biomass bestowed any kind of omniscience to the model-based CMPs, nor did it cause the CMPs to "overfit" to the OM grid. The AMs only fit to biomass from the scaling set of OMs in the historical period, as a method of correctly scaling the catchability parameters for the set of management indices, and any updated operating model information was not fed to the AMs during the projection period. Further, the performance of model-based CMPs on the set of OMs outside of the selected five scaling OMs constitutes a form of hold-out analysis similar to a cross validation, and while CMP performance was acceptable on the full grid, it was still quite variable over combinations of OM grid factors that were not included in the sub-procedure scaling. Further, with the exception of the GOM spawning stock under OM 31, the weighted estimates of spawning stock biomass combined from each sub-procedure at each time step were on average biased on the scaling set of OMs, with mean relative errors around 50% in absolute value, showing that omniscience was not bestowed even for a "self-test" of the CMPs on the scaling OM subset.

There may be some limitations associated with the method used to select the scaling set of OMs. Here, each OM was represented by a concatenated East/West historical biomass in the clustering algorithm, and the number of clusters was chosen to include each OM factor level at least once. The concatenation may have caused problems, where the higher variance in the East required a larger number of clusters than the more concentrated West biomasses. Further, clustering by biomass is only one dimension in which the operating models vary, and another choice may yield a different set. However, based on preliminary results not presented here where we selected particular OMs from the grid to ensure some of the more extreme scenarios were considered, it appears that conservation outcomes (as measured by the Br30 metric) are not too sensitive to scaling set choice, as long as the scaling set of operating models is in some sense representative (either of biomass history or over the OM grid factor levels).

The range of CMP performance across the operating model grid suggests that further development tuning and CMP refinement is required for both empirical and model-based CMPs. Although median performance of most presented CMPs was within an acceptable range, there were some cases where the CMPs were overly conservative (median Br30 > 1), indicating that further tuning is needed to bring the median values closer to Br30 = 1; however, some of those CMPs have a large spread of Br30 values across the OM grid, so development tuning to shift the median Br30 value down would likely induce stock crashes under some OMs, indicating that refinement of the CMP is required to reduce the range of Br30 values. Similarly, some CMPs, like emp_msyCap, had median performance close Br30 = 1, but the East stock crashed under some OMs, indicating that CMP refinement is necessary. Future CMP refinement may replace the TAC weighted average with a trend-based choice of the TACs suggested by each sub-procedure, where a recent negative trend in biomass causes the TAC to be chosen from the lower end of the set of TACs, and vice-versa.

References

N. Begum, L. Ulanova, H. A. Dau, J. Wang, and E. Keogh, 2016, “A General Framework for Density Based Time Series Clustering Exploiting a Novel Admissible Pruning Strategy,” arXiv:1612.00637. https://arxiv.org/abs/1612.00637.

Deriso, R.B., 1980. Harvesting Strategies and Parameter Estimation for an Age-Structured Model. Can. J. Fish. Aquat. Sci. 37, 268–282. doi:10.1139/f80-034

Schnute, J., 1985. A General Theory for Analysis of Catch and Effort Data. Can. J. Fish. Aquat. Sci. 42, 414–429.

Spall J.C., 2003. Introduction to Stochastic Search and Optimization. Wiley, New York.

Table 1. Unfished and fished equilibrium quantities for the delay-difference population dynamics.

Description	***Equation***
Survivorship	$S^{(f)} = e^{-M-f}$
Average Weight	$\overline{w}^{(f)} = \frac{S^{(f)}\alpha + w^{(k)}(1 - S^{(f)})}{1 - \rho S^{(f)}}$
Unfished Numbers	$N = B_0/\overline{w}^{(f=0)}$
Unfished Recruitment	$R_0 = (1 - S^{(f=0)})N_0$
Stock-Recruit	$a = \frac{4hR_0}{B_0(1-h)}$, $b = \frac{5h-1}{B_0(1-h)}$
Biomass	$B^{(f)} = \frac{S^{(f)}(\alpha + \rho\overline{w}^{(f)}) + \overline{w}^{(f)}(aw^{(k)} - 1)}{b(\overline{w}^{(f)} - \rho S^{(f)}\overline{w}^{(f)} - \alpha S^{(f)})}$
Recruitment	$R^{(f)} = \frac{aB^{(f)}}{1 + bB^{(f)}}$
Yield	$Y^{(f)} = \frac{f}{M + f}(1 - e^{-M-f})B^{(f)}$

Table 2. Process and observation model components of the delay difference stock assessment model used in BC Sablefish management procedures. Initialisation values for biomass, numbers, and recruitment are equilibrium unfished values from Table 2.

No.	***Equation***
A2.1	$\theta_s = \{\log F_{s,0}, \log B_{s,0}, q_g, \tau_g, \vec{\omega}_{s,t}\}.$
A2.2	$B_{s,1} = B_s^{(f=F_{s,0})}$
A2.3	$N_{s,1} = N_s^{(f=F_{s,0})}$
A2.4	$R_{s,1} = R_s^{(f=F_{s,0})}$
A2.5	$B_{s,a,t} = p_{s,a} \cdot B_{s,t}$
A2.6	$B_{a,t} = \sum_s B_{s,a,t}$
A2.7	$C_{s,t} = \sum_a \frac{B_{s,a,t}}{B_{a,t}} C_{a,t}$
A2.8	$R_{s,t} = \begin{cases} R_{s,1} e^{\omega_{s,t-k} - \sigma^2/2} & t \le k \\ \frac{aB_{s,t-k}}{1 + bB_{s,t-k}} e^{\omega_{s,t-k} - \sigma^2/2} & t > k \end{cases}$
A2.9	$Z_{s,t-1} = M_s + F_{s,t-1}$
A2.10	$N_{s,t} = e^{-Z_{s,t-1}} N_{s,t-1} + R_{s,t}$
A2.11	$B_{s,t} = e^{-Z_{s,t-1}}(\alpha N_{s,t-1} + \rho B_{s,t-1}) + w_s^{(k)} R_{s,t}$
A2.12	$\hat{I}_{g,t} = q_g \cdot B_{s,t}$
A2.13	$\hat{I}_{g,t} = q_g \cdot B_{a,t}$
A2.14	$\log I_{g,t} \sim N(\log \hat{I}_{g,t}, \tau_g)$
A2.15	$\omega_{s,t} \sim N(0, \sigma)$

Table 3. List of model-based (delay-difference) CMPs tested for Atlantic bluefin tuna.

Name	***TAC Cap (East; West)***	$\boldsymbol{U_{max}}$ ***(East; West)***
DD-NoCap	None; None	$F_{MSY}; F_{MSY}$
DD-MsyCap	MSY; MSY	$F_{MSY}; F_{MSY}$
DD-NoCapFM	None; None	$M; \frac{2}{3}M$
DD-MsyCapFM	MSY; MSY	$M; \frac{2}{3}M$
DD-MixCapFM	None; MSY	$M; \frac{2}{3}M$

Table 4. List of empirical CMPs tested for Atlantic Bluefin Tuna.

Name	$\boldsymbol{U_{max}}$	***UCP***
Emp-MsyCap	F_{MSY}	B_{MSY}
Emp-MsyCapB0	F_{MSY}	$0.4B_0$
Emp-MsyCapFM	M/λ	B_{MSY}
Emp-MsyCapFMB0	M/λ	$0.4B_0$

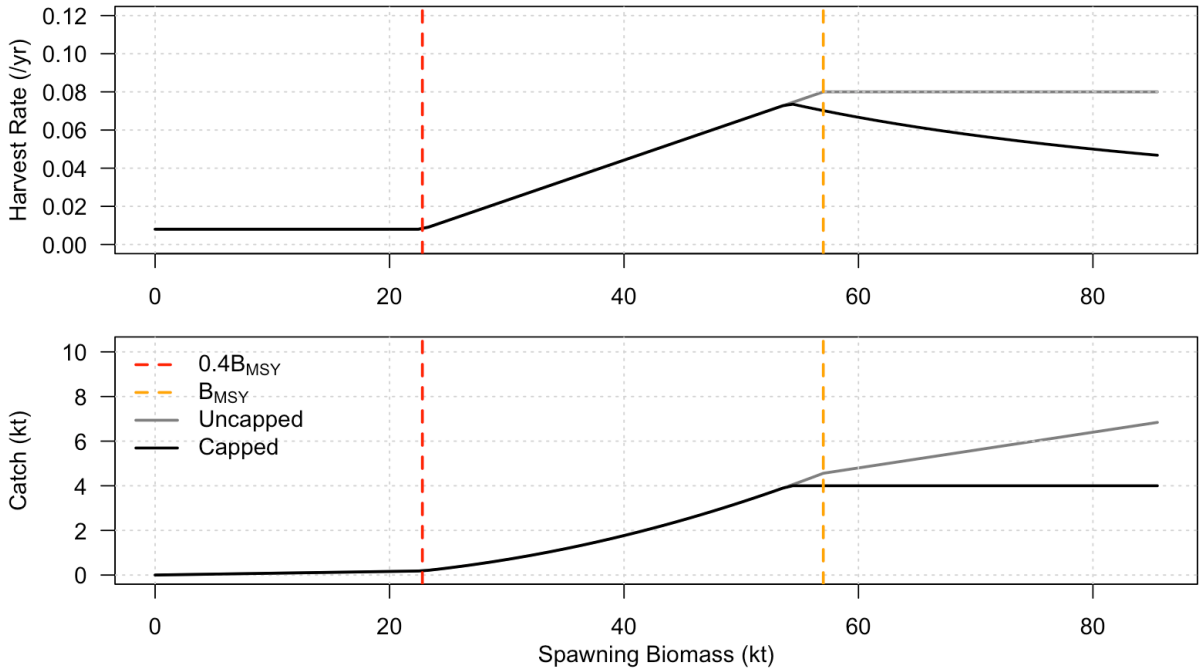

Figure 1. An example harvest control rule with maximum harvest rate $U_{\max}$ = 0.08, an upper control point of B_{MSY}= 57, and a TAC cap at 4 kt.

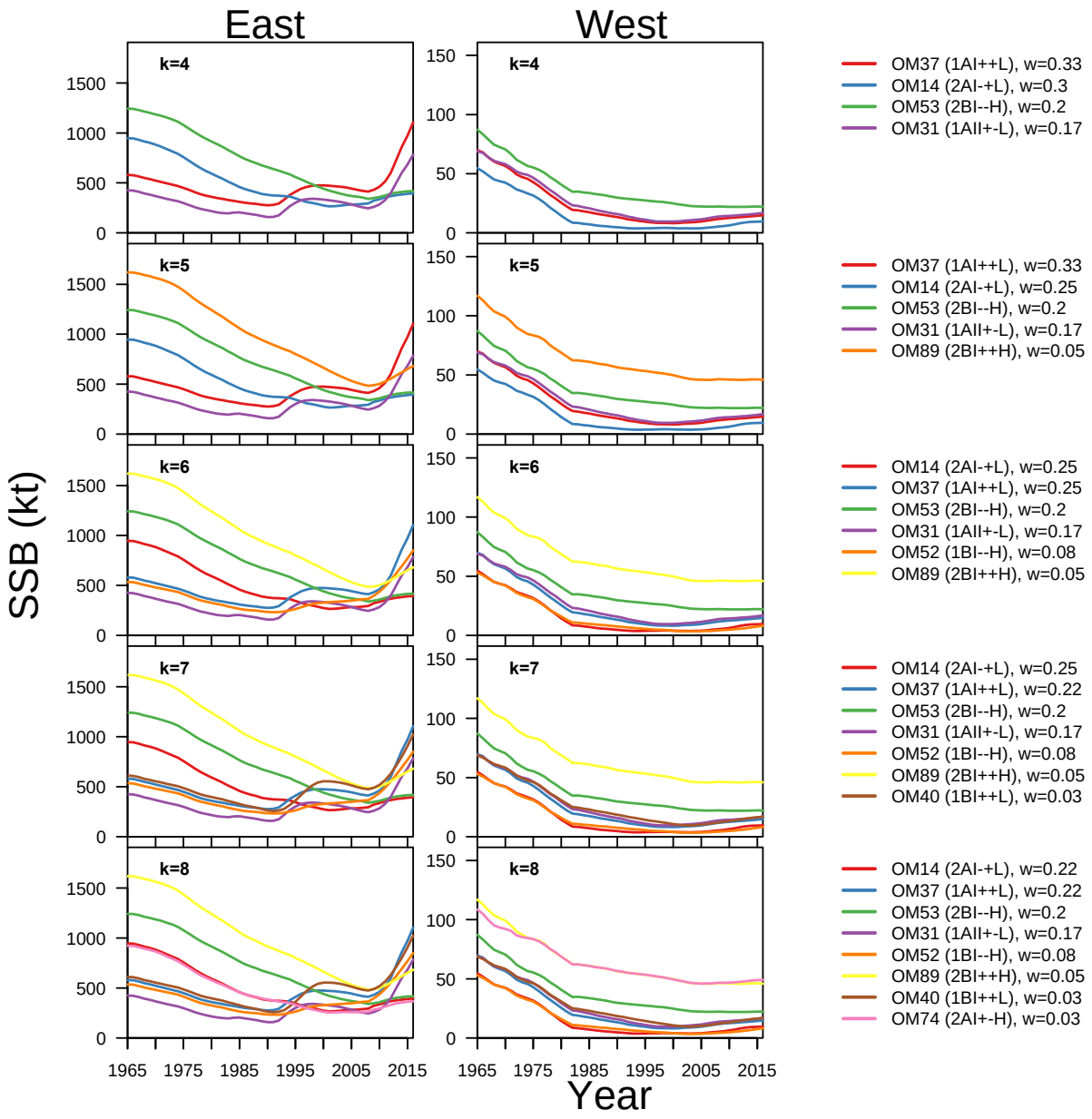

Figure 2. Spawning stock biomasses for the reduced set of 32 operating models for both the Mediterranean stock (East) and Gulf of Mexico stock (West). Rows represent different clustering analyses, each with a different number of clusters (k). Coloured lines represent cluster centers, each of which corresponds to a different OM from the reference grid.

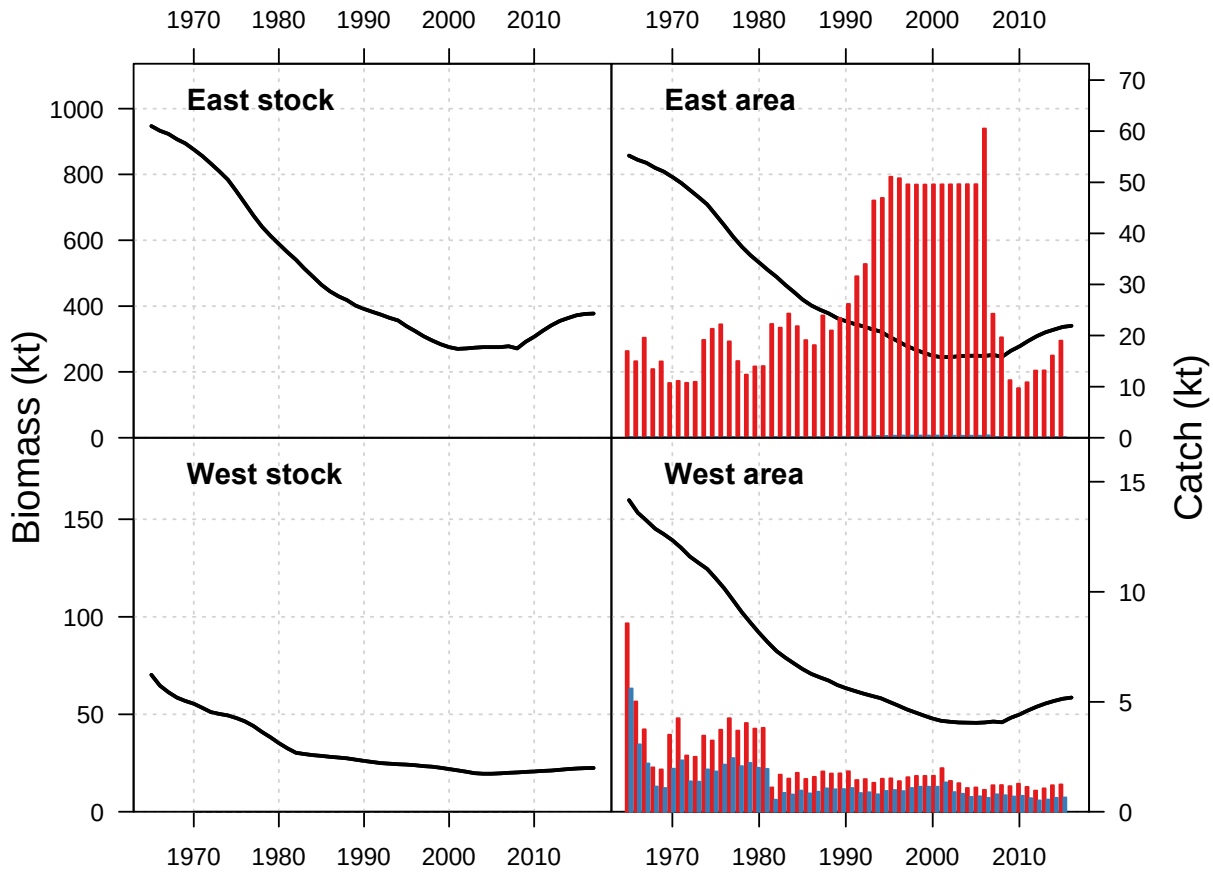

Figure 3. Spawning stock biomass estimates by stock (left hand column) and area (right hand column) from the delay difference stock assessment model fit to OM 50. Total area catch is shown split into East stock (red bars) and West stock (blue bars). Note that the catch scale is exaggerated with respect to the biomass, so that West stock catch is visible in the East area.

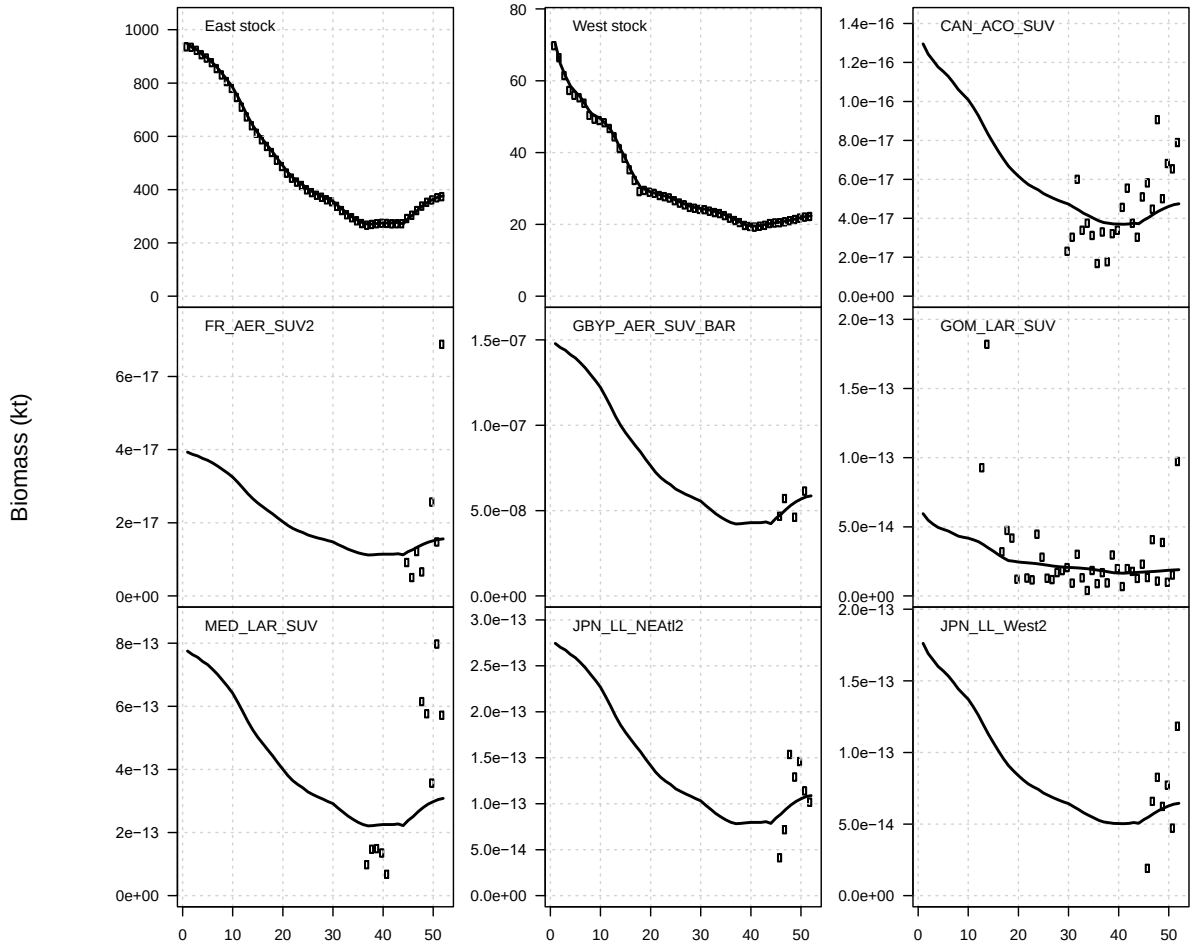

Figure 4. Fits of the delay difference stock assessment model to stock and area management indices, with the associated catchability estimates. Data are shown as circles, while the lines indicate the model biomass scaled by catchability. East Stock and West Stock indices are the spawning stock biomass estimates from operating model 50.

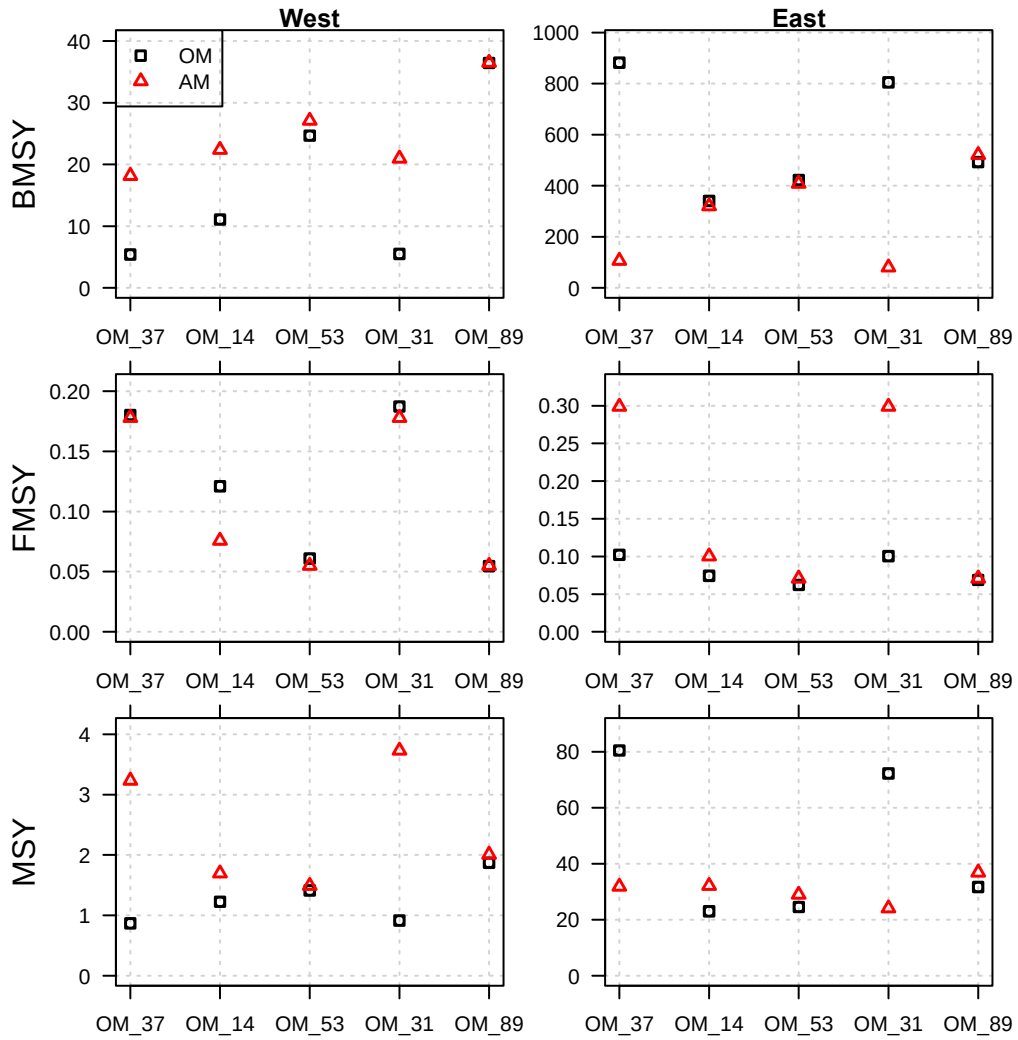

Figure 5. Comparison of B_{MSY} (top row), F_{MSY} (middle row) and MSY (bottom row) values between the operating models (black squares) and delay-difference assessment models (red triangles).

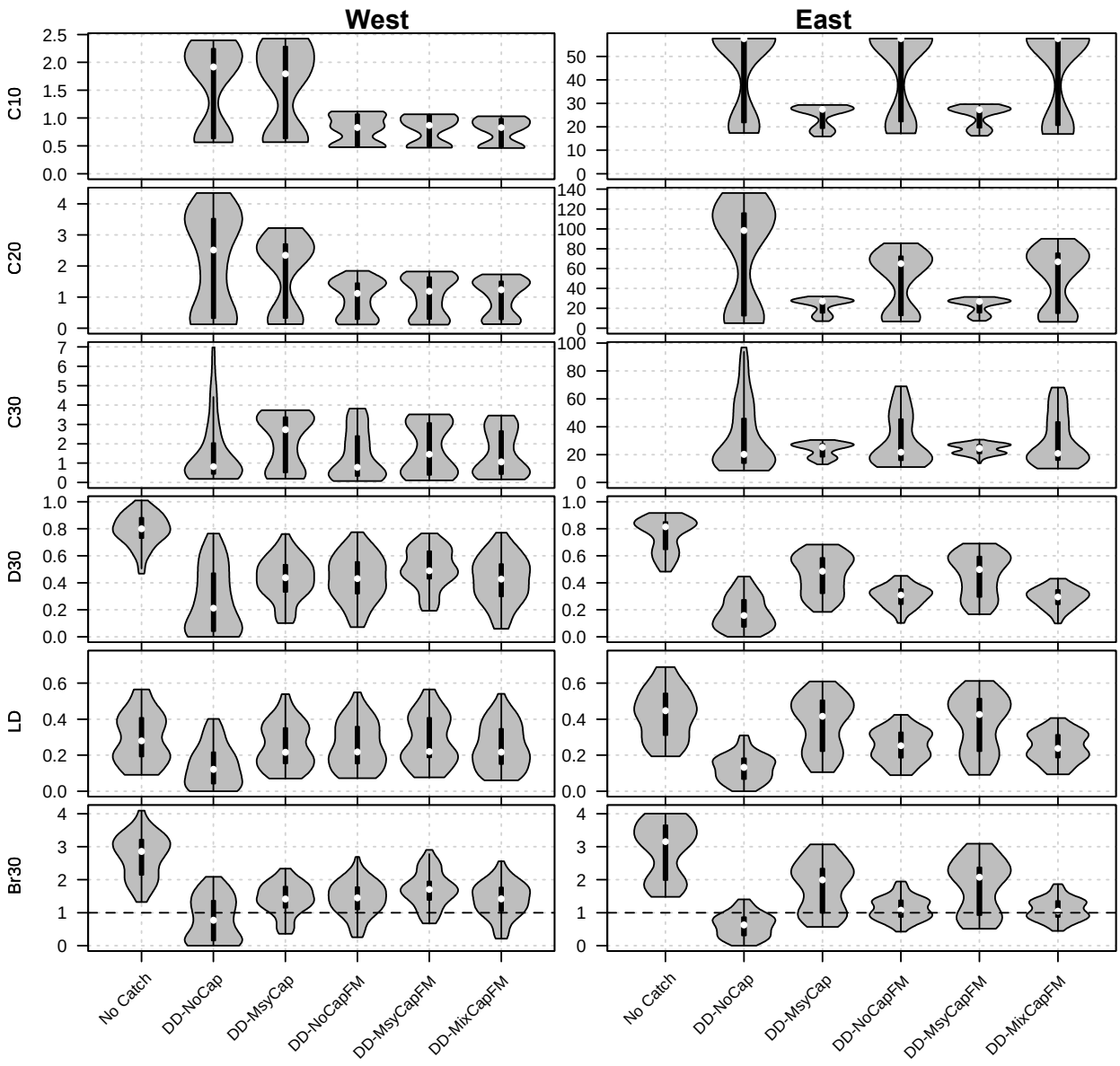

Figure 6. Violin plots of MSE performance metrics for model-based CMPs. The thin line, thick line and white circle within each "violin" represents a boxplot of values across OMs, while either side of boxplot shows a rotated kernel density plot of the distribution of values.

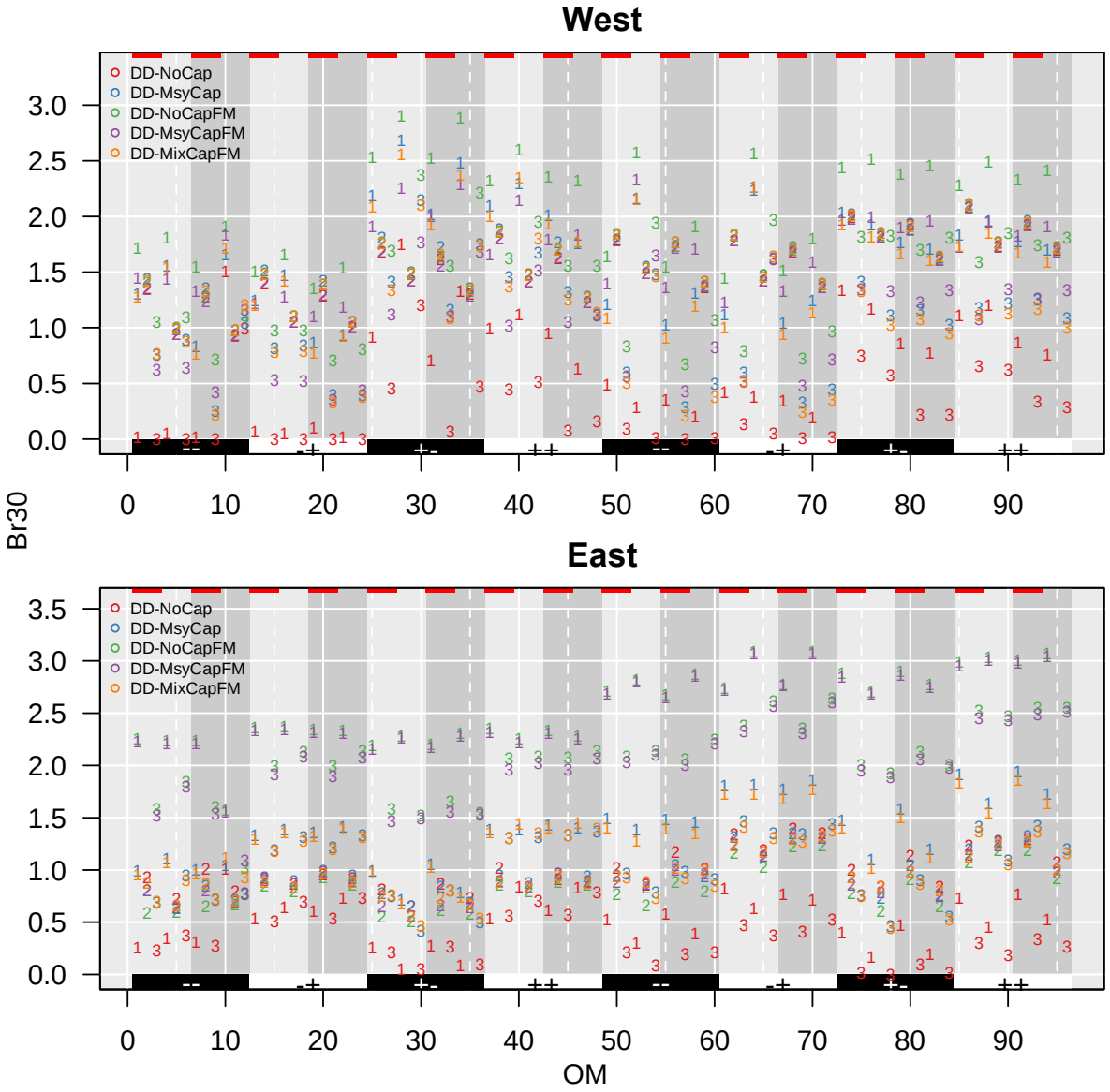

Figure 7. Estimated Br30 values for model-based MPs. Numerical values indicate the steepness factor for each OM (1=high, 2=low, 3=shifting). Younger spawning/higher M scenarios are indicated by the red line at the top of each plot. Migration factor levels are indicated by the background shading (light grey indicates low stock migration, dark grey indicates high). SSB factor levels are indicated by the black and white symbols on the bottom of each plot (-- indicates low West/low East; -+ indicates low/high; +- indicates high/low; ++ indicates high/high). The first 48 OMs have low length composition weighting while the latter 48 have high length composition weighting.

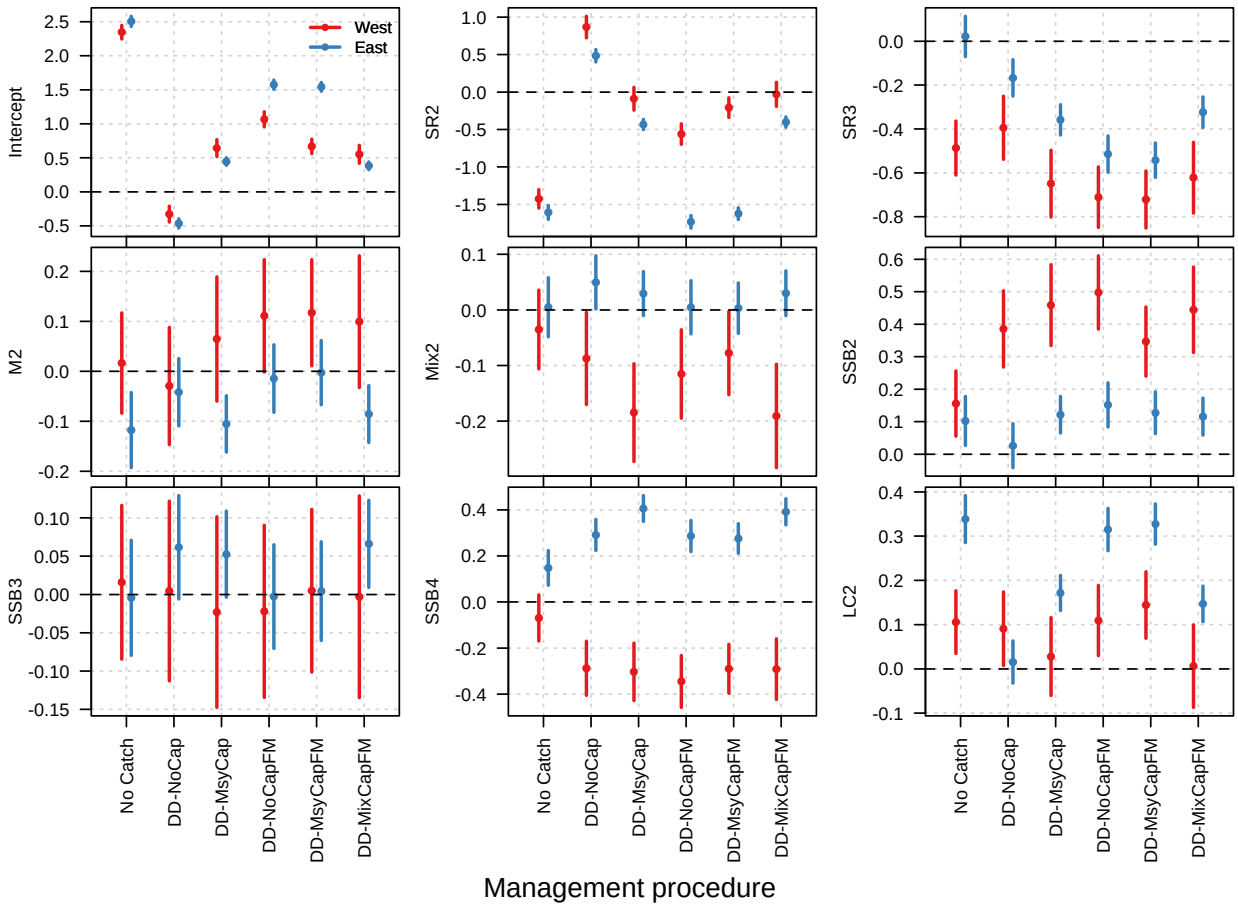

Figure 8. Coefficients from regression of Br30 on OM factor levels for model-based CMPs.

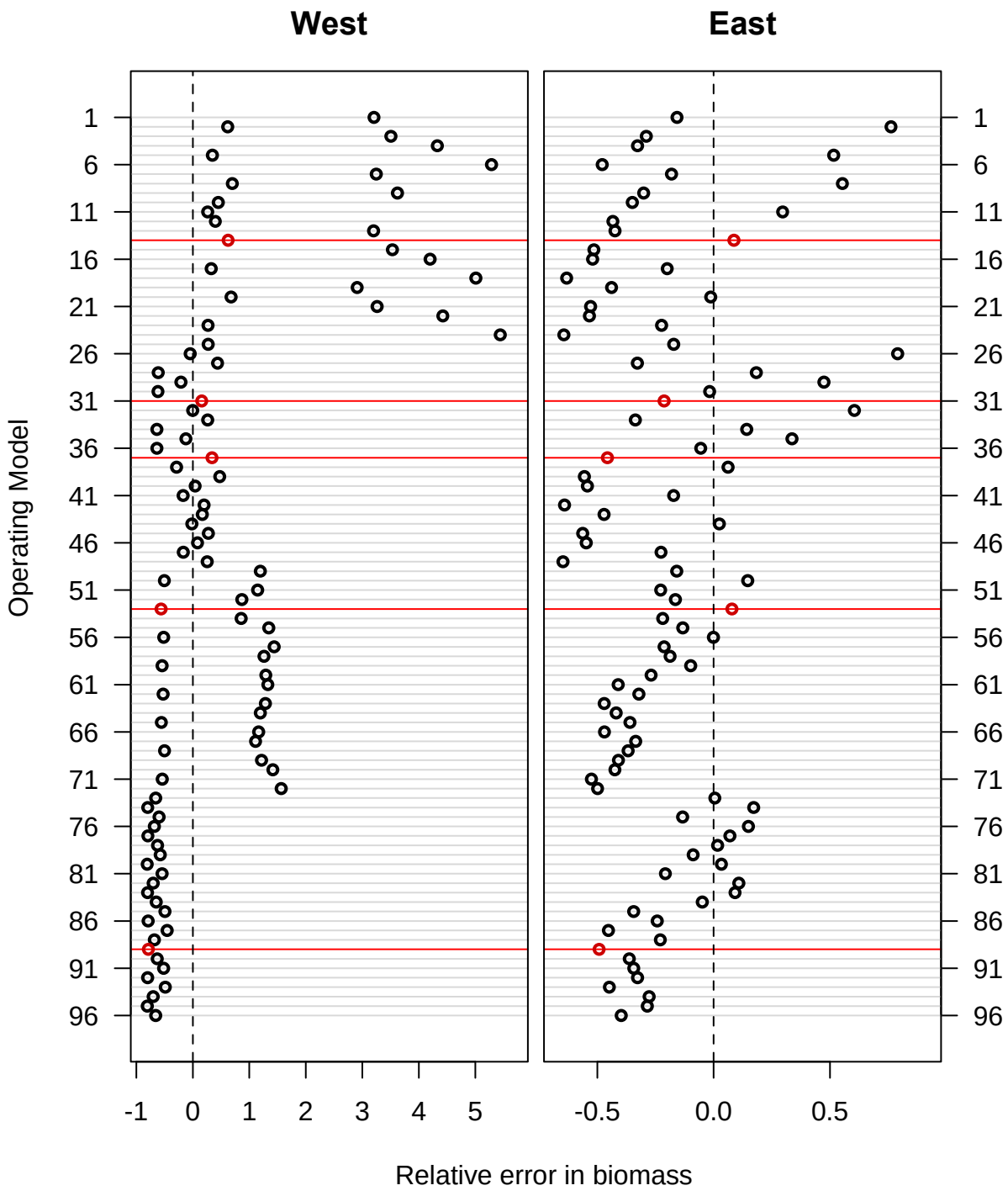

Figure 9. Average bias is biomass estimates from the top AIC-ranked model in each assessment interval for the MP_Mixed procedure. OMs used to condition the MP are coloured red.

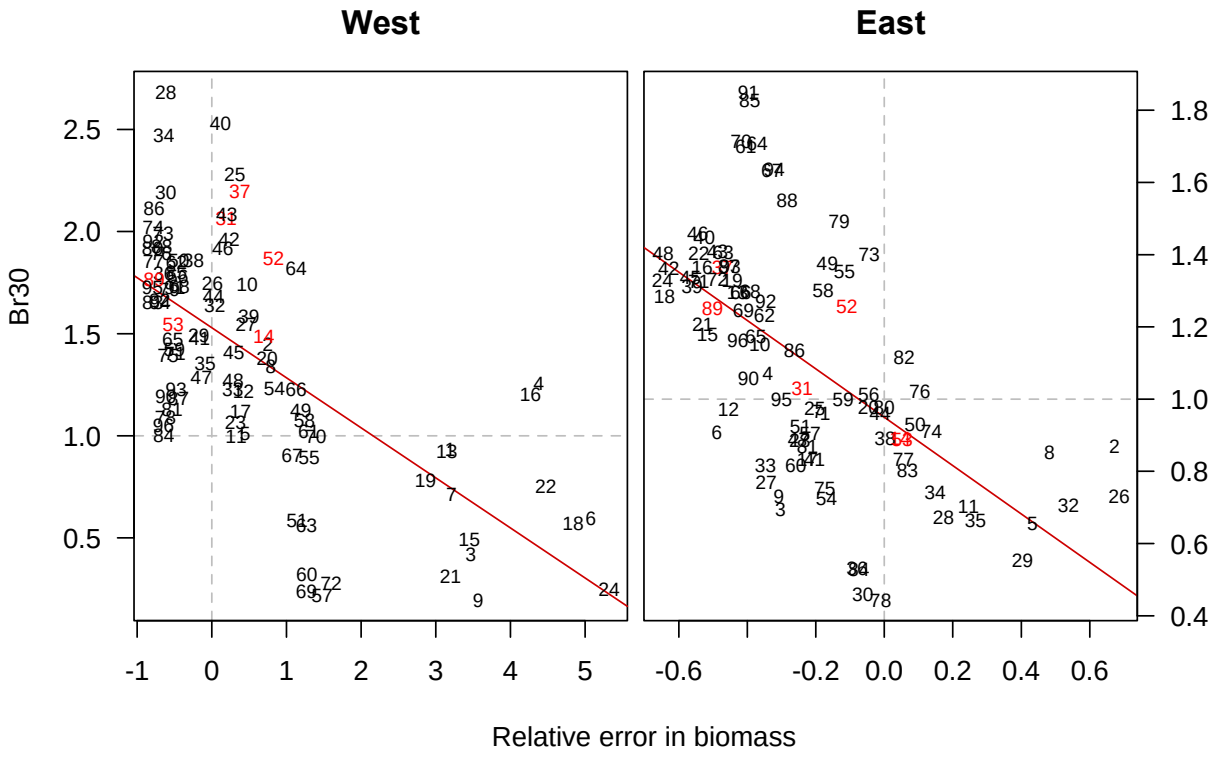

Figure 10. Relationship between Br30 and the average bias is biomass estimates from the top AIC-ranked model in each assessment interval for the MP_Mixed procedure for each of the 96 OMs. OMs used to condition the MP are coloured red.

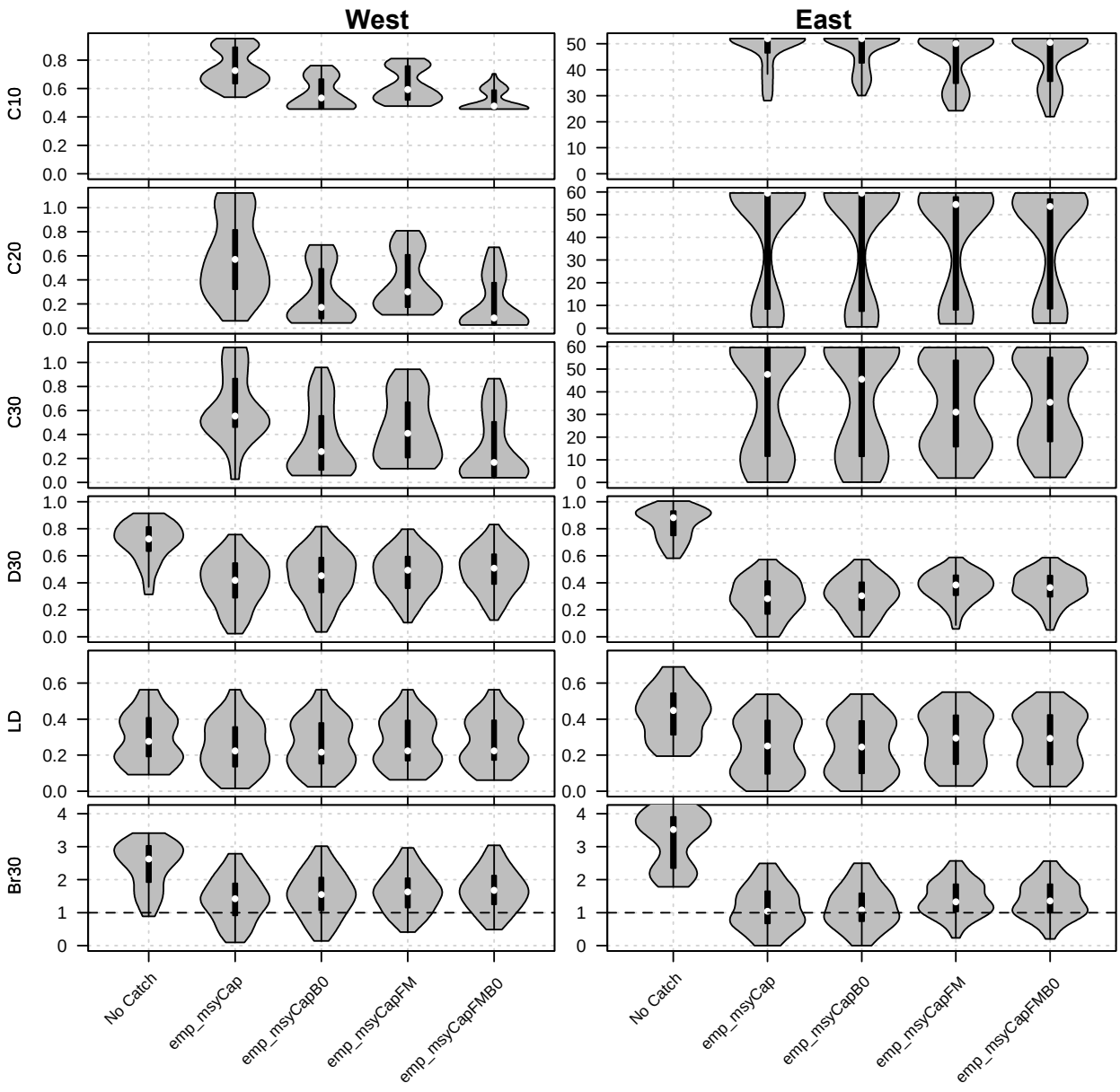

Figure 11. Violin plots of MSE performance metrics for empirical MPs. The thin line, thick line and white circle within each "violin" represents a boxplot of values across OMs, while either side of boxplot shows a rotated kernel density plot of the distribution of values. For the "FM" procedures a maximum harvest rate of F=M/4 was used.

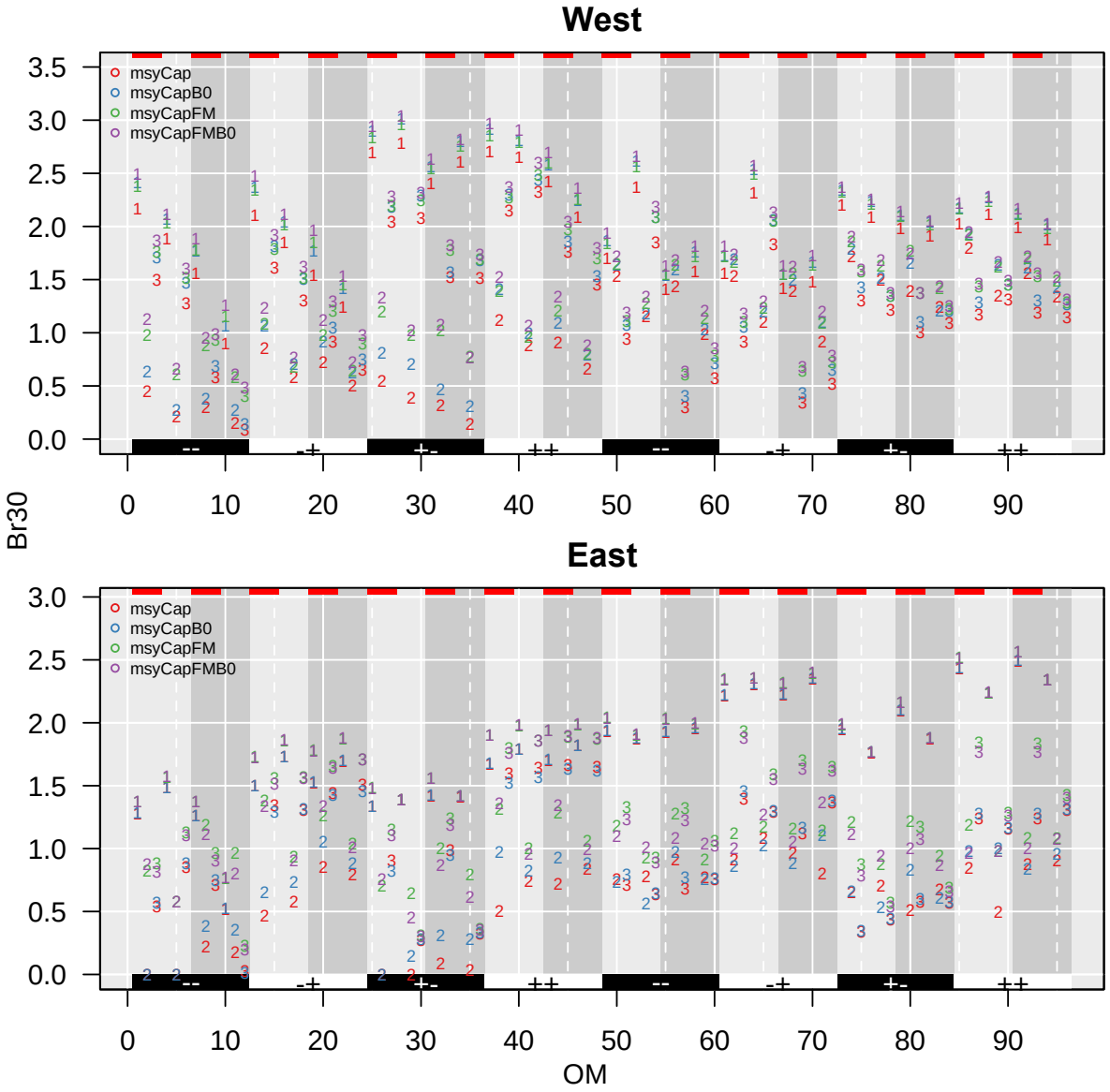

Figure 12. Estimated Br30 values for empirical MPs. Numerical values indicate the steepness factor for each OM (1=high, 2=low, 3=shifting). Younger spawning/higher M scenarios are indicated by the red line at the top of each plot. Migration factor levels are indicated by the background shading (light grey indicates low stock migration, dark grey indicates high). SSB factor levels are indicated by the black and white symbols on the bottom of each plot (-- indicates low West/low East; -+ indicates low/high; +- indicates high/low; ++ indicates high/high). The first 48 OMs have low length composition weighting while the latter 48 have high length composition weighting.

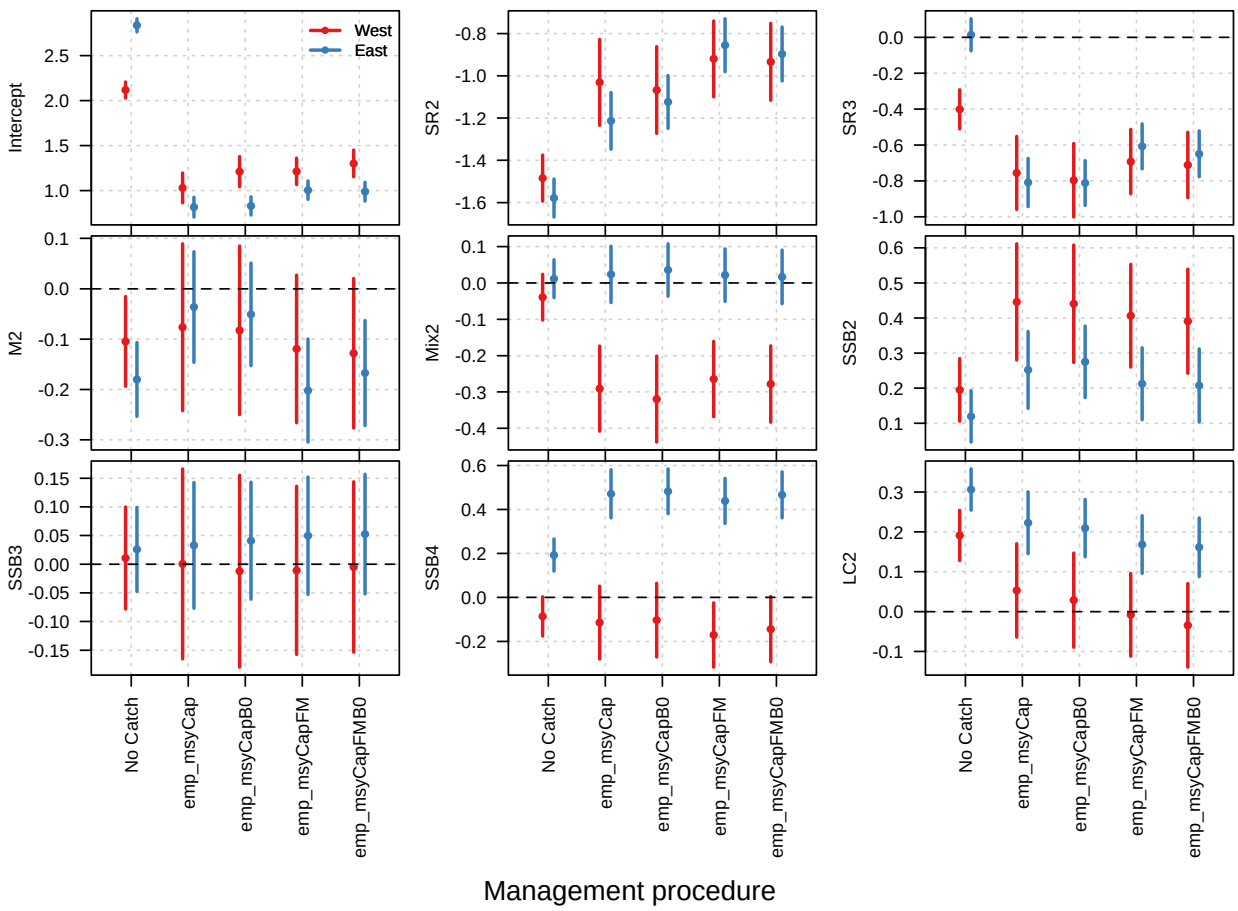

Figure 13. Coefficients from regression of Br30 on OM factor levels for empirical CMPs.