An approach to weighting the various Operating Models in a Reference Set in inverse relation to the similarity of their results

Candysse Vrancken and Doug Butterworth

November 2017

Summary

How to weight results from different OMs in getting a "best" representation across their differing results is a problem not only in fisheries but also in Climate Change. We attempt here to borrow an approach from the latter field and compare it to the conventional likelihood (or AIC) basis sometimes used in fisheries, which gives higher weights to models in a Reference Set (RS) that fit the data better. In contrast, a major problem perceived in Climate Change analyses, when averaging over an ensemble of models, is how to avoid "bias" through including too many models which scarcely differ amongst each other – one therefore downweights models on the basis of "nearness" of their results to each other. Here we apply a multi-dimensional scaling (MDS) approach which has been applied for weighting different Climate Change models to the RS developed for selecting the current OMP used to recommend TACs for the South African hake fishery. It is found that the MDS and AIC weights are very different, which begs the question of how then to "average" across these two distinct bases for model preference to perhaps obtain some combined weight. It was nevertheless found that in all the cases considered, the weighted RS provided a higher spawning biomass projection for *M. paradoxus* than the equally weighted RS used to select the current hake OMP in 2014. This suggests that, had some unequal weighting approach been used in 2014, it might have led to a slightly less conservative OMP, which allowed for greater catches to be taken, being selected.

Introduction

When a multimodel ensemble or Reference Set (RS) is used to provide the Operating models (OMs) for management procedure (MP) testing in fisheries management, the most common method used to provide an overall summary is to find the average over the results of the individual operating models (OMs) in the RS. In many cases, South African hake included, the average is determined with all the models being given equal weights. Problems can occur, however, when this equally weighted model average is used. Results can easily become biased in some sense if there are many rather similar models in the RS.

Climate change scientists experience the same problems with large multimodel ensembles and the need to summarise over their differing results. An investigation was consequently conducted into the methods used in climate sciences to address this problem. An approach used there involves model "nearness". Essentially, it attempts to find an "unbiased" average by downweighting similar "nearer" models.

Data and methods

Some of the RS OM (Appendix A) results differ only in their projections due to the different stock-recruitment relationships used for these. In order to find a comparable way of incorporating consideration of these projections, it was decided to use a fixed catch time series for all of the OMs.

After some trial and error, the catch series was chosen to be 0.7 times the median over the 100 simulations of the future catch series from RS1 under the current hake OMP.

Spawning biomass (B^{sp}) estimates for both hake species combined (median over 100 simulations) from each of the models over the full time period (1917-2038) were used (see plot in Appendix B). The data were standardised to z-scores with a mean of 0 and a standard deviation of 1 in order to remove the effect of the different scales of the data which assists a comparison in which different natural mortality M value choices might otherwise dominate. The Euclidean distances were then found between each model and all the other models as follows:

$$d_{ij} = \sqrt{\sum_{k=1917}^{2038} (B_{ik}^{*} - B_{jk}^{*})^2} \qquad (1)$$

where B^* is the z-score corresponding to spawning biomass B^{sp}, i and j represent the OMs and k years.

These distances form a proximity matrix (Table 1).

In order to visualise the similarities between the 11 models, multidimensional scaling[1] (MDS) was used to represent the proximity matrix in a smaller dimension. First, the values in the proximity matrix were transformed using a simple ratio transformation (transformed proximities are proportional to the original proximities).

MDS models ideally require that each proximity value be mapped exactly into its corresponding distance:

$$\hat{d}_{ij} = d_{ij}(X) \qquad (2)$$

where $\hat{d}_{ij}$ are the transformed proximities and $d_{ij}(X)$ are the Euclidean distances between the object points in the common space (new MDS space).

However, proximities are never free of error. Computerised procedures for finding an MDS representation usually start with some initial configuration (simple random start) and improve this configuration by moving around its points in small steps (iteratively) to approximate the ideal model relation (Equation 2) more and more closely. In order to do this, the following raw stress function (Borg and Groenen, 2005) is minimised by PROXSCAL:

$$\sigma^2 = \sum_{i<j}^{11} [\hat{d}_{ij} - d_{ij}(X)]^2 \qquad (3)$$

with normalised raw stress:

$$\sigma^2 = \frac{\sum_{i<j}^{11} [\hat{d}_{ij} - d_{ij}(X)]^2}{\sum_{i<j}^{11} [d_{ij}(X)]^2} \qquad (4)$$

The normalised stress values in different dimensionalities are given in a scree plot (Figure 1). With the scree plot, the goal is to find the point where the stress decrements become less pronounced. That point corresponds to the dimensionality that should be chosen. This is because it marks the point where MDS uses additional dimensions to essentially better match only the noise in the data, after having

---

[1] Used SPSS software

managed to reasonably represent the proximities in the given dimension. For this reason, based on Figure 1, it was decided to represent the OMs in 3D space.

Once the stress function (evaluated while plotting in the new 3D space) meets the convergence criteria, the final coordinates in the common space are plotted (Figure 2). The distances between the models in the 3D space (Table 2) are used to determine weights for each of the models with the most similar models being downweighted.

For each RS, the average of the distances from that RS to all of the others was calculated. These averages were then normalised, producing the weights (MDS weight (1)) shown in Table 3.

Another, slightly more complicated, method was used to specify weights (MDS weight (2)), as explained in Appendix C. These weights can also be found in Table 3.

A third set of weights (AIC weights, reflecting goodness of fit of the OM to the data) (see Table 3) was determined based on the log-likelihood differences for each of the OMs:

$$w_i = e^{-(\Delta(-lnL))_i} \tag{5}$$

where i represents the OM.

Results and Discussion

Each OM was weighted according to the weights in Table 3. It seems to be that the reason why RS10 and R13 are so “different” from the other OMs (and hence are accorded the greatest weights) is because they both have the centre year for the change from *M. capensis* to *M. paradoxus* preponderance in catch to be 1965 and a Ricker stock-recruitment relationship.

Figure 3 shows the differences between the weights calculated using the two different MDS methods and the AIC weighting based on log-likelihood differences. The uniqueness method described in Appendix C puts more importance on the models (RS10 and RS13) that are “further away” than the distance averaging method described in the main text. The AIC weights are quite different from the MDS ones, spanning a much greater range. RS13 (which fits the data best) remains with a (very) high weight, but RS10 corresponds to the worst fit to the data of all the OMs and gets effectively zero weight.

Figure 4, shows the comparison of the OMP-2014 projections between the equally weighted and unequally weighted averages with MDS weights (1), Figure 5 shows the same comparison but the weights (MDS weights (2)) used were those determined using the method described in Appendix C and Figure 6 uses the AIC weights. These indicate that in all cases, the weighted RS provides a higher spawning biomass projection for *M. paradoxus* than the equally weighted RS. This effect is greater for MDS method 2 than for method 1, and greater still for AIC weights.

Conclusion

In all cases, the weighted RS gives a more optimistic projection forecast than the equally weighted case. This suggests that, had the unequal weighting been used in 2014, it might have led to a slightly less conservative approach allowing for greater catches being selected.

These results are intended only as examples to illustrate this approach “borrowed” from the Climate Change area. They do however indicate a form of conflict between the criteria: the MDS approach which weights “farness” (difference of results) the highest for RS10, compared to the AIC basis of goodness of fit to the data which weights RS10 the lowest. This begs the question of how then to “average” across these two distinct bases for model preference to perhaps obtain some combined weight. Clearly however, there needs to be some “badness” of fit threshold below which OMs are excluded.

References

Borg, Ingwer and Groenen Patrick J.F., 2005. *Modern Multidimensional Scaling Theory and Applications,* Springer, New York.

Rademeyer, R.A. and Butterworth, D.S., 2014. Results leading to a Proposal for a Reference Set of Operating Models for Testing the 2014 OMP Revision for the South African hake resource. FISHERIES/2014/MAR/SWG-DEM/14.

Rademeyer, Rebecca and Butterworth, Doug, 2016. Reference Set results and projections under the current OMP for the South African hake resource. FISHERIES/2016/NOV/SWG-DEM/84.

Sanderson et al., 2015. A Representative Democracy to Reduce Interdependency in a Multimodel Ensemble. *Journal of Climate,* 28(13): 5171-5194.

Table 1: Proximity matrix found using Equation 1

	RS1	RS2	RS3	RS4	RS5	RS6	RS7	RS10	RS13	RS14	RS15
RS1	0										
RS2	0.924	0									
RS3	0.924	0	0								
RS4	1.020	1.684	1.684	0							
RS5	1.385	1.776	1.776	1.240	0						
RS6	1.385	1.777	1.777	1.240	0.001	0					
RS7	0.854	0.886	0.886	1.681	1.899	1.899	0				
RS10	6.446	6.147	6.147	6.812	7.497	7.497	6.074	0			
RS13	4.964	4.665	4.665	5.659	6.148	6.148	4.428	3.407	0		
RS14	1.796	1.138	1.138	2.554	2.748	2.748	1.478	5.639	3.984	0	
RS15	1.796	1.138	1.138	2.554	2.748	2.748	1.478	5.639	3.984	0	0

Table 2: Distances between OMs in 3D space after multidimensional scaling

	RS1	RS2	RS3	RS4	RS5	RS6	RS7	RS10	RS13	RS14	RS15
RS1	0										
RS2	0.217	0									
RS3	0.217	0	0								
RS4	0.286	0.466	0.466	0							
RS5	0.336	0.516	0.516	0.315	0						
RS6	0.336	0.517	0.517	0.314	0.003	0					
RS7	0.248	0.209	0.209	0.531	0.513	0.513	0				
RS10	1.768	1.749	1.749	1.918	2.008	2.007	1.622	0			
RS13	1.391	1.314	1.314	1.507	1.707	1.706	1.303	0.960	0		
RS14	0.479	0.298	0.298	0.686	0.804	0.805	0.400	1.613	1.073	0	
RS15	0.478	0.298	0.298	0.684	0.804	0.805	0.401	1.614	1.073	0.004	0

Table 3: Model weights – where the subscripts show similar models. If each model were equally weighted it would have weight = 0.091. Log-likelihood values relative to the best fitting OM (RS13) are also given with their associated AIC weights. MDS weight (1) refers to the weights determined using the distance averaging method in the main text and MDS weight (2) refers to the weights found using the method described in Appendix C.

RS model	Centre year	Natural mortality	Stock-recruitment	MDS weight (1)	MDS weight (2)	$\Delta(-lnL)$	AIC weight
$\text{RS1}_{\text{RS2,RS3}}$	1958	Mmed	Ricker	0.065	0.043	4.7	0.009
$\text{RS2}_{\text{RS1,RS3}}$	1958	Mmed	BH	0.063	0.044	6.6	0.001
$\text{RS3}_{\text{RS1,RS2}}$	1958	Mmed	BHmod	0.063	0.044	6.6	0.001
$\text{RS4}_{\text{RS5,RS6}}$	1950	Mlow	Ricker	0.081	0.059	11.1	0
$\text{RS5}_{\text{RS4,RS6}}$	1950	Mlow	BH	0.085	0.060	13	0
$\text{RS6}_{\text{RS4,RS5}}$	1950	Mlow	BHmod	0.085	0.060	13	0
RS7	1950	Mhigh	Ricker	0.067	0.049	5	0.007
RS10	1965	Mlow	Ricker	0.192	0.266	14.1	0
$\text{RS13}_{\text{RS14,RS15}}$	1965	Mhigh	Ricker	0.151	0.259	0	0.982
$\text{RS14}_{\text{RS13,RS15}}$	1965	Mhigh	BH	0.073	0.058	10.1	0
$\text{RS15}_{\text{RS13,RS14}}$	1965	Mhigh	BHmod	0.073	0.058	10.1	0

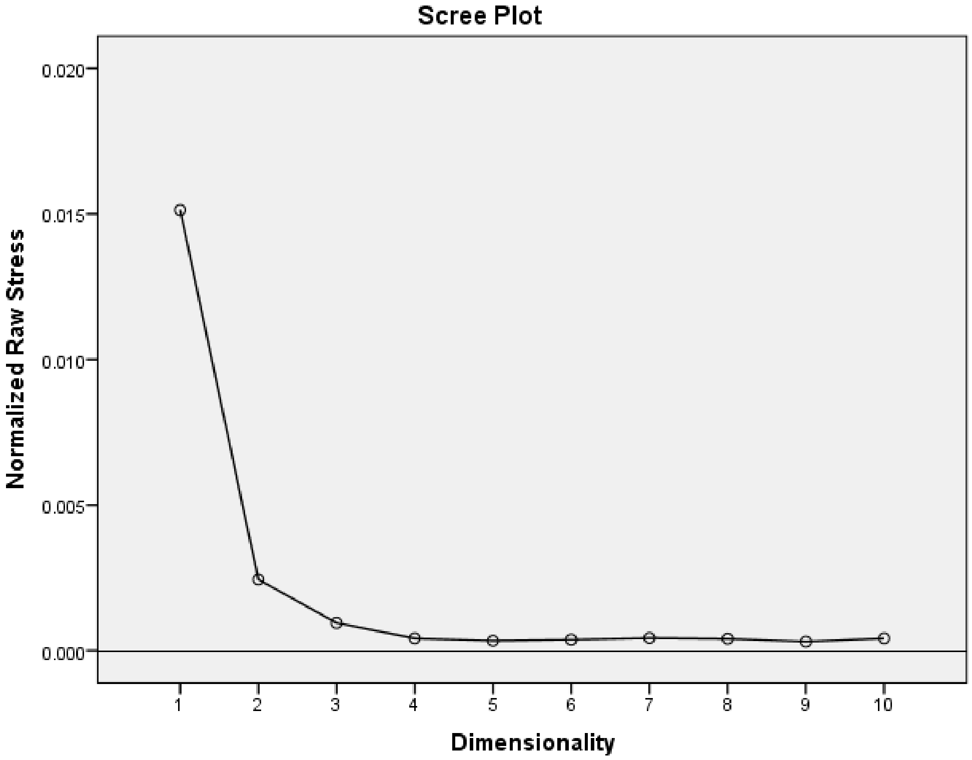

Figure 1: Scree plot representing the normalised raw stress values calculated using Equation 4.

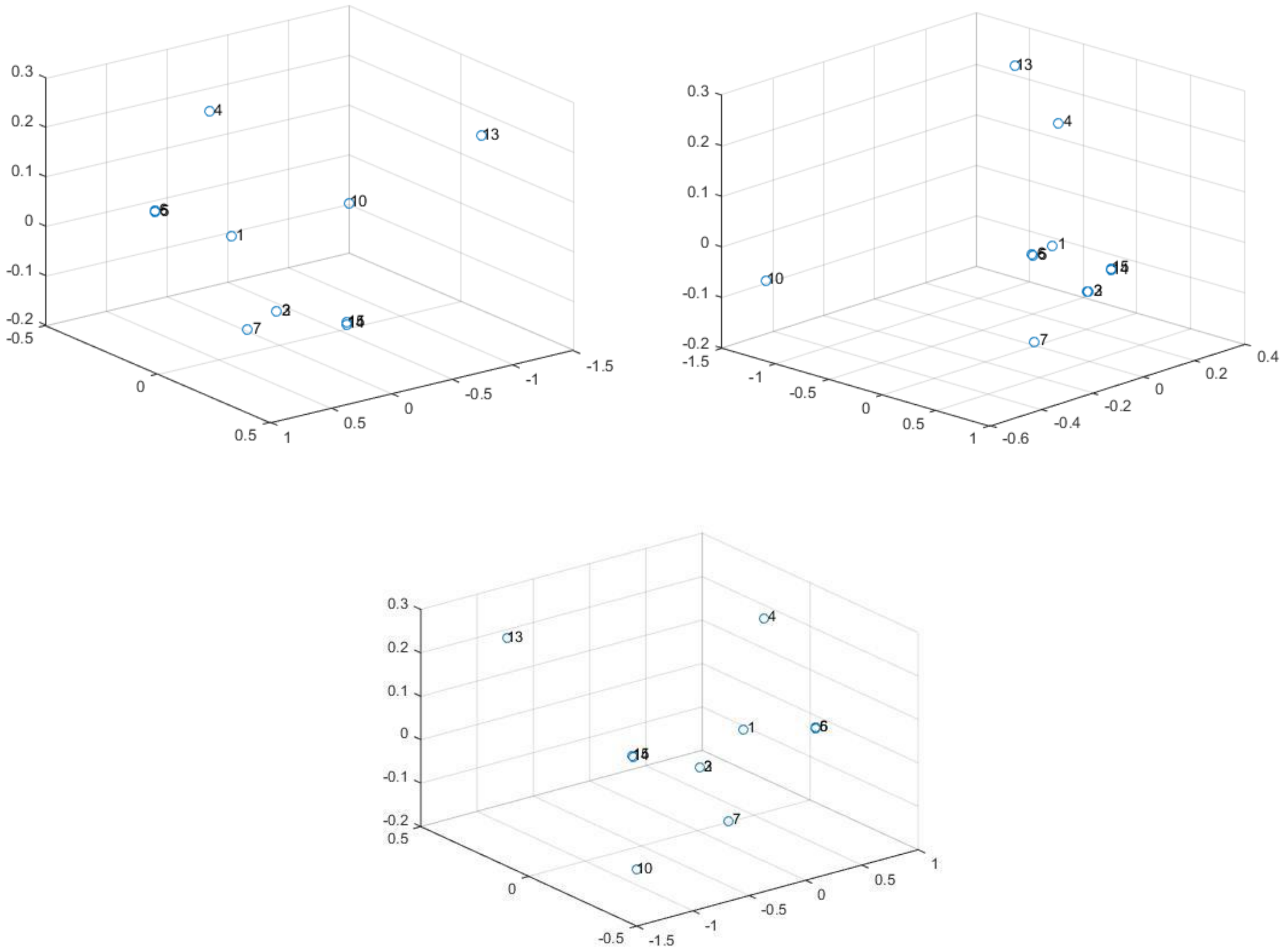

Figure 2: Views from different orientations of the same three-dimensional representation of the proximity matrix for all the RS OMs.

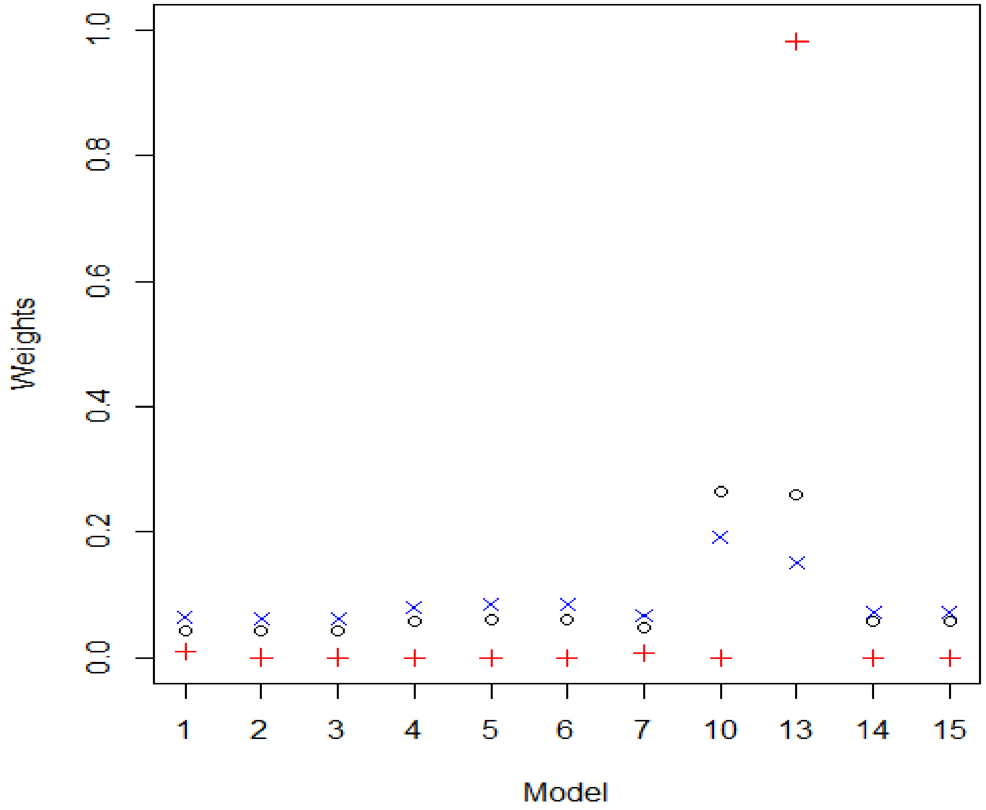

Figure 3a: MDS weights determined using the two different methods – distance averaging described in the main text (blue x) and the uniqueness weighting method described in Appendix C (black diamond) and AIC weights based on log-likelihood differences (red cross).

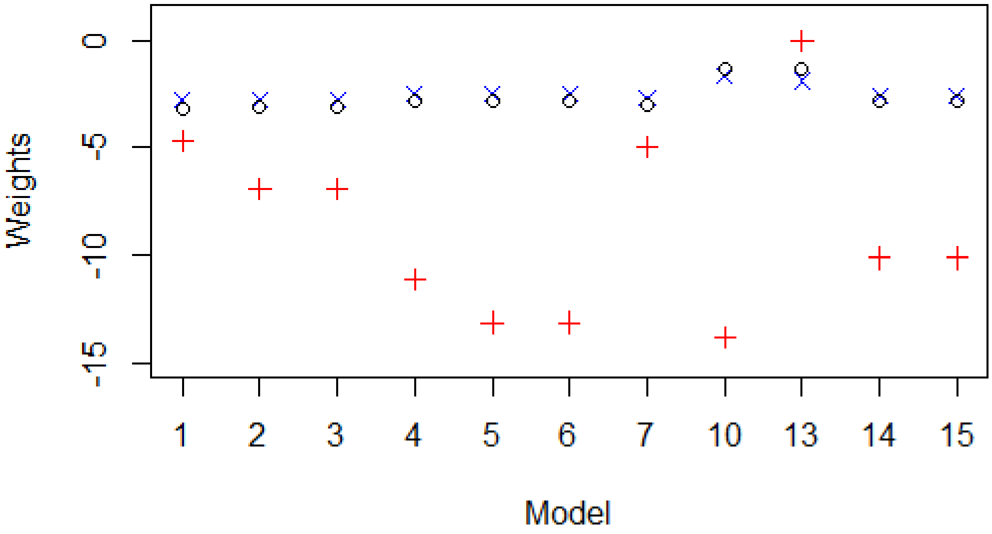

Figure 3b: Log values of MDS weights determined using the two different methods – distance averaging described in the main text (blue x) and the uniqueness weighting method described in Appendix C (black diamond) - and of AIC weights based on log-likelihood differences (red cross).

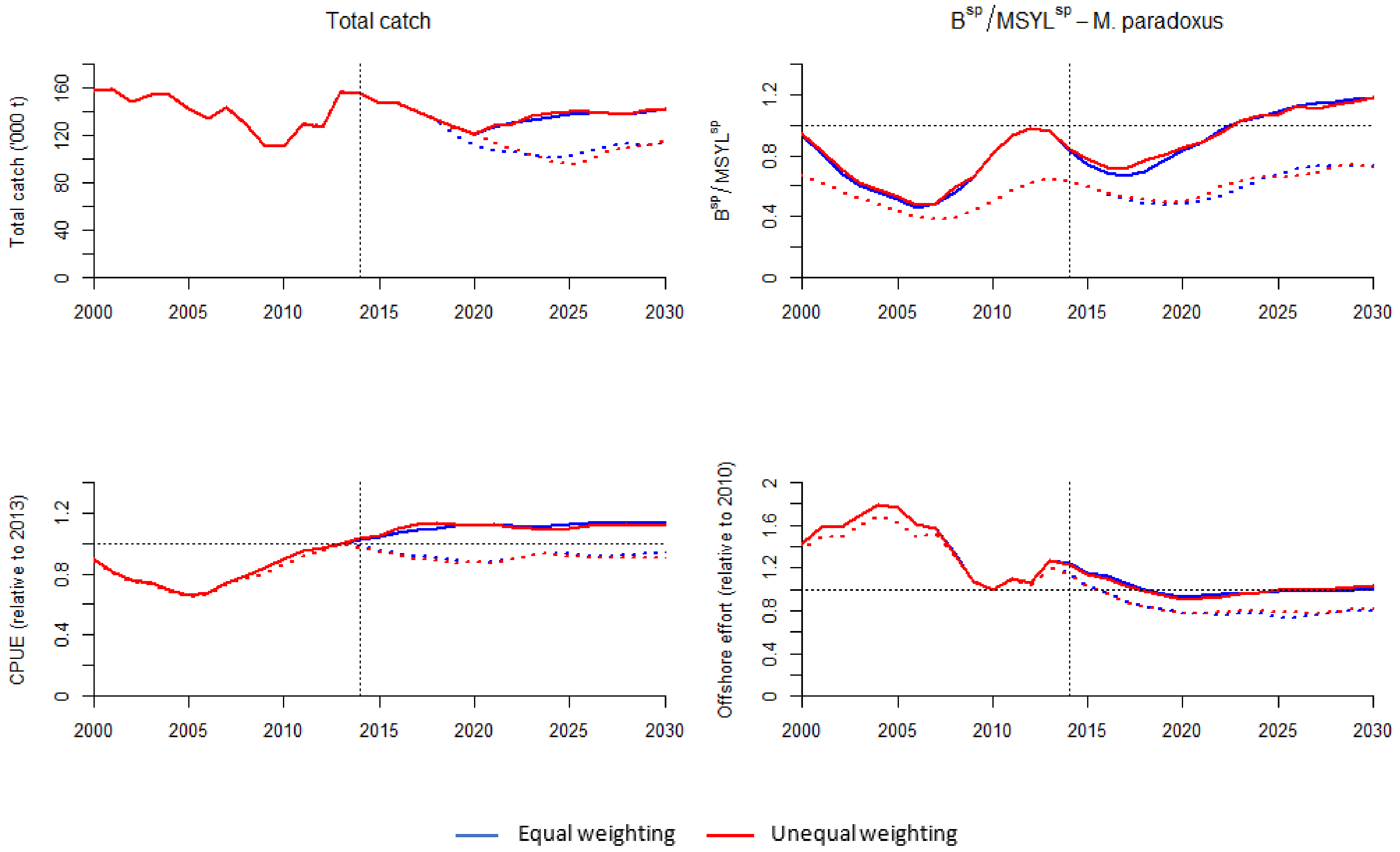

Figure 4: Medians (full lines) and lower 5%iles (dashed lines) for total catch (top row, LHS), *M. paradoxus* spawning biomass (relative to MSYL level – top row, RHS), CPUE (relative to 2013, bottom row LHS) and effort (relative to 2010, bottom row, RHS) for the RS with each OM equally weighted (blue) and with OMs unequally weighted (red) – using MDS weights (1) with a horizontal dotted line at 1 and a vertical dotted line at year 2014 when the future projections begin.

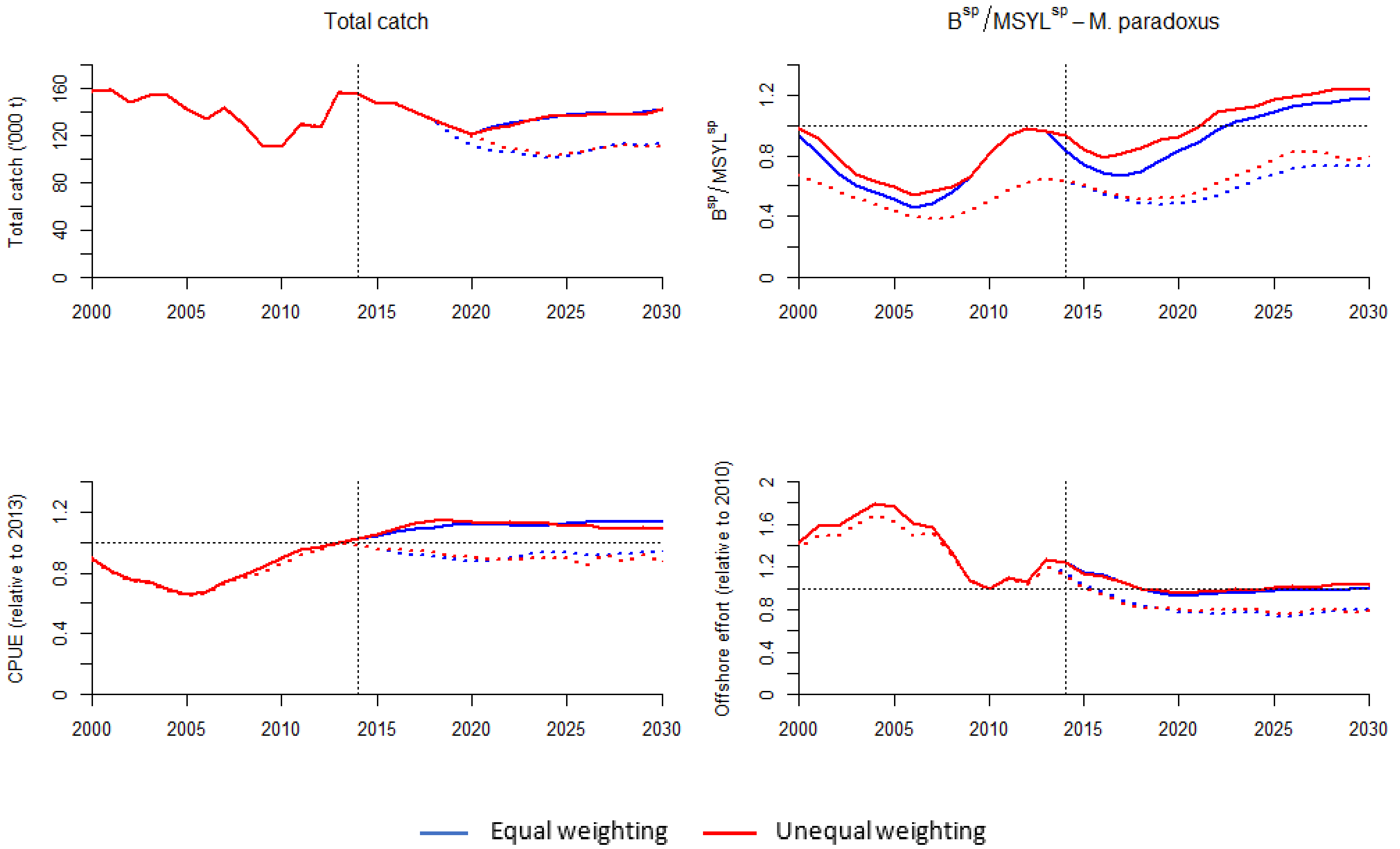

Figure 5: Medians (full lines) and lower 5%iles (dashed lines) for total catch (top row, LHS), *M. paradoxus* spawning biomass (relative to MSYL level – top row, RHS), CPUE (relative to 2013, bottom row LHS) and effort (relative to 2010, bottom row, RHS) for the RS with each OM equally weighted (blue) and with OMs unequally weighted (red) – using MDS weights (2) with a horizontal dotted line at 1 and a vertical dotted line at year 2014 when the future projections begin.

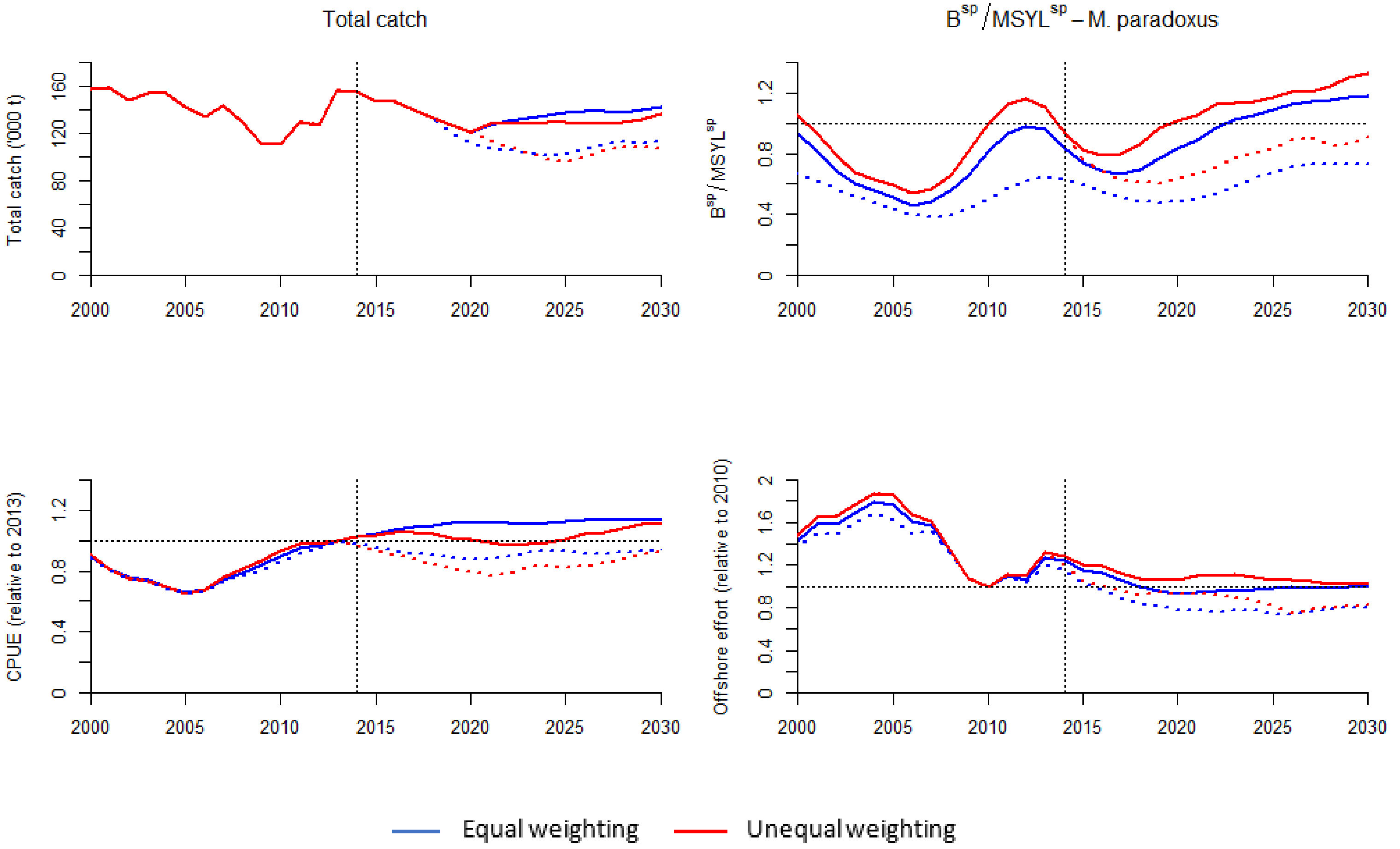

Figure 6: Medians (full lines) and lower 5%iles (dashed lines) for total catch (top row, LHS), *M. paradoxus* spawning biomass (relative to MSYL level – top row, RHS), CPUE (relative to 2013, bottom row LHS) and effort (relative to 2010, bottom row, RHS) for the RS with each OM equally weighted (blue) and with OMs unequally weighted (red) – using AIC weights with a horizontal dotted line at 1 and a vertical dotted line at year 2014 when the future projections begin.

Appendix A

The 11 Operating Models (OMs) in the RS cover three major axes of uncertainty: 3 centre-years for the species preponderance change in the catch, 3 natural mortality vectors and 3 stock-recruitment relationships.

- Centre years for the change from *M. capensis* to *M. paradoxus* preponderance in catch: 1950,1958 and 1965.
- Natural mortality vectors: Mmed (M_{2-}=0.75 and M_{5+}=0.375), Mlow (M_{2-}=0.6 and M_{5+}=0.25) and Mhigh (M_{2-}=0.9 and M_{5+}=0.5).
- Stock-recruitment relations: Ricker (modified Ricker), BH (Beverton-Holt, h estimated) and BHmod (modified Beverton-Holt).

For the modified Ricker relationship: $R_y = \alpha B_y^{sp} \exp\left(-\beta (B_y^{sp})^{\gamma}\right)$

For the modified Beverton-Holt, the curve below Bmin (the lowest previous biomass level) is replaced by the average of that curve and a straight line from the origin to the value of the curve at Bmin.

Table A.1: Reference Set Operating Model specifications

RS model	Centre year	Natural mortality	Stock-recruitment
RS1 (RC)	1958	Mmed	Ricker
RS2	1958	Mmed	BH
RS3	1958	Mmed	BHmod
RS4	1950	Mlow	Ricker
RS5	1950	Mlow	BH
RS6	1950	Mlow	BHmod
RS7	1950	Mhigh	Ricker
RS10	1965	Mlow	Ricker
RS13	1965	Mhigh	Ricker
RS14	1965	Mhigh	BH
RS15	1965	Mhigh	BHmod

Appendix B

Spawning biomasses trajectories and projections for both hake species combined for the various OMs of the Reference Set.

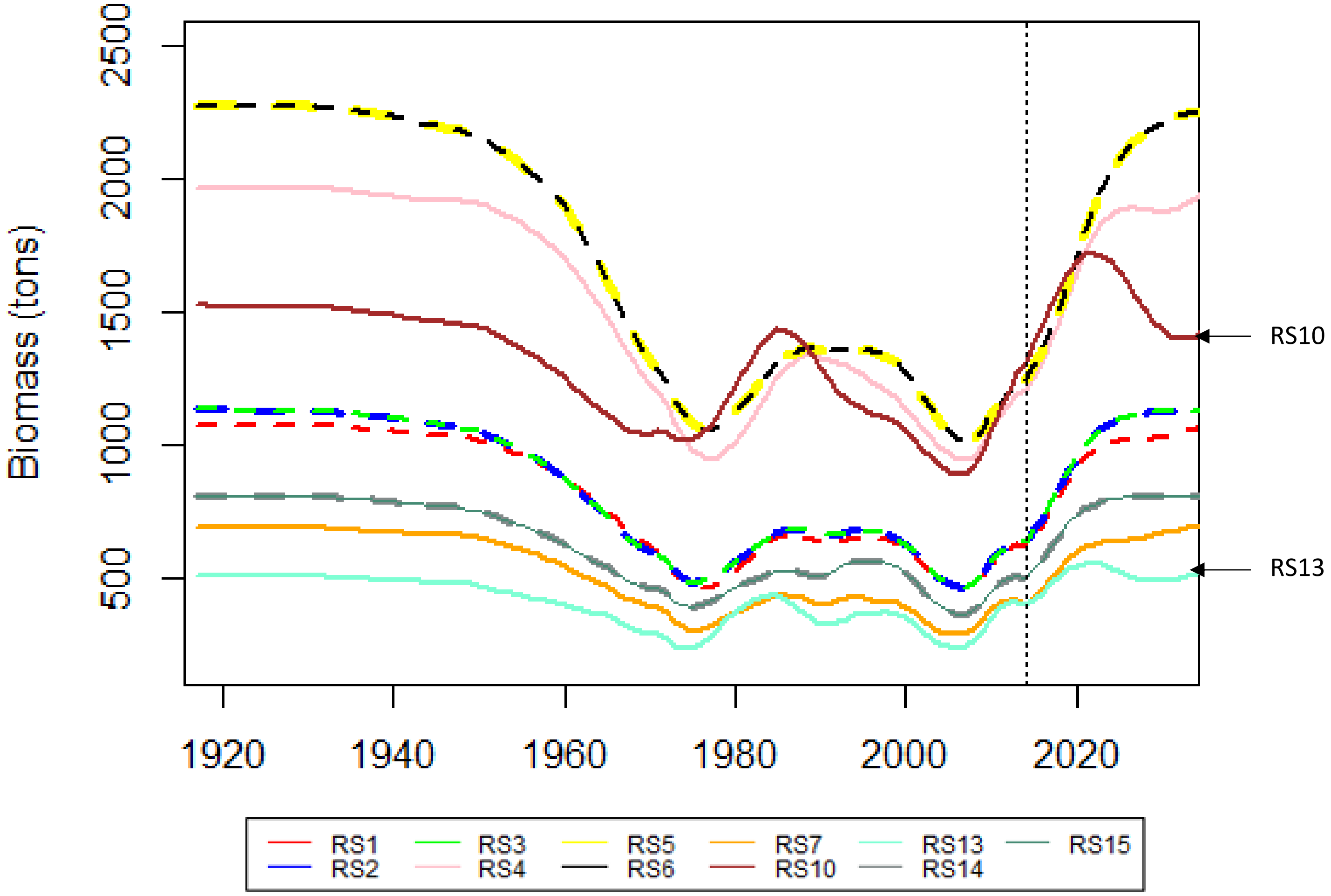

Figure B.1: Spawning biomass trajectories (for both species combined) for each of the operating models, with fixed future catch as detailed in the main text with a vertical line at year 2014 when the future projections begin. RS10 and RS13 are the two "most different" and therefore have the highest MDS weights.

Appendix C

A method described by Sanderson et al. (2015) was used to determine the relative weights for each of the models according to their similarities. A simple functional form for model similarity was determined to be:

$$S(d_{ij}) = \exp\left(-\left(\frac{d_{ij}}{D_u}\right)^2\right) \tag{C.1}$$

where D_u is a free parameter such that model pairs separated by less than this value are considered to be very similar. From this equation, it can be seen that the weights depend on the value of this parameter. A more detailed investigation will be conducted into the "best" value to use for D_u , but in the meantime for this paper $D_u = 0.51$. This choice is near 0.5 is somewhat arbitrary; the exact value used was chosen as it happened to make the weighting process simpler.

Then, a value for the effective repetition of model i in the ensemble is given by:

$$R_u(i) = 1 + \sum_{j \neq i}^{m} S(d_{ij}) \tag{C.2}$$

where m is the total number of models.

A uniqueness weighting (normalised in Table 3) for each model:

$$w_u(i) = [R_u(i)]^{-1} \tag{C.3}$$

Projections using these weights can be seen in Figure 5.