ON COMPARING CMPS ACROSS DIFFERENT DEVELOPMENT TUNINGS AND THE ASSOCIATED PERTINENCE OF OM WEIGHTING

D.S Butterworth, R.A Rademeyer[1] and T.R Carruthers[2]

SUMMARY

Medians and lower %iles for the East and West Br30 and AvC30 performance statistics over the interim grid of OMs are compared for the most recent versions of the Butterworth-Rademeyer (BR) and Carruthers (TC) CMPs. Each CMP has been tuned to the agreed development tuning targets of median values for Br30 West (1.00, 1.25, 1.50), and all to the same value of Br30 East. Performances are very similar, despite rather different CMP structures. The differences are generally independent of the development tuning value for the western stock. This suggests that CMP performance comparisons can proceed before awaiting agreement on OM weightings; these are primarily of consequence for finalization tuning (which will be needed in 2022 when the Commission is to make its final MP choice). The performances of different CMPs will need to be compared over a much wider set of performances statistics than considered in the illustrative example presented. This large task will require the allocation of considerable time for discussion in meetings of the BFT WG during 2021 for the MP development process to remain on schedule.

RÉSUMÉ

Les médianes et les percentiles inférieurs des statistiques de performance Br30 et AvC30 de l'Est et de l'Ouest sur la grille provisoire des OM sont comparés pour les versions les plus récentes des CMP de Butterworth-Rademeyer (BR) et Carruthers (TC). Chaque CMP a été calibrée aux objectifs du calibrage du développement convenus des valeurs médianes de Br30 Ouest (1,00, 1,25, 1,50), et tous sur la même valeur de Br30 Est. Les performances sont très similaires, malgré des structures CMP assez différentes. Les différences sont généralement indépendantes de la valeur du calibrage du développement pour le stock occidental. Cela suggère que les comparaisons des performances de la CMP peuvent être effectuées avant d'attendre un accord sur les pondérations des OM ; celles-ci sont principalement importantes pour le calibrage final (qui sera nécessaire en 2022 lorsque la Commission devra faire son choix final de MP). Les performances des différentes CMP devront être comparées sur un ensemble de statistiques de performances beaucoup plus large que celui considéré dans l'exemple présenté. Cette tâche de grande envergure nécessitera l'allocation d'un temps considérable pour les discussions lors des réunions du Groupe d'espèces sur le thon rouge au cours de l'année 2021, afin que le processus de développement de la MP reste dans les temps.

RESUMEN

Se comparan las medianas y los percentiles inferiores de las estadísticas de desempeño Br30 y AvC30 del este y del oeste en la matriz provisional de OM para las versiones más recientes de los CMP de Butterworth-Rademeyer (BR) y Carruthers (TC). Cada CMP se ha calibrado con los objetivos de calibración de desarrollo acordados de los valores de las medianas de Br30 oeste (1,00, 1,25, 1,50), y todos al mismo valor de Br30 este. Los desempeños son muy similares, a pesar de que las estructuras del CMP son bastante diferentes. Las diferencias son, en general, independientes del valor de calibración de desarrollo para el stock occidental. Esto sugiere que las comparaciones de los desempeños del CMP pueden llevarse a cabo antes de esperar a que se acuerden las ponderaciones de los OM; éstas son principalmente importantes para finalizar la calibración (que será necesaria en 2022, cuando la Comisión deba hacer su elección final de MP). El desempeño de los distintos CMP deberá compararse con un conjunto de estadísticas de desempeño mucho más amplio que el considerado en el ejemplo ilustrativo presentado. Esta gran tarea requerirá la asignación de un tiempo considerable para el debate en las reuniones del Grupo de especies de atún rojo durante 2021 para que el proceso de desarrollo del MP se pueda realizar dentro del calendario.

KEYWORDS

Management Strategy Evaluation, Candidate Management Procedure, Operating Model, Atlantic bluefin tuna, tuning

[1] Marine Resource Assessment and Management Group (MARAM), Department of Mathematics and Applied Mathematics, University of Cape Town, Rondebosch 7701, South Africa
[2] IOF, 2202 Main Mall, University of British Columbia, Vancouver, B.C., Canada, V6T 1Z4

Introduction

The concepts of tuning and plausibility weighting in an MSE/MP development process are complex, and it is understandable if some members of the BFT Working Group are not entirely clear about them, their objectives and how they inter-relate.

First it is important to distinguish two types of tuning:

- "Development" tuning is used in the initial stages of comparing the performances of different Candidate Management Procedures (CMPs). Its primary purpose is to remove the fundamental confounding effect of the negative correlation between catch achieved and final stock(s) abundance over the management period (here 30 years) under consideration. This allows the other performance features of alternative CMPs to be compared on a "level playing field", without this confounding feature adding difficulty to the exercise.

- "Finalisation" tuning comes into play only at the final stages of selection amongst a few final CMPs by the Commission. At that time a choice will need to be made as to the trade-off desired between the broad overall objectives of maximizing catches, and of keeping final resource abundance high so as to limit any risk of undue depletion of the stock(s). This may be achieved through the specification of some performance objective which (the control parameters) of CMPs are to be tuned to achieve.

At this time, the situation is as follows:

- the BFT WG has specified three targets for development tuning: median values of Br30 (dynamic B/Bmsy after 30 years of management) for the western stock of 1.00, 1.25 and 1.50, where that median is for the distribution of results from deterministic runs of CMPs over the 96 OMs of the interim grid;

- finalisation tuning has yet to be considered in any detail; it will come to the fore only after the Commission has been presented with initial CMP results late in 2021, so will likely start to need discussion only early in 2022; and

- the interim grid of 96 OMs has yet to be finalized by the BFT WG, and its component OMs are currently being equally weighted; however, work is in progress, following a poll of WG members' views on OM relative plausibilities, to perhaps use those results to weight the OMs included in a final grid.

Recently Butterworth and Rademeyer (BR) and Carruthers (TC) have tuned their currently preferred CMPs to these Br30 West targets. For enhanced comparability of the results, these CMPs have also been tuned (as near as makes no odds) to an identical median Br30 East target. The outcomes from this process are used here to provide an initial illustrative example to indicate the following.

1) The complexity of the process of comparing CMPs and their performances, so as to make WG members aware that considerable time will need to be allocated to discussion of this during meetings later this year.

2) The role that OM weighting plays at this stage (which is actually fairly limited – this will rather play its primary role in the finalisation tuning process which will come under fuller discussion only early in 2022).

Given these basic objectives, the main text of this document has been kept brief. Only a few performance statistics are reported in the interests of simplicity for what is intended primarily as an illustrative exercise – a later more intensive CMP comparison exercise would certainly need to consider a wider set of performance statistics. Even given these few statistics, however, there still remain many ways that they can be reported, both in Tables and Figures. Again, in the interests of simplicity, a number of such Figures are shown only in the Appendix, with a limited number in the main text. The latter have been singled out as they are considered sufficient to get the main messages from these results across to readers.

Results

For each West median Br30 tuning (1, 1.25 and 1.5), the deterministic Br30 and AvC30 median, lower 5%ile and lowest value are compared in Table 1 for the BR and TC CMPs. The absolute differences are also given.

Figure 1 plots these values (medians and lower 5%iles) by pair based on the median Br30 West tuning, over the interim grid of OMs.

The relative differences in Br30 and AvC30 (medians and lower 5%iles) between BR and TC CMPs are plotted in **Figure 2** for each of the three median Br30 West tunings.

Discussion

Figure 1 immediately makes clear that, despite their rather different structures, there is relatively little difference between the results for the BR and TC CMPs for any specific tuning choice. Clearly by construction, the median Br30 values are the same. Comparison is essentially for the median AvC30 values and the lower 5%iles for Br30 and AvC30 (lower is chosen because such %iles reflect risk, for the resource and to the fishing industry respectively).

In **Figure 2**, where differences are expressed in relative terms to be comparable across East and West, and between resource status and catch, note that positive values suggest BR performs better, and negative values that TC performs better, in terms of the statistic in question. To summarise briefly across the three tunings, with only differences sufficiently large to be meaningful being **bolded:**

Br30 west	median	N/A (tuned)
	5%ile	TC better by 0-2% - note that this is a key measure of robustness to risk
AvC30 west	median	**TC better by 6% for 1.0 tuning switching to BR better by 3% at 1.5 tuning**
	5%ile	**BR better by 4-12%**
Br30 east	median	N/A (tuned)
	5%ile	TC better by 0-3% - note this relates to robustness to risk, but values for BR and TC are sufficiently high that this is of little concern
AvC30 east	median	TC better by 1%
	5%ile	**TC better by 4-5%**

Importantly, only for AvC30 for the West does the tuning choice make any meaningful difference – for the other statistics, the curves across the tunings in Figure 2 are essentially flat. This means that one statistic aside (and it is not surprising, given the complexities introduced by needing to take account of the consequences of stock movements across across the Atlantic, that there is at least one instance of a tuning-related difference), the development tuning value choice will not impact any evaluation of which CMP is to be preferred.

The range for the western stock development tuning targets was chosen by the BFT WG on the basis that it likely included the value on which the Commission might eventually settle for finalization tuning. Changing the plausibility weightings of the OMs in the final grid will affect that ultimate tuning, as medians of performance statistics such as Br30 will change to some (though a likely limited) extent depending on those weightings. But this seems likely not to be of much consequence for judging the comparative performances on CMPs at this stage, given the insensitivity of CMP performance differences to such tuning target values which is indicated by **Figure 2.**

Comparing CMP performance will, however, require going well beyond the few performance statistics shown for illustrative purposes in **Figure 2**. Development tuning values for Br30 East will also need to be agreed and the associated results obtained and presented. As indicated by the various Figures in the Appendix, even for the limited number of performance statistics considered here, there are many ways in which they can be interrogated. These though, are of course not the only performance statistics of importance to stakeholders, and which are available for the interim grid. And furthermore, the comparative performances of CMPs for the various robustness tests will

also need to be considered. These points are listed here to make clear to the BFT WG that it has much work to undertake during the balance of this year in addressing these comparisons, and also to point to the importance of selecting and focusing on a set of more important statistics, as will be impossible to review "everything/output" in detail.

Concluding remarks

Notwithstanding its initial and illustrative nature, this document suggests two important take home messages.

1) Development tuning based comparisons of the performance of different CMPs can proceed usefully **without** first awaiting finalization of OM plausibility weightings, which will rather be of primary relevance to the finalisation tuning process that will come under consideration from early in 2022.

2) Comparing the merits of different CMPs in terms of their multiple performance statistics across many OMs is a large task. This will require the allocation of considerable time for discussion in meetings of the BFT WG during the remainder of 2021 for the MP development process to remain on schedule.

Table 1. Br30 and AvC30 (in kt) medians, lower 5%iles and lowest values for the East and the West, for each of the three median Br30 West tunings for the BR and TC CMPs. The absolute differences between the equivalent CMPs are also given.

		Tuning = 1 (BR10 vs TC10)				Tuning = 1.25 (BR11 vs TC11)				Tuning = 1.5 (BR12 vs TC12)			
		East		West		East		West		East		West	
		Br30	AvC30	Br30	AvC30	Br30	AvC30	Br30	AvC30	Br30	AvC30	Br30	AvC30
BR	Median	1.55	36.44	1.00	2.11	1.56	36.53	1.25	1.63	1.57	36.64	1.50	1.04
	Lower 5%ile	0.85	14.45	0.35	1.50	0.86	14.49	0.48	1.21	0.87	14.51	0.72	0.84
	Lower 0.5%ile	0.73	12.95	0.31	1.28	0.74	13.05	0.44	1.06	0.74	13.08	0.58	0.78
TC	Median	1.54	38.17	1.00	2.24	1.56	38.15	1.26	1.62	1.58	38.13	1.50	1.01
	Lower 5%ile	0.84	14.82	0.36	1.42	0.88	14.77	0.52	1.08	0.92	14.71	0.72	0.72
	Lower 0.5%ile	0.41	14.42	0.27	1.23	0.42	14.36	0.41	0.97	0.43	14.29	0.57	0.69
BR-TC	Median	0.01	-1.73	0.00	-0.13	0.00	-1.62	-0.01	0.01	-0.01	-1.49	0.00	0.03
	Lower 5%ile	0.01	-0.37	-0.01	0.08	-0.02	-0.28	-0.04	0.13	-0.05	-0.20	0.00	0.12
	Lower 0.5%ile	0.32	-1.47	0.04	0.05	0.32	-1.31	0.03	0.09	0.31	-1.21	0.01	0.09

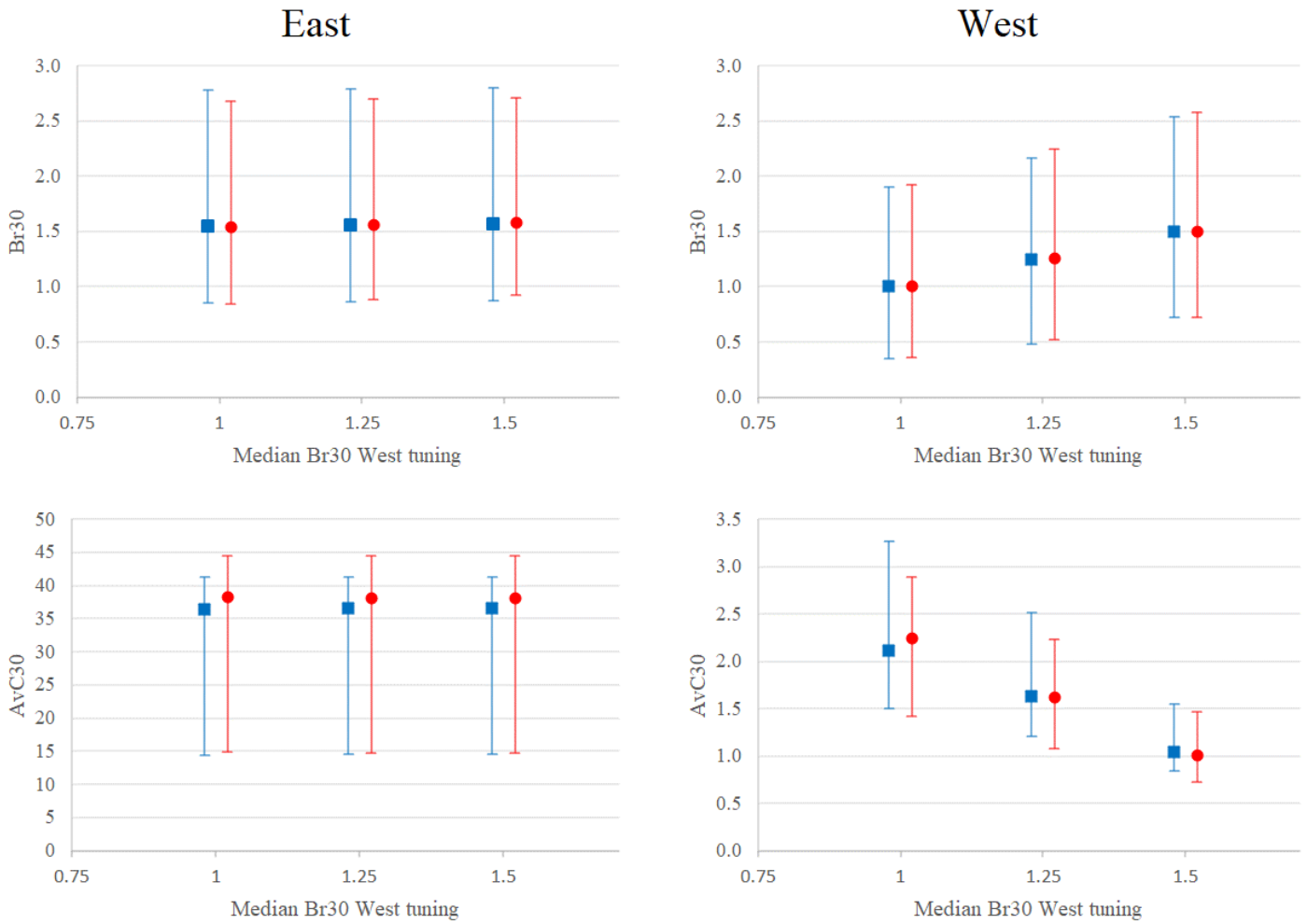


Figure 1: Deterministic Br30 and AvC30 values for the six CMPs considered, shown by pair based on the median Br30 West tuning, over the interim grid of OMs, showing medians and 90%-ile ranges. The **blue** squares are for the **BR** CMPs, while the **red** dots are for the **TC** CMPs.

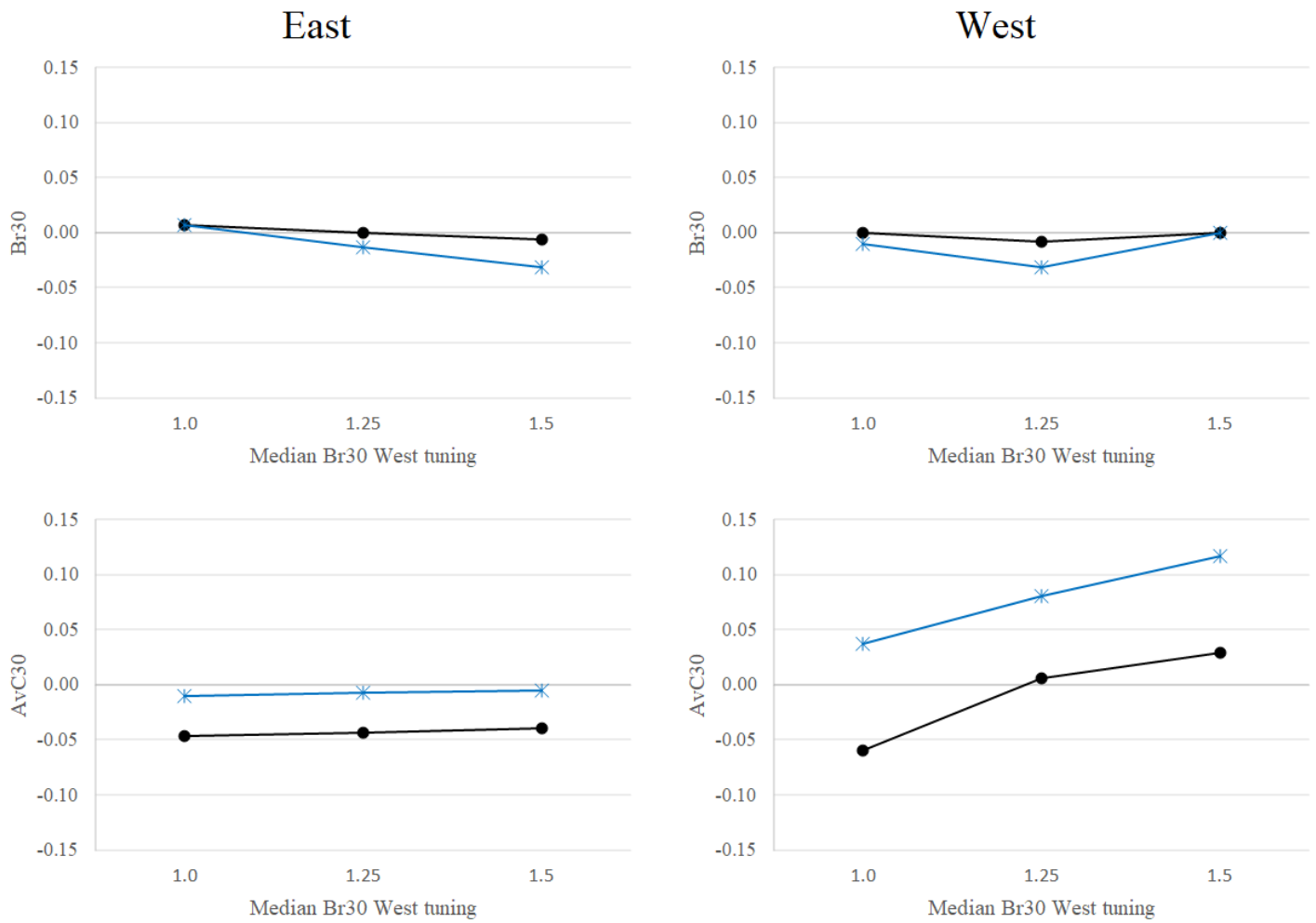


Figure 2: Relative differences in Br30 and AvC30 medians (black lines) and lower 5%iles (blue lines) for the East and the West between BR and TC CMPs (BR-TC) for each of the three median Br30 West tunings.

Appendix A

Figure A1 plots the absolute differences in Br30 and AvC30 medians, lower 5%iles and lowest values for the East and the West between BR and TC CMPs (BR-TC) for each of the three median Br30 West tunings.

The deterministic Br30 and AvC30 values under the BR and TC CMPs are compared in Figure 2 first for all OMs, and then for each recruitment scenario separately. These values are then compared for each of the 96 OMs of the interim grid separately and by pair of BR and TC CMPs, with the same West median Br30 tuning, in Figures A3 to A5.

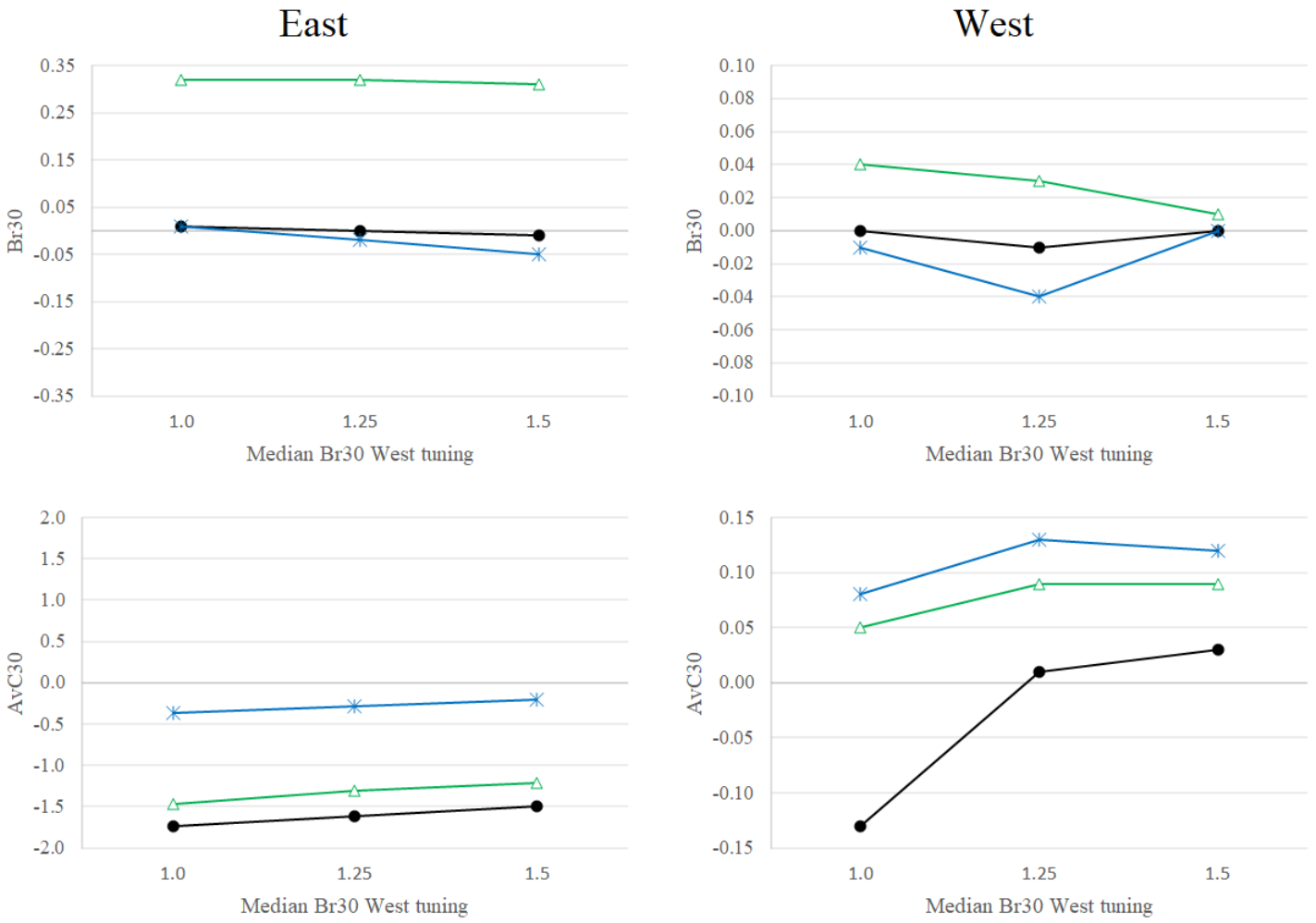
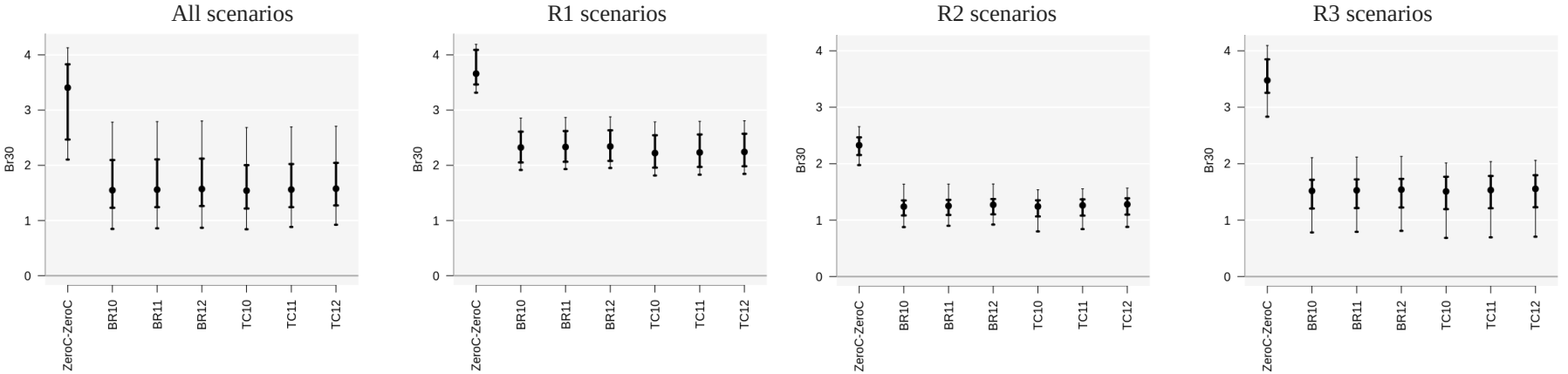


Figure A1. Absolute differences in Br30 and AvC30 medians (black lines), lower 5%iles (blue lines) and lowest values (green lines) for the East and the West between BR and TC CMPs (BR-TC) for each of the three median Br30 West tunings.

EAST



WEST

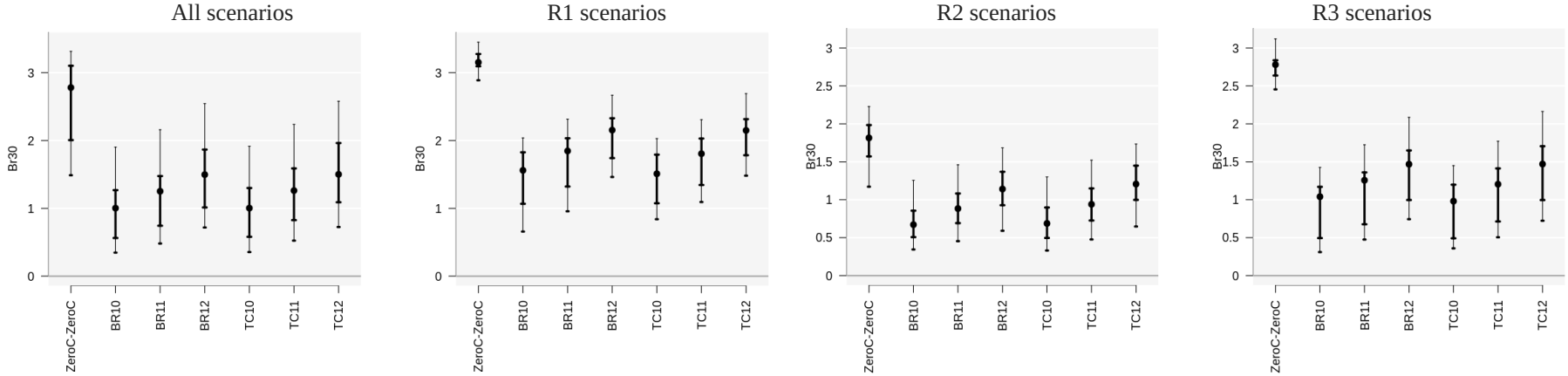
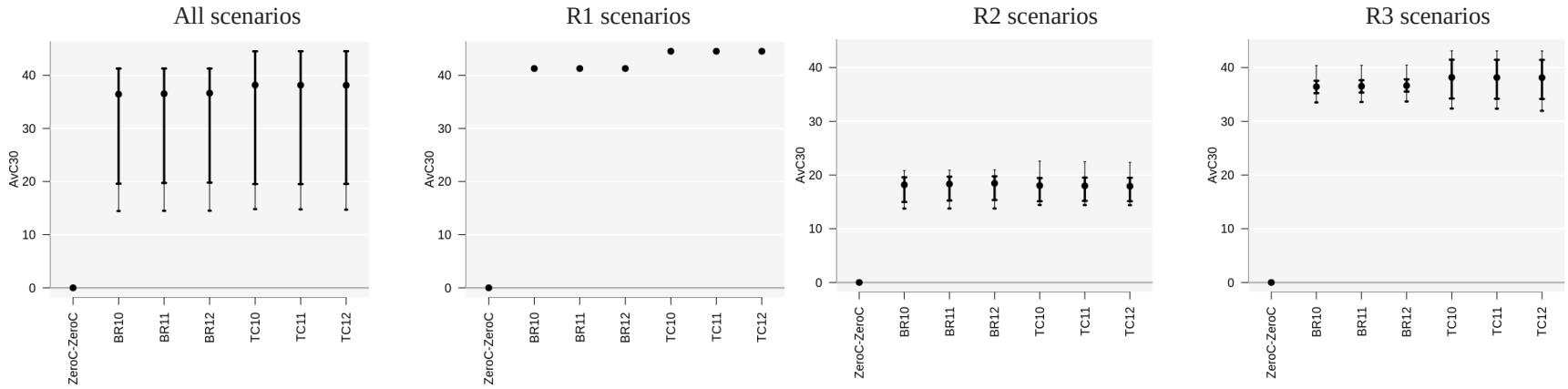


Figure A2a. Deterministic Br30 values for zero catch and the CMPs considered over the interim grid of OMs for CMPs BR10 to BR12 and TC10 to TC12 ("All scenarios"), and then for each of the recruitment scenarios separately, showing medians, and interquartile and 90%-ile ranges.

EAST



WEST

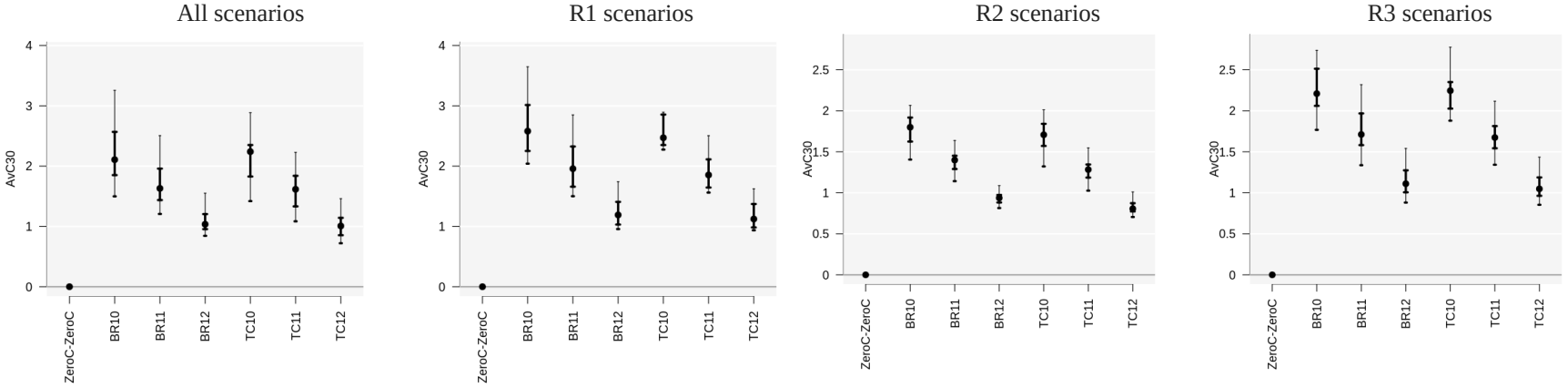
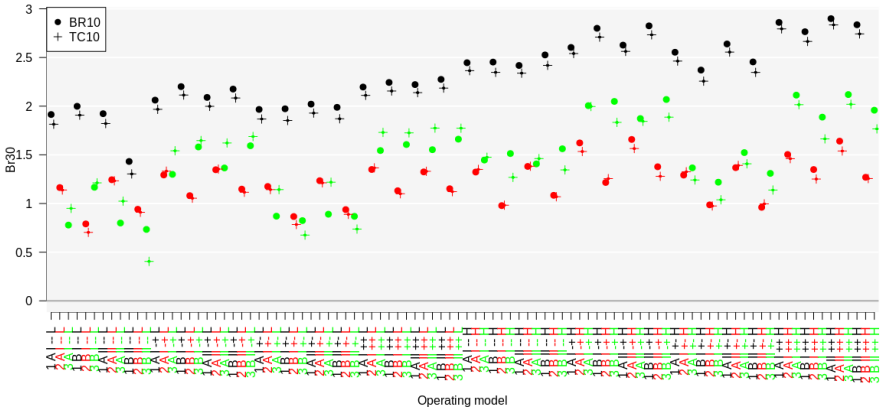


Figure A2b. Deterministic AvC30 values for zero catch and the CMPs considered over the interim grid of OMs for CMPs BR10 to BR12 and TC10 to TC12 ("All scenarios"), and then for each of the recruitment scenarios separately, showing medians, and interquartile and 90%-ile ranges.

Eastern stock



Western stock

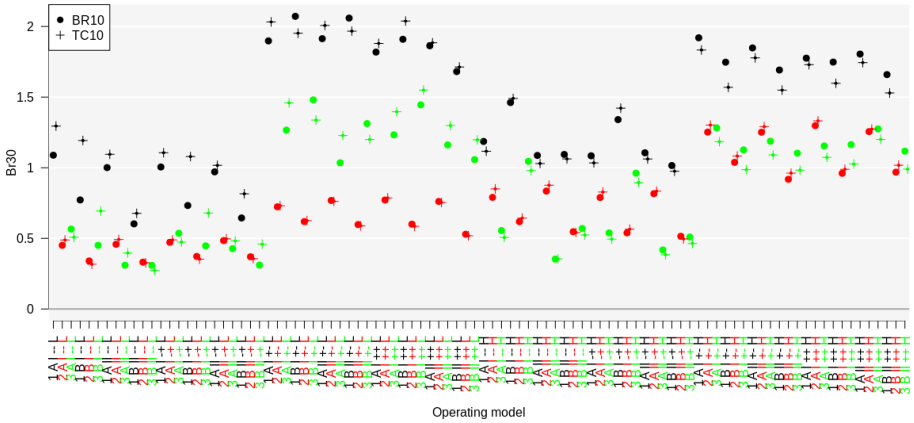
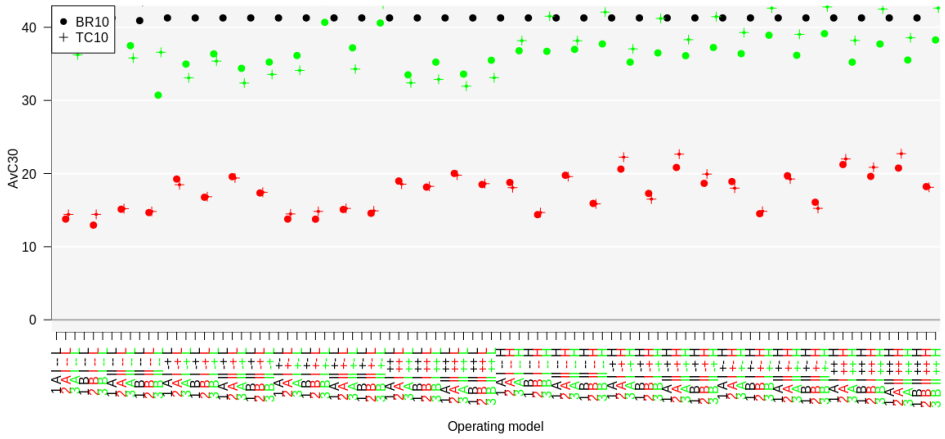


Figure A3a. Deterministic Br30 results for BR10 and TC10. The three colours correspond to the three recruitment scenarios: black, red and green to R1, R2 and R3 respectively.

Eastern stock



Western stock

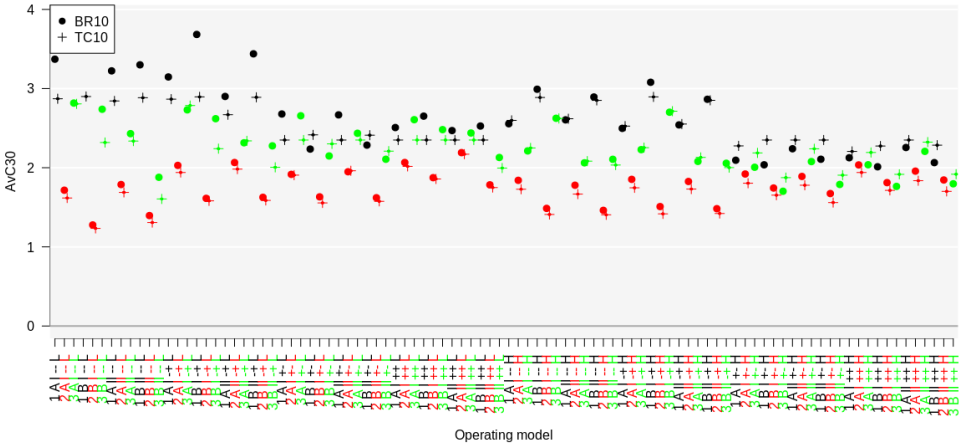
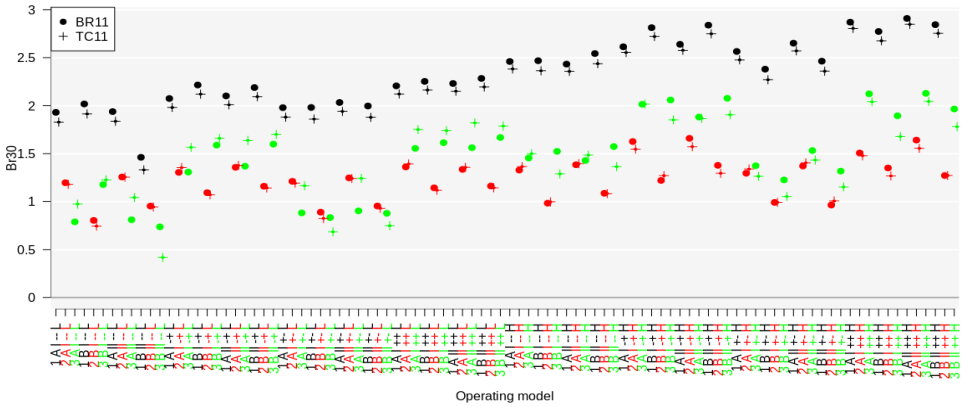


Figure A3b: Deterministic AvC30 results for BR10 and TC10. The three colours correspond to the three recruitment scenarios: black, red and green to R1, R2 and R3 respectively.

Eastern stock



Western stock

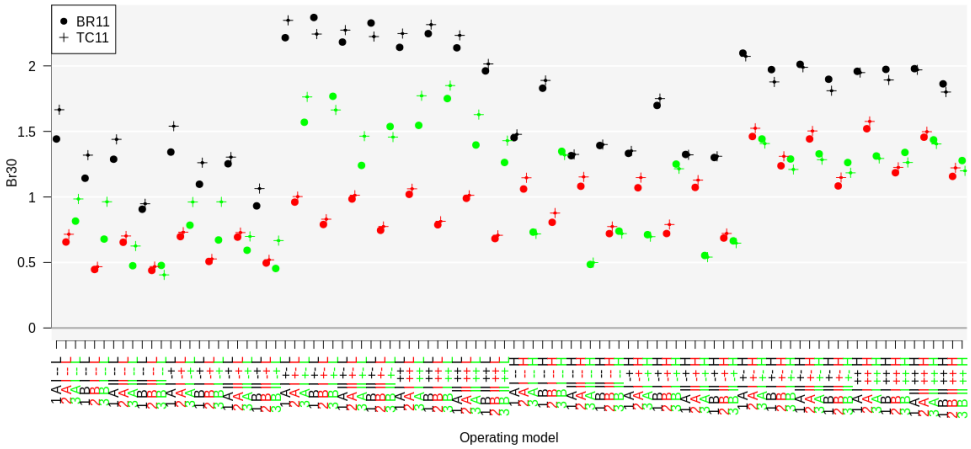
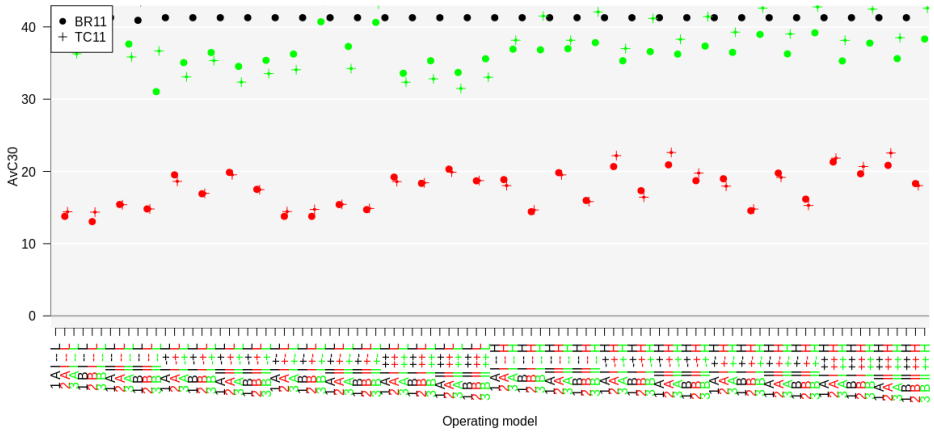


Figure A4a. Deterministic Br30 results for BR11 and TC11. The three colours correspond to the three recruitment scenarios: black, red and green to R1, R2 and R3 respectively.

Eastern stock



Western stock

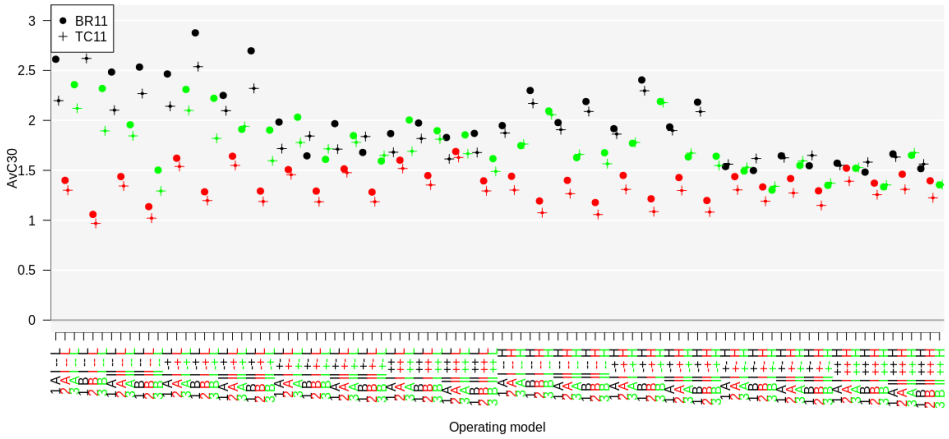
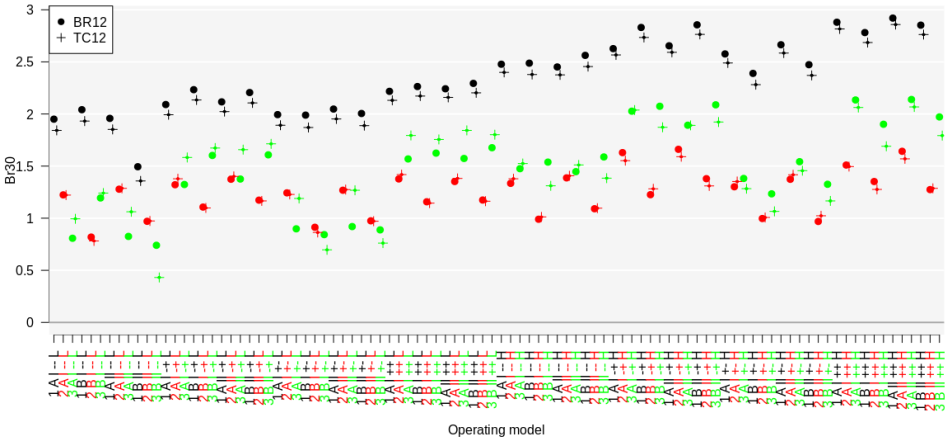


Figure A4b: Deterministic AvC30 results for BR11 and TC11. The three colours correspond to the three recruitment scenarios: black, red and green to R1, R2 and R3 respectively.

Eastern stock



Western stock

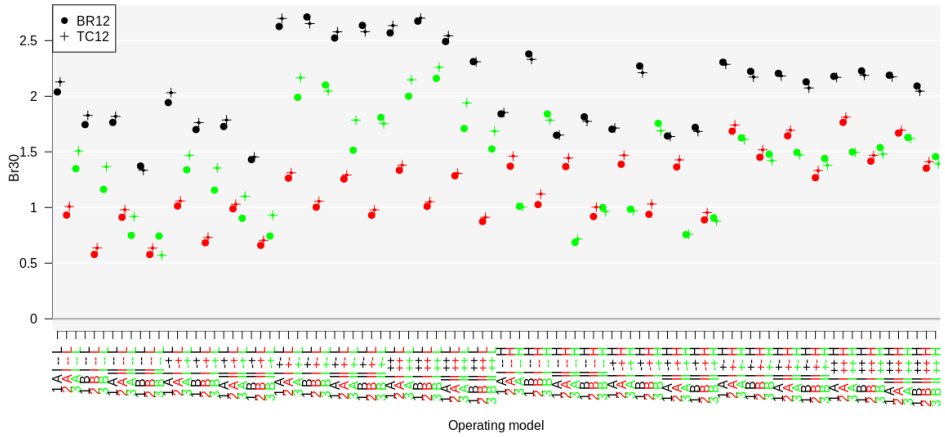
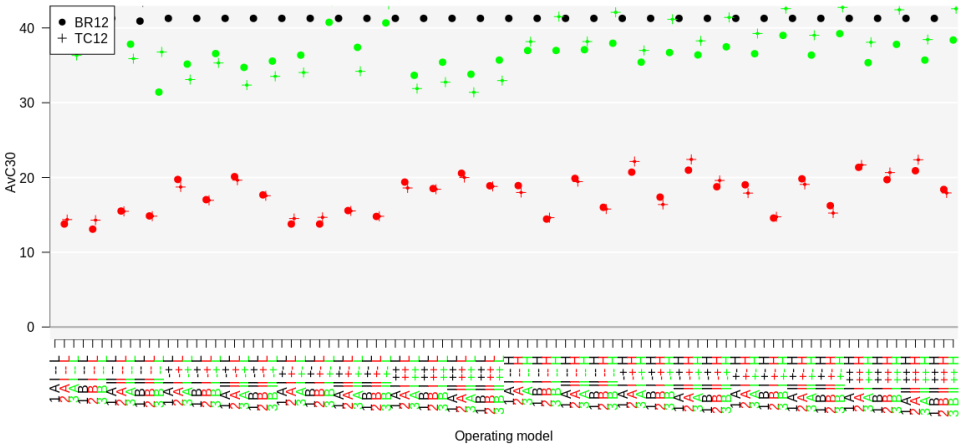


Figure A5a. Deterministic Br30 results for BR12 and TC12. The three colours correspond to the three recruitment scenarios: black, red and green to R1, R2 and R3 respectively.

Eastern stock



Western stock

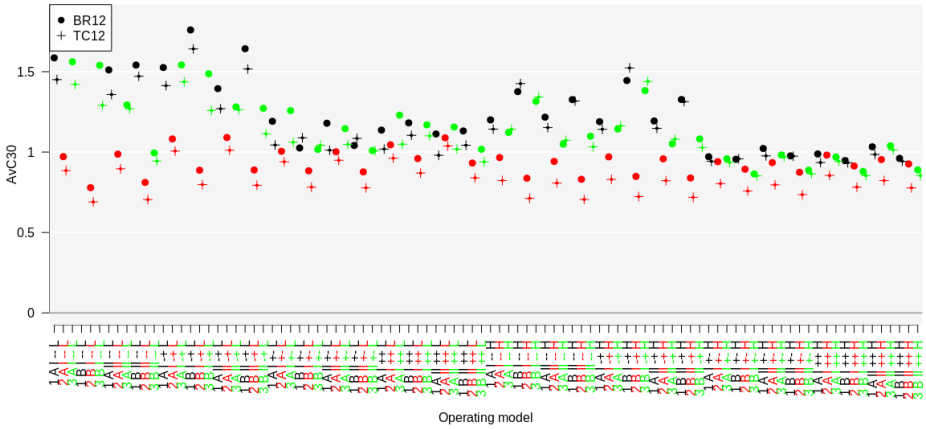


Figure A5b. Deterministic AvC30 results for BR12 and TC12. The three colours correspond to the three recruitment scenarios: black, red and green to R1, R2 and R3 respectively.