Response to Reviewer #1

We thank Reviewer #1 for his/her comments, which lead to valuable improvements in the manuscript. Below you will find the specific answers to the comments. In the following, the points and questions raised by the reviewer are reproduced in italic, text from the second version of the manuscript is shown in red, and changes in the updated version of the manuscript are shown in blue.

Authors have mostly satisfactorily addressed all questions in my previous review, I would like some more in-depth discussion on two points (see below). The extent to which the discussion is done it will depend on the necessary workload. Some points can be addressed with additional discussion; when numerical assessment is required it has to be done if it is easily achievable and not going to excessively delay publication.

Question 1 — "CMIP5 and CMIP6 models are analysed together, is there an impact on the results because of comparing models coming from different CMIPs?"

Authors explain that RCSM improvement is worse/less important when CMIP6 models are analysed. They state that lower mean biases from CMIP6 limits the importance of that improvement. This should be carefully considered as a conclusion that can come up is that RCSMs are maybe not necessary as CMIP6 already provides good guidance on MHW characteristics. If easily achievable, some numbers should be included by separating both CMIPS. Not strictly required but better clarification will strengthen the paper results and add robustness to their conclusions.

Additionally, a discussion on the feasibility of not using RCSM for semi-enclosed seas and coastal areas but only using CMIP6 GCMs would be a valuable addition.

Answer: We thank the reviewer for returning to this point. We would first like to clarify our previous response, which may have suggested a more systematic CMIP5-versus-CMIP6 distinction than our results actually support. We observe that RCSM added value for MHW duration tends to be larger when the driving GCM has a larger bias, which is expected by construction, since AVi also captures the extent to which the RCSM corrects a bias already present in the GCM. This relationship is not observed for MHW intensity, where added value appears instead to be governed by model-specific configuration factors, as discussed in Section 3.2.3. Given that our ensemble includes only 2 CMIP6-driven and 7 CMIP5-driven configurations, we do not think a separate quantitative breakdown by GCM generation would be statistically meaningful. We therefore prefer to frame this relationship in terms of GCM bias magnitude rather than GCM generation, which more accurately reflects what drives the pattern in our results and avoids overstating a CMIP5/CMIP6 distinction that our ensemble is too small to establish robustly. We have revised the corresponding paragraph in Section 4 accordingly (see manuscript change below).

Regarding the reviewer's broader concern that lower GCM bias could imply RCSMs are unnecessary where CMIP6 GCMs already perform well, we agree this deserves clarification. Even where the driving GCM's bias leaves limited room for area-averaged AVi, this does not imply that RCSMs are unnecessary: by design, RCSMs resolve local and regional-scale processes (coastal dynamics, strait circulation, semi-enclosed sea variability, local air-sea interactions) that GCMs, irrespective of generation, simply cannot represent at their native resolution. We have added a sentence to clarify this in Section 4 (see manuscript change below).

Manuscript change: we replaced the following paragraph in the Discussion:

"It is worth noting that CMIP6 GCMs tend to exhibit lower mean biases for both MHW intensity and duration relative to CMIP5 GCMs, which naturally reduces the room for improvement by the corresponding RCSMs in terms of mean bias correction. The reduced magnitude of the driving GCM bias partly conditions the RCSM ability to show added value in MHW duration, as discussed in Section 3.2.2, while no such systematic relation is found for MHW intensity, where added value is governed by model-specific configuration factors."

with:

"We note that RCSM added value for MHW duration tends to be larger when the driving GCM has a larger bias, which is expected by construction. This relationship is not observed for MHW intensity, where added value appears instead to be governed by model-specific configuration factors. This bias-dependence for MHW duration, in any case, concerns basin-averaged metrics and does not reflect an improved representation of the local and regional-scale processes (dynamics of coastal and semi-enclosed areas) that RCSMs are designed to resolve and where their added value remains concentrated."

Question 2 — For the comparison of the GCM vs RCSM, the authors calculate metrics and statistics in coincident grid points and exclude 'empty' points when only one of the GCM or RCSM has data. Have the authors analysed if there is a bias in the results because of excluding points with valid data?

The authors stated that "To prevent artificial negative AVi scores in these cases, those bins contribute zero to DRCSM, ensuring the RCSM adds value for capturing events missed by the GCM. If the RCSM fails to populate a bin where both the GCM and OBS show events, the same rule is applied symmetrically." This seems confusing to me or indeed contradictory with the new statement in their answer "For each model family, OBS and GCM were bilinearly regridded on the correspondent RCSM grid to facilitate a fair comparison between regional and global models. Interpolating GCM values to a finer resolution may produce unrealistic outputs, while interpolating OBS and RCSM values to a coarser resolution can compromise the representation of the PDF (Prein et al., 2016). However, to effectively illustrate the added value of downscaling,

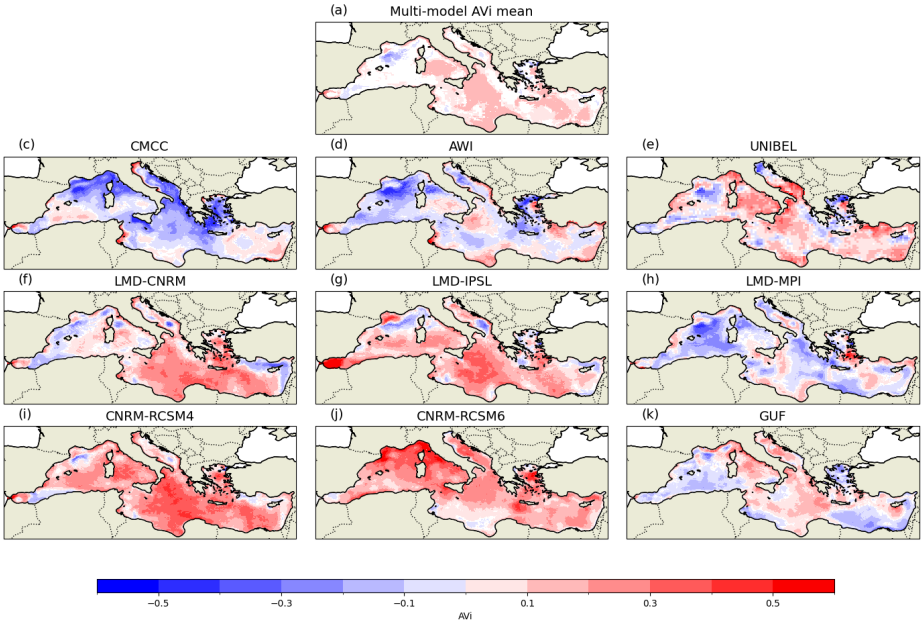
it is crucial to demonstrate that resolving local-scale processes provides more information than simply interpolating a coarse, non-resolving grid onto a finer grid." I still need a clarification on the methodology and procedures used.
I referred to the "empty" wording in lines 115-119, not figure 2.

We apologize for the confusion in our previous response. We had understood the original question as referring to the spatial regridding procedure rather than to the treatment of empty histogram bins in the AVi calculation (Section 2.2.2, lines 115–119), which is why our earlier answer focused on the interpolation methodology rather than on this specific point.

As a reminder, the correction addresses two symmetric cases: (1) when a bin is empty in the GCM but populated in both OBS and the RCSM, that bin contributes zero to D_{RCSM}, so the RCSM is not penalized for correctly capturing an event that the GCM entirely missed, while the GCM's own error for that bin is still counted normally; and (2) when a bin is empty in the RCSM but populated in both OBS and the GCM, that bin instead contributes zero to D_{GCM}, so the GCM is not penalized for a bin the RCSM failed to capture, while the RCSM's own error for that bin is still counted normally.

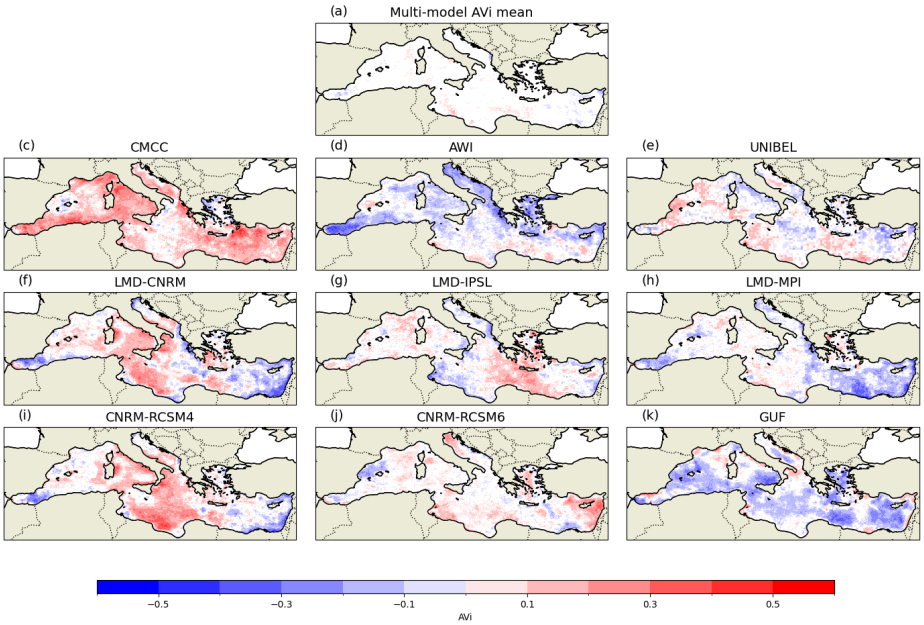
To directly address the reviewer's question, we compared AVi maps and area-averaged values computed with and without this correction applied, across all four variables considered in this study: MHW duration, MHW intensity, median-removed SST, and upper-tail SST. We can confirm that the correction does not affect the interpretation of our results in any way for any of these variables: it produces only small changes in AVi magnitude, irrespective of sign, and leaves model rankings and the sign of AVi unchanged at every location.

SST (median-removed and upper-tail), version with no correction:



l)

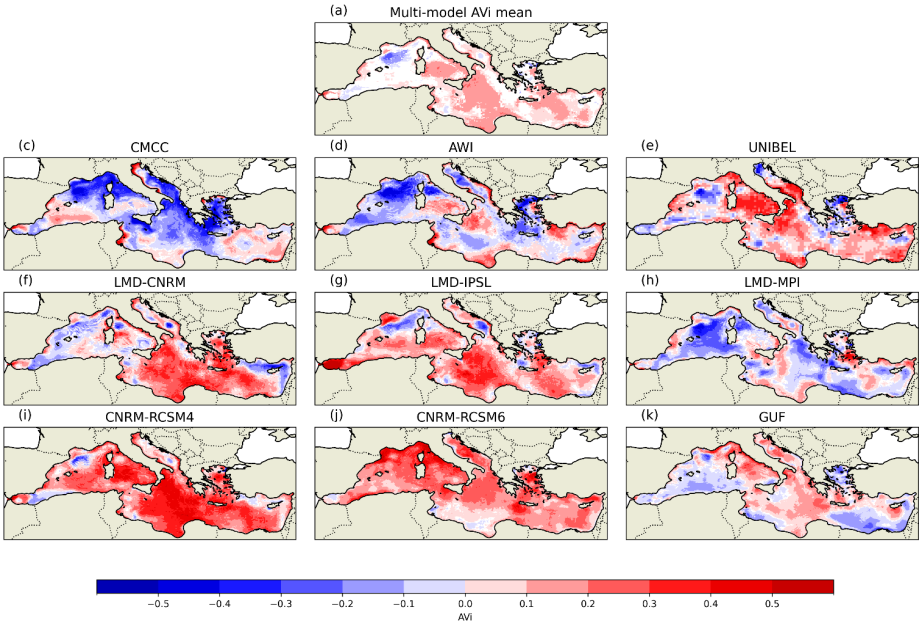
Model	Area-avg AVi	% pos (% neg)
CNRM-RCSM4	+0.20	91% (6%)
CNRM-RCSM6	+0.19	93% (4%)
LMD-IPSL	+0.14	83% (12%)
LMD-CNRM	+0.11	73% (20%)
UNIBEL	+0.08	69% (24%)
GUF	+0.02	51% (39%)
AWI	-0.02	32% (59%)
LMD-MPI	-0.04	30% (59%)
CMCC	-0.09	25% (66%)



l)

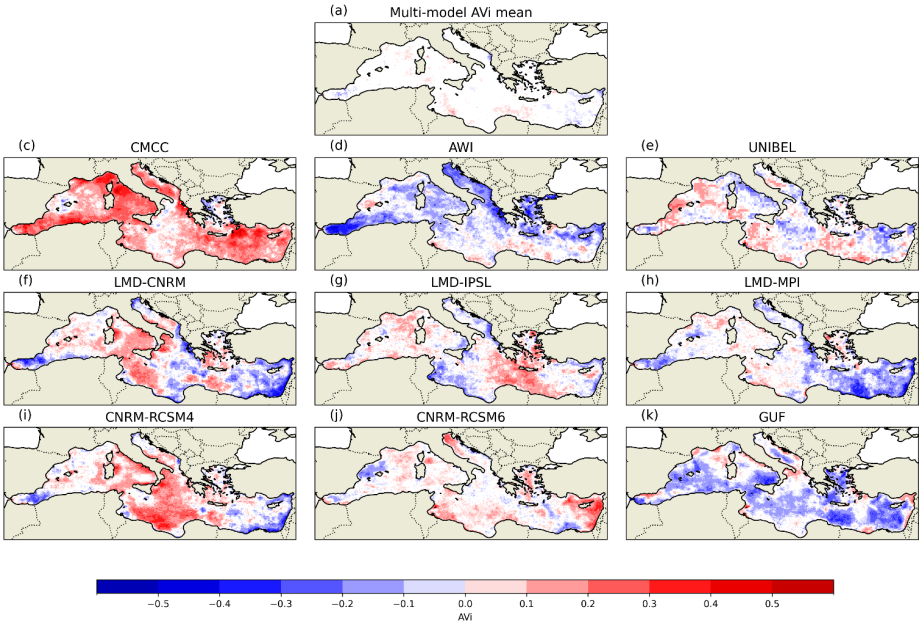
Model	Area-avg AVi	% pos (% neg)
CMCC	+0.12	77% (4%)
CNRM-RCSM4	+0.05	49% (16%)
LMD-CNRM	+0.02	45% (26%)
CNRM-RCSM6	+0.02	35% (15%)
LMD-IPSL	+0.02	37% (20%)
UNIBEL	+0.01	30% (25%)
LMD-MPI	-0.04	15% (38%)
AWI	-0.06	9% (53%)
GUF	-0.07	14% (56%)

SST (median-removed and upper-tail), version <u>with correction</u>:



l)

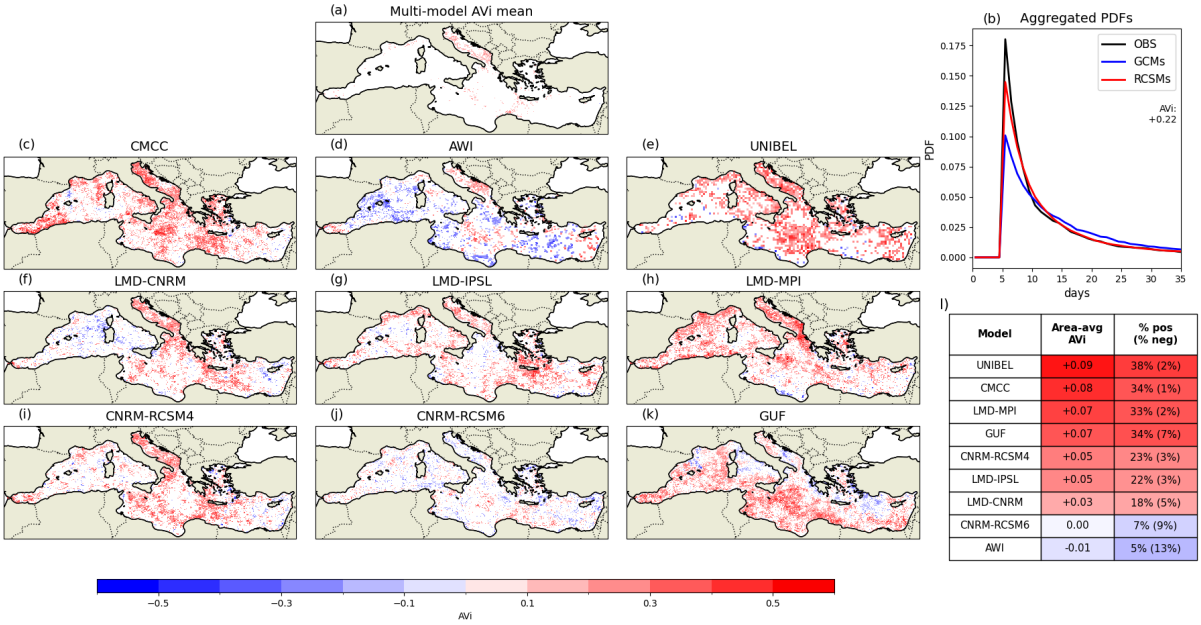
Model	Area-avg AVi	% pos (% neg)
CNRM-RCSM4	+0.22	91% (6%)
CNRM-RCSM6	+0.18	93% (3%)
LMD-IPSL	+0.13	80% (14%)
LMD-CNRM	+0.11	68% (24%)
UNIBEL	+0.10	71% (23%)
GUF	+0.03	54% (37%)
AWI	-0.02	34% (57%)
LMD-MPI	-0.04	32% (58%)
CMCC	-0.08	30% (62%)



l)

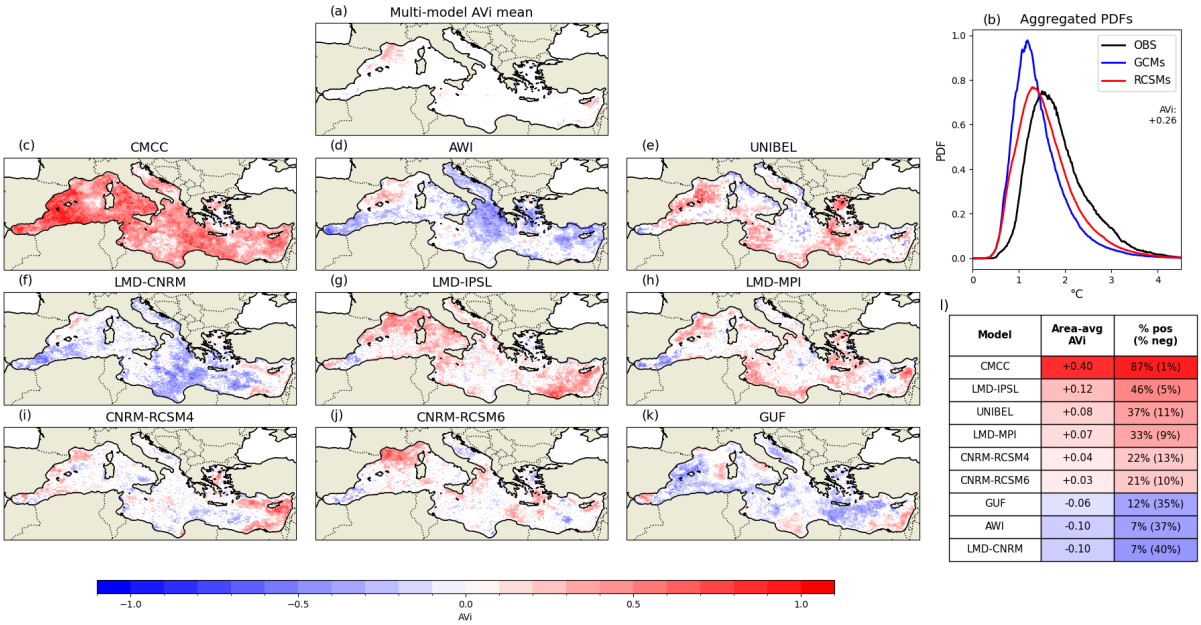
Model	Area-avg AVi	% pos (% neg)
CMCC	+0.14	81% (3%)
CNRM-RCSM4	+0.05	48% (17%)
CNRM-RCSM6	+0.02	36% (15%)
LMD-IPSL	+0.02	37% (21%)
UNIBEL	0.00	30% (27%)
LMD-CNRM	0.00	38% (31%)
LMD-MPI	-0.04	15% (38%)
GUF	-0.07	14% (56%)
AWI	-0.07	8% (57%)

MHW duration and intensity, version with <u>no correction</u>:



l)

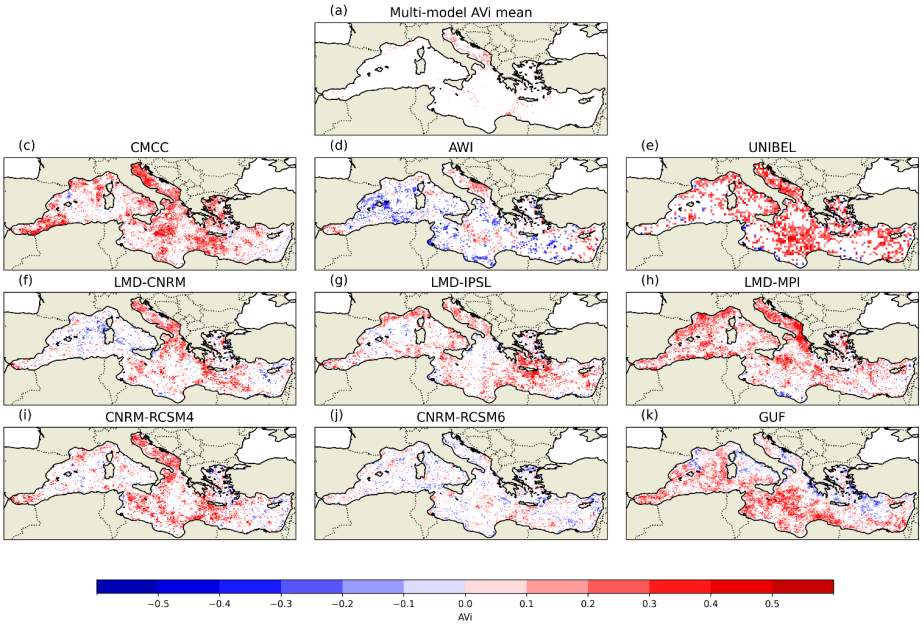
Model	Area-avg AVi	% pos (% neg)
UNIBEL	+0.09	38% (2%)
CMCC	+0.08	34% (1%)
LMD-MPI	+0.07	33% (2%)
GUF	+0.07	34% (7%)
CNRM-RCSM4	+0.05	23% (3%)
LMD-IPSL	+0.05	22% (3%)
LMD-CNRM	+0.03	18% (5%)
CNRM-RCSM6	0.00	7% (9%)
AWI	-0.01	5% (13%)



l)

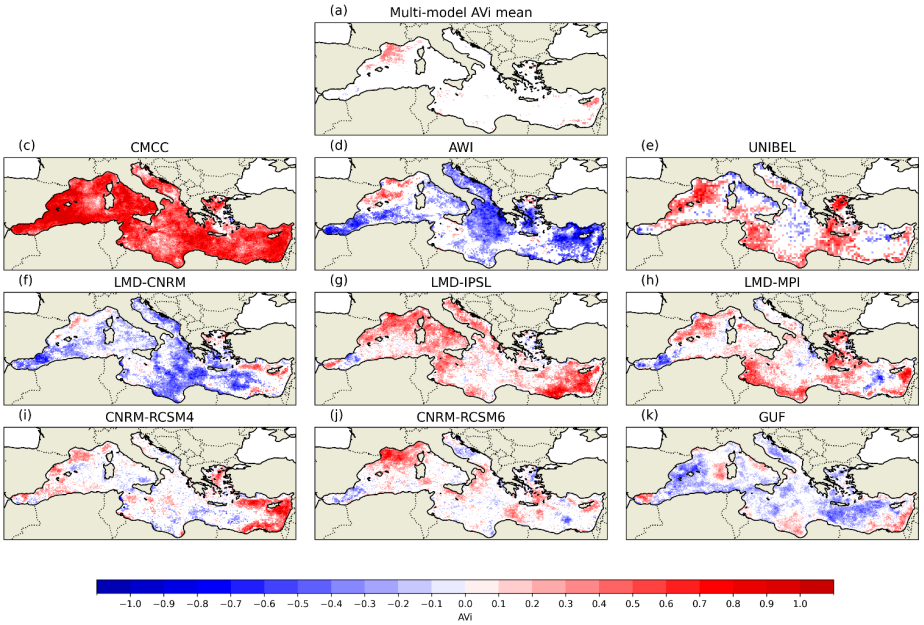
Model	Area-avg AVi	% pos (% neg)
CMCC	+0.40	87% (1%)
LMD-IPSL	+0.12	46% (5%)
UNIBEL	+0.08	37% (11%)
LMD-MPI	+0.07	33% (9%)
CNRM-RCSM4	+0.04	22% (13%)
CNRM-RCSM6	+0.03	21% (10%)
GUF	-0.06	12% (35%)
AWI	-0.10	7% (37%)
LMD-CNRM	-0.10	7% (40%)

MHW duration and intensity, version with correction:



l)

Model	Area-avg AVi	% pos (% neg)
UNIBEL	+0.09	36% (2%)
CMCC	+0.08	32% (1%)
LMD-MPI	+0.07	32% (2%)
GUF	+0.05	29% (8%)
CNRM-RCSM4	+0.05	22% (3%)
LMD-IPSL	+0.04	21% (3%)
LMD-CNRM	+0.03	17% (5%)
CNRM-RCSM6	0.00	7% (8%)
AWI	-0.01	6% (12%)



l)

Model	Area-avg AVi	% pos (% neg)
CMCC	+0.57	94% (0%)
LMD-IPSL	+0.17	50% (4%)
UNIBEL	+0.12	38% (10%)
LMD-MPI	+0.10	37% (8%)
CNRM-RCSM4	+0.06	23% (10%)
CNRM-RCSM6	+0.04	22% (10%)
GUF	-0.06	12% (35%)
LMD-CNRM	-0.13	6% (44%)
AWI	-0.18	6% (43%)

While carrying out this comparison, we also identified and corrected a minor inconsistency in our implementation: case (2) above had not been consistently applied in our original calculation. We corrected the implementation accordingly, and the figures above reflect this corrected version. We have updated the relevant numbers in the main text and figures accordingly, but no changes to the scientific interpretation, conclusions, or discussion were necessary.

We have accordingly added a sentence in Section 2.2.2 to report this robustness check:

Manuscript change:
"Following Ciarlo et al. (2021), we apply a correction for bins that are empty in the GCM but populated in OBS and RCSM. To prevent artificial negative AVi scores in these cases, those bins contribute zero to D_{RCSM} ensuring the RCSM adds value for capturing events missed by the GCM. If the RCSM fails to populate a bin where both the GCM and OBS show events, the same rule is applied symmetrically. Sensitivity tests confirmed that AVi values are not substantially affected by this correction; in particular, area-averaged AVi rankings and signs across models are preserved for all metrics."

Suggested (not requested) references for the state of the art of MHW in the Mediterranean and on related atmospheric processes

- *Martínez, J. et al. Evolution of marine heatwaves in warming seas: the Mediterranean Sea case study. Front Mar Sci 10, (2023).*
- *Rosselló, P., Pascual, A. & Combes, V. Assessing marine heat waves in the Mediterranean Sea: a comparison of fixed and moving baseline methods. Front Mar Sci 10, (2023).*
- *Paredes-Fortuny, L., Pastor, F. & Khodayar, S. Concurrent atmospheric heatwaves intensify marine heatwaves through air-sea heat flux change in the Mediterranean Sea. Commun Earth Environ 6, (2025).*
- *Pastor, F., Paredes-Fortuny, L. & Khodayar, S. Mediterranean marine heatwaves intensify in the presence of concurrent atmospheric heatwaves. Commun Earth Environ 5, (2024).*

We thank the reviewer for these suggested references. We reviewed all four suggested references and found Martínez et al. (2023) and Paredes-Fortuny et al. (2025) to be the most directly relevant to specific points already discussed in the manuscript. We have added Martínez et al. (2023) to the sentence discussing the importance of mean-state representation for MHW characterization, in Section 3.2.3 (lines 316-319):

"Accurate representation of mean-state evolution is indeed essential for correctly characterizing present and future MHW properties (Frölicher et al., 2018; Schlegel et al., 2019; Oliver et al., 2021; Martínez et al., 2023)."

We have also added Paredes-Fortuny et al. (2025) to the paragraph discussing atmospheric drivers of MHWs in Section 4:

"Recent work has shown that Mediterranean MHWs are primarily driven by anomalies in surface heat fluxes, in particular reduced latent heat loss associated with weak winds, enhanced near-surface humidity, and reduced cloud cover associated with enhanced surface downwelling radiation (Bonino et al., 2025; Garfinkel et al., 2026; Paredes-Fortuny et al., 2025)."