Responses to Reviewer 1

Review of the manuscript titled "Sensitivity of marine heatwaves metrics to SST products, focusing on the Tropical Pacific" by Carla Chevillard and co-authors.

General comments:

MHW detection and characterization (metrics) still have some open issues in their settings and criteria, such as the use or not of detrended SST data, baseline climatology, spatio-temporal constraints… In this work, the authors run an extensive analysis of the impact of the SST database selection on MHW detection and metrics in the Tropical Pacific Area.

The authors analyse a complete set of MHW metrics over different areas of the Pacific Ocean and different satellite and reanalysis SST datasets. They find differences between calculated metrics characteristics, regional differences and metrics dispersion depending on the selected dataset. They also analyse the temporal evolution of regional averaged MHW metrics.

The results in the manuscript show that for different metrics, the best results are observed with different databases. In the same direction, different MHW sizes yield different results depending on the SST used. They also observe that the variability of results between different databases has been decreasing in recent years. No clear distinction is obtained between one database and the others, nor is there one that obtains a better result in most metrics.

Work shown in the manuscript is methodologically consistent and provides interesting results on the impact of SST databases in MHW analysis. My recommendation is to publish the manuscript with minor revisions.

>> We thank the reviewer very much for his/her careful reading, valuable comments and suggestions that helped us to improve the manuscript. All comments have been considered, as detailed below. Numbers of lines will refer to the revised version of the manuscript with tracked changes.

Main comments:

- My main concern comes from the methodology choice in detecting all pixels constituting a MHW in the 2.2.3- Filtering MHWs by size section. This is an important issue, as the authors separate MHWs in micro and macro scales, which needs to be better explained. Please, provide more details and the rationale on how "joint pixels" are detected.

>> We agree that this point needs to be more detailed. Joint pixels are pixels connected along either north–south or west–east directions. They are detected thanks to the label function from python package scipy.ndimage, which calculates the connectivity of features to their neighbors, based on a structuring element matrix establishing the directions in which the connectivity is defined (in our case, along north–south or west–east directions).

For each day, we define a 2D map of pixels associated with 0 (non MHW state) or 1 (MHW state), and then we compute the spatially connected pixels thanks to this label function. This method is inspired from Bonino et al. 2023.

We consequently obtain snapshots of MHWs spatial structures at each time step. In a given pixel, a MHW is thus associated with a daily time series of sizes over its duration. We worked with the maximum size of a MHW reached over its duration to separate the micro and macro scale events.

Some of these precisions were added in the manuscript L.158-159 to complete section 2.2.3:

"The sensitivity analysis was carried out for micro (maximum area ≤ 5°lonx5°lat) and macro (maximum area > 5°lonx5°lat) events, separately. MHWs spatial extent was characterised as follows : for each day, all pixels where MHWs were detected were assigned a MHW area, defined as the number of contiguous pixels to the studied pixel also experiencing a MHW. These joint pixels are connected along either north-south or west-east directions and were detected thanks to the label function from python package scipy.ndimage (method inspired from Bonino et al., 2023)."

- Which is the impact of the methodology (based on Bonino) on MHW detection? Have you tried any other methodology? Please, see the references below (global and Mediterranean scales) and discuss why you chose the methodology in Bonino, used in the Mediterranean where scales are much smaller than in the Pacific.

Sun, D., Jing, Z., Li, F. & Wu, L. Characterizing global marine heatwaves under a spatio-temporal framework. Prog. Oceanogr. 211, 102947 (2023).

Pastor, F., Paredes-Fortuny, L. & Khodayar, S. Mediterranean marine heatwaves intensify in the presence of concurrent atmospheric heatwaves. Communications Earth & Environment 5, 797 (2024).

>> We thank the reviewer for this relevant remark.

First, methods inspired from Bonino et al. (2023) were used in other papers studying MHWs in the Pacific such as in Lal et al. (2025), where the authors focus on the impacts of MHWs in the PICTs (South Pacific Island countries). Similar methods where the spatial extent of MHWs is computed by ad hoc aggregating adjacent grid cells with simultaneous SST above a prescribed threshold were also used in Frolicher et al. (2018) and Sen Gupta et al. (2020). We thus decided to follow the method from Bonino et al. 2023 to be in line with these references and to allow a better comparison with our results.

Then, trying another methodology like the ones presented in Sun et al. (2023) or Pastor et al. (2024), which also better address the evolution in time of MHWs, would be interesting and add valuable information for the MHW community. Yet, the focus of our paper was on the sensitivity of MHW metrics to SST products, so we decided to keep our focus on this variability and didn't test any other methodology for calculating MHW spatial extent. Yet, the reviewer's remark was added in section 4.4 as a perspective to our study L.642-643 :

"Choosing other thresholds for MHW detection to focus on the most extreme events (e.g 98th percentile) might also affect the observed inter-product differences. In the same way, other methodologies for characterising MHWs spatial extent (Sun et al. 2023, Pastor et al. 2024) might influence our results."

- Although you mention a possible impact of re-gridding in the MHW analysis. Have you checked the impact of re-gridding in the dataset characteristics? Some simple statistics, correlations… of this impact would be interesting to be included in the manuscript, maybe as supplementary material.

>> We fully agree with the reviewer and added complementary analyses in the present document as well as in the Supplementary Information of the manuscript to answer his/her comments. Also, the reviewer remark being in line with Reviewer 2 specific comment, this paragraph aims to answer both reviewers comments on the impact of re-gridding in our study.

The impact of re-gridding on our results was investigated for C3S, CRW and OSTIA, which were all three re-gridded from 0.05° to 0.25°, and for GLORYS which was re-gridded from 0.08° to 0.25°. To investigate the impact of re-gridding, we started to analyse simple statistics on SST time series between raw datasets (0.05° and 0.08° for GLORYS) and re-gridded datasets (0.25°), as suggested by Reviewer 1. The main characteristics of SST time series standard deviation - the spatial minimum, maximum, mean and percentile 90 - were computed inside the seven regions studied, for the raw and the re-gridded datasets (Table R1.1). Very little differences were observed between the two datasets for these characteristics, except for the maximum of standard deviation in the NPTG and PNEC. Yet, if the re-gridding induced a slight decrease in the maximum of standard deviation (certainly due to the smoothing of time series when computing the spatial mean of SST within a 0.25°x0.25° pixel), the order of magnitude in the products remained the same (meaning the largest standard deviation maximum or minimum remains the same for both raw and re-gridded datasets). This analysis suggests that the re-gridding has little impact on our results since we focus on interproducts comparison.

RAW / REGRID	STD MIN	STD MAX	STD MEAN	STD PERC 90
NPSW				
C3S	0.77 / 0.78	2.81 /2.79	1.44 / 1.43	2.13 / 2.13
CRW	0.73 / 0.73	2.80 / 2.78	1.42 / 1.42	2.13 / 2.13
OSTIA	0.74 / 0.75	2.81 / 2.79	1.42 / 1.42	2.13 / 2.13
GLORYS	0.69 / 0.70	2.93 / 2.82	1.41 / 1.41	2.14 / 2.14
NPTG				
C3S	0.79 / 0.78	4.74 / 4.64	1.29 / 1.29	1.71 / 1.71
CRW	0.75 / 0.75	4.46 / 4.35	1.26 / 1.26	1.70 / 1.70
OSTIA	0.76 / 0.76	4.66 / 4.54	1.28 / 1.28	1.71 / 1.71
GLORYS	0.71 / 0.71	4.43 / 4.43	1.30 / 1.30	1.73 / 7.73

WARM				
C3S	0.41 / 0.41	1.21 / 1.20	0.74 / 0.73	0.94 / 0.93
CRW	0.45 / 0.45	1.16 / 1.16	0.71 / 0.71	0.90 / 0.90
OSTIA	0.33 / 0.36	1.15 / 1.15	0.72 / 0.71	0.91 / 0.91
GLORYS	0.88 / 0.88	1.25 / 1.19	0.67 / 0.67	0.88 / 0.88
PNEC				
C3S	0.68 / 0.69	4.20 / 4.04	1.06 / 0.06	1.43 / 1.45
CRW	0.62 / 0.62	3.92 / 3.85	1.02 / 1.02	1.33 / 1.33
OSTIA	0.51 / 0.51	4.07/3.97	1.03 / 1.03	1.38 / 1.39
GLORYS	0.67 / 0.68	4.16 / 3.96	1.04 / 1.04	1.37 / 1.38
PEQD				
C3S	0.56 / 0.56	3.10 / 3.10	1.56 / 1.56	2.63 / 2.63
CRW	0.56 / 0.56	3.10 / 3.10	1.53 / 1.53	2.55 / 2.55
OSTIA	0.53 / 0.53	3.07 / 3.06	1.53 / 1.53	2.56 / 2.56
GLORYS	0.51 / 0.51	3.06 / 2.94	1.48 / 1.48	2.47 / 2.47
SPSG				
C3S	0.77 / 0.76	3.16 / 3.15	1.56 / 1.56	2.13 / 2.12
CRW	0.77 / 0.77	3.05 / 3.05	1.54 / 1.54	2.12 / 2.12
OSTIA	0.71 / 0.73	3.08 / 3.07	1.54 / 1.54	2.11 / 2.11
GLORYS	0.75 / 0.75	2.94 / 2.93	1.53 / 1.53	2.10 / 2.10
ARC				
C3S	0.58 / 0.59	3.31 / 3.30	1.60 / 1.60	1.92 / 1.92
CRW	0.57 / 0.57	3.27 / 3.27	1.58 / 1.58	1.93 / 1.94
OSTIA	0.44 / 0.45	3.27 / 3.23	1.58 / 1.57	1.93 / 1.93
GLORYS	0.59 / 0.60	3.44 / 3.41	1.62 / 1.62	1.97 / 1.97

Table R1.1 : spatial minimum, maximum, mean and 90th percentile of the standard deviation of SST time series inside each region for the raw (0.05° and 0.08° for GLORYS) and re-gridded (0.25°) datasets of C3S, CRW, OSTIA and GLORYS.

Then, MHW statistics were computed from the raw datasets at 0.05° resolution (0.08° for GLORYS) and compared to the results of the 0.25° dataset for C3S, CRW, OSTIA and GLORYS, inside two small areas of study : one in the ARC subregion corresponding to the area around New Caledonia (155E - 175E ; 14S - 27S) and one in the Warm Pool subregion (150E-170E ; 0-10N). We focused on small areas as computing MHW statistics at 0.05° resolution over the Tropical Pacific was very costly in terms of computation. Results of MHW detection for the raw and re-gridded datasets are illustrated in Fig. R1.1 (a-h) for the area around New Caledonia. A zoom over the southern coast of New Caledonia was also illustrated in Fig. R1.2 to better illustrate the differences between the datasets. Fig. R1.1 and R1.2 illustrate the results for the metric duration only.

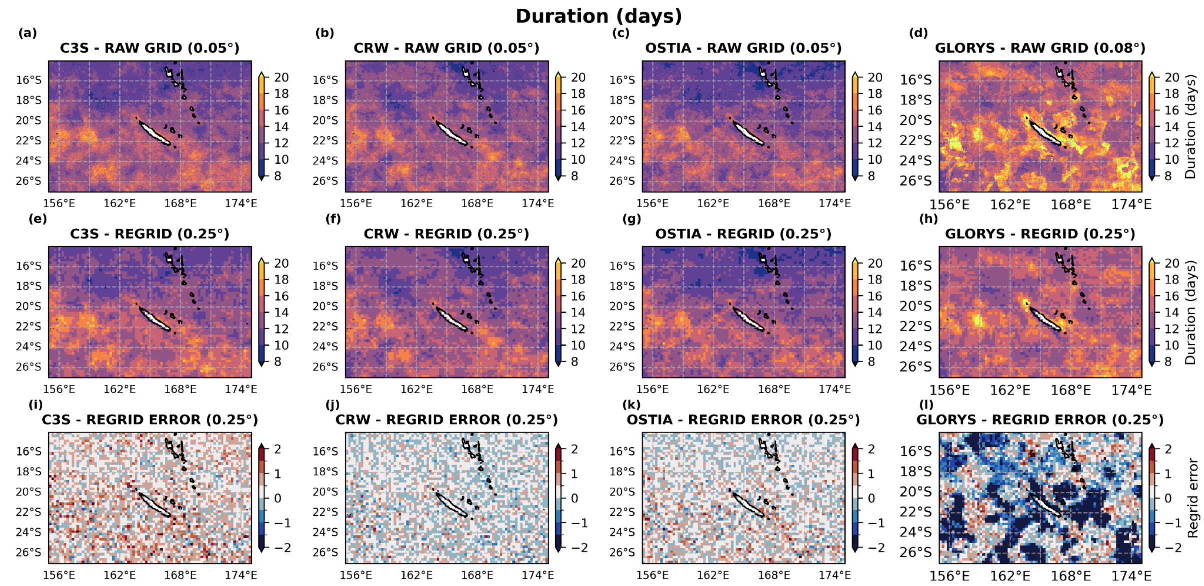

Figure R1.1: (a-d) Mean duration of MHWs over 1993-2021 for C3S (a), CRW (b), OSTIA (c) and GLORYS (d) at 0.05° resolution and 0.08° resolution for GLORYS. (e-h) Mean duration of MHWs over 1993-2021 for C3S (e), CRW (f), OSTIA (g) and GLORYS (h) at 0.25° resolution (i-l) Re-gridding error as defined in the text for C3S (i), CRW (j), OSTIA (k) and GLORYS (l).

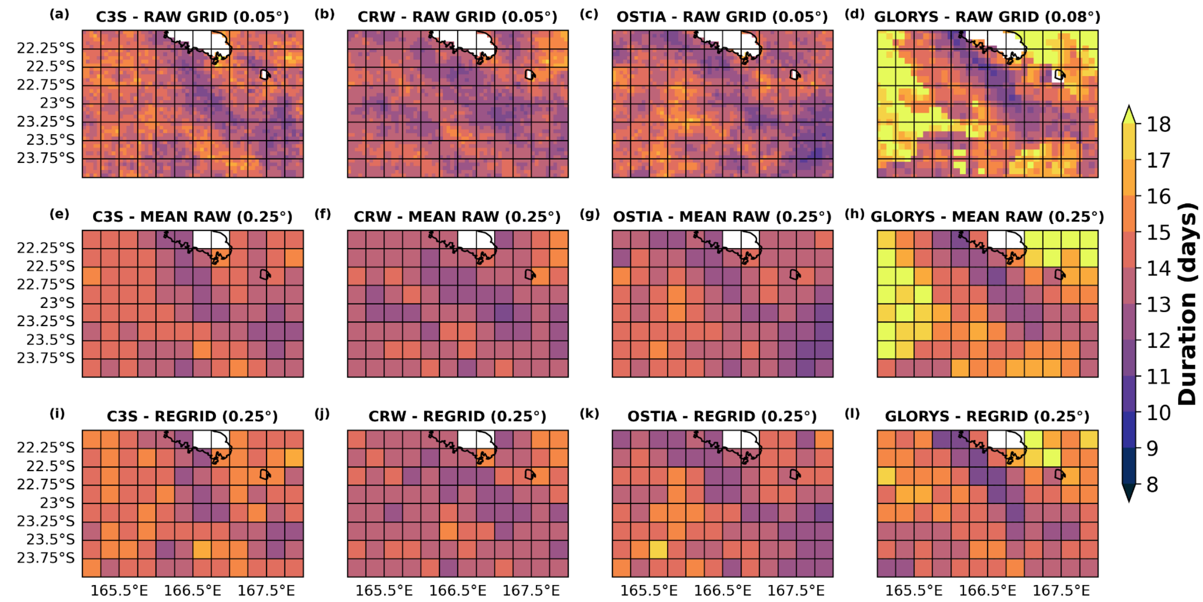

Figure R1.2: Zoom of Fig. R1.1 on the southern coast of New Caledonia. (a-d) Mean duration of MHWs over 1993-2021 for C3S (a), CRW (b), OSTIA (c) at 0.05° resolution and at 0.08° resolution for GLORYS (d). (e-h) Mean of MHW statistics of the RAW grid inside each pixel of the 0.25° grid (i-l) Mean duration of MHWs over 1993-2021 for C3S (i), CRW (j), OSTIA (k) and GLORYS (l) computed from the re-gridded datasets (0.25° resolution).

As one pixel of the re-gridded datasets corresponds to 25 pixels of the raw datasets, we associated each pixel of the re-gridded dataset to its 25 pixels in the raw grid and computed the mean and standard deviation of MHWs statistics inside these 25 pixels. The mean statistics of the 0.05° grid inside the 0.25° grid were represented in Fig. R1.2 (e-h).

To quantify the impact of re-gridding, we computed the re-gridding error as well as the information lost in the re-gridding process.

The re-gridding error was defined as the difference between MHWs statistics computed from the SST regridded datasets and the mean of the 25 associated MHW statistics of the raw grid inside the 0.25° grid as defined in Fig. R1.2 (e-h). It is illustrated in panels (i-l) of Fig. R1.1 for the metric duration. These panels actually represent the difference between Fig. R1.2 (i-l) and Fig. R1.2 (e-h), over the whole area of New Caledonia.

If there is no clear overestimation or underestimation due to the re-gridding for the metric duration for C3S, CRW and OSTIA, the re-gridding appears to underestimate the duration of MHWs for GLORYS. The metrics maximum intensity and onset rate clearly show lower values in the re-gridded datasets compared to the raw dataset for all products (negative values over the whole area, not shown), and the MHW days per year shows higher values (especially for C3S, not shown) in the re-gridded dataset. This result is in line with the fact that the re-gridding induces a smoothing of the SST signal and thus will detect more MHW days and lower maximum intensities and onset rates. The re-gridding error is also more spatially variable for GLORYS compared to the other products, probably due to the nature of the product (GLORYS is a reanalysis and includes finer scale processes).

Then, the re-gridding error was expressed in percentage by dividing it by the mean value of the statistics of the 0.05° grid inside the 0.25° grid, to give the re-gridding error in percentage (Fig. R1.3 a,b,c,d). In the same way we computed the mean of the 25 values of MHW statistics of the raw grid inside the 0.25° grid to define the re-gridding error, we computed the standard deviation of the 25 values of MHW statistics of the raw grid inside the 0.25° grid. Then, we divided it by the mean value of the statistics of the 0.05° grid inside the 0.25° grid. This quantity was defined as the information lost during the re-gridding process and is illustrated in Fig. R1.3 (e,f,g,h) for the metric duration.

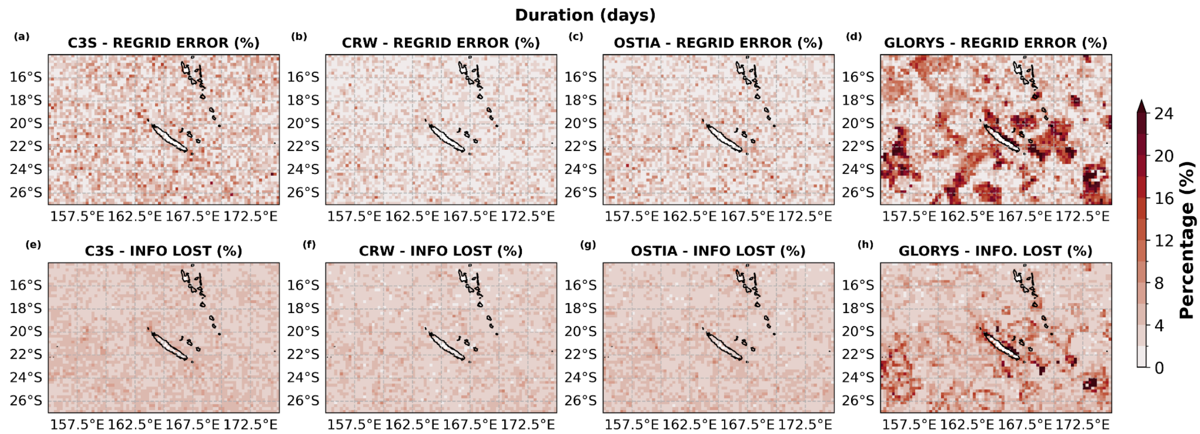

Figure R1.3: (a-d) Re-gridding error (%) as defined in the text for C3S (a), CRW (b), OSTIA (c) and GLORYS (d) for the metric duration. (e-h) Information loss (%) as defined in the text for C3S (e), CRW (f), OSTIA (g) and GLORYS (h) for the metric duration.

To sum up the information of Fig. R1.3, the spatial mean over the area of study of the re-gridding error in percentage and the information lost in percentage was computed for each dataset and variable, as shown in Fig. R1.4.

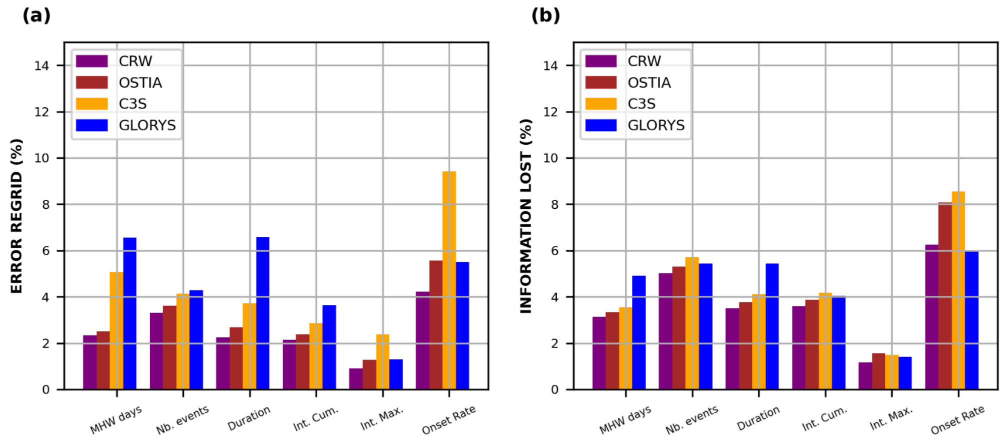

Figure R1.4: spatial mean over the area of study (New Caledonia, Fig. R1.1) of the regridding error expressed in percentage (a) and Information lost in the re-gridding process (b) as defined in the text.

For C3S, CRW and OSTIA, the re-gridding error is lower than 4% for all metrics except the onset rate for which it ranges between 4 and almost 10% (Fig. R1.4a). GLORYS appears to be more impacted than the other products by the re-gridding, except for the onset rate and maximum intensity metrics for which C3S shows the largest regridding errors (Fig. R1.4a). Similar observations can be made for the percentage of information lost (Fig. R1.4b), which is lower than 5/6% for all metrics and products except for the onset rate for which it ranges between 6 and 8/9%. If Fig R1.4 shows that the re-gridding process has an influence on MHW statistics, these values of re-gridding error and information loss are rather low and suggest that the re-gridding process has little impact on our results. The higher re-gridding error and information loss for the onset rate, as well as the difference of impact between products might play a role in the higher inter-products dispersion observed for this metric. If the impact of re-gridding is negligible for most MHW metrics, it might not be the case for the onset rate. Let's also note that for all metrics, the re-gridding error is rather similar between C3S and OSTIA while it is quite different from C3S and GLORYS, for which it is higher for all metrics. The higher impact on C3S and GLORYS might be due to a higher spatial variability at finer scales for these products, thereby generating more impacts of the re-gridding.

The same analysis was conducted over a small area of the Warm Pool, showing very similar results (not shown).

This analysis over New Caledonia was added in Supplementary Information in the 'Impact of re-gridding' section, and a few lines were added in 4.2 Potential explanations of these differences (L. 588-592), to be in line with Reviewer 1 and Reviewer 2 comments :

"A preliminary work on the impact of re-gridding on the observed inter-products differences was conducted inside two small areas of the tropical Pacific, and results for the area of New Caledonia are presented in Supplementary Information (Table S1 and Fig. S8,S9). The analysis suggests that the re-gridding has little impact on the results of our study but might have an influence on inter-products dispersion for the onset rate metric. C3S and GLORYS also seem to be more impacted by the re-gridding than CRW and OSTIA."

A few lines were also added in 4.4 Perspectives L.645-649:

" In addition, the re-gridding of SST datasets onto a common 0.25° grid might also have influenced our results (computing spatial means for re-gridding tends to smooth SST time series), as mentioned in 4.2 and illustrated in Supplementary Table S1 and Fig. S8. Further investigation over the whole area would be needed to thoroughly answer the question of the impact of re-gridding and how much information is lost in the process. Since re-gridding is a common practice in MHW studies, more investigation on the impact of the chosen target resolution could also help to advance MHW research."

- Is the climatology period 1993-2021 the same as the whole period studied? I understand that the full study period is the period analysed but it has to be clearly stated in the manuscript.

>> Yes, this is correct. We added clarifications in section 2.2.2 L.135 and L.137.

"MHW detection was performed for each pixel of the six datasets presented in section 2.1 on the 0.25° common grid over the tropical Pacific and the period 1993-2021, following the Hobday et al. (2016) method."

" The full study period 1993-2021 also served as the common climatological baseline across the SST products (Table 1, common period to all products)."

- The authors separate MHW events in micro and macro scales, greater or smaller than 5x5 degree. How is this size threshold determined? Have you checked and compared results for other thresholds? A MHW of 4x4º occupies an extensive area, especially in the case of marginal seas. I would like to see some figures about mean size of micro-events, dispersion, percentiles that justify the 5x5 is a good choice. Some micro events can be almost as big as some macro events.
Maybe your threshold is appropriated for the open ocean, but this election needs to be better justified. Check methodology in Pastor (2023) to identify MHW area.

Pastor, F. & Khodayar, S. Marine heat waves: Characterizing a major climate impact in the Mediterranean. Science of the Total Environment 861, (2023).

>> We thank the reviewer for this relevant remark and will answer his/her comment with analyses which were not shown in the manuscript.

First, the threshold of 5°x5° was used in other studies (Lal et al. 2025, Sen Gupta et al. 2020, Sun et al. 2023) and this choice allowed us to be in line with these works. This was motivated by the fact that such threshold filters out MHW linked to large mesoscale eddies.

To confirm this threshold was an appropriate choice, we computed the distribution of MHW maximum size for each product over all events detected (Fig. R1.5). The largest inter-product differences are observed for the 0-50 square degrees sizes, which was also observed by Lal et al. 2026. When zooming inside these smaller MHWs, results highlight that differences mainly come from the smallest events between 0 and 20 square degrees (Fig. R1.6). To justify our choice, we computed the 25th percentile of MHW maximum size distributions for each product (vertical lines in Fig. R1.6). The highest value being observed for the COMPOSITE and reaching 22 square degrees (Fig. R1.6), we decided that the threshold of 25 square degrees was appropriate for our study and at the same time allowed us to be aligned with other studies (Lal et al. 2026, Sen Gupta et al. 2020, Sun et al. 2023, Bonino et al. 2023).

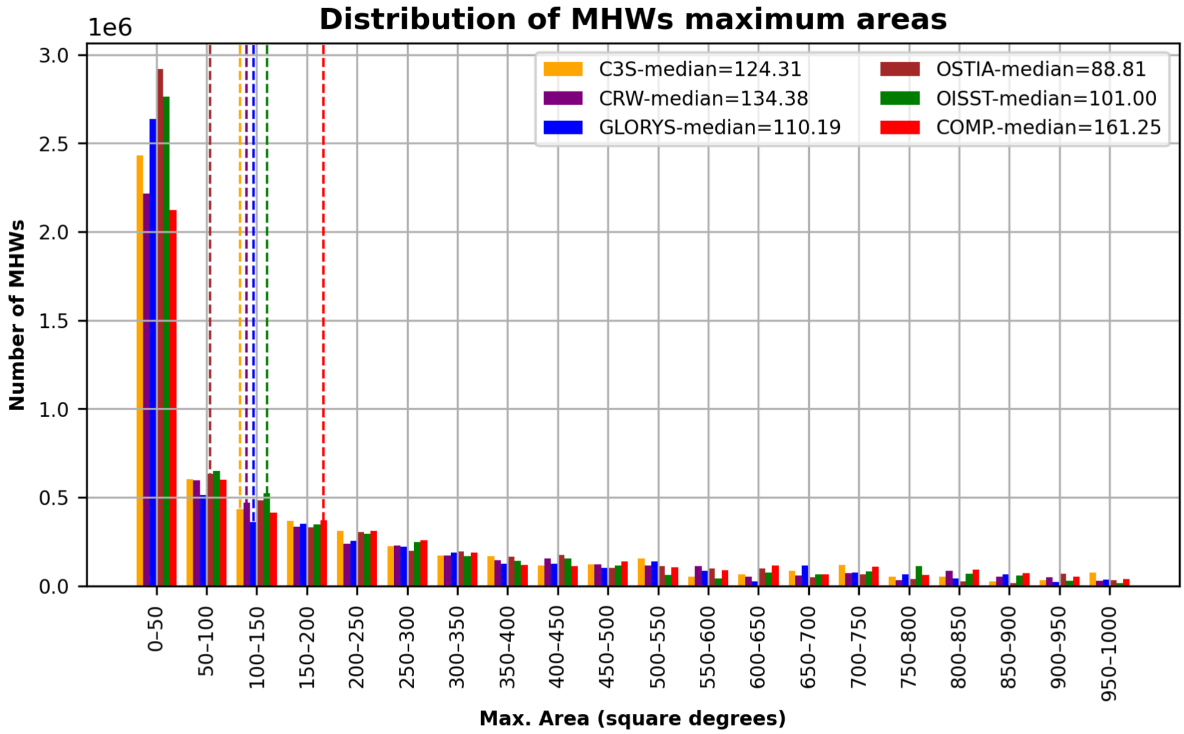

Figure R1.5: distribution of MHWs maximum areas for all MHWs detected over 1993-2021 in the tropical Pacific for each studied product. The median value of distributions is indicated in the legend and shown by the vertical dashed line.

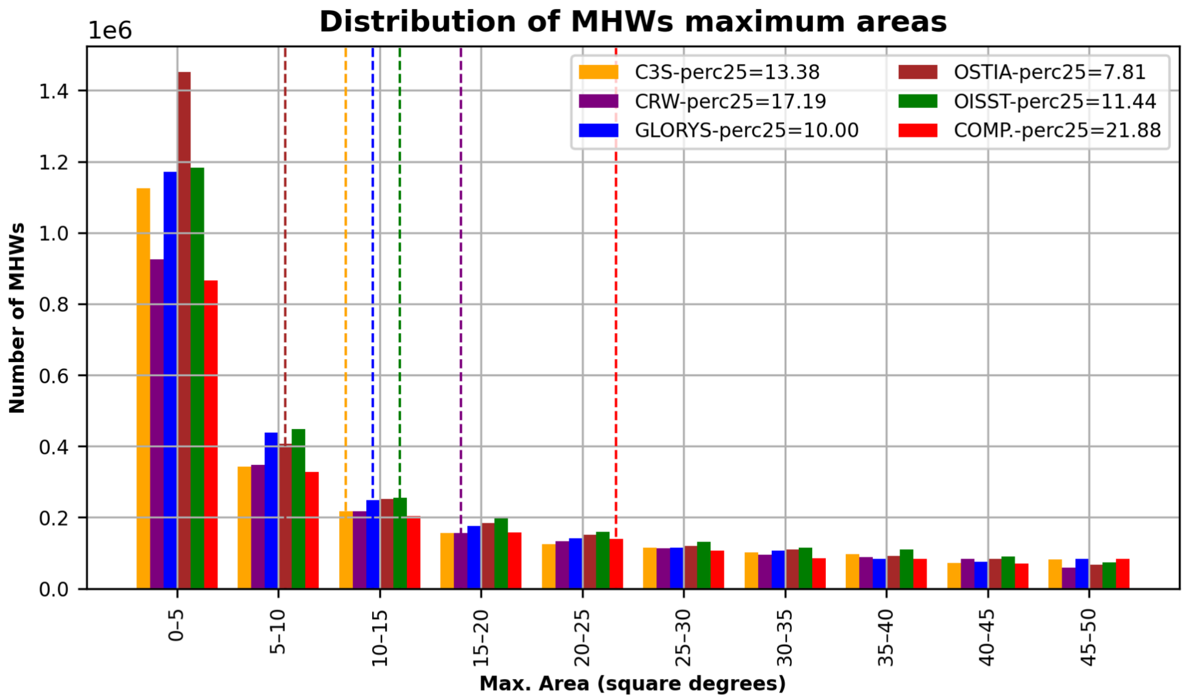

Figure R1.6: Same as Figure 1 but zoomed in for the smallest MHWs. The vertical dashed lines here indicate the 25th percentile of distributions.

To better answer the reviewer comment, we also computed the boxplots of MHWs maximum sizes for macro events and micro events, after being filtered at 5°x5°.

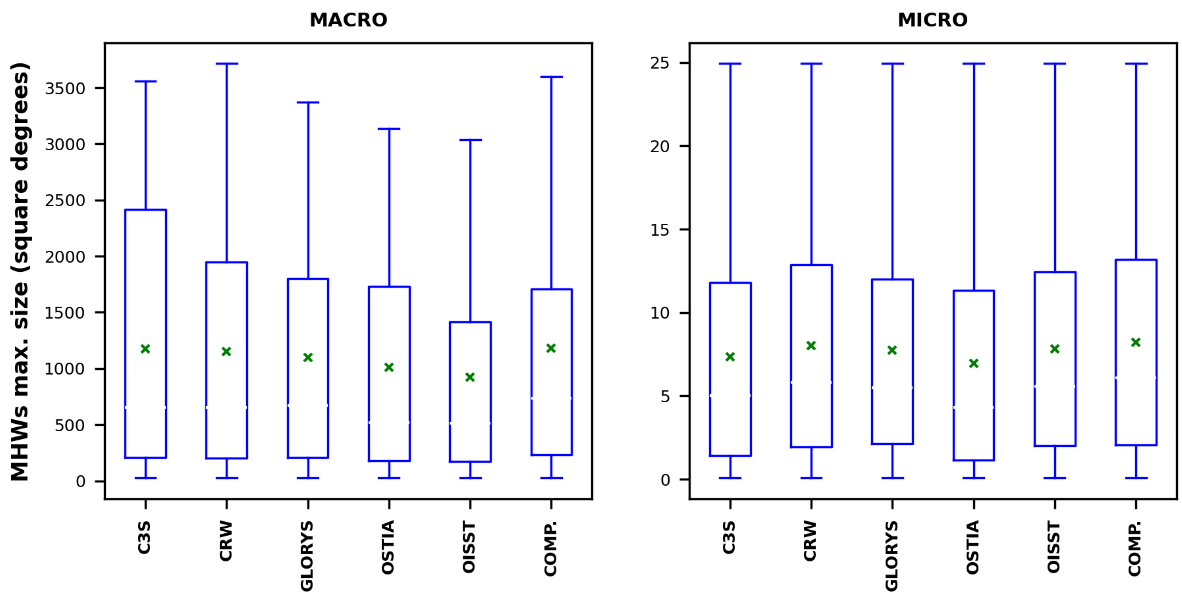

Figure R1.7: Distribution of MHW maximum sizes for macro MHWs (left panel) and micro MHWs (right panel) for each studied product.

As also suggested by the reviewer, we compared our results with the 4°x4° threshold. Fig. 6 of the manuscript was reproduced for this threshold, as illustrated in Fig. R1.8. Results were almost unchanged, showing that our analysis is robust and not highly dependent on the threshold chosen.

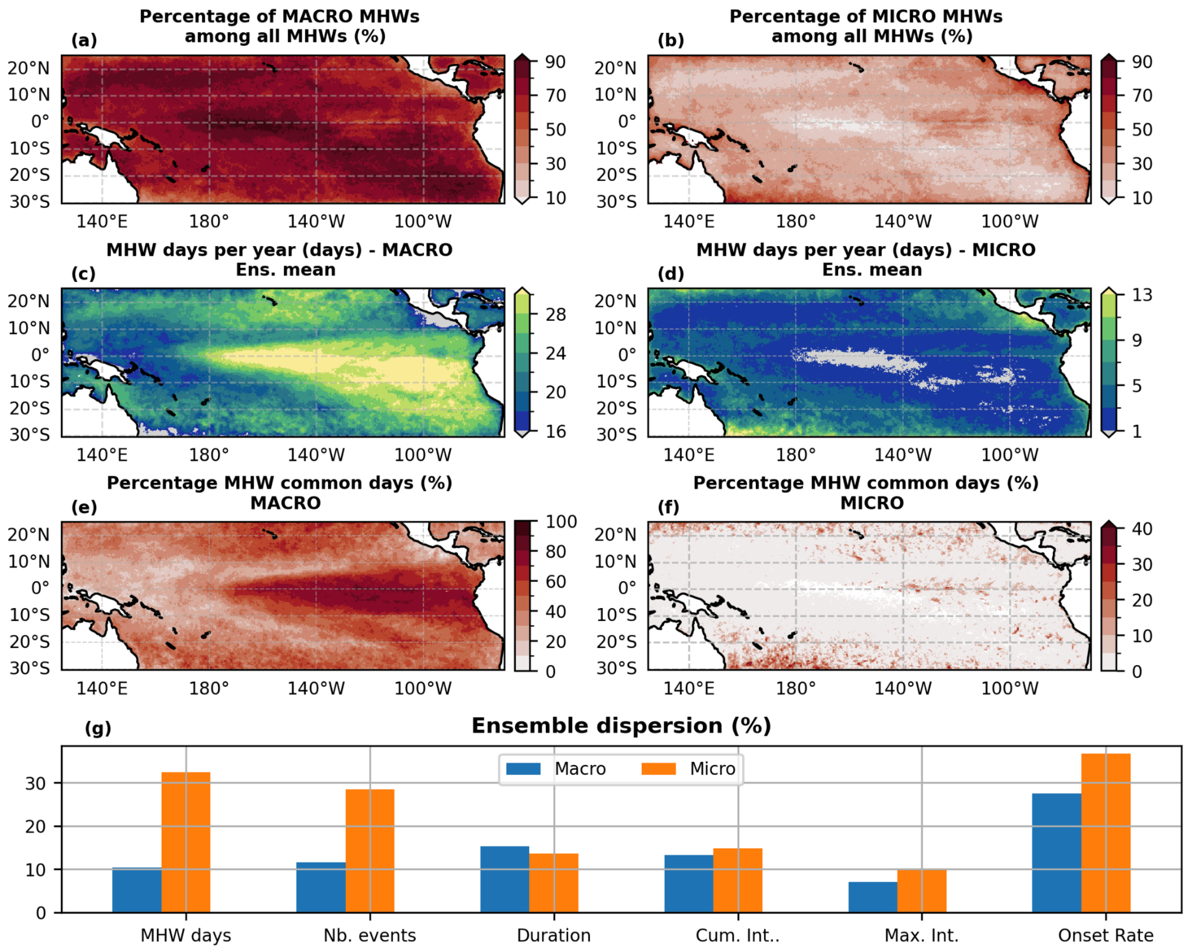

Figure R1.8: Same as Fig. 6 of the manuscript for micro and macro MHWs detected with a 4°x4° threshold (instead of 5°x5°).

Some elements of answer to the reviewer's comment were added in methodology section 2.2.3 L.163-167 :

"The maximum area reached during the event was associated with each MHW in the evaluated pixel. Events with a maximum area smaller than 25 square degrees were classified as micro-scale, whereas those with a maximum area exceeding 25 square degrees were defined as macro-scale. The threshold of 25 square degrees was chosen as it filters out MHWs linked to large mesoscale eddies and it is in line with other MHW studies (Lal et al., 2025; Sen Gupta et al., 2020; Sun et al., 2023). The distribution of MHW maximum sizes in the Tropical Pacific for all studied products confirmed that this threshold was appropriate for our study (not shown). The 4°x4° threshold was also tested and our results remained almost unchanged, showing that our analysis is robust and not highly dependent on the threshold chosen (not shown)."

Minor comments:

Line 294 “for the maximum intensity (total MHW days) (Fig. 4b,f)”. Correct if necessary.

>> Corrected.

2.3.2 Temporal evolution : “The year attribution of a MHW was based on its time start”. Why do not use central date? Have you checked how many MHWs start and end on different years? And how many days of this event correspond to the end year?

>> Indeed, the time of the MHW start was used for assigning the year. Following the reviewer remark, we highlighted that less than 4% of all MHWs detected in all pixels over 1993-2021 in the Tropical Pacific start and end on different years (4.0% for C3S, 3.9% for CRW, 3.6% for OSTIA, 4.6% for GLORYS, 3.5% for OISST and 4.2% for the COMPOSITE). Such events thus represent a very small fraction of the studied MHWs. To answer the reviewer's question on how many days of these events correspond to the end year, we computed for each of these events overlapping over two years the number of days associated to the year start and the number of days associated to the year end, expressed in percentage of the whole duration of the event. On average, slightly less than 50% of the duration of these events correspond to the end year (more precisely, 48.87% for C3S, 49.22% for CRW, 49.50% for OSTIA, 49.93% for GLORYS, 50.02% for OISST, 48.71% for the COMPOSITE). This result suggests that such MHWs really overlap equally over the two years and thus attributing the start year, end year or central date won’t change the results significantly.