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.

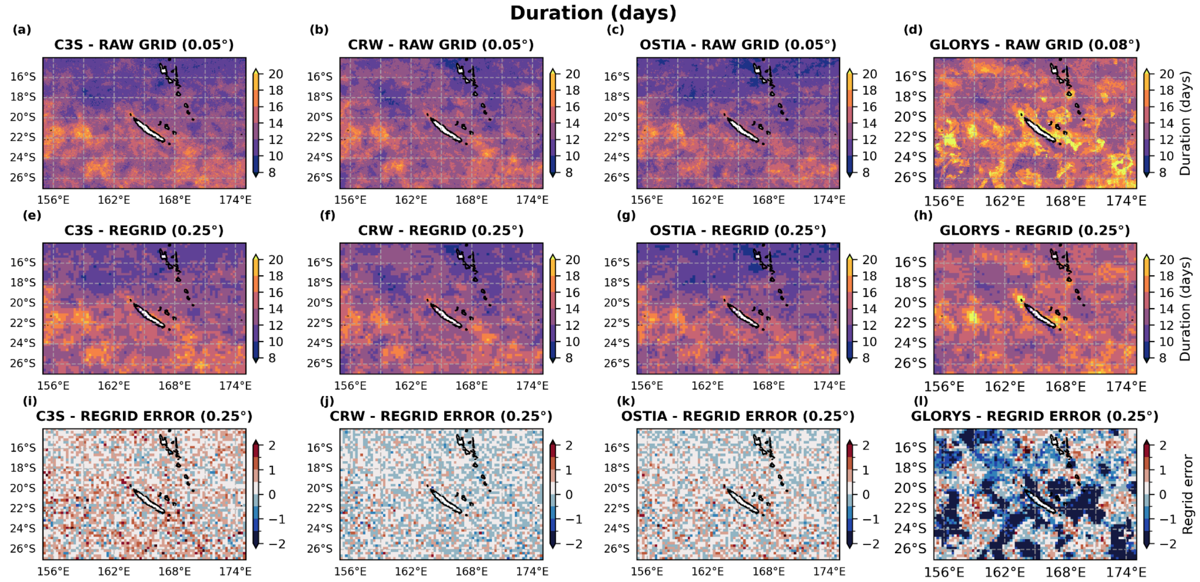

Figure R1.1: (a-d) Mean duration of MHWs over 1993-2021 for C3S (a), CRW (b), OSTIA (c) and GLORYS (d) at 0.05° resolution and 0.08° resolution for GLORYS. (e-h) Mean duration of MHWs over 1993-2021 for C3S (e), CRW (f), OSTIA (g) and GLORYS (h) at 0.25° resolution (i-l) Re-gridding error as defined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

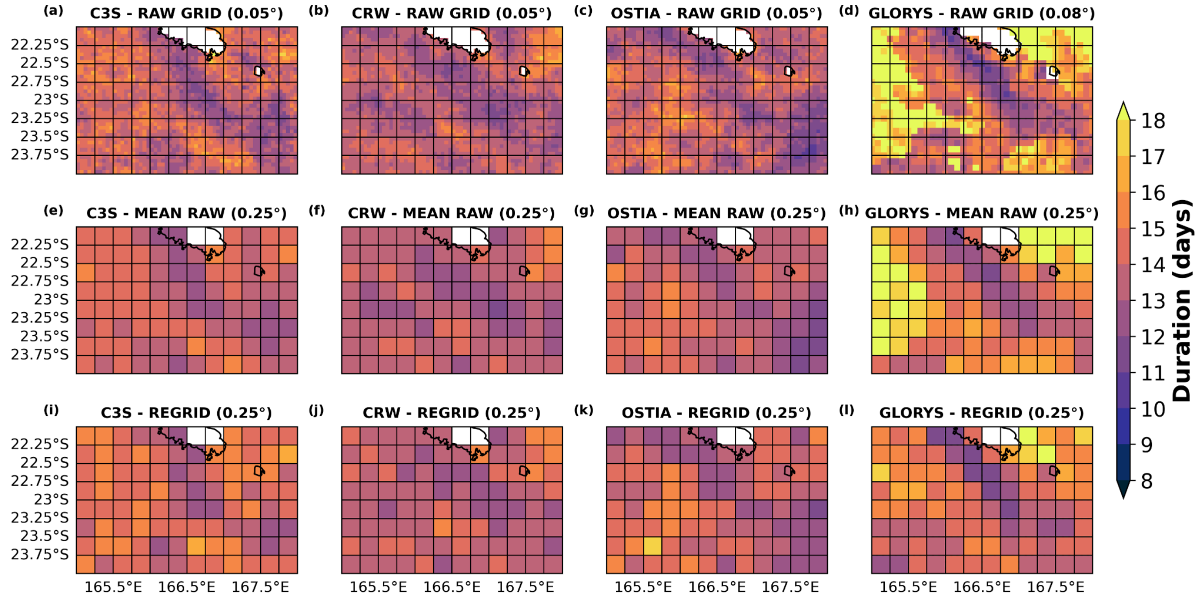

Figure R1.2: Zoom of Fig. R1.1 on the southern coast of New Caledonia. (a-d) Mean duration of MHWs over 1993-2021 for C3S (a), CRW (b), OSTIA (c) at 0.05° resolution and at 0.08° resolution for GLORYS (d). (e-h) Mean of MHW statistics of the RAW grid inside each pixel of the 0.25° grid (i-l) Mean duration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

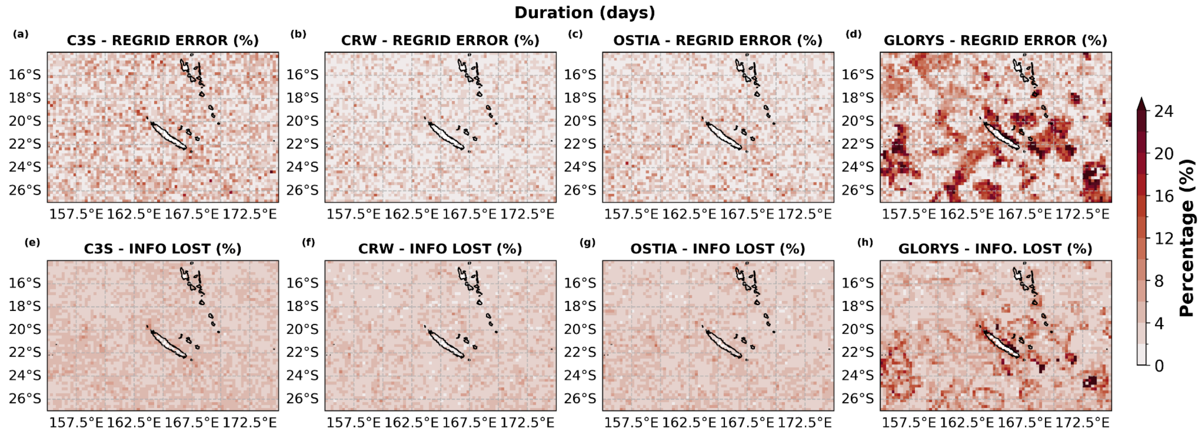

Figure R1.3: (a-d) Re-gridding error (%) as defined in the text for C3S (a), CRW (b), OSTIA (c) and GLORYS (d) for the metric duration. (e-h) Information loss (%) as defined in the text for C3S (e), CRW (f), OSTIA (g) and GLORYS (h) for the metric duration.

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.

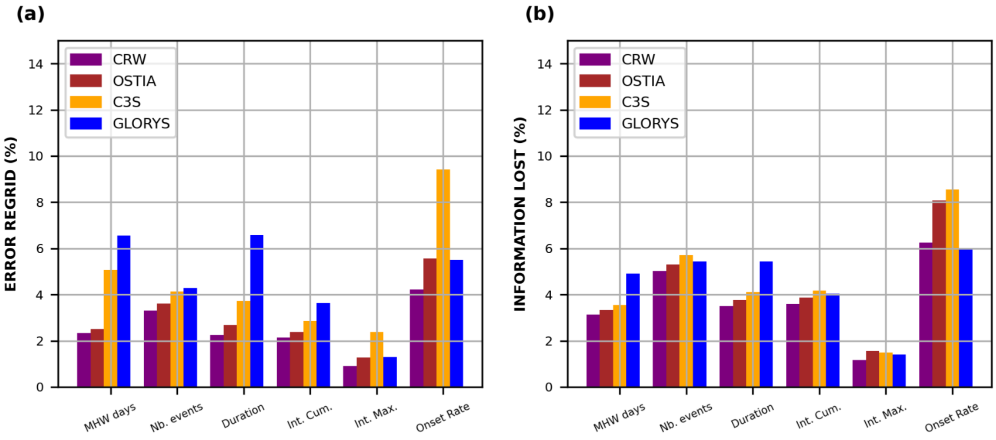

Figure R1.4: spatial mean over the area of study (New Caledonia, Fig. R1.1) of the regridding error expressed in percentage (a) and Information lost in the re-gridding process (b) as defined in the text.

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 4x4° 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).

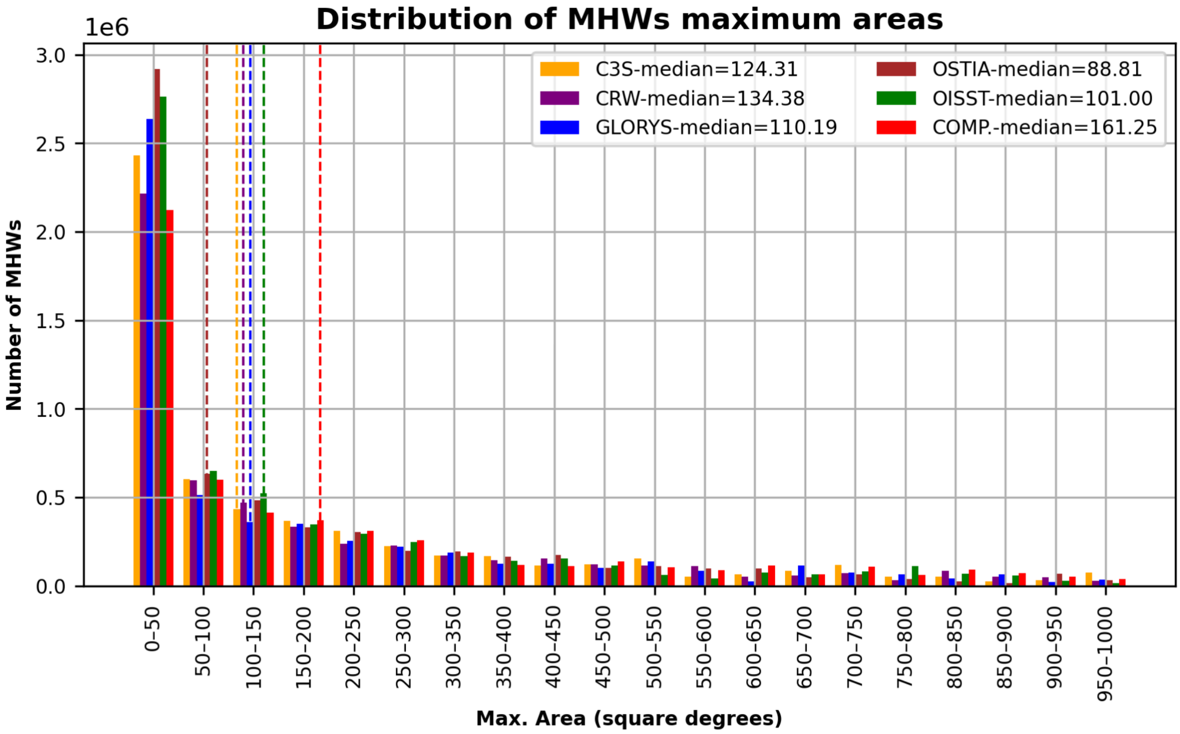

Figure R1.5: distribution of MHWs maximum areas for all MHWs detected over 1993-2021 in the tropical Pacific for each studied product. The median value of distributions is indicated in the legend and shown by the vertical dashed line.

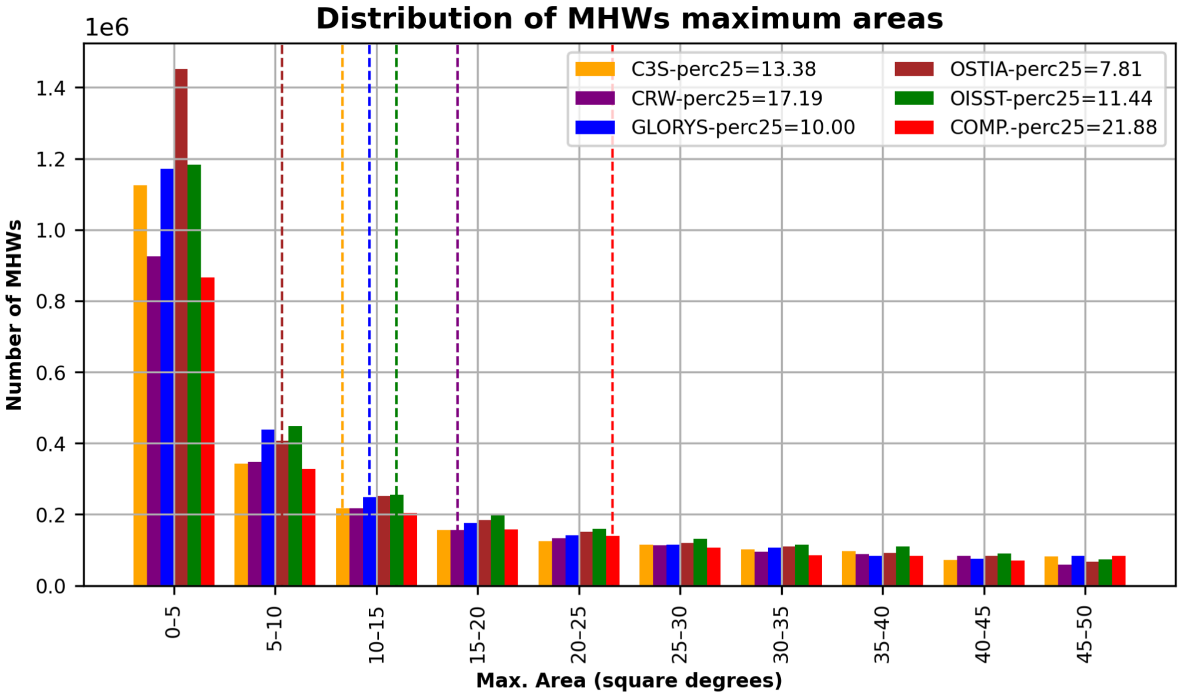

Figure R1.6: Same as Figure 1 but zoomed in for the smallest MHWs. The vertical dashed lines here indicate the 25th percentile of distributions.

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°.

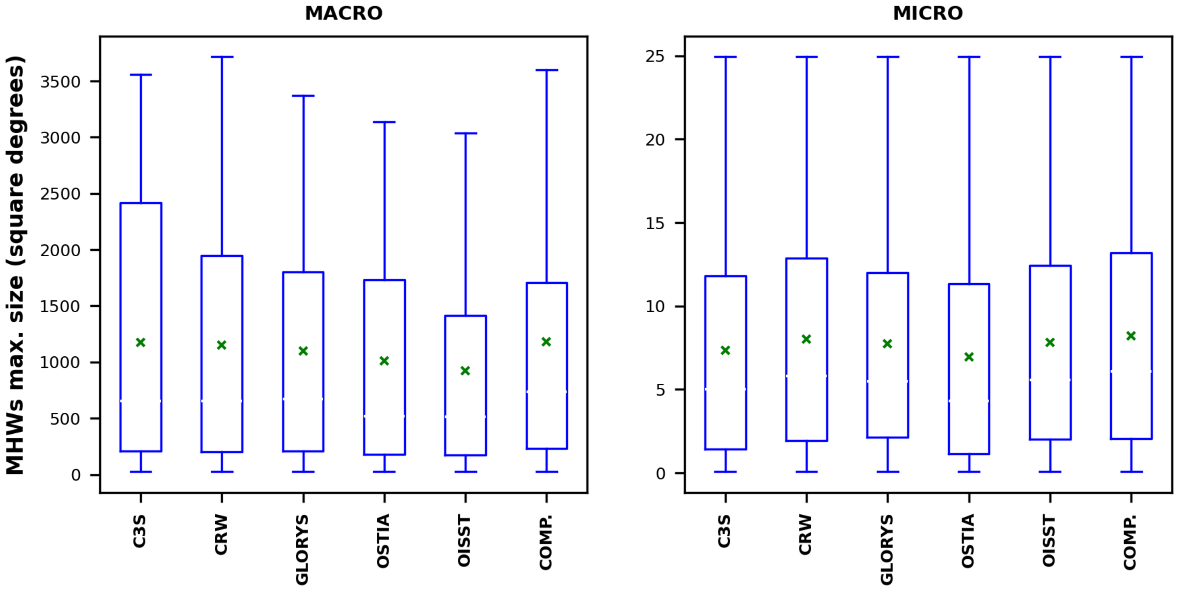

Figure R1.7: Distribution of MHW maximum sizes for macro MHWs (left panel) and micro MHWs (right panel) for each studied product.

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.

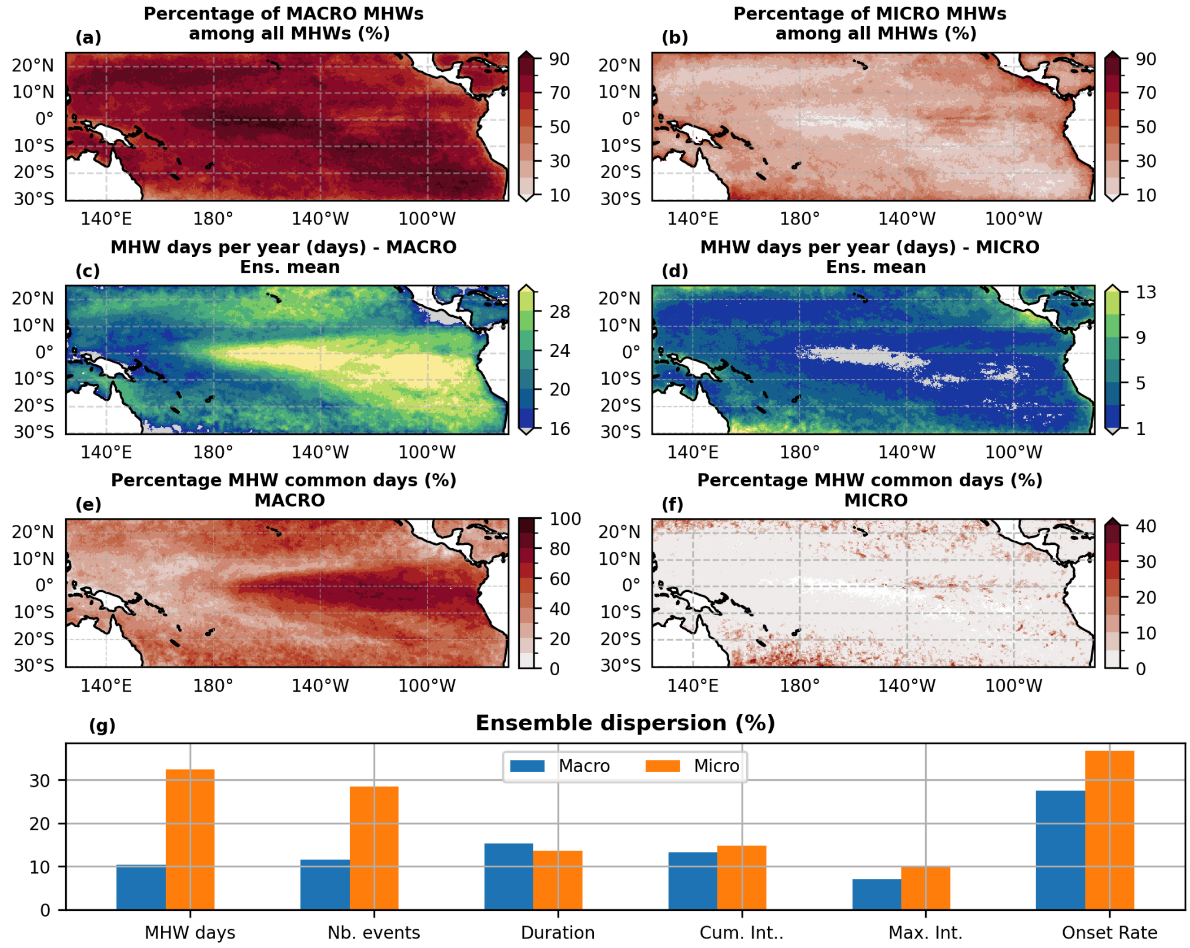

Figure R1.8: Same as Fig. 6 of the manuscript for micro and macro MHWs detected with a 4°x4° threshold (instead of 5°x5°).

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.

Responses to Reviewer 2

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:

This manuscript evaluates how the choice of observational or observation-based sea surface temperature datasets influences the comparison of marine heatwave metrics. The topic is timely and of relevance to the marine heatwave community, as the results demonstrate that commonly used metrics can be sensitive to dataset selection. This is an important issue, given that many marine heatwaves studies rely on only one observational dataset. The study is generally well designed and the analysis of good quality, but some aspects of the methodology and clarity of presentation would benefit from revision. Addressing the specific and technical comments below would substantially strengthen the manuscript. Although the individual comments may not necessitate extensive changes, taken together they represent substantial revisions. On this basis, it is recommended that the manuscript be reconsidered after major 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.

Specific comments:

- Given the focus on the tropical Pacific, where ENSO is a dominant driver of sea surface temperature variability, the role of ENSO in shaping the identified marine heatwaves warrants clearer and more explicit consideration. Additional analysis separating marine heatwaves that are associated with El Niño or La Niña from those that are independent of ENSO would be a valuable extension. While such an analysis is not essential for publication, it would help place the dataset-dependent differences in marine heatwave metrics into a clearer physical context. Some ENSO-related relationships are already evident in Section 3.3.3, and building on these would enhance the manuscript.

>> We agree with the reviewer that ENSO is indeed a dominant driver of MHW in the region considered, and that the dependence of our results on the ENSO phase is worth exploring. Following his/her suggestion, we divided MHWs into three groups : El Niño MHWs, La Niña MHWs, and neutral MHWs. El Niño MHWs were defined as MHWs starting during a month where the Oceanic Niño Index (ONI) is higher than 0.5 and La Niña MHWs were defined as MHWs starting during a month where the ONI is lower than -0.5. All other MHWs were classified in the neutral group. This analysis aimed to answer the reviewer comment and added interesting information to our study. Yet, the relationship between MHWs and ENSO was not the focus of the paper, is complex and could be a study on its own. The method presented here to classify MHWs should be regarded as a first step and does not account for the substantial diversity among ENSO events. In particular, Eastern Pacific and Central Pacific events (Capotondi et al., 2020) can exert different influences on marine heatwave characteristics (Gregory et al., 2024; Pagli et al., 2025).

First, MHW mean statistics for the three groups defined previously were computed and some of them are illustrated in Fig. R2.1 (MHW days per year, duration and maximum intensity). The spatial means of these metrics and of the onset rate inside the seven regions of study were also computed (Fig. R2.5).

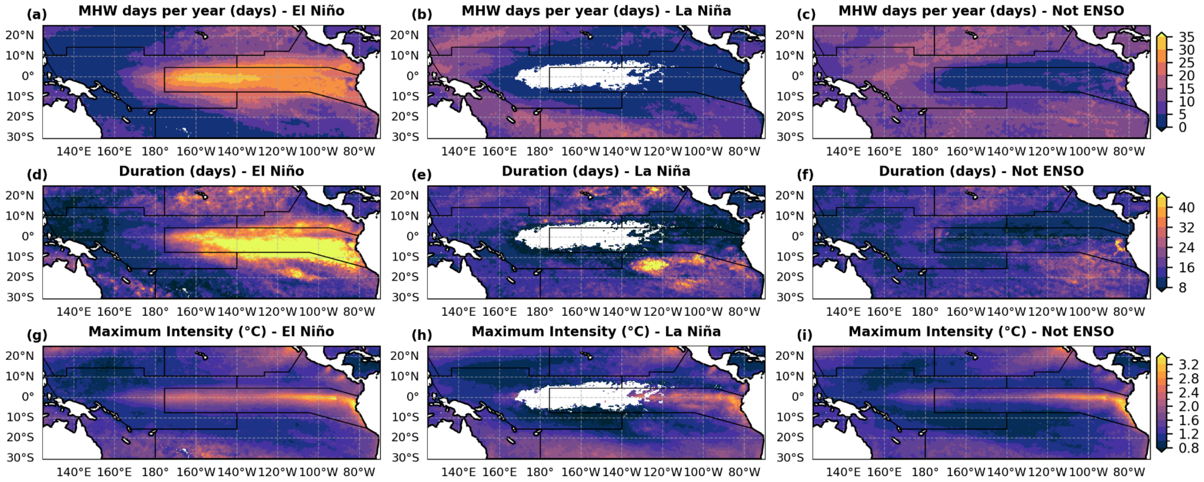

Figure R2.1: ensemble mean of MHWs mean statistics for El Niño MHWs (a,d,g), La Niña MHWs (b,e,h) and neutral MHWs (c,f,i).

Then, dispersion maps as shown in Fig. 4 of the manuscript were computed for El Niño MHWs, La Niña MHWs and neutral MHWs (Fig. R2.2, R2.3, R2.4, respectively), as well as the spatial means of dispersion in the regions of study for the metrics MHW days per year, duration, maximum intensity and onset rate (Fig. R2.5 e,f,g,h).

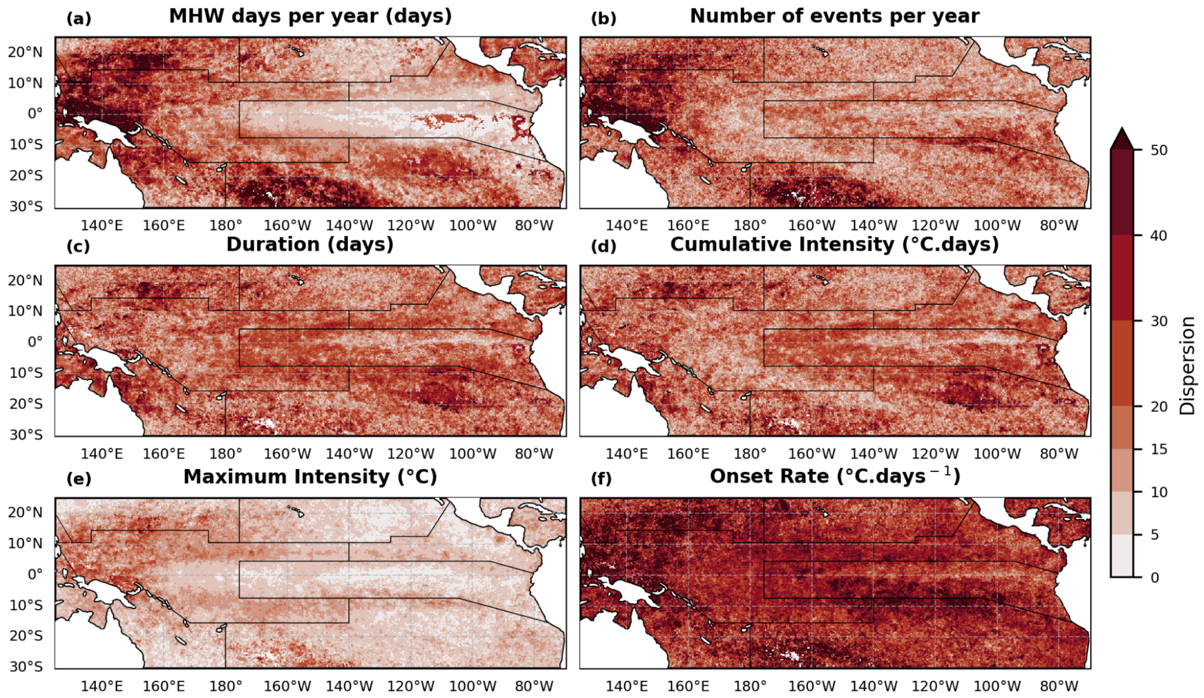

Figure R2.2: ensemble dispersion as in Fig. 4 of the manuscript for El Niño MHWs.

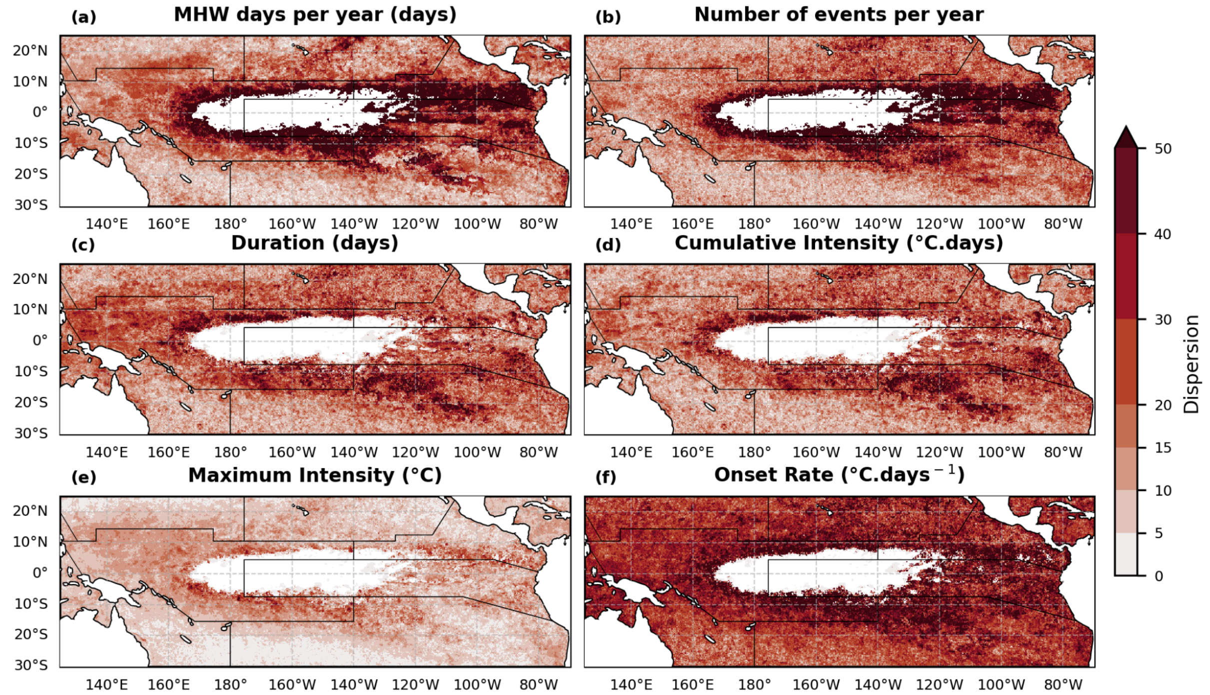

Figure R2.3: ensemble dispersion as in Fig. 4 of the manuscript for La Niña MHWs.

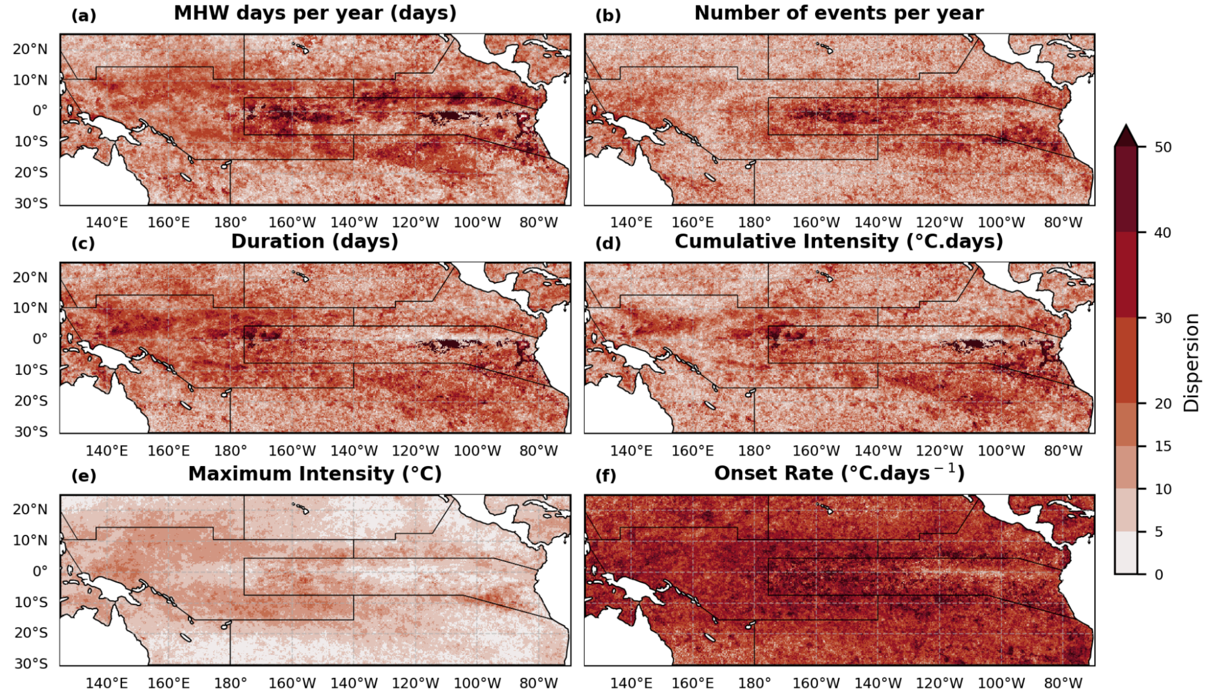

Figure R2.4: ensemble dispersion as in Fig. 4 of the manuscript for neutral MHWs.

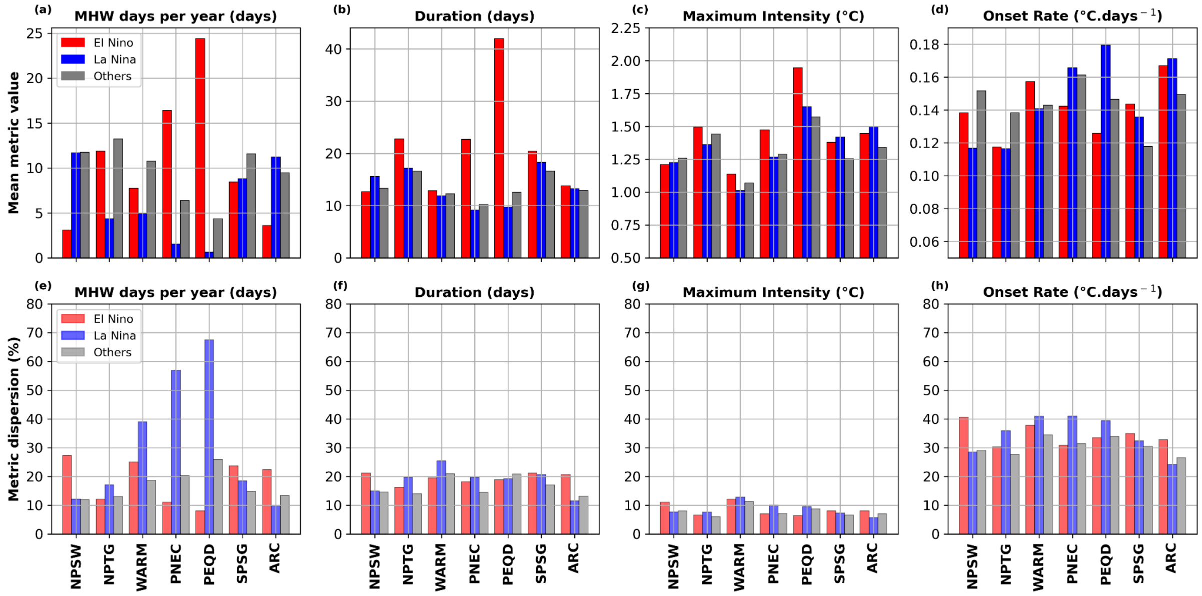

Figure R2.5: (a-d) Histograms of the spatial means of MHWs metrics (ensemble mean) inside the seven regions of study for the MHW days per year (a), duration (b), maximum intensity (c) and onset rate (d) for El Niño MHWs (red), La Niña MHWs (blue) and neutral (gray). (e-h) Histograms of the spatial mean of ensemble dispersion inside the seven regions of study for the MHW days per year (e), duration (f), maximum intensity (g) and onset rate (h) for El Niño MHWs (red), La Niña MHWs (blue) and neutral (gray).

As can be also seen in the temporal plots of section 3.3.3, these results highlight that the PNEC and PEQD are highly influenced by El Niño events while the NPSW and ARC are influenced by La Niña events (in these regions, most MHW days per year are attributed to El Niño and La Niña, respectively). Histogram of ensemble dispersion in Fig. R2.5 highlights that El Niño leads to lower inter-products dispersion in the regions where it has an influence, while it is not the case for La Niña. The decrease in dispersion in regions influenced by El Niño could be due to the presence of macro scale MHWs during El Niño events, which show lower inter-products dispersion as shown in the results of section 3.3.2.

Fig. R2.1 was added in Supplementary Information and Fig. R2.5 was added in the manuscript in section 3.3.3.

The methodology used to classify MHWs into these three groups was thus added in the manuscript L.171-177:

"2.2.4- Filtering MHWs according to El Niño Southern Oscillation

Since ENSO is a dominant driver of MHWs in the Tropical Pacific (Holbrook et al., 2019; Sen Gupta et al., 2020; Pagli et al., 2025), the dependence of our results on the ENSO phase was worth exploring. For this purpose, MHWs were divided into three groups : El Niño MHWs, La Niña MHWs, and neutral. El Niño MHWs were defined as MHWs starting during a month where the Oceanic Niño Index (ONI) was higher than 0.5 and La Niña MHWs were defined as MHWs starting during a month where the ONI was lower than -0.5. The monthly time series of the ONI were used (available at : https://psl.noaa.gov/data/timeseries/month/). All other MHWs were classified in the neutral group."

The following lines were also added in section 3.3.3 along with Fig.R2.5 (L.397-410):

"… The particularly high values in 2015 compared to 2016 for the metrics duration, cumulative intensity, maximum intensity and rate of onset is explained by the year attribution of MHWs to its time start and the starting of the strong El Niño event in 2015.

As Fig. 7 reveals a strong influence of ENSO events on MHW metrics, MHW mean metrics and ensemble dispersion were computed for El Niño MHWs, La Niña MHWs and neutral MHWs as defined in section 2.2.4, in the seven subregions studied (Fig. 8). Our results being similar between MHW days per year and number of events per year as well as between duration and cumulative intensity, only four metrics (MHW days per year, duration, maximum intensity and onset rate) were illustrated in Fig. 8 to make it more readable. Regional maps of MHW mean metrics for the three groups are also shown in Fig. S6 to better understand Fig. 8. Let's note that the method presented here to classify MHWs should be regarded as a first step and does not account for the substantial diversity among ENSO events. In particular, Eastern Pacific and Central Pacific events (Capotondi et al., 2020) can exert different influences on marine heatwave characteristics (Gregory et al., 2024; Pagli et al., 2025). Fig. 8a-d highlights that the PNEC and PEQD are highly influenced by El Niño events while the NPSW and ARC are influenced by La Niña events (in these regions, most MHW days per year are attributed to El Niño and La Niña, respectively). Spatial means of ensemble dispersion inside the regions (Fig. 8e-h) highlight that El Niño leads to lower inter-products dispersion in the regions where it has most influence, while it is not the case for La Niña. The decrease in dispersion in regions influenced by El Niño could be due to the presence of macro scale MHWs during El Niño events, which show lower inter-products dispersion as shown in the results of section 3.3.2.

The yearly time series of the ensemble dispersion and spatial correlation within each region (section 2.3.2, Fig. 9 and Fig. S6) provide more insights on the temporal evolution of inter-product differences, and highlight four main points. … "

- The datasets are regridded to a common 0.25° grid prior to the calculation of marine heatwave metrics. While this approach allows the products to be more readily compared, further discussion of the potential implications of the regridding step would strengthen the methodological transparency of the study. In particular, regridding higher-resolution products to a coarser grid may smooth small-scale temperature features, which could in turn reduce peak marine heatwave intensities or alter event characteristics such as duration and spatial extent. It would be helpful to discuss whether this smoothing could contribute to some of the inter-dataset differences reported, and whether the results are sensitive to the chosen target resolution. An alternative approach, such as regridding coarser products to a finer grid, or computing marine heatwave metrics at native resolution prior to regridding, could be briefly considered, even if not implemented. Clarifying the rationale for the chosen approach, and its potential impact on the results, would improve confidence in the robustness of the conclusions.

>> We thank the reviewer for this relevant remark. In response to a similar comment of Reviewer 1, we analysed the impact of re-gridding in the response document to Reviewer 1 where we tried to answer both reviewers comments. About the alternative approach of

regridding from coarser resolution to a finer grid, this method would be very costly in terms of computation time and amount of data.

Common response with Reviewer1:

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.

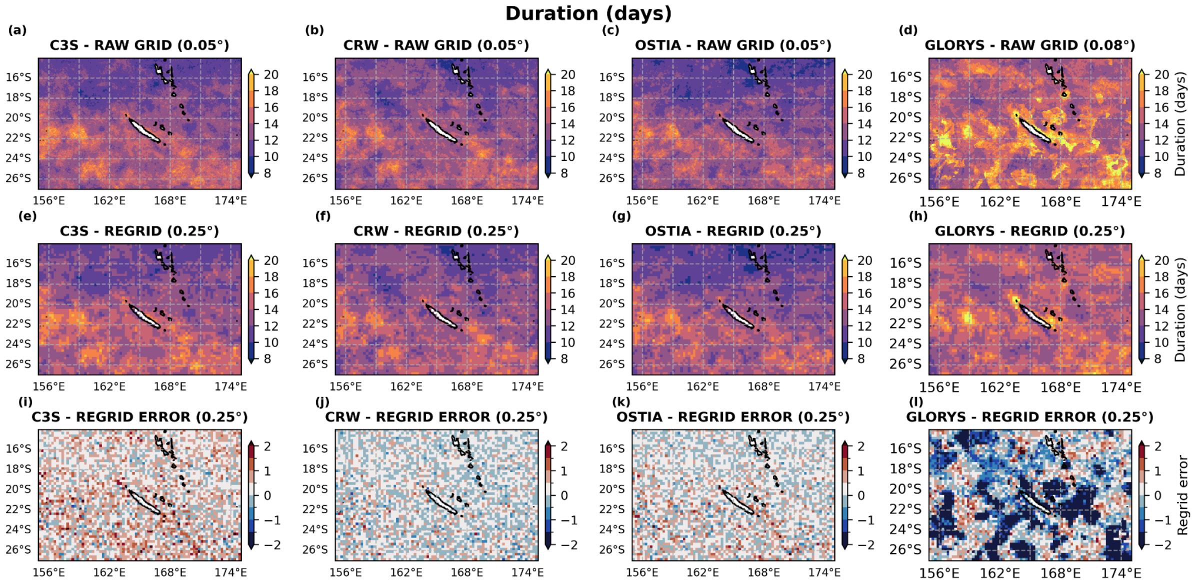

Figure R1.1: (a-d) Mean duration of MHWs over 1993-2021 for C3S (a), CRW (b), OSTIA (c) and GLORYS (d) at 0.05° resolution and 0.08° resolution for GLORYS. (e-h) Mean duration of MHWs over 1993-2021 for C3S (e), CRW (f), OSTIA (g) and GLORYS (h) at 0.25° resolution (i-l) Re-gridding error as defined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

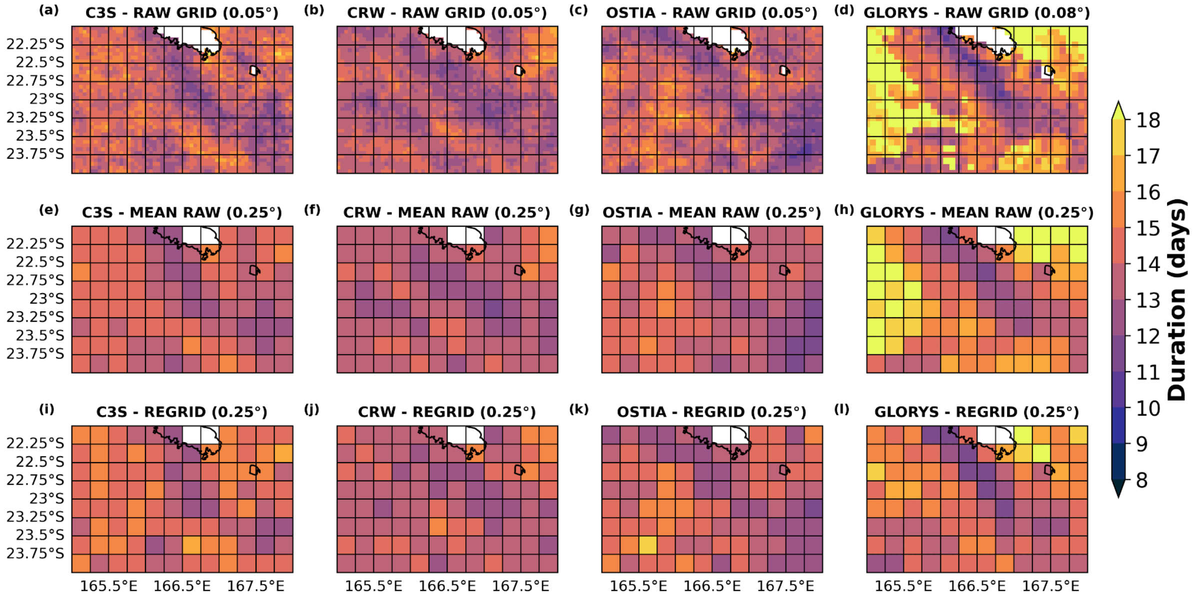

Figure R1.2: Zoom of Fig. R1.1 on the southern coast of New Caledonia. (a-d) Mean duration of MHWs over 1993-2021 for C3S (a), CRW (b), OSTIA (c) at 0.05° resolution and at 0.08° resolution for GLORYS (d). (e-h) Mean of MHW statistics of the RAW grid inside each pixel of the 0.25° grid (i-l) Mean duration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

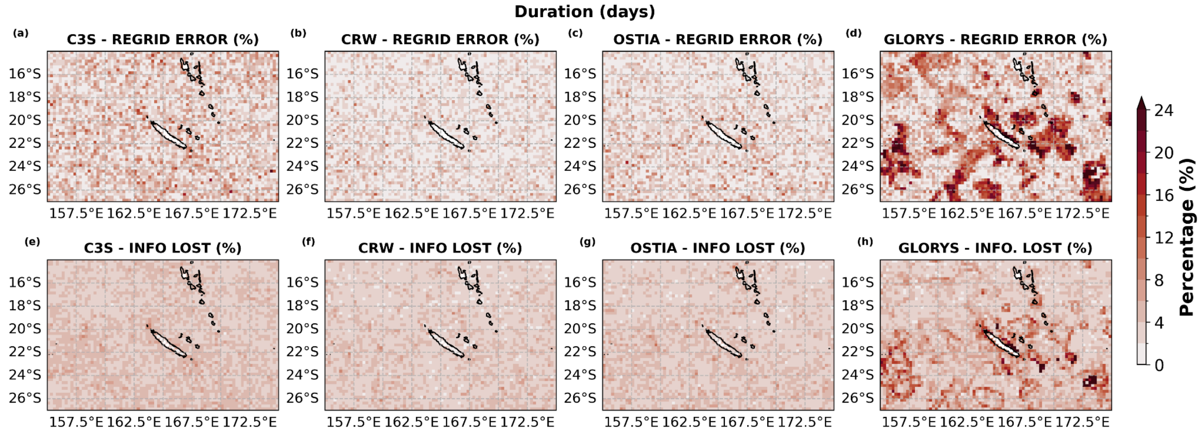

Figure R1.3: (a-d) Re-gridding error (%) as defined in the text for C3S (a), CRW (b), OSTIA (c) and GLORYS (d) for the metric duration. (e-h) Information loss (%) as defined in the text for C3S (e), CRW (f), OSTIA (g) and GLORYS (h) for the metric duration.

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.

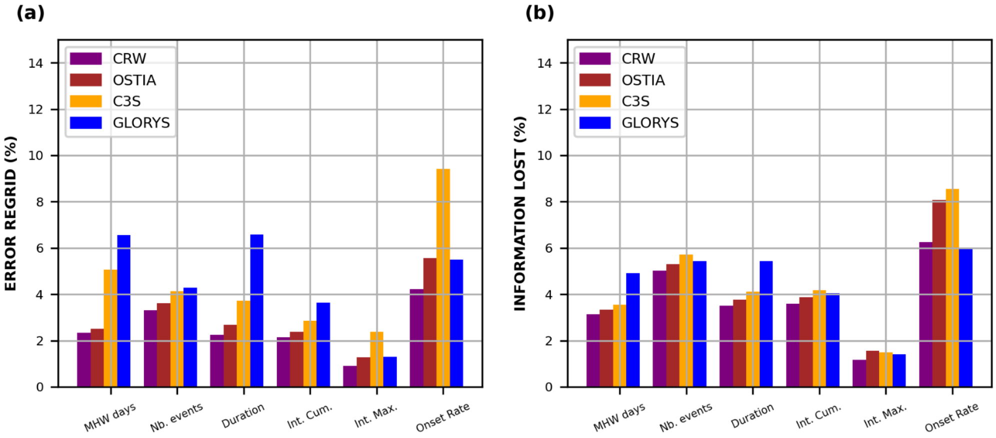

Figure R1.4: spatial mean over the area of study (New Caledonia, Fig. R1.1) of the regridding error expressed in percentage (a) and Information lost in the re-gridding process (b) as defined in the text.

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.”

- The inclusion of a “composite” dataset, defined as the mean of the four observational products, is potentially informative as a diagnostic or reference product. However, its subsequent inclusion in ensemble-based analyses raises concerns. In particular, incorporating the composite alongside the original observational datasets in the calculation of the ensemble mean effectively results in the four observational products being counted twice, thereby skewing the ensemble statistics. A more statistically consistent approach would be to exclude the composite from ensemble calculations, while retaining it as a separate product for comparison.

>> We thank the reviewer for his/her relevant comment and we agree with his/her remark. The COMPOSITE was removed from ensemble based statistics and Figures 1, 2, 4, 5, 6, 9 and 11 of the manuscript were consequently modified, as well as numerical values in the manuscript. Yet, results are very similar and remain almost unchanged.

The methodology in part 2.3.1 was slightly modified to specify that the composite was excluded from ensemble based statistics L. 186; 190-202; 206:

“To better highlight the inter-product differences on MHW metrics, anomalies were also mapped over the tropical Pacific. For each product and metric, the product anomaly at each pixel was defined as the difference between the metric value for that product (as defined above) and the ensemble mean value of the metric over all products except the COMPOSITE (i.e. the mean of five values, hereafter designated as “ensemble mean” metric) (Eq. 1):

$$anomaly(metric_i,\ product_j,\ pixel_k) = metric_i(product_j,\ pixel_k) - ensemble_{mean}(metric_i,\ pixel_k)$$

(1)

where $ensemble_{mean}(metric_i,\ pixel_k) = (\sum_{j=1}^{6} metric_i(product_j, pixel_k))\ /\ 6$, with i varying from 1 to 6 and representing the six evaluated metrics, j varying from 1 to 5 and representing the 5 evaluated products C3S, CRW, OSTIA, OISST and GLORYS12v1, and k representing the pixel number in the domain. The same maps were produced for the temporal mean of DHW annual maxima (section 2.2.2). Mean metrics and anomalies were computed for the COMPOSITE but the latter was removed from the ensemble-based statistics to avoid counting the observational products twice, since the composite is derived from them.

The sensitivity of each MHW metric to SST product choice was evaluated at each pixel by computing the dispersion of the metric across all SST products excluding the COMPOSITE (hereafter designated as “ensemble dispersion”, Eq. 2).

$$\sigma = ensemble_{dispersion}(metric_i, pixel_k) = \sqrt{\frac{1}{5}\sum_{j=1}^{5}[(metric_i(product_j, pixel_k) - ensemble_{mean}(metric_i, pixel_k))^2]}$$

(2)

The ensemble dispersion was also computed for DHW. Maps of dispersion over the tropical Pacific were produced for each metric. These values of ensemble dispersion were defined as the "uncertainty" of the metric with respect to SST product choice. In order to quantify and compare metrics sensitivity, the ensemble dispersion at each pixel was also expressed as a percentage by dividing the ensemble dispersion by the ensemble mean value of the metric at that pixel, and then multiplying by 100.

The co-occurrence of MHWs across SST products was also assessed by computing the percentage of common MHW days over the basin. At each pixel, this percentage was defined as the number of days detected simultaneously as a MHW in all products excluding the COMPOSITE divided by the ensemble mean number of total MHW days at that pixel. Similar analysis was conducted for DHW, by computing the percentage of common bleaching alerts of level 1 (days for which DHW≥4°C.weeks^{-1}) across products."

Other small modifications were made to L.310 and 371.

However, following the results of Fig. 3, we decided not to retain the COMPOSITE product as a product for comparison since we showed it behaved as an outlier among other products.

Technical Corrections:

(Note: the notation "L16" below refers to "line 16", as an example)

- *L15: Here and throughout, "sea surface temperature" does not need to be capitalised. Likewise for "degree heating weeks".*

>> Corrected.

- *L31: The phrase "methodological choices" may be misleading here, as the analysis appears to focus primarily on differences from dataset selection. Clarification of rephrasing may be helpful.*

>> The reviewer is right, we replaced "methodological choices" by "SST product choice" L.32 : "These findings contribute to a better understanding of how SST product choice affects the characterization of MHWs and DHW, and their associated uncertainties."

- *L49: "requires" might be a better word than "implies".*

>> Corrected.

- *L49: Not just the length (i.e., duration) of the baseline, but also the start year.*

>> Corrected and replaced by "the definition of the climatological baseline" L.49:

"MHW detection requires several methodological choices: the choice of the SST product, the definition of the climatological baseline, whether or not to detrend the SST time series, and the definition of the MHW threshold (Farchadi et al., 2025)."

- *L50-53: The structure of this sentence is unclear. Rephrasing is recommended to distinguish between the agreed methodology and the remaining sources of variability.*

>> We agree with the reviewer. The sentence was changed L. 50-54 :

"For better agreement across studies, the scientific community agrees on a common methodology (use of 30-year climatology, no detrending and seasonally varying 90th percentile, Hobday et al., 2026) even though other options can lead to significantly different results in MHW metrics evaluation potentially leading to different policy responses to MHWs (Capotondi et al., 2024)."

- *L56-58: Please rephrase. It seems this sentence should not begin with "if".*

>> Corrected L.57-58 : "This step appears crucial for MHW analysis (Farchadi et al., 2025), yet most MHW studies rely on a single blended SST product, either satellite-only or satellite combined with in situ data, or an ocean reanalysis."

- *L61: Please rephrase "show variability between themselves". Perhaps: "and consequently exhibit differences".*

>> Corrected L.62.

- *L67: "basinwide" or "oceanwide" scales might be more accurate than "regional", given the text that follows.*

>> Corrected.

- *L71: How is it that understanding of SST product differences would improve MHW forecasts?*

>> We removed "forecast" from the sentence and kept only "information" as it was not really appropriate.

- *L93: It may be misleading to state that six datasets were used, since one of these is the composite derived from the four observational products. For transparency, this could be rephrased (see the relevant Specific Comment above).*

>> The reviewer is right. We rephrased L.94 : "In this study, four observation-based products, their ensemble-mean, and one reanalysis were analysed."

- *L106: It seems (from L134) that the re-gridding was performed before calculating the MHW metrics, but it would be helpful to discuss the potential impact of the re-gridding (see the relevant Specific Comment above). For example, regridding to a coarser resolution may smooth small-scale features and potentially reduce maximum intensities. Could an alternative be to regrid the coarser products to a finer scale?*

>> We thank the reviewer for this remark and answered him in his/her specific comment above.

- *L114: Please rephrase the description for the composite product in the table for clarity, e.g., Mean of the four SST analysis products, having regridded C3S, CRW and OSTIA to the OISST 0.25° grid.*

>> Corrected.

- *L118: “not symmetric” -> about the equator.*

>> Corrected.

- *L120: “detailed” -> conducted?*

>> Corrected.

- *L131: This figure shows the composite in panel (a) and the ensemble mean in panels (b-g). It might be better to make the panels consistent.*

>> The reviewer is right. We modified panels b-g of Fig. 1 to represent the ensemble mean without the COMPOSITE. Yet, the mean SST in panel (a) actually corresponds to the COMPOSITE since it is the mean of the four observational products. The figure was modified in the new version of the manuscript.

- *L136: “Gaps of less than two days… were ignored”. Do you really mean that they were ignored? I.e., not taken into consideration for calculating the duration or cumulative intensity? Or do you mean that they did not split the event into two (as is usually done)?*

>> We meant that they did not split the event into two, as usually done. The sentence was corrected L. 136-137 to make this point clearer : “Events separated by fewer than two days were considered as a single continuous event”.

- *L151: Should the units of DHW be °C.weeks, and not °C.weeks^{-1}?*

>> The reviewer is right. Units were corrected in the whole manuscipt.

- *L166 and L168: “maps were defined at each pixel”. Please rephrase, since the wording implies that maps are created at each pixel, rather than pixel values within a map.*

>> Corrected L.181-182 : “For duration, maximum intensity, cumulative intensity and onset rate, each pixel of these maps was defined as the mean value across all MHWs detected between 1993 and 2021 in this pixel.”

- *L169-176: Including the composite dataset in ensemble mean effectively results in the four observational products being counted twice. For statistical consistency, it would be fairer to exclude the composite from the ensemble mean (see the relevant Specific Comment above).*

>> The reviewer is right. We made the appropriate changes and answered with details the reviewer's comment in the Specific Comments section.

- *L187-188: As above, the counting of simultaneous events should exclude the composite, otherwise the statistics are skewed.*

>> The composite was excluded from the counting in the new version of the manuscript.

- *L194: "time start" -> onset?*

>> Corrected L.213.

- *L204: "defined in Eq. 3" -> defined as, since the equation follows directly after.*

>> Corrected L.223.

- *L207: "p_value" -> p-value. And elsewhere throughout the text.*

>> Corrected L.226.

- *L207: "inferior" -> less than. Alternatively, all of this sentence could be rephrased in terms of the statistical significance at the 99% confidence level.*

>> Corrected L.226.

- *L214: In the PEQD region, the MHWs are typically El Niño events. ENSO is mentioned in this paragraph, but not El Niño specifically.*

>> The reviewer is right. ENSO was here replaced by El Niño L.234.

- *L243-250: The large differences in counts of MHW days per year across products is somewhat surprising. Given the use of a 90th percentile threshold, one might expect around 10% of days to qualify as MHW days, with deviations arising from the persistence criterion and asymmetries in temperature variability. However, the magnitude of the differences shown here appears larger than might be expected. It would be helpful to explore this further, for example by discussing whether differences in long-term trends, variability, or dataset characteristics could contribute to the tendency for some products (i.e., CRW, OSTIA, OISST) to underestimate, and others (i.e., C3S and GLORYS) to overestimate, MHW days. A brief discussion that considers recent findings (Brunner, L., Voigt, A. Pitfalls in diagnosing temperature extremes. Nat Commun 15, 2087 (2024). https://doi.org/10.1038/s41467-024-46349-x) could help to contextualise these results.*

>> Indeed, this point is interesting and was further analysed in section 4.2 "Potential explanations of these differences". The maps of Appendix A1 showing the standard deviations of the high frequency SST signals for all products might explain these differences, as explained in 4.2 : GLORYS and C3S show lower high frequency standard deviations compared to CRW, OSTIA and OISST. Such characteristics of the SST signal might explain these differences and result in a lower number of MHWs days per year for GLORYS and

C3S compared with the other products. This point is further illustrated by the time series of Fig. 12.

- *L262-265: The methodology used to derive the “ranking” is not entirely clear, and would benefit from further clarification. Taking MHW days as an example, GLORYS appears to have the highest mean value and is therefore assigned a value of 1, but it is not clear how the corresponding value for OISST (approximately 0.17) is determined. Is this based on a domain-averaged quantity? Furthermore, the term “ranking” may be misleading here, as it typically implies an ordinal ordering (e.g. 1 to 6), whereas the values shown appear to represent a continuous or scaled metric. Clarifying both the calculation and the terminology would improve transparency.*

>> The spider chart of Fig. 3 is indeed based on a domain-averaged quantity. More precisely, for each product, we compute the average of the MHW metric evaluated for all events detected in the whole domain over the period 1993-2021. We then get the minimum and maximum values across products. Then, for each product, we subtract the minimum value across products to its average value and divide it by the difference between the maximum and minimum values so that the product reaching the highest average value will get a value of 1 and the product reaching the lowest minimum value will get a value of 0.

We decided to keep this radar chart with decimal numbers so it can help to check if some products are closer to the maximum or minimum values. If we take the example of the maximum intensity, in terms of ranking OISST is first, followed by OSTIA, CRW, C3S, GLORYS and the COMPOSITE. Yet, the radar chart adds the information that OISST really shows higher maximum intensity values and rather behave as an outlier as OSTIA, CRW and C3S which are closer and around 0.4.

Calculation and terminology were clarified in L.283-289 :

“To summarize inter-product differences, the average of MHW metrics for all events detected in the whole domain over 1993-2021 were computed for each product. For a given metric, the minima and maxima across products of these average values were defined so that values on Fig. 3 represent the average value of the given product minus the minimum across products, divided by the difference between maximum and minimum. This normalization standardizes the results between 0 (product ranking last) and 1 (product ranking first) allowing an easier comparison between products. The distance to the center also gives the information on whether the averaged values of some products are closer to others and if some products behave differently.”

- *L288: Some care may be needed in the interpretation that the “onset rate is the most sensitive metric”. In this context, “sensitivity” could be interpreted as having a physical meaning, whereas the result appears to indicate that the onset rate exhibits the largest normalised dispersion across products. Clarifying this distinction, or rephrasing the statement, would help avoid potential ambiguity.*

>> Corrected L.316 : “The ensemble dispersion normalized by the ensemble mean for each MHW metric (Fig. 4b-g, section 2.3.1) highlights that the onset rate exhibits the largest normalised dispersion across products.”

- *L388: What does "SP7" refer to?*

>> Corrected : mistake, replaced by Fig. S7 (Supplementary) L.417.

- *L451: "go" -> be?*

>> Corrected.

- *L459-464: this single pixel analysis is a good example of where the effects of regridding (as noted in the Specific Comments) should be considered.*

>> Yes, that's why we conducted our analysis on the impact of regridding in one small area of the Warm Pool subregion, and the area around New Caledonia.