Reply to Reviewer 1:

We would like to thank the reviewer for their careful reading and constructive suggestions, which have significantly improved the presentation of our manuscript. We have revised the paper in line with the provided comments, and have also taken the opportunity to present some additional important insights following the provided recommendations. These actions have actually improved the quality of our manuscript.

Please, find below a point-by-point response to all comments. For ease of reference, we have assigned a number to each comment. The reviewer's comments are presented in blue font, while the authors' responses are presented in black font.

General review

C1. In this manuscript, the authors introduce the new (to my knowledge) and interesting concept of extreme marine summers. The manuscript presents a huge amount of work from the analysis of EMS to the different drivers/mechanisms involved in such EMS with interesting results.

But, in my opinion, there is a fundamental problem regarding the definition of EMS. When reading the manuscript, I could not get a clear idea of how EMS are defined.

Thank you very much for pointing out the need for a better explained definition of EMSs. It is essential that the reader understands the proposed concept.

We identify extreme summers within 1950-2020, separately at each grid point, on the basis of mean seasonal values. In particular, we define EMSs at each grid point as the summers (Jul-Aug-Sep) presenting an average summer SST above the locally computed 95th percentile of the mean summer SST values within 1950-2020. Considering the 71 summers of the study period, there are 4 summers exceeding this percentile threshold at each location.

An improved introduction of the new term has been included in the updated Introduction section, and a clearer and detailed definition has been included in the updated Methods of the revised manuscript, according to the clarifications provided also in our answers below.

C2. First, the authors define EMS as "define EMSs, separately at each grid point, as the four summers with the highest average JAS SST, i.e., exceeding the 95th percentile of the 71 available summer periods from 1950 to 2020". I would expect a basin definition for EMS, maybe coming after the grid point analysis. As we are talking about a season, I would expect that the categorization as extreme would affect the whole region (not with the same intensity in all grid points but with a big enough extent).

Thank you for your comment. The definition we propose applies to grid points, so the categorization of a summer as extreme does not affect the whole basin. We define EMSs locally, on the basis of mean summer SST values at each grid point. We basically treat extreme seasons as events, which may occur at a specific location. In this way, each grid point experiences its own locally detected extreme seasons. Of course, like with marine heatwaves (MHW), we expect some spatial continuity. For instance, summer 2015 (example included in the manuscript) is identified as extreme in a great part of the basin. Neighboring areas that do not experience this specific summer as extreme (grey sea areas in Fig. 10) present a mean summer SST in 2015 that does not exceed the local (per grid point) threshold. These areas have potentially been warmer than usual, but not enough to satisfy the selected criterion, as expected with any local threshold-exceedance approach.

We agree that taking into account spatial extent is interesting and useful in the context of understanding extreme conditions affecting the basin. This is why we discussed an example case, summer 2015, being the EMS with the greatest spatial extent (i.e., with the greatest number of locations experiencing the same summer as extreme). However, this study primarily aims to treat EMSs in a statistical way rather than focus on specific summer cases. The current approach allows for investigating SST variability patterns during EMSs throughout the basin following the methodology of Röthlisberger et al. (2020). By applying a spatial extent criterion on top of the existent threshold, a different number or EMSs would be detected at different locations. In addition, locations experiencing extreme summer conditions that do not present a significant spatial extent, would then be excluded. The followed approach includes extreme seasons experienced at local scale allowing for a statistical analysis of their characteristics in a consistent way across the basin.

C3. Later in the manuscript the authors analyse different variables for the whole basin, calculate mean summer values,… maybe suggesting a basin-wide concept. But later there is a section named “Differences among local extreme marine summers”.

Thank you for this comment. We do not suggest a basin-wide EMS concept. Each grid point experiences 4 locally identified EMSs. Therefore, the set of 4 EMS years for every location is not the same throughout the basin. The concept of extremity in this study relies on local SST variability. For the EMS detection we use a percentile-based threshold calculated at each location, rather than an absolute threshold value, or a threshold calculated for the whole basin. As a consequence, there are no locations with a different number of assigned EMS-years, or not experiencing EMSs at all. For this reason, there are no spatial gaps in the figures depicting variables during EMSs, except when we show results for a single summer (example EMS 2015; Fig. 10), as it was not experienced as extreme by every single location.

The Section “Differences among extreme marine summers” examines how robust is the concept of a mean EMS state, i.e., of averaging quantities over the 4 local EMSs at each grid point, not averaging over the basin. This is clearly stated in the beginning of this section in the revised manuscript.

C4. Another question that is not clarified is the choice of the 95% percentile.

Thank you for your comment. The 95th percentile of mean summer SST values (at each grid point separately) was selected after performing sensitivity tests changing the threshold value. These tests aimed to check if important changes are observed in our results when using different thresholds, due to different characteristics of specific summers. These tests showed a consistent spatial distribution of SST anomalies and SST substructures in EMSs (as in Fig. 3a). Of course, when higher thresholds are used, the magnitude of the SST anomalies in EMSs increases, and vice versa.

Following these tests, we considered the 95th percentile (top 4 summers out of the 71 available summers within the study period) as a good compromise between the "extremity" of EMSs and a sufficient number of EMSs to be analyzed per grid point. A lower (higher) threshold would lead to the detection of less (more) "extreme" summers. On our concerns on the extreme character of EMSs (in relation to the threshold selection and the dataset's length), a larger dataset based on ensemble model data providing several realizations for the same period would certainly increase statistical confidence in our analysis (as Röthlisberger et al. (2020) did, in addition to the use of reanalysis data). For the moment, this remains a challenging idea for future work.

C5. Why not 90% as for extreme air and sea temperature to define extreme events?

The 90th percentile commonly used in MHW detection (also in this work) is applied on daily values. On the other hand, EMSs here are defined on the basis of mean seasonal values, i.e., without taking into account how daily SSTs are distributed during the season (see also our answer on C4).

C6. Why 4 summers? Are there only 4 summers exceeding the percentile? Is this needed for any single grid-point or as a basin mean? Are the 4 summers the same for the whole basin?

Thank you for this comment. As noted in C4, the selected 95th percentile threshold (top 5 out of 100 mean summer SST values) corresponds to the 4 warmest summers out of the 71 available summers of our study period. It is calculated separately for each location, based on the time-series of mean summer SST values of each grid point (we do not use mean basin values). Therefore, for each grid point, the 4 highest mean summer (JAS) SST values correspond to the EMS-years at this specific location. Consequently, the 4 EMS-years are not the same for the whole basin.

C7. Figure 4 shows results for three grid points where EMS occur in different years. I don't feel comfortable with a year being extreme in WMED but not in CMED. Or having a pixel suffering an EMS but not the closest one (figure 10 grey areas).

There is, of course, the possibility I did not properly understand the proposed definition. In any case, better explanation of the actual EMS definition is needed.

Thank you for pointing out the need for a better explanation. Please, see our answer in C2. We explain therein the proposed EMS definition and our choice to treat the occurrence of an extreme summer as a local event.

C8. I also have some major concerns about the analysis of MHWs. Please, see comments below.

Thank you very much for your detailed comments on the MHW analysis. Our answers are provided following your specific comments later in this document.

C9. The authors have done extensive, very solid and interesting work on the analysis of EMS drivers. A review of the above will add value to all this work and may lead to a really interesting publication. So, I encourage the authors to send a revised version of the manuscript and please understand these criticisms as a way to improve an already good work.

From the above comments, my recommendation is that the manuscript needs major revision before it is accepted for publication.

Thank you very much for this review. We believe your comments have greatly contributed in improving the paper. We have put our effort to provide a well revised version incorporating your recommendations, primarily focusing on providing a) a clearer understanding of the proposed concept of EMS and related methods and b) an improved presentation of the MHW analysis.

2.2 Methods

2.2.1 Extreme marine summer definition and associated SST substructures

C10. In lines 142-144 you define “as the four summers with the highest average JAS SST, i.e., exceeding the 95th percentile of the 71 available summer periods from 1950 to 2020”. Why four? Are they the only ones exceeding the 95th percentile?

Thank you for your comment. Please, check our answer in C4. As the 95th percentile is computed separately at every grid point, the corresponding SST threshold value varies across the basin. By definition though, the same number of threshold exceedances exist at each location (4 in our case, considering the 71-year dataset).

C11. Why not another percentile?

Please, check our answer in C4.

C12. Have you checked if this percentile excludes some major MHW events?

EMSs are identified based on mean seasonal values, without taking into account how SST evolves within the season. If the SST anomalies during a major MHW are not large enough to make mean summer SST exceed the EMS detection threshold, this summer will not be identified as extreme at the examined location. In that case, such an event will not be included in the events occurring within EMSs.

We generally expect to meet higher MHW activity during EMSs compared to mean summer MHW conditions. This is simply because EMSs are defined on the basis of mean seasonal SST, and the elevated SSTs during MHWs are expected to contribute to the increase of the mean seasonal SST. Results in the MHW section generally confirm this assumption, also through the addition of a short analysis for MHW days following your relevant suggestion (please, see C22 for these results).

Going through the review, we understand that the presentation of methods needs improvements to increase clarity. It is important to clarify that only after detecting EMSs we investigate how they have become extreme. After detecting EMSs, the analysis of SST substructures provides information on SST distribution patterns during EMSs at different locations. Later on, the comparison of mean MHW conditions during EMSs against mean summer MHW conditions aims to further inform on the SST distribution during EMSs from an event-occurrence perspective. We have put our effort to update the relevant parts with a clearer explanation of the EMS concept and methods that depend on their detection.

C13. Please, explain clearly the process to select/exclude EMS/non EMS years. Do you choose the percentile first and then there are only four? Do you choose 4 years and look for a common percentile? I think I understand the methodology, and mostly agree, but please explain better to avoid confusion. EMS definition needs to be more robustly justified.

Thank you for your comment. Please, see our relevant answers in C4 and C6.

C14. What will happen when 2022 and 2023 are included in the study? 95th percentile will rise and maybe (probably) some of the 4 years will not meet this condition. Which impact can this have in the definition or the subsequent analysis. Please, if the data for 2022 are available, check this question.

Thank you for this comment. Having similar concerns when first applying the methodology, we had performed an additional test using different sub-periods (within 1950-2020) for the EMS detection and analysis of SST substructures. This was done to examine specific years significantly alter our results. These tests showed that, despite the expected differences in magnitude of SST anomalies, the spatial distribution of SST anomalies in EMSs and results for the revealed SST substructures are, to a great extent, independent of the study period.

In the same context, having concerns on the impact of summer 2003 (being particularly warm and associated with severe MHW conditions in a significant portion of the basin) on our results, we had repeated the analysis excluding this year. This naturally resulted in a lower SST threshold. However, results again showed very similar SST substructures, despite resulting from a different combination of locally identified EMSs.

On these grounds, we conclude that the SST substructures during EMSs primarily stem from fundamental characteristics of the climatological variability of summer SST in the basin, rather than being attributable to exceptional cases. This is further supported by our results based on detrended SST (already included in the manuscript). Specifically, despite the different combination of 4 EMS years and the absence of warming trend in the detrended dataset, results highly resemble the ones based on the original one. Therefore, we believe that the 71-year study period, even if not updated with the very latest data, is able to capture the SST substructures during anomalously warm summer conditions in the Mediterranean Sea in a present climate context.

2.2.2 Detrending time series

C15. In this section Figure 2a is mentioned. In this figure time series for sub-basins are shown and sub-basins are defined in the figure caption. It would be better to explain the sub-basins, if they are used throughout the paper for analysis, in the methodology section and it would be good to add (up to you) a figure showing them or some indication in one of the actual figures. You should also explain how you decided the limits of the basins (22º, Sicily strait,…) Is the Thyrrenian sea included in the WMED or CMED? Please explain.

Figure 2a is included in the manuscript to provide the reader an idea on the SST temporal evolution within the study period. Neither the mean basin values nor the ones for the considered sub-basins in this figure are used in our analysis, as the entire work is based on analysis per grid point.

Time-series for the different sub-basins derived from the ERA5 dataset are included only to provide an idea on the non-uniform warming trend across the basin based on the dataset in use, also showing consistency with literature. The choice of the 3 sub-regions is only indicative, and area-averaged values for these sub-regions are not used in any part of the paper. As regards the Thyrrenian Sea, it is included in WMED. By stating *"Strait of Sicily up to 22°"* in the legend of Fig. 2, we meant the strait itself being the boundary separating the 2 sub-regions, as shown in Fig. Rev 1. This has been rephrased in the revised document for clarity.

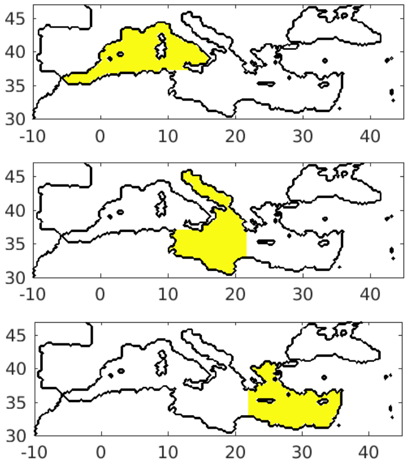


Figure Rev 1 Western, Central and Eastern Med sub-regions used to compute the example time-series of Fig. 2 in the manuscript

2.2.3 Marine heatwave identification

C16. For the detection of MHW events you use the fixed climatology 1983-2012, which selection is adequately explained. In a recent work Rosselló et al suggest the use of "a moving baseline covering the 20 years prior to the year under study". As you try to remove the effect of multidecadal trend, the use of a moving climatology it could also be useful. This is just a suggestion you can try or at least keep in mind for future work.

Rosselló, P., Pascual, A., & Combes, V. (2023). Assessing marine heat waves in the Mediterranean Sea: a comparison of fixed and moving baseline methods. Frontiers in Marine Science, 10. https://doi.org/10.3389/fmars.2023.1168368

Thank you for this suggestion. The use of shifting climatologies in MHW detection is indeed the most recently suggested approach for investigating MHW properties in a changing climate. A 20-year length moving baseline would remove the effect of long-term trends and would not require selecting a specific climatological period which is, to some extent, arbitrary. However, for consistency reasons, we used the same detrending method (Empirical Mode Decomposition) in any task of this work where we aimed to remove the multi-decadal trend. For MHWs, we aimed to perform a direct inter-comparison of event properties derived from the original vs the detrended dataset. For this reason, minimizing the methodological discrepancies between the two runs was considered adequate. We certainly keep in mind this robust method for removing the effect of long-term warming trends in MHW detection, for future works.

C17. Lines 195-196: “Finally, to focus on summer MHWs, we isolated events with their onset and end day falling within the JAS summer period.” Have you checked how many events are removed by using this condition? How many MHW days are removed? In the case an event started on June 28 and lasted for 30 days, would it be considered? Do you apply any spatial extent filter to detect MHWs? Just to avoid single pixel heat spikes.

Thank you for this comment. The issues you raise have also been the authors’ concerns. The fixed temporal length of a season inevitably leads either to the underestimation of the number of MHWs by including only events beginning and ending within the season (approach 1), or its overestimation by including also those that fall partly within the season (approach 2). We understand that mean frequency values computed based on both approaches generate similar concerns (related to under- and over-estimation of MHWs count, respectively). However, the aim of this section is to inter-compare MHW properties between extreme and non-extreme summers over the 71 years. For this reason, we believe that the followed approach (1) is able to serve this objective despite the discussed uncertainty.

We performed two tasks in order to a) examine how many events are excluded using the followed approach and b) provide an additional way for assessing the role of MHWs in EMSs, that does not suffer from this uncertainty. The latter was done by using MHW days, and results are included in C22 (and in the revised manuscript) following your relevant suggestion thereinx.

As regards (a), we took into account also events starting before 01 Jul or ending after 30 Sep, with at least half of their duration falling within the JAS period (to avoid including events occurring mainly during spring/autumn). The number of additional events when following approach 2, is mapped in the following fig. Rev 2:

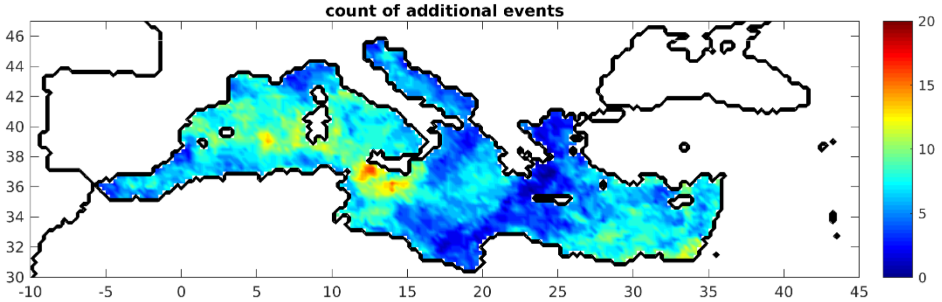


Figure Rev 2 Additional summer events within 1950-2020 when using approach 2 (including also events with at least half of their duration within JAS period)

The total MHW count within 1950-2020 and the corresponding annual frequency for the 2 approaches are presented in Fig. Rev 3: left, same approach as in the manuscript; right, for the new test including extra events:

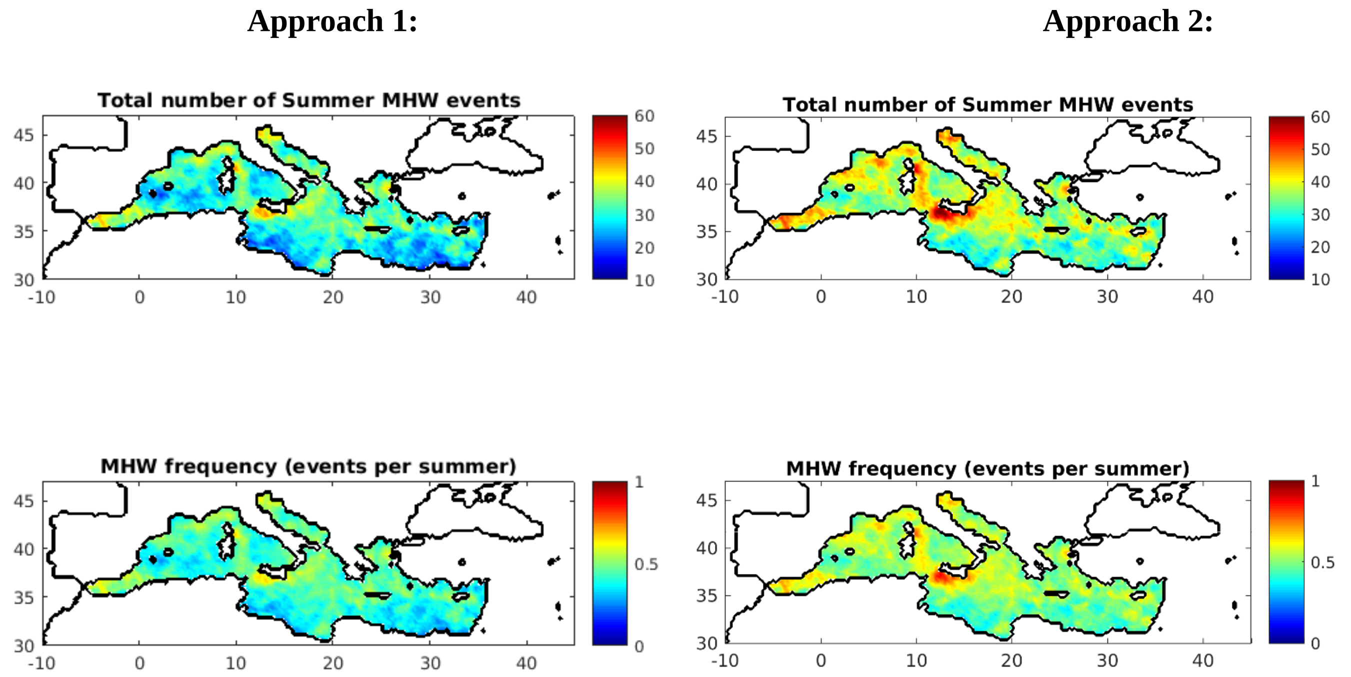


Figure Rev 3 Left column: Total number of summer MHWs within 1950-2020 and mean frequency (events/year) using Approach 1; Right: Same as left but using Approach 2

The additional summer events within 1950-2020 when using approach 2 are less than 8 in most of the basin (Fig. Rev 2). The increased number of events (and the corresponding increased frequency) present a quite similar spatial distribution compared to the previous results (Fig. Rev 3).

We consider that either of the two approaches is appropriate in the context of this analysis, as it solely aims to compare events among extreme and non-extreme summers. However, as intensity and duration of events (apart from event frequency) are also inter-compared between extreme and non-extreme summers, we believe that the followed approach (1) may be preferable because it incorporates information derived strictly from JAS periods.

Based on the above, we believe that it is adequate to keep the current approach, enriching the corresponding part of the methodology. In particular, we will discuss the uncertainty associated with counting summer events and computing mean summer frequency in the revised text, briefly explaining that results from different approaches do not significantly alter our conclusion on whether and where increased MHW occurrence is met. Importantly, we will also include an additional way for assessing the role of MHWs in EMSs, free of this limitation, based on MHW days. In this respect, please see our answer including relevant results in C22.

As regards spatial perspective in MHW detection, we have not applied any filter for spatial extent in this work. The entire paper is based on analysis per grid point, as explained in the clarifications provided for the proposed EMS definition. Similarly, we do not advance the Hobday et al. (2016) per pixel methodology for detecting MHWs, as the initially identified EMSs at each grid point (without any spatial filtering) are then used in the MHW analysis. We acknowledge studies suggesting spatial limitations to retain the most important MHW cases and exclude potential spikes. However, considering the objectives of this section, along with the fact that we are not targeting impacts, which would make spatial limitations of great relevance (e.g., Pastor and Khodayar, 2023), we consider that potential downsides of the employed per-pixel approach are not major in the context of this task.

3.1.1 Decomposing the extreme marine summers

C18. Lines 242-243

“the largest contribution to the mean EMS RDA at this location comes from the warmest part of the SST distribution (higher rank days), with the exception of EMS 1999 where the middle part contributes the most.” Has this been calculated? Is there a numerical value? I can understand from looking at figure 4a but, what exactly is “the middle part”? Do you divide the rank daily anomaly in 3 periods of 30 days? Do you divide the RDA by some anomaly threshold? Line 244: Why “the 60 warmest rank days”?

I suppose you are following the methods of Roethlisberger (2020) but a clearer explanation would be good.

In this section you comment figure 3 with contributions of the warmest/coldest half to the EMS but in the analysis of the three locations you refer to “the middle part”. It would be better to use the same process for EMS analysis/characterization. Or explain what you mean as “the middle part”? Is it a percentile interval of the rank days? Is there an anomaly threshold to define the RDA parts? You also mention “the 60 warmest rank days” for 2003. Please clearly explain the methodology/process to analyse rank days series.

Thank you for pointing this out. We agree that mentioning other than the 2 parts introduced in Methods (coldest/warmest half of 92 ranked SST values) confuses the reader. Although basic characteristics of other parts (“*the middle part”, “the 60 warmest rank days”*) may be visually easy to observe in the discussed examples, these sentences should be rephrased for clarity. We have modified them without referring to finer parts of the distribution, to be consistent within the text and avoid confusions.

Regarding your question on how we divide the SST distribution, we do not employ any anomaly threshold or SST percentile-based intervals. SST values for each summer have been ranked before their division into different parts. Therefore, by dividing the distribution in any count of equal parts (two in our case) from the lowest to the highest rank days, we get colder towards warmer parts, of equal temporal length.

Then, following Röthlisberger et al. (2020), the contribution from the two parts (three in their study) of the SST distribution to the seasonal SST anomaly is calculated for a specific EMS *s* as follows:

$$Contr_{coldest}(s) = \left(\frac{1}{D} \sum_{d=1}^{\frac{D}{2}} RDA_{d,s} \right) / SA_s \quad (Eq. 1)$$

$$Contr_{warmest}(s) = \left(\frac{1}{D} \sum_{d=\frac{D}{2}+1}^{D} RDA_{d,s} \right) / SA_s \quad (Eq. 2)$$

, where D=92 summer days, $RDA_{d,s}$ is the SST anomaly of the ranked day *d* of the EMS *s*, with respect to the mean daily SST of the specific rank day over the 71 summers, and SA_s is the seasonal SST anomaly of the EMS *s* with respect to the mean SST over the 71 summers:

$$SA_{EMS} = \frac{1}{92} \sum_{d=1}^{92} SST_{d,EMS} - \frac{1}{71} \sum_{s=1}^{71} \frac{1}{92} \sum_{d=1}^{92} SST_{d,s}$$

Figures 3b-c in our manuscript map these contribution percentages considering all 4 EMSs at each grid point, as follows:

$$Contr_{coldest} = \frac{1}{4} \sum_{EMS=1}^{4} \frac{1}{92} \sum_{d=1}^{46} RDA_{d,EMS} / MA \quad (Eq. 3)$$

$$Contr_{warmest} = \frac{1}{4} \sum_{EMS=1}^{4} \frac{1}{92} \sum_{d=47}^{92} RDA_{d,EMS} / MA \quad (Eq. 4)$$

, where MA is the mean SST anomaly of the 4 EMS*:*

$$MA = \frac{1}{4} \sum_{EMS=1}^{4} \frac{1}{92} \sum_{d=1}^{92} SST_{d,EMS} - \frac{1}{71} \sum_{s=1}^{71} \frac{1}{92} \sum_{d=1}^{92} SST_{d,s}$$

Similarly, we could split the ranked SST distribution in more than 2 parts, e.g. in 3 parts of equal length (coldest-middle-warmest) by changing the rank day summation limits (d = 1 -> D/3, d = D/3+1 -> 2D/3, d = 2D/3+1 -> D, respectively). This was actually our initial approach. However, after checking that the same main conclusion is obtained through the two-part analysis, we chose to present the latter, being probably more easy to follow.

Regarding the quantification of the relative contributions, numerical values (%) for the mean EMS state are already provided in Fig. 3b-c, (computed for the 2 parts based on Eq. 3-4). Discussing SST substructures at the example locations of Fig. 4 mainly aims to present the methodology by showing in more detail (looking into the different local EMSs) what Fig. 3b-c provides for the mean EMS state. We have however included in the revised text these values (contribution percentages of Eq. 3-4) for the three example locations.

In addition to the aforementioned modifications, we have adjusted Methods (Sect. 2.2.1) including a clearer explanation on the process to apply the Röthlisberger et al. (2020) methodology in our case (including also the Eq. 3-4).

3.1.3 Line 300 SST substructures using detrended data

C19. If I understood well, there are no relevant differences in the analysis from detrended SST data.

Yes, despite the smaller magnitude of SST anomalies in EMSs resulting from the detrended dataset due to the removal of any long-term warming trend, SST substructures in EMSs throughout the basin are found to be remarkably similar with the ones based on the original one.

3.2.2 Marine heatwave properties during extreme marine summers

C20. Line 384 “EMSs exhibit lower intensity relative to the mean summer conditions”. Do you mean “mean summer MHW conditions”?

Yes, thank you for pointing this out. Anomalies presented in this section are mean values for each MHW property among the 4 locally detected EMSs, with respect to the mean value of the MHW property over all 71 summers (always per grid point).

C21. Lines 386-387 “This suggests either the suppression of MHWs in EMSs or the existence of alternative mechanisms potentially enhancing the MHW conditions in EMSs”. Please, rephrase in a clearer way (can’t understand).

Thank you for this comment. We agree that this sentence is not clear. It refers to the negative MHW intensity anomalies in the southern parts of the basin. These anomalies suggest that MHWs are less intense during EMSs compared to mean MHW conditions in these areas. However, mean MHW intensity is only one way to describe MHW conditions. Considering the other two examined properties (frequency and duration), we see that although being less intense, MHWs in these areas occur more frequently and last longer during EMSs compared to mean MHW conditions.

We have removed this sentence, along with the previous one referring to mean-basin values. The aforementioned findings are supported in the revised manuscript without any reference to quantities averaged over the basin. For this reason, we have also removed the few extra sentences mentioning mean-basin metrics along with Table 1 that includes such values, as explained in our answer in C26. Reformulation of this section also involves the addition of our results for MHW days (according to our answer in C22), as an additional means to compare MHW conditions between extreme and non-extreme summers.

C22. Line 389: You say MHWs in EMS are more frequent in a just 4 summers series. I’m not sure this can be statistically sound. Please, consider the count of summer days meeting MHW conditions as an alternative variable to analyse.

Thank you for this comment. This sentence means to report that MHWs occur more frequently during EMSs compared to non-extreme summers.

In case your concerns originate from an unclear explanation of the way we compute EMS anomalies based on the 4 EMSs, please consider the following clarification. Our step-by-step process is as follows: Firstly, we detect all summer MHWs and compute their properties (intensity, duration, frequency) at each grid point. Secondly, we compute the mean value of each property over all summers. This is what we consider as mean MHW conditions (in terms of this property) for the summer season. Then, we compute the mean value of this property over the 4 EMSs. This is what we consider as mean MHW conditions in EMSs. EMS anomaly values for the examined MHW quantities are computed at each grid point, by subtracting the mean value over the 71 summers from the mean value over the 4 locally detected EMSs (mean EMS value – mean summer value). Importantly, to compute EMS anomalies of frequency, we do not subtract number of events, but events per year. Therefore, the sentence in Line 389 reporting that MHWs occur more frequently during EMSs compared to non-extreme summers, means that the number of events per EMS is greater than the normally expected number of events per summer.

In case your concerns are related to the limitations of the way we count summer events and compute mean frequency for EMSs, please see our clarifications on the relevant comments C17 and C26.

In any case, we agree on the added value of using MHW days. Following this request, MHW days during each summer within 1950-2020 were counted, taking into account also events that fall partly within the JAS period. The relative contribution of MHWs during EMSs with respect to mean summer MHW conditions, can be expressed in terms of MHW days, through the percentage of MHW summer days occurring within EMSs (Fig. Rev 4). Results show greater contribution percentages in the central and eastern Mediterranean, suggesting a more pronounced role of MHWs during EMSs in these areas, this time by means of temporal coverage of MHW conditions.

Based on the above, we have added Fig. Rev 4 in the Appendix, and have updated the MHW methods and relevant discussion part accordingly. This addition complements the existent results of Fig. 6 providing an additional metric for the role of MHWs in EMSs, free from the previously discussed (C17) uncertainty associated with counting events within a fixed period.

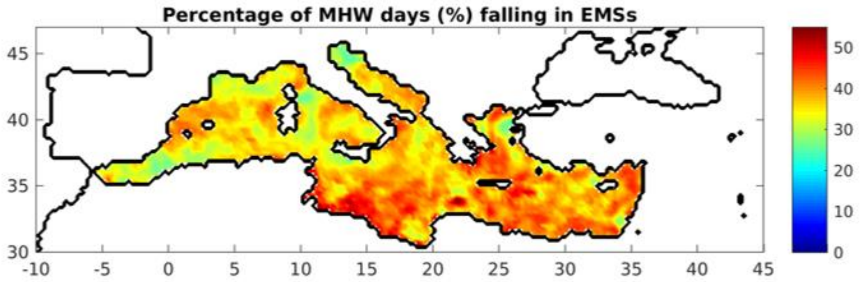


Figure Rev 4 Percentage of MHW days falling within EMSs (with respect to total count of MHWs within 1950-2020)

C23. Lines 395-397: Please, reconsider the frequency analysis.

Thank you for this comment. Please, see our answers in C17 and C22.

C24. Line 400: Have you checked that MHWs occur in the "warmer summer days" during the 4 EMS? You can indeed have MHW events in the "colder summer days". How do you attribute MHWs to the different parts of the RDA?

Thank you very much for this important comment. It motivated an actually fruitful analysis with interesting results. This paragraph (Line 400 onwards) is indeed not sufficiently supported. It relates results from the SST substructures and MHW analysis by attributing MHWs to specific parts of the SST distribution based on assumptions.

We conducted an additional analysis using MHW days, as follows: Before ranking SST values, we created a binary dataset storing if there is an active MHW at each summer day, or not. After ranking SST for each summer, we paired the ranked SST values with the above information (MHW day or not), creating a 2nd binary dataset free of temporal order. The count of MHW days during the 1st and 2nd half of this dataset corresponds to the MHW days falling within the warmest and the coldest half of the rank day distribution, respectively. In this way, for each summer we were able to compute the relative contribution (%) of MHW days of the warmest and the coldest half with respect to the total count of MHW days during the summer. Figure Rev 5 shows the percentage of MHW days falling in the warmest part of the SST distribution, considering all (4 locally identified) EMSs. The corresponding percentage for the coldest part is simply the remaining percentage.

Results show that the majority of MHW days during EMSs take place within the warmest half of the SST distribution in the entire basin (Fig. Rev 5). Particularly in northern regions (northwestern Med, Adriatic, Aegean Seas), very large percentages (locally exceeding 95%) of MHW days fall within the warmest part of the SST distribution, while high percentages are also encountered in the Alboran and the Ionian Sea to the southwest of Crete. The lowest values, of about 50%, are observed in certain spots in the eastern Levantine Sea, showing that MHW days during EMSs tend to be uniformly distributed over the warmest and the coldest part of the SST distribution at these locations.

Importantly, the northern sub-regions and areas in the southeastern Ionian Sea where the vast majority of MHW days in EMSs occur within the warmest part of the summer, are also the areas presenting the higher EMS anomalies for all MHW properties (see Fig. 6d-e in the manuscript). This finding suggests that the more intense, more frequent and longer lasting events observed during EMSs occur within the warmest part of the SST distribution.

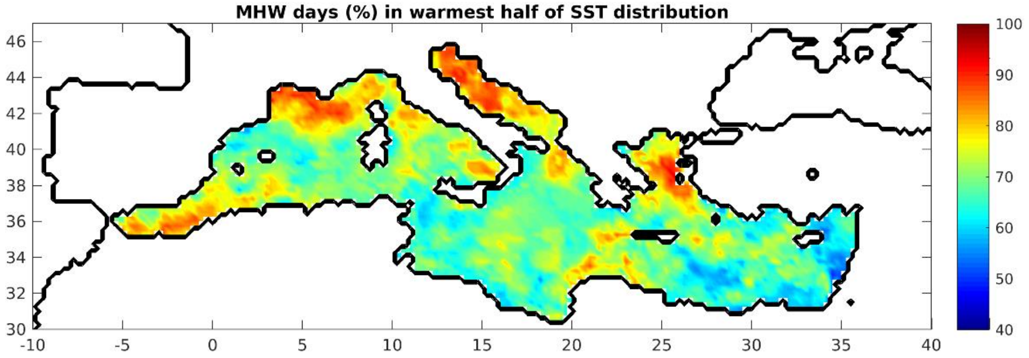
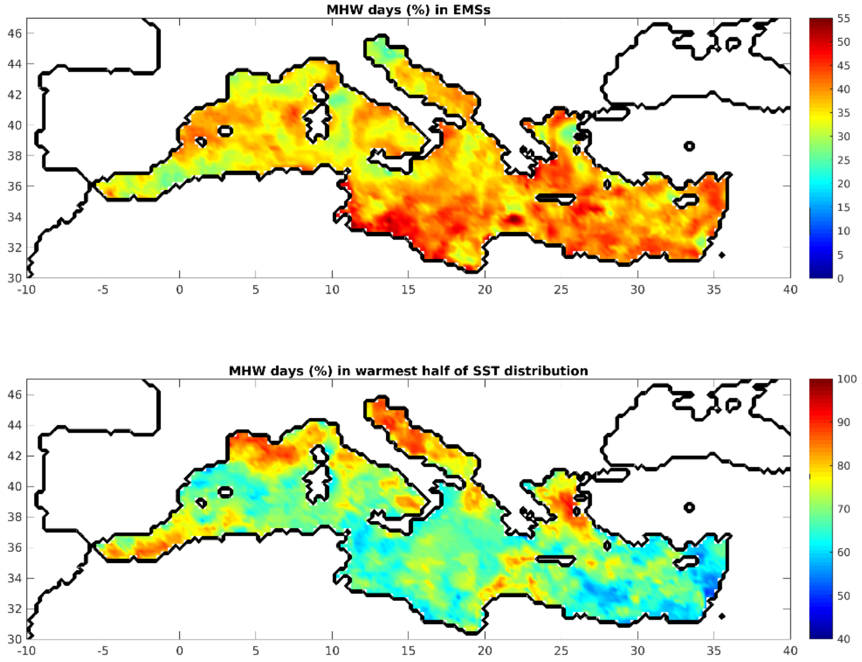


Figure Rev 5 Percentage of MHW days falling within the warmest part of the SST distribution during EMSs

Based on the above, we have reformulated this part of the manuscript, including basic results from this additional analysis and Fig. Rev5. Specifically, we suggest the inclusion of this figure (Fig. Rev5) in the Appendix, merged with the relevant Fig. Rev4 presented in C22, as follows:



Suggested figure for Appendix: *Percentage of MHW days falling within EMSs, with respect to total count of MHWs within 1950-2020 (top); Percentage of MHW days falling within the warmest part of the SST distribution during EMSs (bottom)*

C25. Lines 414-415 “Given, however, the smaller variability of daily SST values during summers in the EMED, we conclude that the relative role of MHW events in the formation of EMSs is more pronounced in the EMED”. I’m not sure this is strongly supported from the previous lines. It seems logical but it needs a more robust support. Please, extend this paragraph to support this affirmation.

Thank you very much for noting this. This part should have been more thoroughly examined. To do so, here we examine the percentage difference of MHW intensity in EMSs relative to mean MHW conditions.

The produced normalized anomalies (Fig. Rev6-right) are very similar to the absolute anomalies of Fig. 6d of the manuscript (included herein in Fig. Rev6-left). The absolute increase in mean intensity during EMSs in the Adriatic, Aegean and parts of the Ionian Seas is slightly smaller than the increase observed in the northwestern basin (Fig. Rev6-left). However, normalized anomalies (Fig. Rev6-right) suggest that these intensity changes in EMSs are of approximately equal importance across the aforementioned areas. For instance, mean MHW intensity in the NW Med reaches 2.5 °C, being significantly larger than in the Aegean Sea (1.7 °C). The corresponding mean intensity anomalies in EMSs are 0.35 °C and 0.25 °C, for the NW Med and the Aegean, respectively. These anomaly values correspond to a relative increase in MHW intensity of about 14% and 14.7% for the two areas, respectively. Such small differences reveal that our initial statement regarding the relative importance of the observed changes is not supported by these results and thus we modified this part accordingly.

Nevertheless, from the perspective of temporal coverage of MHW conditions, the relative contribution of MHW days in EMSs reveals a more pronounced role of MHWs in EMSs in areas of the central and eastern Mediterranean (C22, new Fig. Rev4).

Based on the above, we have modified this sentence considering the new results from the MHW days analysis included in C22.

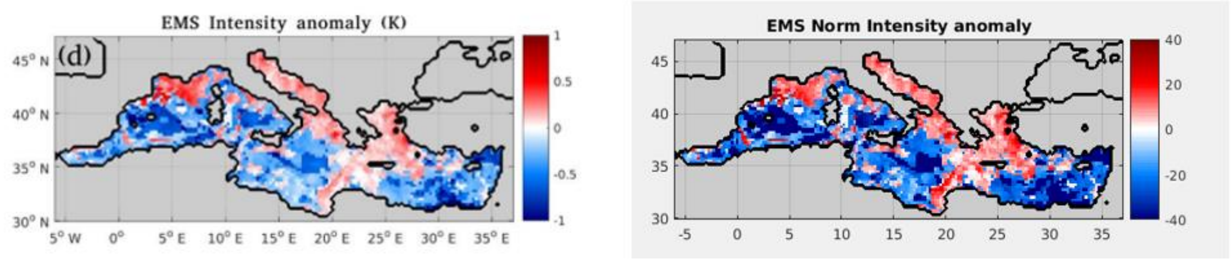


Figure Rev 6 Left: MHW intensity anomaly in EMSs; Right: Normalized intensity anomaly (percentage difference (%) of MHW intensity in EMSs relative to mean intensity)

C26. Figure 6 (d,e,f). These figures show some MHW anomalies in the EMS. I see all grid points in the Mediterranean present MHW anomalies. Do you mean that any place in the Mediterranean experienced an MHW in EMS, at least one time? I’m not sure if these maps make sense, unless you can state that there have been MHWs in all grid points during EMS.

Thank you for this comment. We agree that the methodology or the discussion of results currently do not explain why all grid points in the basin present MHW anomalies. MHWs do not occur every summer at each grid point. However, they occur at every single grid point during EMSs, as shown in the following figure (Fig. Rev 7) mapping their count. A relevant note has been included in the revised manuscript.

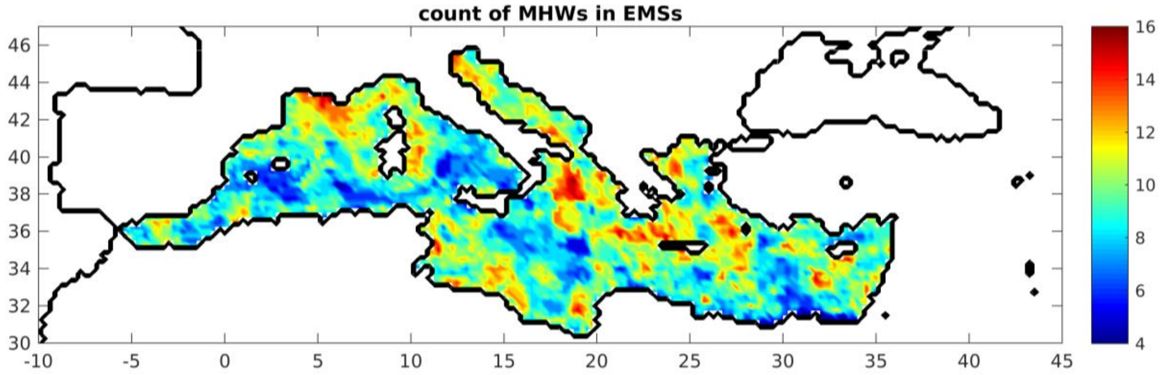


Figure Rev 7 Number of MHWs falling within the locally identified EMSs

C27. Table 1: In caption you say "EMS identified in 1950-2020". Please, state that this are only 4 years. This whole section needs to be reconsidered. I don't think that the way the four EMS are compared to the mean 1950-2020 series is appropriate (see previous comments and figure 6) comments. Consider that the four EMS experienced strong and/or long events. You should also analyse which Med areas were affected to properly decide if basin averages are adequate. It is possible that MHW events did not affect the majority of the basin so spatial MHW averages could not be appropriate, please check and justify your assumptions.

Thank you for this comment.

Regarding the comparison of extreme vs non-extreme conditions:

For any variable in the paper, we consider that the mean EMS conditions are represented by the mean value of that variable over the 4 EMSs, at each location. The corresponding EMS anomalies are computed by subtracting the mean value of the variable over the 71 summers from the EMS-mean. This is how EMS anomalies are computed per grid point throughout the paper. In the same way we compute the EMS anomalies also for MHW properties, as explained in C22.

On any concerns about the event occurrences across the basin in relation to the computation of EMS anomalies, we show in C26 that all grid points experience MHWs during EMSs. This has been noted in the revised text when introducing EMS anomalies for MHW properties. Clarifications on relevant concerns provided in C17 and C22, as well as the new analysis of MHW days, have been included in the revised manuscript (please, see our answers in the above comments).

In case your concerns are related to the count of EMS-years (4), please consider our explanations in C4, C6, C22. In the context of comparing extreme and mean conditions, we consider that 4 out of 71 summers is an adequate sample to represent extreme cases.

Regarding the averaging over the basin:

Using mean basin values in this Table may indeed not be appropriate, considering the spatial differences "masked" in a single value (such as the smaller than usual intensity found over southern regions during EMSs in Fig. 6d). We also understand that no valuable insight is added through this Table, as the principal conclusions, in the context of our objectives for this section, can be derived through the maps of Fig. 6. Separating the basin into different sub-regions based on similar MHW behaviour during EMSs and then compute regional means would be adequate, still not providing additional information on top of Fig. 6. Considering also that no mean-basin approach is used in any other part of this paper, we have removed this table and the short discussion parts that refer to these mean-basin values.

3.3.4 Illustrative example of extreme marine summer 2015

C28. Lines 658-659: What does "Summer 2015 has been the most widely experienced EMS in the basin" mean? Is it possible that part of the basin experiences EMS but not the whole basin? Does EMS refer to extreme summer for the whole basin or not?

Thank you for your comment. Please, check our clarifications on the definition of EMSs provided for C2. "Summer 2015 has been the most widely experienced EMS in the basin" means that this was the summer with the greatest number of locations experiencing it as extreme. Although its spatial extent is the greatest among all detected EMSs, EMS 2015 (like every EMS) does not refer to (and does not imply extreme conditions for) the whole basin. Following the proposed definition, it is possible that only a part of the basin experiences a specific EMS.

C29. Line 659: "non-grey locations in figure 8". Does this refer to figure 10?

Thank you very much for this correction. Yes, it is a typo, the correct reference is Fig. 10.

C30. Figure 10: What are the grey areas? Non-EMS points?

Yes, grey areas are the ones that have not experienced summer 2015 as extreme.

C31. Figure 10: "Non-coloured sea grid points in the Mediterranean Sea stand for locations that did not experience summer 2015 as extreme". Does your definition apply only to grid points? Is there any EMS for the whole basin? Is there an spatial extent threshold to declare a summer as an EMS for the basin?

Thank you for your comment. Please, see our answers to these questions in the relevant comments C2 and C28. A clearer and more detailed explanation of the definition has been included in the revised version of the manuscript, based on these answers.

3.3.5 Air-sea heat fluxes using detrended data

C32. Do you detrend all datasets? Wind, SWR,… ?

Yes, we detrended all datasets following the same methodology as with SST to remove the multi-decadal trend (mentioned in Methods, lines 172-174).