Reply to Reviewer 2:

We would like to thank the reviewer for the valuable feedback on our manuscript. We have revised the paper in line with the provided comments and recommendations.
Please, find below a point-by-point response to all comments. For ease of reference, the reviewer's comments are presented in blue font, while the authors' responses are presented in black font.

The paper untitled "Investigating extreme marine summers in the Mediterranean Sea" - by Dimitra Denaxa Gerasimos Korres, Emmanouil Flaounas, and Maria Hatzaki – investigates Marine Heat Waves (MHW) during "extreme marine summers" in the Mediterranean Sea. The authors use a reanalysis to identify MHW characteristics and understand the main drivers explaining regional differences during particularly hot summers.

In general, I found the paper is well-written with a straightforward general structure. I really appreciated efforts of authors to clearly describe their results, even if some parts could be shortened to reduce the long length of the manuscript.

However, I have several concerns from the beginning of the paper regarding methodological choices:

1. I do not understand why the authors decide to only consider 4 events as "extreme marine summers". Even if Figure 4 gives an idea about the distribution of summer SST, this is only at three locations. The authors need to give a more precise definition of what they call "extreme marine summer" regarding the distribution.

Thank you for your comment.

As regards our choice to consider 4 extreme marine summers (EMSs):

EMSs at a particular location are the summers presenting a mean summer (Jul-Aug-Sep) SST above the local 95th percentile of mean summer values within 1950-2020. This percentile threshold (top 5 out of 100 mean summer SST values) corresponds to the <u>4 warmest summers out of the 71 available summers of our study period</u>. It is calculated separately for each location, based on the time-series of mean summer SST values of each grid point. It should be noted here that it is after detecting EMSs on the basis of mean seasonal SST, that we investigate daily SST distributions and MHW characteristics during these extreme summers.

The 95% percentile 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). 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) detection threshold would increase (decrease) the number of EMSs, leading to less (more) "extreme" summer conditions.

As regards Figure 4:

Discussing SST substructures at the example locations of Fig. 4 aims to present the methodology by showing in more detail what Fig. 3 shows for the entire basin using averages over the 4 EMSs. In this way we obtain a closer look into EMS conditions, as we may see the SST distributions during the different EMSs at the same location, and how the different parts of each distribution may contribute to the EMS anomalies. Please, see also our answer in minor comment for L239.

As regards the *"more precise definition of what they call "extreme marine summer" regarding the distribution":*

EMSs are identified based on mean seasonal values, without taking into account how SST evolves during the season. It is after detecting EMSs that we investigate how they have become extreme. Then, it is the analysis of SST substructures that provides information on SST distribution patterns during EMSs at different locations. Later on, the MHW analysis aims to further inform on the SST distribution during EMSs from an event-occurrence perspective.

To conclude, a summer may get unusually warm (in terms of mean seasonal SST) under different daily SST distributions during the season. But for the identification of EMSs, we do not consider any criterion relevant to the shape of the SST distribution.

We agree that the presentation of the proposed definition for EMSs needs improvements to increase clarity. We have updated the relevant parts with a clearer explanation of the EMS concept and detection method in the revised manuscript, based on the above clarifications.

2. My second concern is a lack of justification about applying methodologies for each "grid point". Results gives a very statistical point of view. Is there any criterion about spatial scales? This is confusing as they give physical interpretation associated with regional patterns. I would at least explain why they are not considering EMS composites.

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 of each grid point. We treat extreme seasons as events, which may occur at a specific location. In this way, each grid point experiences its own locally detected EMSs. Of course, like with marine heatwaves (MHW), we expect some continuity across the basin. For instance, summer 2015 (example included in the manuscript) is identified as extreme in a great part of the basin.

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 (or continuity) criterion, or using composites, a different number or EMSs would be detected at different locations.

Moreover, locations experiencing extreme summer conditions that do not present a significant spatial extent, would then be excluded. The followed approach captures extreme seasons experienced at local scale allowing for a statistical analysis of their characteristics in a consistent way across the basin. We have included a brief explanation of this choice in the revised Methods.

On your concerns on the physical interpretation associated with regional patterns, please see our answer in the following comment (3). We explain therein the pre-processing tests we initially performed to ensure that the per-grid-point analysis and the fixed number of considered EMSs do not significantly alter the resulting spatial patterns.

3. If I understand Figure 3 correctly, it needs more explanation since patterns might results from one "very extreme" event for instance, and might not be representative of the 4 events. This also might be an issue in other figures.

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 if 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 EMS detection 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 (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.

4. It misses a discussion about the interest for the community of this new term "ESM".

Thank you for pointing this out. Extreme marine seasons constitute a new concept and thus require a clear definition as well as explanation of our motivation and scientific interest.

The idea of studying extreme seasons was based on the need to better understand elevated SST conditions on the seasonal timescale. As a lot of research has focused on long term warming trends and warm ocean extreme events, we believe that investigating extreme seasons in the Mediterranean Sea provides interesting insights from a different perspective.

First of all, the summer season was selected due to the greater summer SST trends (in past and future periods) compared to other seasons. After defining extreme marine seasons, on the basis of mean seasonal SST, we aimed to investigate their characteristics. The first question we begun with, is how the warmest marine summers in the basin have become such. Here comes the analysis of SST substructures during extreme summers, providing information on these summers in relation to climatological summer conditions. Such analysis of extreme seasons has not been performed (to our knowledge) for the Med. Sea and is potentially relevant to impacts on marine life, as marine species do not experience the same thermal stress in any case of elevated summer temperatures, e.g., when the coldest, or the warmest summer days are warmer than normal.

We understand (thanks to all reviewers) that it is important to prevent potential confusion between EMSs and MHWs, as EMSs are obviously related to warm events. We naturally expect a higher MHW activity during EMSs compared to mean summer MHW conditions, because MHWs contribute to the increase of the mean seasonal SST. However, MHWs are events that typically last for days and may reach several weeks according to their causing factors and the concurrent atmospheric and oceanic conditions regulating their duration. On the other hand, an extreme marine season is, by definition, of fixed duration. Extreme summer analysis therefore constitutes a different perspective for investigating anomalously warm surface conditions in the basin.

The revised manuscript explains the scientific interest behind the concept of EMS, based on the above. Relevant additions/updates (on the EMS definition, difference with MHWs, motivation and scientific interest) have been included in the Introduction and Methods of the revised manuscript.

I think these points need to be addressed before introducing this new concept. A major revision is required to clarify these points.

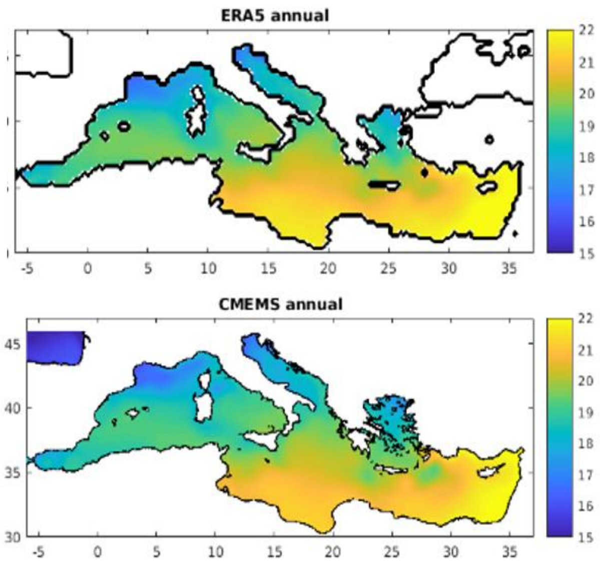
Minor comments:

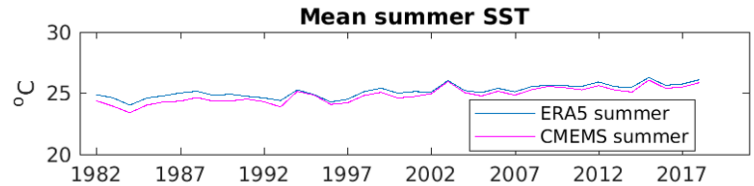
L128 – "derives from a combination of HadISST2 and OSTIA datasets" – ERA5 SST is based on a reanalysis which includes model and observations datasets. I would reformulate.

Thank you for this suggestion. This sentence has been rephrased in the revised manuscript.

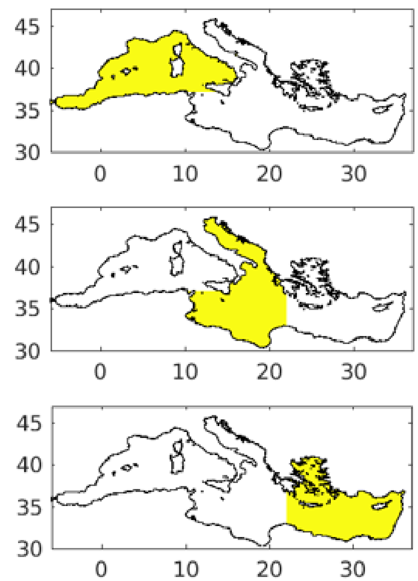
L133 – "In addition to ERA5 (reference SST dataset), we use the CMEMS L4 satellite SST product (EU Copernicus Marine Service Product, 2022c) for the period 1982-2019, at 0.05°x0.05° grid spacing, to cross check the quality of the reference dataset against a high-resolution observational SST dataset." What did you compare? What did you conclude?

Thank you for this comment. We inter-compared mean annual/seasonal SST values and corresponding linear trends for the Med. Sea and sub-basins using the 2 datasets (ERA5 vs CMEMS SST) for a common period. Indicatively, we show some basic inter-comparison results:



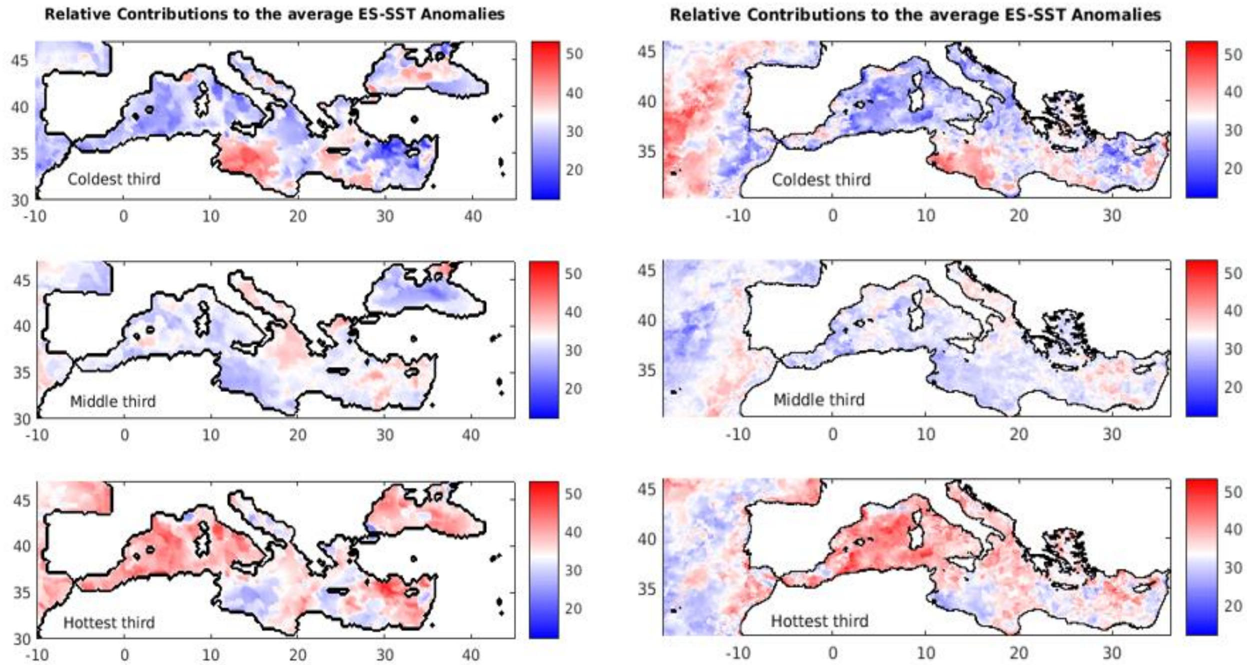


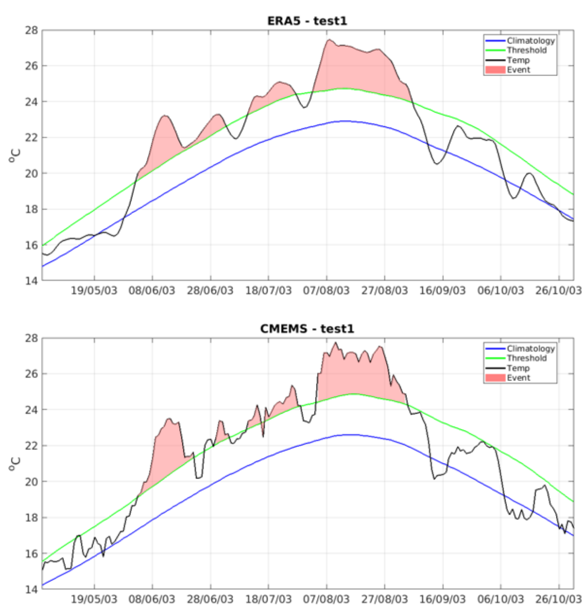
Linear trends in Med regions (deg/year) 1982-2018	ERA5	CMEMS
Summer JAS	0.039	0.048
WMED	0.031	0.040
CMED	0.036	0.044
EMED	0.049	0.060



The ERA5 dataset slightly overestimates SST against CMEMS observations, with a larger overestimation observed during the first years of the common period (1982-2018) and an improved match over the most recent years, as shown through the mean summer time-series for the 2 datasets presented above (top-right). ERA5 SST presents smaller increasing trends across the basin (see indicative trend values in the Table for the 3 yellow-colored sub-regions).

Satellite SST was additionally used in many tasks performed to investigate EMSs (e.g., rank day anomalies methodology, MHW detection), in the context of ensuring that the reanalysis produces similar results with the observations. As examples of such inter-comparisons, we show below the contributions from 3 different parts of the SST distribution (coldest-middle-warmest) to the rank day SST anomalies of EMSs (top left for ERA5; top right for CMEMS), and how an extreme MHW is captured by the 2 datasets (bottom).





Differences fall within acceptable ranges, considering that we inter-compare modeled against observational SST. Although we considered such tests a necessary pre-processing task, we opted not to include any relevant results as they fall out of the paper's objectives and also to avoid increasing the length of the manuscript. We agree however on reporting in Methods that inter-comparison of ERA5 against CMEMS SST revealed a satisfactory level of similarity among the two datasets. This information has been included in the revised manuscript.

L143 – “We then 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. “ As the authors want to introduce a new concept, it needs more explanation and justification.

Thank you for pointing out the need for a clearer definition for extreme marine seasons, as they constitute a new concept.

We define EMSs, at a particular location, as the summers presenting a mean seasonal SST above the 95th percentile of mean summer SST values within 1950-2020. The selected 95th percentile threshold 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. The period July-August-September (JAS) has been considered as the marine summer season, being the warmest part of the SST annual cycle in the Med. Sea (Pastor et al., 2020).

The EMS definition in Introduction and Methods has been rephrased in a clearer and more detailed manner in the revised manuscript, based on the above explanation and including clarifications requested by the three reviewers.

L239 – “we provide examples of different patterns for three grid points of the domain ». Why these points? Are they representative a regional patterns?

Thank you for your comment. The main spatial pattern of SST substructures is the increased contribution from the higher rank days in the greatest part of the basin, with main exception an area in the central-southern basin along the African coasts (as appears in Fig. 3). The reason we included Fig. 4 is to provide examples of SST distributions for different EMSs, as such information is “hidden” in the fields of Fig. 3 that include averages over the 4 local EMSs.

In particular, the location in the Levantine Sea shows a greater contribution of the warmest part to the seasonal anomaly, with SST distributions similar among the local EMSs. At the location in WMED the warmest half again contributes the most to the seasonal SST anomaly in all EMSs, but the shape of the SST distribution here varies significantly among the EMSs. This difference is related to the greater summer SST variability in the WMED (as discussed in the manuscript based on variance and contributions from different parts to variance). In contrast, the location in the southern-central basin shows a greater contribution from the lower rank days of the SST distribution during every local EMSs.

In other words, the examples of Fig. 4 confirm the information on SST substructures presented in Fig. 3, and serve to illustrate the rank anomalies methodology, by showing local rank day distributions of SST for every EMS along with the local rank day variability of SST over the study period.

Figure 4 – This figure is interesting, but this arises questions. I would imagine with such "Extreme Marine Summer" concept, that the spatial scales would be close to those of the basin. Maybe, the authors should show several plot with seasonal SST anomaly during two or three typical ESM.

Thank you for this comment. Please, see our answer in Major Comment (2). We support therein our methodological choice to define and study extreme seasons separately at each grid point of the domain. The followed approach aims to capture extreme summers at local scale. Such EMSs may not always be significant from a basinwide perspective. Although there are EMSs experienced by a great portion of the basin (as the example provided for EMS 2015), this is not a given.

L627 – "These values indicate that oceanic processes are primarily responsible for the observed EMS SSTs in these areas " – You often refer to oceanic processes, but can you give references that support this result.

Thank you for your comment. It is important to first clarify that when we refer to the role of oceanic processes in driving EMSs we refer to the contribution of the residual term R in Eq. A1. The metric we constructed to quantify the role of air-sea heat fluxes in the formation of EMSs provides a percentage contribution at each location. The remaining percentage is therefore attributed to non-heat flux (i.e., oceanic) factors. The role of oceanic factors is therefore supported by this equation, despite any uncertainties relevant to our methodological approach. It is clear that this topic is neither investigated nor discussed based on literature. However, we doubt on the added value of the latter, as this metric is first introduced in this paper and our findings cannot be easily inter-compared with literature. Investigating the role of horizontal advection in the formation of EMSs would provide valuable insights into the relevant role of oceanic processes supporting our analysis. Our choice however for this paper has been to focus only on the role of air-sea heat flux. For this reason, future work recommended by the authors within the manuscript includes a broader assessment of physical mechanisms behind the development of Mediterranean EMSs.