**Reply to Reviewer 1:**

We would like to thank the reviewer for their careful reading, suggestions and encouraging comments. 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.

## Summary

The authors use a bevy of remotely sensed, reanalysis, and modelled products to determine the interplay between SST, Qnet, and MLD in the Mediterranean Sea from 1993-2022. More specifically, the authors investigate what the physical drivers are of MHWs during the onset and decline portions of the these events by decomposing the contribution of specific components of the mixed-layer heat budget equation o the anomalous SST observed during a MHW. These results are also parsed out by season, region, length of the event, and category (e.g. Moderate, Strong, etc.). The methodology used here is previously published, but has not yet been applied to the Med. The results presented here are an important contribution to the MHW literature and I think this paper is going to be very well cited over the coming years.

I have no substantive comments on the intro, methods (which is rare), or conclusions. I only have two minor issues (repeated in the specific comments below). The first is that the authors appear to not have accounted for the loss of shortwave radiation out of the bottom of the MLD in their mixed-layer heat budget equation. This will likely have little impact on the results, but it should either be accounted for, or the authors should argue for why they aren't doing it. The second is that while I found the authors criticism of the methodology for how to account for the onset/decline periods, they unfortunately stop short of providing any sort of advice on what could be done instead.

All the best,

-anonymous

## Title

- Consider "heat flux" rather than "heat fluxes". Heat flux is generally used as a non-count plural noun. As the authors prefer.

Thank you for your comment. We have incorporated this suggestion in the revised manuscript.

## Abstract

- ln 9: "research studies" is a tautology. Rather choose one word or the other.

Thank you, this has been corrected.

- ln 10: "Marine Heatwaves" -> "Marine heatwaves" Or as the editor prefers.

Thank you for this suggestion, we have included it in the revised manuscript (leaving the final decision to the editor).

- ln 16: “Moreover...” Which direction is the relationship? Is their more or less input of heat flux with increased MHW severity?

We observe a smaller contribution of heat flux for events of higher severity, and vice versa. For this reason, we report an "inverse" relationship between the two quantities in the sentence: “*Moreover, an inverse relationship between MHW severity and the contribution of heat fluxes is observed*”, i.e., the greater the severity, the lesser the contribution of heat flux.

ln 21: “have key” -> “have a key”

Thank you, this has been corrected.

- ln 23: “emphasizes the need” Almost all studies of surface MHWs end by saying there is a need for the consideration of subsurface information. The criticism of the limitations caused by the onset/decline phase definition is interesting.

Thank you for you comment. You are right in noting that most recent studies focus or suggest focusing on sub-surface events, which is mainly due to higher relevance to marine life. While this work treats only SST-based events, examining the mixed layer evolution (being the sub-surface information considered here) provided useful insights relevant to the surface events. It is taking this into account that we included this note on the importance of sub-surface information in our concluding remarks.

**Introduction**

- ln 49: “Darmaraki et al. (2023)” Hmm... I’m not sure a talk at a conference can be referenced here. This is not peer reviewed work. I don’t doubt the accuracy of the work though. The authors may want to edit the sentence slightly to account for the non-peer reviewed nature of these findings.

Thank you for your suggestion, we have removed this reference.

**Data and Methods**

- ln 58: It would be nice to list the horizontal resolution of the other products in this paragraph as well.

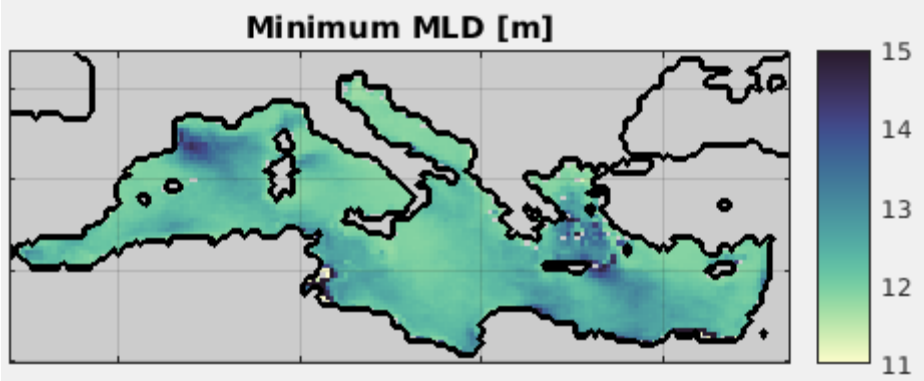
Thank you for your suggestion, we have added this info in this paragraph.

- ln 61: “ref. no. 05” Why not list the products in the text and the table in the same order?

Thank you for noting this, the order has been corrected.

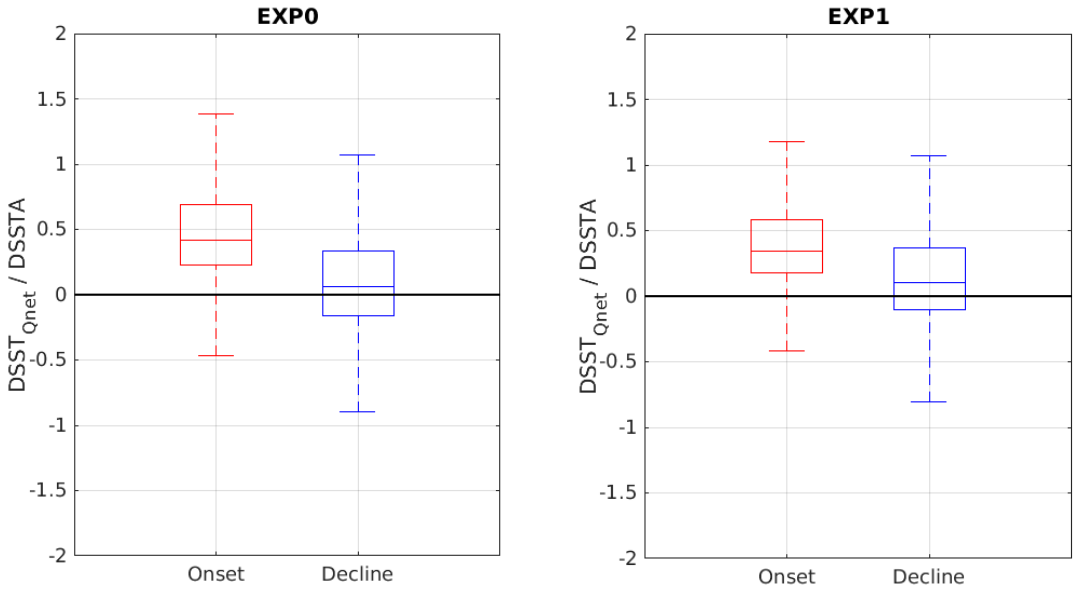
- ln 85: Did you account for the amount of shortwave radiation that passes out of the upper mixed layer?

We thank the reviewer for this comment. We have used a simplified approach that does not account for the penetration of solar radiation below the mixed layer. Considering the Jerlov Water Type IA for relatively clear sea water, 77% of the solar radiation is expected to be absorbed within the upper 10 meters of the ocean (as computed based on the solar radiation attenuation equation in Paulson and Simpson,1977). On these grounds, and taking into account the minimum mixed layer depth values over the study period from the utilized reanalysis dataset (deeper than 11m as presented in the figure below), we have assumed that the followed approach does not significantly affect our conclusions.




Minimum mixed layer depth values during the study period derived from the utilized Med-Physics reanalysis dataset

To verify this, preliminary results have been produced incorporating the discussed parameterization (presented in the boxplots below). These results show the expected effect of the parameterized heat loss, i.e., a slight suppression (enhancement) of the contribution of air-sea heat flux in driving the SST onset (decline). No significant differences are observed and results are in line with the existing ones concluding on the dominant role of oceanic processes during both phases.




Box plots for the contribution of heat flux during onset (red) and decline (blue) of Mediterranean MHWs during 1993-2022. Left: current results without considering the penetration of solar radiation below the mixed layer (EXP0), Right: including this parameterization (EXP1).

## Results

- ln 121: Reporting on MHW frequency is a bit of a tricky thing. This is because, particularly in the Med, we are beginning to approach the time when we have fewer MHWs as they start to last long enough to merge into one long MHW per year. I therefore always recommend to report on total MHW days per year, rather than frequency.

Thank you very much for this suggestion. We totally acknowledge this concern and we can certainly follow the suggested approach in future work, as MHW days are free of this limitation. Nonetheless, we find a statistically significant positive trend for the count of events per year almost over the entire basin (Fig. 1d in the manuscript). The very few exceptions (e.g., to the southeast of Sicily) do not present negative frequency trends, they simply lack statistically significant frequency trends. Therefore, we believe a clear message regarding the increasing count of events per year can confidently be derived from the analysis of this dataset for the considered period.

- ln 136: This sentence is confusing. After reading it a couple of times I understand what the authors are trying to say. It is unfortunately the nature of trying to explain this concept using words. It's not easy to do and I don't have a better recommendation.

- ln 249: "definition..." I agree with the authors to not give the full category definition here.

- ln 243: "events of higher severity categories..." I'm not convinced this is factually accurate.

We thank the reviewer for bringing their attention to this. First, let it be noted that the index we use to address the MHW severity here is the following:

$$S_{i,j,t} = \frac{T_{i,j,t} - T_{i,j,d}^{clim}}{T_{i,j,d}^{90th} - T_{i,j,d}^{clim}}$$

where T is the maximum SST during the event at the i, j location at day t, Tclim is the mean climatological SST at the i, j location for the climatological day-d (that corresponds to the actual day-t), and T90th is the threshold calculated for that location and climatological day. This severity index does not account for the event duration. It represents the event severity in terms of extreme temperatures with respect to local climatological variability. The relationship between the role of heat flux and severity could potentially differ if an alternative index, such as the cumulative intensity during each event, had been employed.

In the figure below we map the used severity index (S) against the duration of onset and decline phases (top and bottom panel, respectively), alongside the count of events. This figure shows that the *severe* and *extreme* cases (corresponding to S>3 according to the categorization scheme of Hobday et al, 2018) actually exhibit shorter durations. We do not claim that the smaller heat flux contribution observed for events of higher S (Fig. 4d-f in the manuscript) arises from the relationship between S and duration as observed in this preliminary test. However, results from this test are potentially associated with our findings for the heat flux contribution during events of different S and were thus considered worth noting. Importantly, as noted in the manuscript, further investigations are needed to unravel how our methodological choices affect these findings.

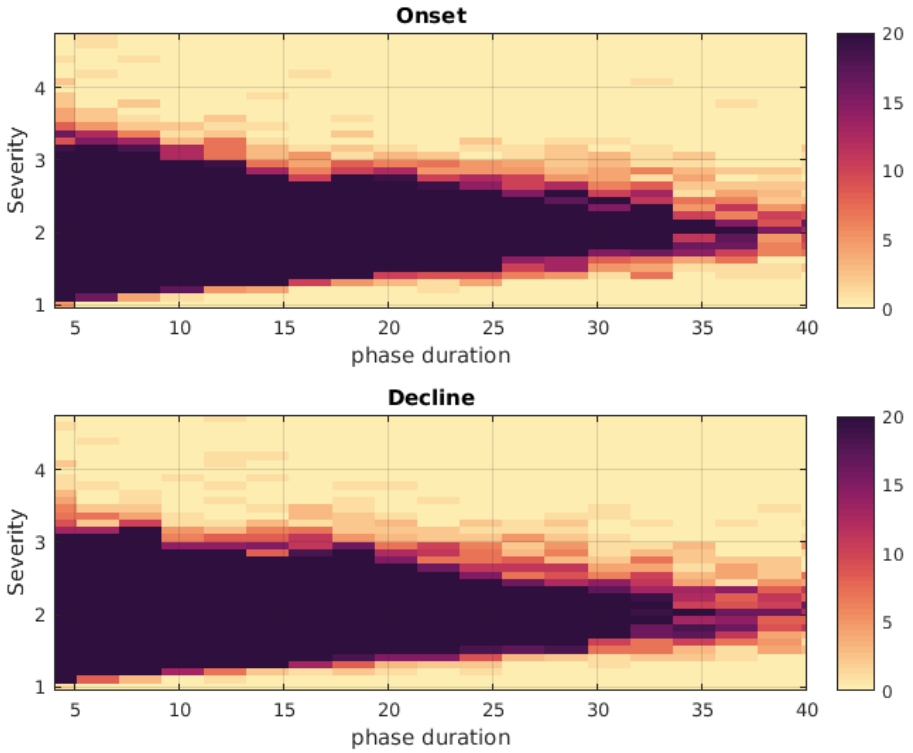



*Figure: Distribution of MHWs in relation to their severity based on the continuous severity index (S) and the duration of the onset and decline phase (top and bottom panel respectively). S ranges (1-2], (2,3], (3,4] and higher than 4 correspond to Moderate, Strong, Severe and Extreme categories respectively. Note: the colorbar is adjusted to allow for visualizing bins with fewer events (a much higher upper limit corresponds to the actual count of events within 1993-2022 throughout the basin).*

## Conclusions

- ln 248: "This..." Delete or merge this sentence with the next one. Not necessary to make this statement.

Thank you for this comment. These sentences have been merged as follows: "*This study investigates the relative role of air-sea heat exchange during MHWs in the Mediterranean Sea, using satellite and reanalysis data within 1993-2022.*"

- ln 265: "These findings..." This is interesting, and makes me think of the basin-wide 2022 event. If oceanic advection is one of the primary drivers of the decline of events, but the whole basin is anomalously warm, and no wind exists to cause vertical mixing, then that would explain how a basin scale event can persist for months. That could be important for a future Med that will be much warmer than now.

Thank you for sharing this thought. Indeed, we believe that this mechanism is most likely responsible for a large part of the spatial and temporal extent of such events. However, the role of oceanic advection in this work is inferred from our analysis and has not been directly investigated. For this reason, we briefly discuss this factor solely in terms of its potential role right after the peak-intensity day, as implied from our actually quantified results (i.e., the contribution of surface heat flux, and the enhanced/suppressed vertical mixing as indicated from mixed layer deepening/shoaling).

- ln 727: "These..." I agree. But what do we do instead? It would be nice if the researchers could propose one or two ideas based on their experiences.

Thank you for this comment. You are right in noting that our suggestion is limited to considering MHW evolution periods aligned with the objectives and specific characteristics of a study. We share this concern and believe that it lies within the wider discussion on challenges associated with the lack of a standardized framework for analyzing MHWs and their drivers in particular. While methodological choices such as the selection of temperature thresholds or the reference climatology (stable or moving) are often supported in relation to impacts on marine species, a series of different approaches have recently been used in quantifying the contribution of physical drivers in absence of a relative discussion. A comparison of methods for quantifying the contribution of potential driving factors has not been performed yet. However, we believe that diverse methodological approaches, such as the definition for MHW phases, can influence findings on MHW drivers. For instance, the continuation of the mixed layer shoaling observed during most decline periods in our study initially surprised us; these results prompted us to examine post-decline periods, revealing delayed (in relation to the considered MHW end day) deepening of the mixed layer. This task highlighted the influence of the methods used to analyze surface vs subsurface information. Similarly, the use of different integration depths when studying drivers is expected to yield varying findings. Therefore, alongside recommending the use of definitions and methods aligned with the specific contexts of individual studies (e.g. SST-based events or events detected based on integrated depths), we underscore the importance of accounting for the associated limitations when interpreting and discussing results. Although we do not propose specific methodological choices, we aim to highlight that clearly articulating the employed ones within a study is vital both for precise interpretation of the corresponding results and for meaningful comparisons across different studies on MHW drivers. This has been stated in a clearer way in the revised manuscript.

**Table 1**

I think it would be useful to give columns for the horizontal resolution, time step, and time series start to end date.

Thank you for this suggestion. We agree, however the content of this Table follows the specific guidelines provided in the context of the Ocean State Report. To our understanding, these

guidelines aim to reduce the datasets-related information within the papers to be included in this special issue, while ensuring that such info can be easily accessed through the references to documentation provided in a table of identical format for all papers. Following your previous comment on products' information in Section Data and Methods, we have enriched this section (but not the Table) adding the spatial resolution of the datasets.

**Figure 1**

Change labels for "[deg]" to "[degC]" or somehow indicate it is degrees Celsius. Same for "[deg/year]". If the longitude labels are only used for the bottom panels, then the latitude labels only need to be used for the left-hand panels. As the authors prefer. Why is a Mann Kendall test used and not a simple linear regression? Are the data significantly non-normally distributed? Generally speaking it is no longer preferable to use the rainbow colour palette. Rather used one of the viridis colour palettes instead (e.g. https://www.mathworks.com/matlabcentral/fileexchange/51986-perceptually-uniform-colormaps).

I like how the Atlantic coastline provides a natural little border for the panel labels :)

Thank you for your suggestions. Units and coordinates in labels have been updated accordingly and the colour palette has been changed as suggested. As regards the computation of trends, linear regression has been used to estimate the trend values. The Mann Kendall test has been applied to assess the statistical significance of the observed trends. The use of linear regression allows us to quantify the magnitude and direction of trends while the Mann Kendall test, being a non-parametric method, allows for the assessment of statistical significance without relying on the assumption of normality.

**Figure 2**

Panel a, colours should be switched for onset/decline to match the rest of the panels. Panel b, I think the comparison of onset vs decline will be communicated better if the bar plots are next to each other, rather than being stacked. Panels d-i, consider filling in, rather than line colouring, for the boxplots. It will make them look more substantial. I agree with the authors choice to limit the y-axis range for the boxplots to +-2 values, allowing longer tales to stretch outside of the plotting range.

Thank you for your comments and suggestions. Colors have been switched (same in Fig. 4). Figure 2c has also been replaced to correct a typo in the latitude label.

**Figure 3**

Same consideration for boxplots as figure 2. Otherwise I like the layout of the panels. I am wondering if there are any spatial patterns across the Med that may explain some of the onset during MLD deepening MHWs, and vice versa?

Thank you for your comments. Analysis of spatial differences of these specific findings across the basin has not been implemented in the context of this paper. However, we are planning to proceed with a more detailed analysis to address this question also combining other parameters (such as wind speed in relation to the observed mixed layer evolution). Such an analysis is expected to allow for a deeper interpretation of the findings presented in Fig. 3 for the entire basin, and an improved understanding of physical mechanisms during onset/decline at sub-regional scale.

**Figure 4**

Swap colours for onset and decline to match other figures. This is an interesting way of visualising these complex results.

Thank you for your suggestion. Figure 4 has been updated accordingly.