Response to reviewers

Review RC1

This paper is a nice analysis of marine heatwaves with depth from an ocean model, which emphasises the role of model drift, baseline choice and resolution on marine heatwaves detection and statistics. Vertical coherence and drivers are also discussed as part of a case study.

The paper is interesting, and a good contribution to the field. However, I think that recommendations around baselines and spin up periods could be clarified, and believe that the heat budget analysis, while not incorrect, may not be fit for purpose.

Thank you for your positive feedback and helpful suggestions to improve our manuscript. As suggested, the revised manuscript contains additional and more clear recommendations and an updated heat budget analysis. Please find a detailed response to the review comments below.

Major Comments

Heat budget

I have concerns about the application of the heat budget. As used in the paper, it allows the diagnosis of the drivers of changes in heat content. However, MHWs are defined as discrete threshold exceedances relative to a local climatology, not by absolute temperature tendencies. Thus, areas of persistent heat convergence are not necessarily directly comparable to discrete MHW events. The manuscript currently suggests a causal link between sustained heat convergence and MHW occurrence or vertical coherence, but does not address this fundamental distinction. In order to prove this causal link, I believe that the heat budget would have to be performed on an event-by-event basis, and vertical coherence would have to be considered very carefully in terms of the boundary conditions for the budget for each event.

While this would be a very interesting analysis, I think it would be beyond the scope of the paper in its current form. In fact, I do not think that the heat budget analysis adds much to the outcomes of this work, and so my recommendation would be to restrict the analysis to that of vertical coherence, and remove, or at least strongly tone down and place the heat budget in the context of warming, and not of MHWs.

Thank you for this suggestion. We fully agree that the link between individual MHWs and the heat budget analysis was too vague. As a consequence, we have replaced these results by an event based approach, as suggested. For that we detected MHWs based on the spatially averaged temperature of the Cape Verde archipelago. In contrast to the area covered by MHWs this allows to obtain a well defined start and end date of individual MHWs events. We then calculate the contribution of the different heat budget terms to the detected MHWs for

each depth level to identify the dominant drivers of MHWs (new figure 10). In a last step we explain how and why the main drivers of individual events are different in certain depth ranges (new figure 11). Corresponding changes can be found in the method section (2.4) and section 3.4

Recommendations

The paper makes recommendations around the use of sufficient model spin up periods, as well as the resolution at which MHW statistics should be calculated. I think that these recommendations will be very useful to researchers planning experiments for MHW use. However, I think the recommendations could be clarified in the text as they are sometimes not clearly laid out.

We thank you for this suggestion and extended our discussion of spin-up strategies and the resolution in the discussion section. We have added more specific recommendations although for the spin-up strategy there is no universally applicable strategy (see our response to your comment below).

Is it possible to include more detail about the spin up process required? How much drift is removed in, for example the 4th cycle as opposed to the 6th? Would a repeat year forcing spin up be sufficient?

We chose to use the two extreme cases here, the shortest and longest spin-ups available for the 1980-2022 time period. The exact spin-up time needed depends on depth, model configuration and forcing. For example, the drift in mid-depth water masses is already small in the 3rd cycle, but as mentioned in the manuscript, the bottom waters may not be in equilibrium even in the 6th cycle. A repeat year spin-up could be sufficient and would allow for distinguishing between intrinsic model drift and forced trends. Since we have no experiment with a repeat year forcing any in-depth comment about this would however be pure speculation. As a consequence, we can only conclude that an adequate spin-up is needed. It should be monitored whether the deep water mass properties stabilise over the course of the spin-up, but we can not provide a specific time or procedure that is valid beyond the procedure we have tested in our VIKING20X configuration here. Nevertheless, we have added more recommendations regarding the spin-up required at different depths in the discussion section (lines 578-597).

I find the finding about resolution very compelling, i.e. that while high resolution is needed to resolve mesoscale processes, MHW statistics can, in most cases be calculated on a coarser resolution grid. I note that this finding is not emphasised in either the abstract or conclusions of the paper and would suggest that it should be.

We thank you for your positive feedback on our study. We have added this result to the abstract. We think it is already mentioned in the Summary and Conclusion section, but we have put more emphasis on this aspect (lines 539-567).

Minor Comments

I found the terminology and methology related baselines a little confusing at times. Smith et al (2025) recently published a detailed investigation of the effects of different baselines. For consistency with future literature, I would suggest citing this paper, and adopting their terminology (e.g. ‘detrended baseline’ instead of ‘linear baseline’)

Thank you for suggesting this study. We have changed the terminology to follow Smith et al. (2025) throughout the manuscript.

I’m confused about the difference between model drift and real temperature trends – using a detrended baseline will remove both indiscriminately, while using a long enough spin up will remove model drift, but preserve natural temperature trends – is this correct? If so, it should be made clearer in the manuscript.

Yes, this is correct. There are “real” trends that are for example caused by the surface forcing (warming, increase in wind stress, …). Additionally, there are trends that only arise from the model adjusting to the initial state, which is not the model's equilibrium state. With a sufficiently long spin-up the forcing related trends are still simulated, but the model is closer to its equilibrium state and therefore model drift is reduced.

References

Smith, K. E., Gupta, A. S., Amaya, D., Benthuysen, J. A., Burrows, M. T., Capotondi, A., ... & Wernberg, T. (2025). Baseline matters: Challenges and implications of different marine heatwave baselines. *Progress in Oceanography*, *231*, 103404.