We thank both referees for their careful and constructive reviews. The comments have improved the manuscript considerably and, in particular, prompted three substantial changes: full compliance with the Hobday (2016) detection methodology, the addition of marine cold spell products, and the deployment of an ERDDAP server that lets users subset and download the data on demand. Together these directly address the central concerns raised about methodological consistency and practical usability. We respond to each comment individually below, with the referees' comments shown in italics and our responses beneath.

Reviewers comments

Reviewer 1

The authors have created an extensive database of MHW results based on the ESA SST CCI v3.0 product. They have applied a numeric algorithm, similar to the Hobday et al. 2016, 2018 definition, to detect MHWs based on a range of different factors. Specifically they have detected events based on different climatology periods, raw vs de-trended data, different minimum lengths, and different percentile thresholds. All of these outputs are saved in a series of NetCDF files and available via HTTPS. While the database appears to be technically sound, I am not convinced that it will see much use in its current form … the Hobday definition has existed for almost a decade now, and databases of global results using this analysis are not new … Now that the files exist online and are publicly accessible it would be best to publish this work here in ESSD.

Thank you for this thoughtful and constructive feedback. We agree that in its previous form the database would have been difficult for users to take advantage of effectively, and the addition of an ERDDAP server, detailed in our responses below, directly addresses this.

The novelty we are offering is not the definition(s) but the pre-computation. To our knowledge there is no global, satellite-derived MHW database that spans multiple definitions and baselines in a single, consistently processed product at this resolution. At present a researcher wishing to work across definitions must run the detection themselves, which is computationally expensive at the resolution of the SST CCI (~4 km, against ~25 km for OISST). Processing a 40-plus-year daily record at that resolution, across many baseline and threshold combinations, is not feasible on a single workstation and requires HPC or dedicated server time.

For institutions and researchers without that capacity, this is a large bottleneck. A centralised database from which MHWs, and now marine cold spells, can be subset under varying definitions therefore has clear value, and we have reworked both the manuscript and the data service to make that value far easier to realise.

If the authors want to spell marine heat waves as three words, 'waves' should be capitalised in the title.

Thank you, a little embarrassing to have a typo in the title! This has now been corrected.

If there are plans to update the database on an annual basis (or more regularly), it would be good to mention that somewhere in the abstract. The other plans for future development, like adding other satellite products, should also be mentioned.

Great point. We have now added a sentence to the abstract noting that we have the ambition to roll this out into the future with new data releases.

L51 : km → km²

Corrected.

L64 : 'MHW events'. Here and throughout the text. A MHW is an event, so writing 'MHW event' is tautology. This is a common grammatical error in the MHW literature.

I had not considered this, having adopted it as nomenclature, but I completely agree. Now that it has been pointed out, its hard to unsee. We have removed the redundant use of 'MHW event' throughout the manuscript.

L91 : v.30 → v3.0

Corrected.

L94 : 'dataset'. Here and throughout. I think the size of this output, and its standardised structure as a series of NetCDF files, could warrant the title of 'database' instead of a dataset. As the authors prefer.

Agreed. Given the size of the output, 'database' is the more accurate term, and we have adopted it throughout.

L113 : extra space after 'x'

Corrected.

L161-168 : I'm not certain that would result in a trend-free time series. I could see how de-trending based on each DOY could still allow for an overall long-term trend in the full time series. Did the authors test the difference in output for their de-trending method vs a simple linear model fit to the full time series at a given pixel?

We did initially compute the trend over the full time series. However, because detection operates on daily exceedances relative to a day-of-year climatology, removing the trend on a per-DOY basis was the more internally consistent choice in this case as it ensures the anomaly that feeds detection is referenced to the same calendar-day baseline used everywhere else in the pipeline. Similarly, as the linear trend is estimated and removed separately for each calendar day, the long-term linear trend is removed from the series at every grid cell. Doing this per calendar day rather than fitting one trend to the whole series also lets the rate of warming differ between seasons, which we felt was more appropriate for marine heatwave detection. We are however open to testing the difference, however this may be more suitable for a comlimentary paper we are working towards assessing how MHWs change globally under varying definitions.

L179-195 : The choice of the moving smooth window is non-negligible … Did the authors perform any sensitivity studies for how the choice of the moving window may create different basic results from the standard Hobday et al. 2016 script?

Note that the basic Hobday methodology performs two smooths, not one. The first is an 11 day window used to calculate the DOY percentiles. A second smooth of a 31 day total window is then applied to the outputs of the first, so that outsized effects from a few extreme years do not overly bias the thresholds. Therefore the methodology employed by the authors is arguably a large departure from the initial methodology, and it would be preferable to provide at least a basic comparison of how this differs.

Thank you for setting this out clearly. We have now reprocessed the entire database to be fully compliant with the Hobday methodology, implementing both smoothing passes (the 11-day percentile window followed by the 31-day climatology smooth) and using the mean rather than the 50th percentile for the climatology. The effect is visible in Figure 4, where the climatology and threshold fields are now considerably smoother.

So does the authors methodology insert an NA value on DOY 60 for non-leap years? … The Hobday algorithm creates DOY 60 from DOY 59 and 61 via a linear interpolation before calculating the seasonal and 90th percentile signals. This may be another source of varying outputs.

Thank you for highlighting this. We have now adopted the same approach as the Hobday algorithm and linearly interpolate DOY 60 from DOY 59 and 61 in non-leap years, before computing the seasonal and percentile signals, removing this as a source of divergence.

So the authors have not provided metrics like mean, maximum, or cumulative intensity? These would be very relevant for a number of applications. That a MHW lasted 10 days with an average anomaly of 0.1°C is very different from an event with an average anomaly of 1.0°C.

This is an interesting point. The reviewer is right that, since we have already worked out the MHW categories for every definition, we could go on to calculate these metrics, and we simply haven't done so in this first version. The slight complication is that things like duration and cumulative intensity belong to a whole event rather than to a single day, so before providing them we would need to decide how best to present them, for example one entry per event, or totals summed over each year. We agree they would be useful, and may add them as a derived product in a future version.

Why add the 10th and 50th percentiles to every 90th, 95th, and 99th percentile climatology file? Wouldn't it be much more space efficient on the server to host these as separate files?

Following this comment, and a related point from Referee #2, we have removed the 50th percentile from the climatology files and replaced it with the mean, consistent with the Hobday method. We have retained the 10th percentile and added the 5th and 1st percentiles, which now support marine cold spell detection, with a corresponding cold-spell variable in the category files.

On the question of separate files, we kept these percentile levels within the same climatology file because they share identical grid, time, and baseline coordinates. Storing them together avoids duplicating those (large) coordinate arrays across many small files and both reduces total storage and made ERDDAP subsetting faster for the user. We are happy to revisit this if the editor would prefer the levels split out.

As the files are organised now, a researcher wanting all years of data for a small study site would need to download all the files on the server, open each one, extract the portion they want, and recombine them. This is generally outside the competence of most non-physical-oceanography marine researchers. Have the authors considered rectangling their data, or providing files organised along latitude or longitude lines? One can create an impressive database, but if it is too complex or inconvenient to use, it will not be used.

This was by far the single biggest limitation of the first iteration of this work. We have addressed it directly by deploying an ERDDAP server, which lets users subset the data spatially and temporally on the server side and download only the slice they need, in the format they need, without ever handling the full files. The data-access section now walks users through several routes: R, Python, the command-line interface, the ERDDAP web interface for point-and-click graphing and subsetting, and direct URL-based download. The ERDDAP service is available here:

https://erddap.dmi.dk/erddap/info/index.html?page=1&itemsPerPage=1000

Users can also graph and subset data interactively:

https://erddap.dmi.dk/erddap/griddap/mhw_raw_1982_2011_climatology_90p_5d_anom.graph

Or download a defined region directly through a constructed URL, for example the North Sea, with the data-access section explaining how to build these links:

https://erddap.dmi.dk/erddap/griddap/mhw_raw_reanalClim_90p_5d_anom.nc?anom[(2023-06-01):(2023-08-31)][(54.0):(60.0)][(2.0):(12.0)]

This effectively delivers the on-demand rectangling the reviewer suggests, without our having to pre-store every possible spatial arrangement of the data.

Additionally, the lack of standard MHW metrics (e.g. duration, cumulative intensity) provided for discrete events will likely prove problematic for most researchers … any use of this database will inherently require a lengthy post-processing analysis by the user. Why not provide files that give these summary metrics, e.g. the standard Hobday-ish output listing each discrete event, its duration, max intensity, and so on?

We agree this is important. However, the heavy lifting, building the climatologies, ranking the percentiles, and classifying every pixel-day into MHW categories at 4 km, has been done. The processing left is far lighter. Once users have subset the daily category fields via ERDDAP, counting an event's duration, its cumulative days, or its mean and maximum intensity is a straightforward tally over time, not a recomputation of the categories themselves. That said, we take the point that ready-made event summaries would lower the barrier further, and we are open to providing these as a derived product in a future version of the database and welcome feedback on whether these should be monthly, or yearly.

Ultimately, who is the target user for this database? In what circumstance would a researcher with a limited study area not simply calculate the results for exactly what they need? Considering that there are already MHW toolboxes for Python, R, and MATLAB, why would they not just use one of these? Most researchers tackling global-scale investigations will already have the necessary competence and their own peculiarities to pursue.

This is a good question and allows us to to state the target users more clearly, The database is built first for researchers who want to study an event or a period, regionally or globally, but who do not have the computing capacity to hold the full SST CCI record in RAM whilst processing, build climatologies and anomalies, rank percentiles, and assign categories themselves. That processing is expensive, particularly at 4 km resolution and across multiple definitions.

As an example, a user wishing to examine how the global proportion of ocean under MHW conditions is changing can now simply grid-average our pre-computed global category fields. As not every user / insitution has access to large servers or HPC, having MHWs pre-computed under varying definitions also helps researchers working at regional or smaller scales, who would otherwise face the same intensive processing for a comparatively small area of interest.

With the addition of ERDDAP we believe both the accessibility and the practical utility of the database are substantially improved, and we have revised Section 6 to make these user groups and use cases explicit.

L403 : The point about climate model evaluation is a tenuous one. Unless the model is in some way related to the ESA CCI SST product, the modelling of temperature extremes tends to be so incongruous between different models that comparing them directly is rarely a fruitful endeavour.

Agreed. This has now been removed.

L411 : The original python code, written in 2016, already allows a user to make all of the changes detailed in this manuscript, including the changing of the DOY moving window. Minus of course the de-trending of the raw data.

L415-420 : A problem with this statement is that the authors themselves have deviated from the base Hobday algorithm, meaning that the results in the database may already differ from *an analysis that uses the exact same climatological baseline, duration limit, etc. Let alone comparisons for results with different baselines.*

The database is now fully compliant with the Hobday algorithm for the baseline definition (two-pass smoothing, mean climatology, leap-year interpolation), so results under the default definition are directly comparable with analyses using the standard method, and the comparisons across baselines are made on that same compliant footing. We have added a sentence to the Summary noting this compliance explicitly.

Reviewer 2

This manuscript introduces a dataset of marine heatwaves identified at each location globally using different baselines and thresholds … While the data product is novel and could make certain analyses more efficient, I agree with Referee #1 that it is unlikely to see much use. The dataset is complete, but I am not convinced it is unique or useful … there are several software packages available that make MHW identification a relatively straightforward and customizable process … A standard data product should present some clear benefit over other satellite and in situ SST data, like OISST or OSTIA, which many MHW studies use as the observational reference.

Thank you. As set out in our responses to Referee #1, the database is aimed primarily at users who lack the computational capacity to derive these products themselves, and at users who want a single, consistently processed reference spanning multiple definitions.

On the choice of SST product, OISST is on a ~25 km grid, which makes fine coastal and submesoscale structure difficult to resolve, whereas the SST CCI is a ~4 km independent climate data record developed to climate-quality standards specifically for long-term consistency.

The presentation quality of the manuscript is good and the writing is clear. The manuscript could be improved by providing more summary statistics for how the different definitions perform over the whole time period … A demonstrated use case for this specific data product would warrant publication. Alternatively, this paper may be better presented as a scientific analysis of the effects of different MHW definitions.

Thank you, and we agree that a demonstrated use case strengthens the paper. We are preparing a complimentary study that uses this database to examine how MHW characteristics change over time under varying definitions, which is exactly the scientific analysis the reviewer describes.

We would prefer to keep the data description and that analysis as two separate papers, in keeping with ESSD's data-focused scope, so that this manuscript documents the product and the companion paper carries the interpretation.

L184 : Why not also compute the 1st and 5th percentiles to facilitate marine cold spell analyses? In section 5, it states the daily climatology files record the 10th, 50th, and 90th/95th/99th percentiles, so why not also include the same levels for MCSs?

An excellent point. We have now added the 1st and 5th percentiles and computed marine cold spells under the same set of varying definitions, with a corresponding cold-spell variable in the category files.

L187 : I agree with Referee #1 that this change from the Hobday definition and methodology should be carefully explained and justified.

We agree. As described in our responses to Referee #1 we have reprocessed the database to be fully Hobday-compliant and have added a clear explanation and justification of the methodology, including the two-pass smoothing and the use of the mean climatology, to Section 3.

L216 : Using the median climatology instead of the mean climatology is another slight departure from the Hobday definition.

Thank you. We have replaced the median (50th percentile) climatology with the mean throughout, so the product is now consistent with the Hobday definition on this point.

Section 3.5 : Mixing multiple thresholds and severity categories may obfuscate the severity of MHWs by requiring both to be reported. For example, a 95th percentile category 2 MHW may also be a 90th percentile category 4 MHW and a 99th percentile category 1 MHW. Multiple thresholds could be used as another way to categorise MHW severity.

We agree, and we think this is best left to the user. Different percentile thresholds give genuinely different and complementary views of severity, and rather than collapse them into a single categorisation we provide each so that users can select the percentile and category scheme most appropriate to their study

Figs. 6, 7 : Consider just showing one of these, as Fig. 7 is just an inset of Fig. 6.

Section 4 : I appreciate choosing a large and well-studied MHW here, but this specific MHW certainly does not show 'typical' summary statistics … there is almost no difference between Fig. 7g and Fig. 7h. The Blob is clearly defined across MHW definitions, but how does that generalise to less severe MHWs?

Agreed that figure 7 is essentially a inset of Figure 6. However it we included it alongside the Blob case, so that the behaviour of less severe and more typical MHWs across definitions could be seen, not only that of a single large extreme event.