**General Comments**

This paper presents an observational dataset documenting extreme ocean temperatures observed at four coastal stations in the vicinity of Australia, referred to as AMDOT-EXT. This dataset serves as an extension of the previously published AMDOT dataset, encompassing both surface and subsurface temperature records.

The study adheres to established standards and utilizes publicly available algorithms to identify and characterize heatwaves.

Given the global concern regarding the warming of the Earth's oceans, the dataset presented herein bears potential significance for climate impact studies and warrants consideration for publication, notwithstanding its inherent limitations. However, a substantial area in need of improvement lies in the discussion of uncertainties.

Thank you for reviewing our paper.

**Specific Comments**

The unique and pertinent attributes of this dataset, despite its limited coverage of only four stations, could be emphasized, particularly in Section 2.

We have added the following text to Section 2:

*"These sites are unique in Australia and the southern hemisphere in that they have been occupied weekly to monthly since the 1940s/50s ..."*

*"The west coast site is close to Rottnest Island (ROT, ~32°) offshore from Perth, Western Australia in approximately 55 m of water and temperatures here are influenced by the Leeuwin current. The southern site is close to Maria Island (MAI, ~42.6°), Tasmania in approximately 90 m of water and is situated in the southern extension region of the EAC. An additional two east coast sites are located close to Sydney (PHA and PHB/PH100, ~34.1°), New South Wales in approximately 50 and 100 m of water, respectively, downstream of the typical EAC separation zone in what are usually eddy dominated waters."*

*"Unlike most long-term ocean temperature data sets , ~~Ocean temperatures~~ measurements have been collected at multiple depths through the water column at PHA, PHB/PH100, MAI and ROT since the 1940/50s."*

Notably, Table 2 reveals a significant discrepancy in the "relative % record available" between station PHA and the other stations (PHB/PH100, MAI, and ROT). For PHA, data availability remains relatively uniform throughout the water column and generally falls below 100% below the surface. Conversely, for the other stations, data availability increases significantly at greater depths compared to the surface, with ROT, for instance, exhibiting 172% data availability at a depth of 29 meters relative to the surface. An explanation for this peculiar behavior would be valuable.

At PHA, we have bottle and CTD profiles only. This means that data availability is consistent throughout the water column over time at the optimal depths. At PHB, MAI, and ROT, we have a mix of bottle, CTD, mooring and satellite data sets. Percentage wise, most of the data contained within the AMDOT-EXT data products were collected after 2008/2009 when we have high-frequency (5 minute) mooring data, however most (and at ROT all) of these data are below the surface. Between 2012 and 2022 at times when we don't have in situ (mooring or CTD data) at the surface, we use satellite data. However, satellite data can be gappy (due to the impact of clouds) which means we have less data at the surface when compared with sub-surface depths.

We have added the following text in Section 3.1:

"*Data availability (Table 1 and Fig. 2) is consistent through the water column at PHA because the temperature record consists of bottle and CTD profiles only. However, at PHB/PH100, MAI, and ROT there is increased sub-surface data availability relative to the surface. This is because the satellite-derived SST data that we use starts in 2012 and has some gaps (dependent on clear skies).*"

Please note that we updated the AMDOT-EXT data products because the Australian Ocean Data Network has provided an improved SST data product, that now provides SST measurements at quality level FV02 between 2019 and 2022, instead of quality level FV01. In the process, it was noticed that the domains selected for area-averaging SST data were approximately 17 to 26 km away from the in situ locations, which has been rectified, with the SST domains now immediately encompassing the in situ locations.

This means that the results in Tables 1 and 3, and figures 2 and 4 have changed slightly, and hence have been updated. Overall, data availability at the surface has increased at PHB/PH100 (~ +15% relative to data availability in the submitted manuscript) but has decreased at MAI (~ -4%) and ROT (~ -8%). There are now more MHWs and HSs at the surface at PHB/PH100. There are also minor changes in statistics at depths below the surface at some sites because of the interpolation technique used to produce the AMDOT climatologies. We have modified the text where necessary to reflect these changes (see manuscript tracked changed document). We now also have a short MCS at MAI, which is shown in figure 2 and in figure 4 along with its characteristics.

The PHA dataset primarily comprises monthly and weekly data which is not consistent with the methodology adopted for the identification of heat waves. It is also located relatively close to the PHB/PH100. It is advisable to provide a more detailed discussion if there is a specific rationale for including this record in the dataset.

We have added the following rationale for including PHA in Section 2:

*"Although we only have bottle and CTD profiles at the PHA site at weekly to monthly sampling intervals, this site has the longest continuous temperature record in Australian waters (1942-present), as well as concurrent biogeochemical sampling. Hence, while we are not able to calculate MHWs (due to the 5 day duration required by the definition used here) we provide extreme temperature statistics at PHA as it*

*may be useful for investigating the potential impacts of extremes temperatures on biogeochemical processes."*

In reference to HS and CS (Page 6, Lines 87-88), my interpretation aligns with temperature exceeding the 90th percentile from above and the 10th percentile from below. Consequently, Table 2's columns pertaining to HSs and CSs appear non-uniform. For PHA, the number of events appears to be associated with a timescale of weeks or months (>5 days), while at the other sites, peaks are expected to encompass events lasting less than 5 days.

Thank you for bringing this to our attention. We incorrectly used the HS/CS time periods for MHW/MCSs and Vice versa. We have corrected the time periods in the updated Table 1 which will improve clarity.

The reviewer is correct that the number of HSs and CSs depend on data availability and site PHA consists of profiles collected typically weekly to monthly. The difference at the other sites is that a mooring with high frequency sampling (5 min – hourly) has also been deployed enabling the detection of MHWs and MCSs as per the Hobday et al. definition. HSs and CSs are days that exceed their temperature thresholds ($90^{th}$ and $10^{th}$ percentiles, respectively), and do not include days counted as a MHW or MCS. We have clarified this in the edited caption for Table 1 copied below. The different data availability in time and space at the sites, as well as different ocean characteristics, explain the non-uniformity that the reviewer refers to, which we have discussed in Section 3.4.

Furthermore, it remains unclear whether "#Days Sampled" for PHS refers to the number of (weekly?) samples or the total number of days.

In Table 1 "#Days Sampled" refers to the unique number of days sampled. For example, at 29m depth at ROT 5199 unique days (or daily data points) are available between 1951 and 2022. To further clarify this, we have edited the caption for Table 1 as follows:

*Table 1: "Table showing the number of Marine Heatwaves (MHWs), Marine Cold-Spells (MCSs), Heat Spikes (HSs) and Cold Spikes (CSs) identified at sites Port Hacking A and B (PHA and PHB/PH100, respectively), Maria Island (MAI), and Rottnest Island (ROT) at multiple depths. HSs and CSs include days when temperatures were higher than the $90^{th}$ percentile, and lower than the $10^{th}$ percentile, respectively, and do not include days identified during MHWs and MCSs. The time period used for MHW and MCS event detection is shown separately to the time period used to detect HSs and CSs.* The total record *available at 2 m depth as a percentage of the time period between the first and last sampled temperatures is listed for each site, and sub-surface data availability relative to 2 m depth is shown at the other depths. The unique number of days (or daily-binned data points) available over the site's record (e.g. between 1951 and 2022 at ROT) are also shown (# Days Sampled)."*

Additionally, we now describe variables 'TEMP_HEAT_SPIKE' and 'TEMP_COLD_SPIKE' in Table 2 to avoid confusion. These variables include all temperatures that exceed the percentile thresholds (clarified in the table). However, the #HSs / CSs in Table 1 are the number of temperatures that exceed the percentile thresholds when there are no MHWs / MCSs.

While the paper acknowledges the impact of employing climatology periods of varying durations (Page 8), it does not provide an analysis of the consequences of selecting different climatology periods. Such an analysis is particularly crucial for this dataset since continuous daily records may not always be available (Figure 2), and the presence of long-term trends cannot be disregarded. In this context, the evaluation of designating the entire 80 year period as a reference climatology remains lacking. A plausible approach could involve generating different realizations of MHW and MCS to reflect data discontinuities as errors, for instance, in the count of heatwaves. Consequently, the statement asserting that the "data products highlight the challenges encountered when detecting extreme temperature events using sporadic long time series" receives weak support from the limited uncertainty analysis presented in the paper.

This is indeed an area of ongoing research, and there is much debate in the community on which way is best (e.g. Amaya et al., 2023; Sen Gupta, 2023; Rosselló et al., 2023), and we believe the answer is dependent on the question being asked. In our case we have chosen to use the full record and have been very clear about the time period used, as stated in a dedicated section 3.4 'Effects of Data Availability and Choice of Baseline'. We do this as it would be challenging to compare events using a 30-year climatology below the surface due to a lack of data at our sites.

We have edited Section 3.4 for additional clarity:

*"We acknowledge that using a climatology of approximately 30 years is recommended (Hobday et al.,2016; WMO, 2018), and is commonly used for surface MHW studies (e.g. Frölicher et al. (2018); Oliver et al. (2018); Elzahaby and Schaeffer (2019)) where such data exists. Schlegel et al. (2019) showed that using a climatology period > 30 years may affect MHW detection in a similar way as using a climatology period < 30 years, dependent on environmental multi-decadal variability. Further, Rosello et al. (2023) highlight the impact of long-term trends on MHWs and argue that a moving baseline is more adequate for detecting consistently rare extremes, although presently there is much debate on this subject (Amaya et al. (2023); Sen Gupta (2023)). We use daily climatologies calculated using the entire temperature record at the long-term sites because 30 year daily sub-surface data sets do not yet exist. At this time, high-frequency sub-surface data at the 4 sites only extend back approximately 13 to 14 years. We acknowledge that the number of extreme temperature events and their metrics will differ in the future when a 30-year daily climatology becomes available, and that it will vary depending on which 30-year period is used (as suggested by Schlegel et al 2019)."*

Amaya, Dillon J., Michael G. Jacox, Melanie R. Fewings, Vincent S. Saba, Malte F. Stuecker, Ryan R. Rykaczewski, Andrew C. Ross, et al. "Marine Heatwaves Need Clear Definitions so Coastal Communities Can Adapt." Nature 616, no. 7955 (April 2023): 29–32. https://doi.org/10.1038/d41586-023-00924-2.

Sen Gupta, Alex. "Marine Heatwaves: Definition Duel Heats Up." Nature 617, no. 7961 (May 16, 2023): 465–465. https://doi.org/10.1038/d41586-023-01619-4.

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, Volume 10, Page 1168368

Analysis quantifying the consequences of selecting different climatology periods is out of scope for this paper as our focus is to describe and explore the new AMDOT-EXT data products. This additional analysis would warrant a whole new paper, when considering work undertaken by e.g. Rosello *et al*., (2023) and Schlegel *et al*., (2019). These studies suggest that time series of different lengths will have different trends, and hence the climatologies will differ.

We have hence not included additional text on this in the manuscript.

In the context of uncertainty assessment, it is also important to account for the influence of secular trends, as exemplified in recent work on marine heatwaves in the Mediterranean by Rosselló, P., Pascual, A., & Combes, V. (2023), titled "Assessing marine heat waves in the Mediterranean Sea: a comparison of fixed and moving baseline methods," published in Frontiers in Marine Science, Volume 10, Page 1168368.

As discussed in our response above, we think that an analysis of how trends affect extreme temperature statistics is out of scope for this paper. Notwithstanding this, the data are available should the reader choose to do their own analysis. However, as mentioned we had already included discussion on this subject in a dedicated section 'Effects of Data Availability and Choice of Baseline'.

In addition to this, we have added the following text in Section 3.4:

"*Further, Rosselló et al., (2023) highlight the impact of long-term trends on MHWs and argue that a moving baseline is more adequate for detecting consistently rare extremes, although presently there is much debate on this subject (Amaya et al., 2023; Sen gupta, 2023)."*

Information regarding potential future updates to the dataset is not provided. Even the acknowledgment of an absence of clear plans for future updates constitutes valuable information.

We have added the following text in the Abstract:

*"The 4 data products are provided as CF-compliant NetCDF files and it is the intention that they be updated periodically. Please refer to the URL in the Code and Data Availability Section for the latest version."*

**Technical Corrections**

On Page 6, Lines 87-88, for clarity, it is advisable to specify that HS and CS refer to temperature exceeding the 90th percentile from above and the 10th percentile from below.

We have updated the sentence as follows:

*"We identify HSs as temperatures that are higher than the daily-varying 90$^{th}$ percentiles and CSs as temperatures that are lower than the daily-varying 10$^{th}$ percentiles."*

The final paragraph on Page 6 (Lines 102-105) is not sufficiently clear and warrants rephrasing.

We have rephrased this final paragraph as follows:

"*The AMDOT-EXT data products sometimes contain extreme events that are not consistent over depth (Fig. 3). For example, event start and end dates, and their duration, can vary as a function of depth. This is partially because data availability over time is often non-uniform through the water column, but also because the depth structure of MHW events can vary due to ocean dynamics. For example in the coastal*

*Sydney region, Schaeffer et al., 2023  showed that MHWs can be shallow, can extend across the whole water column, or can be sub-surface only. For example, during atmospherically-forced shallow MHWs we might expect there to be fewer days when temperatures exceed the 90$^{th}$ percentile at depths below the mixed layer depth when compared with temperatures at the surface (Schaeffer et al., 2023).*"

The presence of a question mark at Line 105 is likely a typographical error.

Thank you.