<u>Responses to Anonymous Referee #2</u>

Referee #2:

The authors present a newly constructed database of temperature-related disaster impacts, which combines meteorological and demographic inputs to provide detailed hazard and exposure data, which is usefully output on different levels (i.e., country, region and also grid point level for heat/cold wave indicators) to satisfy the requirements of a diverse range of end-users.

I was impressed with the rigorous methodology used to provide such a comprehensive dataset that will surely prove invaluable for stakeholders involved in risk mitigation and adaptation (e.g., design of resilience strategies). I think the effort put in to remove inconsistencies in for example EM-DAT will be especially appreciated.

The paper is thorough, the methodology is mostly clear, and the results usefully showcase outputs from the dataset. I think the paper is well suited for publication in ESSD and does not require much in the way of changes for the final version.

I have just a few comments concerning the ‘Data and methods’ section which I think the authors should clarify and provide some additional detail in the manuscript to ensure the approach is easily followed. These specific (line-by-line) comments are provided below:

Authors: We sincerely thank Referee #2 for their positive and constructive feedback. We are delighted that our dataset is perceived as useful for a wide range of users. We also appreciate the specific comments, which will help us improve the clarity and transparency of the “Data and Methods” section. Our detailed responses are outlined below.

Specific Comments

- **Referee #2:** L108: What does GADM stand for? Should this acronym be spelled out here?
  Authors: Thank you for noting this. GADM is the official name of the database and stands for *the Database of Global Administrative Areas*. We will clarify this in the revised manuscript.

- **Referee #2:** L118-119: I find this difficult to conceptualise. Perhaps it could be explained first what the difference between GADM polygons and GDIS subdivisions is? And how the former yields the latter?
  Authors: We agree that this needs to be explained in a more clear way. The polygons of the administrative subdivisions are obtained from GADM. GDIS is a database that links EM-DAT records (impact information) to the corresponding GADM polygons (administrative regions/units) – to create a database of administrative regions that have been impacted by the disasters. We will revise the manuscript to explain this relationship more clearly.

- **Referee #2:** Using France as an example, how does the administrative subdivisions (demarcated in red; https://gadm.org/download_country.html) differ to those in Fig. 1C? They look much the same to me.
  Authors: You are correct, they are – and should be – the same. As noted in the comment above, we retain the official boundaries of the administrative units, but quantify hazard and exposure information only for those units identified as impacted by EM-DAT and GDIS.

- **Referee #2:** L210-211: If I understood correct, linear detrending is only performed on a copy of the data for percentile estimation. These values are then carried across when detecting heat/cold waves on the non-detrended data.
  I wondered what the sensitivity is in detecting extremes if no detrending was instead performed when calculating the daily percentiles? Presumably if there is a statistically significant upward trend, the values for earlier and later years would largely cancel out anyway when calculating a mean over all years.
  If the rate of warming is however not uniform over time (e.g., no trend over first 20 years but strong upward trend for the last 10 years), then I could envisage a greater impact on the resultant percentile values (detrended v non-detrended). Perhaps the justification for detrending the data for percentile estimation could be expanded to point this out, assuming this thinking is correct?
  Authors: You have understood our methodology correctly. One point to clarify (which we will explain in more detail in the revised manuscript) is that detrending was performed on a copy of the full time series (1979–2018), and the percentiles were subsequently calculated using 30 years of this detrended copy (1981–2010). We agree you raise an important point here, and in the revised manuscript we will test and report the sensitivity of using detrended versus non-detrended data for detecting extremes.

- **Referee #2:** L213-214: Another sentence I'm not sure I understood. 'Adding back the temporal mean of the daily maximum and minimum time series to the detrended values' → Is this to help preserve the shape of the seasonal cycle which might be lost when detrending? The temporal mean being the daily averages over 30 years (and not the length of the moving window I assume?). Further to my last point that the earlier and later years might often largely cancel out, adding these (non-detrended) values back to the detrended values is therefore presumed not to lead to any inconsistencies?
  Some additional clarification I think is warranted here, as I think its very difficult to follow what the authors have done exactly.
  Authors: Thank you for highlighting this. When detrending, we use the *detrend*-function in CDO (Climate Data Operators), which removes the linear trend from the time series. After detrending, the series has a mean near zero (since they are the remaining residuals), but we want anomalies to remain relative to the original mean. Therefore, we add back the original temporal mean (calculated over the full time series). This restores the series so that it fluctuates around its original mean while removing the linear trend. We will make this more explicit in the revised version of the manuscript.