Comment on nhess-2022-192
Anonymous Referee #2


This manuscript discusses the uncertainties related to the use of reanalysis datasets for the monitoring and recent evolution of heat wave indices over West Africa, with a focus on several urban areas.

Three types of uncertainties are addressed, that stemming from the dataset itself, the choice of the threshold for heat wave definition, and the type of indicator (minimum or maximum temperature, including or not the influence of other meteorological variables such as surface winds or humidity).

The results presented are consistent with several past works on the topic, adding recent years and in some aspects providing different diagnostics than the analyses previously published (e.g. Cecchirini et al. (2017), Moron et al. (2015), Barbier et al. (2018) - to cite only a few focused on West Africa). However, the manuscript in its present form has a number of shortcomings and fails to present results in a clear and concise way. I found some parts of the manuscript difficult to follow, and the number of figures (including all those in the supplemental) leave me with the impression that extracting key conclusions from the numerous statistics computed was a challenging, but uncompleted, task for the authors. Also missing from the manuscript, with respect to the title of the submission, is a clear description of what the authors are aiming for in using reanalysis data for city or district-level monitoring. The conclusions in terms of reanalysis uncertainties clearly point major caveats to such an approach, which then weakens the key messages of the paper.

That said, I am confident that the authors can revise their submission, taking into account major comments listed below, and turn this manuscript into a valuable contribution.
Major comments

The manuscript dwells quite some time on the description of differences between the MERRA and ERA5 reanalyses, evidencing large uncertainties between both. Then, ERA5 is kept as “truth” for the rest of the paper (from section 3.3). Differences between heat indices over West Africa computed from reanalyses and other sources of data and related uncertainties have been discussed in a number of past publications (Barbier et al. 2018, Batté et al. 2018, Engdaw et al. 2022...). Why is ERA5 thought to be better a reference than MERRA, and why is reanalysis kept as a suitable source of data for the rest of the analysis? Did you find similar caveats in the MERRA dataset than those highlighted by Engdaw et al. 2022? You (very briefly) mention station data in Supplemental Fig S8, but did you perform some assessment of ERA5 versus MERRA with respect to station data corresponding to your cities of interest? How adequate is ERA5 to represent heatwaves in these cities? All of this is left to the reader to guess or infer, which is a bit confusing given the title of the manuscript.

I’m a bit puzzled by the mention of station data and the separation of the region of interest in climatic regions, and the use of gridded reanalysis data in the study. It wasn’t fully clear to me upon reading the manuscript whether the approach used was completely validated. In the supplement, there is a figure (S8) which appears to tackle this question, but it is only very briefly mentioned in the manuscript. The authors furthermore say they find high levels of correlation, but I would argue this is only the case for Tmax over Dakar out of the four results shown.

The classification of cities should be described in more detail, and justified. Indeed the classes found are used to compute composites of characteristics in section 3.4, but this approach will be valid only if there is indeed some level of consistency between the cities. Given the spatial distribution of cities of interest, some will likely be characterized by neighboring gridpoints from the reanalysis, whereas others are much more distant. I’m missing a clear justification of why this approach is valid.

A final point more related to editing is the numbering and order of figures. The figures (including supplemental figures) should be numbered according to the order of appearance in the text. If not, the reader has to go back and forth between figures and this makes the paper tedious to read.

Specific comments

Abstract

What is the main goal of the manuscript? Already from reading the (quite long) abstract, it appears that the scientific questions are not very specific.
**Introduction**

The first two sentences are already a bit confusing to someone not familiar with the reference period chosen in IPCC reports to assess temperature evolution with climate change. The first sentence refers to temperature changes since the industrial revolution whereas the second also states a change of 1.5°C, not yet reached, which must be with respect to the IPCC baseline 1850-1900 reference period. This is not central to the manuscript but I would suggest rephrasing.

l. 65: Either explain more how this result is important (if relevant for your work) or shorten the paragraph.

**Region of interest, data and methods**

l. 110: “The choice of these regions has been validated by conducting some analyses over the cities belonging to each region (not shown).”

This is a shame, since it clearly is a key aspect in your use of this regional scale in the analyses that follow, and links to the title of the manuscript (see one of my major comments above).

l. 124: The authors restrict their analysis to 1993-2020. Both MERRA and ERA5 data are available before 1993, and statistics would likely be more robust by including more years. Is there a specific reason for this?

Section 2.3.1: As stated earlier, I think this section leaves a lot of crucial points of the study partially hidden to the reader, which weakens the conclusions.

l. 157: Did you compare the nearest neighbor strategy with lsm > 0.5 to the station data? Of course station data will be representative of temperature at a very local scale, but on the other hand, resolution of the reanalyses is quite coarse when compared to cities.

Section 2.3.3 and heat wave duration computation
I was confused by the equation 1.200 and the explanation. In lines 196-199 you explain that heat wave duration is computed as the mean over the number of heat waves of the total number of hot days in heat waves (I agree with this definition). But then when describing the equation terms, it appears you count all of the hot days whether belonging to a heat wave or not. If $d$ is the number of hot days, then shouldn’t $\delta_j$ in the sum be an indicator of the day belonging to a heat wave rather than the corresponding temperature exceeding the 90th percentile (this condition is already fulfilled for a hot day…)?

Later in the manuscript, it wasn’t clear to me why mean duration could be lower than 3 days (for instance in Fig. 7), since your criteria to define a heat wave is for having at least 3 consecutive days above the given threshold.

I may have missed something here, but in any case, this needs clarification in the methods section.

Section 2.3.4

You define POD but then refer to “hit rate” when discussing the results and in Fig. S1.

More generally, in your definitions of the statistical metrics, you use the terms “forecast system” and “observations”. Implicitly, later in your discussion of results, ERA5 is often the “observation” and MERRA2 the “forecast system”, but I would argue that these terms are quite misleading and suggest you rather use terms like “evaluated dataset” and “reference”.

Results

Fig. 3: The blue/red color scale for figures a) and b) isn’t the best choice.

l. 263: It would be worth specifying either here on in section 2.3.4 for what event the scores are computed (hot days).

l. 281: “changes of heat waves occurrence”: What do you call occurrence? The total number of events over the period of study?

l. 308: “we use the 90th for heat wave analyses” □ you mean the 90th percentile?
I. 321: “Tw takes in account the effect of humidity on the temperature” □ I would argue this is also the case for AT, which includes this influence through the term related to water vapor pressure.

I. 332-334: These sentences introduce a new aspect of results, I would therefore recommend moving this to section 3.4. By the way, the numbering of Fig. 12 should be Fig. 7.

I. 336: “CONT, AT and GU see section “region of interest” for more details” □ As a reader I was frustrated at this stage since the details in the section to which you refer doesn’t provide these details (it is even stated “not shown”).

I. 344: “The heat waves detected in the GU region have a short duration and a weak intensity [Fig 7]” □ As mentioned earlier, I was surprised that duration is lower than 3 whereas by your definition heatwaves should last a minimum of three days to be considered as such. Maybe the values are divided by the number of cities? This is a clear blind spot in your methodology. Please clarify this (also in the figure legend).

I. 364-374: Splitting the (already short) period into yet shorter sub-periods calls for some comment on the robustness of the analysis, especially since other factors may influence the occurrence of heat waves (e.g. El Nino, decadal variability, ...)

Discussion

Regarding the differences between ERA5 and MERRA2, Engdaw et al. (2021) identify striking differences between MERRA and other reanalysis and observational datasets in the 2000s for heatwave indices. MERRA appears to be a clear outlier. Did you look into this and draw similar conclusions?

I. 412: The correspondence between heatwaves and El Nino events was suggested in Moron et al. 2016 which you could include in your introduction and at this stage of the discussion.

Typos and editing suggestions

I. 64: CRNM □ CNRM
AT is used as an abbreviation both for apparent temperature and the Atlantic cities.

I. 280: “see [Fig S3] Fig. S3 in the supplemental material”

Please harmonize the notations used and specify carefully each notation: for example Ws is wind speed in the AT equation, this is never specified. What is Ta in the same equation?

Table 2, C2: typo persistent □ persistent

Figure 6: Top row figure titles are missing

Figure S16: “incertitude” □ you mean uncertainty?

Overall the manuscript requires careful proofreading (watch out for missing parentheses and brackets).

The figure captions should also be revised carefully, and include information on the datasets used (the reader shouldn’t have to dig for this information in the text).

Suggested reference: