EGUsphere-2023-2643 reviewer comments

This paper documents the ClimaMeter online platform in detail, outlining the methodology used to carry out the analysis but also describing the graphical elements used to visualise the results, and the protocol used to write the reports. Examples of previous reports are included, with more detailed analysis provided in the supplementary material (although these are not discussed in the main text), along with templates for writing the reports.

The paper is well written throughout, and the methodology and protocol are clearly defined. However, it is not entirely clear to me what the purpose of the paper is: is the intention merely to publish the protocol template to allow readers to produce their own ClimaMeter-style analyses, or to establish the scientific basis for the project? The level of detail sometimes suggests the former, but I think that the paper would benefit from offering more insight into the reasoning behind the modelling and reporting choices made, both to provide the scientific foundations for the work, and also as a guide to understanding and interpreting the official ClimaMeter output.

Overall the ClimaMeter approach is a valuable addition to the D&A toolkit. I anticipate that the paper will be suitable for publication after relatively minor changes: no new analysis is required, but more discussion of the analyses presented and reflection on the rationale behind some of the modelling choices is needed.

General issues

My main criticism of this paper would be that some of the most interesting information has been relegated to figures in the appendices. The reports reproduced in section 5 are, in essence, publicly available elsewhere: rather than reproducing them here, it would be useful to see more detailed discussion of the elements of the additional plots, particularly the diagnostic plots showing analogue quality, predictability and persistence, and an explanation of how those plots should be interpreted. This should be done for at least one of the studies in detail, and preferably for all (although in this case I would just highlight the most interesting features).

Similarly, there needs to be some discussion in the main text of how much the results might be altered if using the ERA5 dataset, rather than MSWX: and if the results are very different, is this because of the weather conditions on the analogue days, or because entirely different analogues are found? Are the results more stable for some hazards than others? An understanding of these potential limitations of the method is vital to understand when the method will be of most use, and when other methods may be more appropriate.

Specific points

40-44. Storyline-based (or reforecast-based) approaches to attribution do consider extreme weather events in terms of the weather system: while I appreciate the distinction between that and the

ClimaMeter approach, the existence of such methods should at least be acknowledged here. Also suggest changing ‘classical’ to ‘probabilistic’ to highlight exactly what is meant by the term.
73-76. A more thorough discussion of potential limitations of the method is needed. How sensitive is the analogue quality to the domain used, or to the choice of dataset? How sensitive are the results of the analysis to these factors? You could also add that, while there is a risk of underestimating the role of climate change due to the warming state of the climate during the reference period, the comparison is at least well defined and avoids extrapolation beyond the available data.

77. ‘Data download and pre-processing’ doesn’t quite cover all of the steps in this section - perhaps ‘data pre-processing and analysis’?

87-88. If MSWX did provide mid-tropospheric fields, would it be preferable to use those? Has any testing been done to understand how much difference this would make to the results?

95. The method actually used to find the analogues gets a bit lost here - I’d suggest moving ‘that is, the surface pressure fields minimizing the Euclidean distance to the event itself’ after ‘in terms of the event’s surface pressure pattern only’ so that the analogues are defined right away. Similarly, ‘once analogues are found… best 15 in each period’ seems to belong at the start of point 5.

140-145. First, please clarify in the text how ‘the dial points 95% to the right’ should be interpreted (for readers unfamiliar with the format it’s confusing to keep having to refer to the plot to check). Secondly, It should be highlighted that the relative sizes of the effects of climate change and natural variability are not actually estimated, which could be considered a limitation of the method when compared to the standard probabilistic approach.

Finally, I think the reasoning behind this dial needs to be explained in more detail. The argument here is that if the 15 analogues in the past are consistently in a different phase to the 15 analogues in the present, then this constitutes evidence that natural variability contributed significantly to the difference from past to present. This seems (partly) counter-intuitive to me, in that I would expect some circulation patterns to be more likely to appear in particular phases of ENSO in both periods and the phases of ENSO to be independent of the period chosen: so I wouldn’t expect the best analogues in the past to be systematically associated with a different phase of ENSO (I can see the argument more clearly in the case of lower-frequency modes such as the AMO, where each period may be dominated by a different phase). Can you elaborate on this? Or is it actually the case that ENSO is less often found to have a significant effect? This is perhaps something that could be discussed with reference to Figure 2 (I note that it’s very unusual for all three modes of variability to have a significant effect, but can only speculate as to why).

148: Suggest rephrasing as ‘Q is simply the average Euclidean distance of each analogue from all other analogues’ or similar. Otherwise this almost seems to suggest that 15 more analogues are found for each of the analogues, and Q computed for those.

149-155. Which category is assigned if Q is below the 75th percentile in one period and between the 75th and 95th in the other? (it may be easier to rephrase these categories in terms of ‘below the

95th percentile' to ensure an exhaustive definition, and to capture any cases where one period is above the 95th and the other is below the 75th?)

170. Why are the significance maps only included for surface pressure changes but not the hazard variable? I would have thought (perhaps naively) that significant changes in the hazard would be of more interest.

173. Please add a line explaining why the Cramer-von Mises test was used (presumably to compare the two distributions nonparametrically - is this a more stringent test than just comparing means, and therefore more likely to assign differences to indices of natural variability than to climate change?)

179-181. The choice to stick to a pre-specified report format may seem like an odd one to many scientists reading this paper, so I think it would be worth adding a line or two here explaining the benefits (and limitations) of doing so. It would be particularly interesting to hear of any feedback from the media on this - do journalists find it easier to digest this kind of complex information when it is presented in a format that they have become familiar with?

199. 'As soon as possible' can mean very different things, so I think it would be useful to highlight just how quickly this analysis can be produced - please add the typical/target timeframe for release.

201. What kind of feedback prompts an update of the report?

205-223. I don't think this description of the website structure adds anything to the paper, and would recommend removing it. However, a discussion of the details behind Figure 2 (perhaps just before the conclusions) could be informative - for example, discussing the relative frequency with which each mode of variability is found to be significant, and commenting on the fact that a high proportion of events studied have no close analogues: is this because the core team are choosing to study the more meteorologically unusual events, so we shouldn't expect to find any close analogues? Or is there some factor that could be affecting the quality of the analogues identified, such as sensitivity to the domain size?

224-399. It's not clear to me what the purpose is of simply reproducing these four studies here without any further discussion or analysis. For readers hoping to replicate the analysis for another event, it would be far more useful to choose one study to focus on, and to walk through the process of defining the event and interpreting the analysis. The more detailed figure for that study should be moved from the Appendix to the main text, and the elements that are not already discussed in the text commented on (in particular, the violinplots showing the predictability and persistence, and the plots of trends in the distribution of analogues). The differences and similarities between the MSWX and ERA5 results should also be addressed here. If the four separate case studies are retained in the main text, there needs to be some discussion in the conclusion of what they illustrate about the method.

231. Broken Wikipedia link/citation.

282. Should be ‘Haikui’.

402. Typo: ‘this critical issue’.

401-404. This paragraph rather implies that no other methods or groups exist to communicate the changing risks of extreme events, which is simply not true (there are now several operational met services carrying out rapid attribution studies and communicating them to the media, as well as WWA). It would be more accurate to say that ClimaMeter is *part of* a continuing effort to contextualise extreme and hazardous events: to me, the innovation here is that, rather than using statistical methods that analyse time series of events that may arise from different processes, ClimaMeter considers changes in extreme weather arising from the same circulation patterns, which allows a more nuanced understanding of the spatial and multivariate changes associated with the weather type of interest. (This is particularly important for wind and precipitation, because unlike in univariate probabilistic attribution, the event definition doesn’t require averaging over the spatial domain and so doesn’t smooth out the local extremes)

410. Rather than referring to the four case studies, it would be more relevant to refer to the map in Figure 2.

411-412. I’m not sure what ‘These underscore the significance of contextualising extreme events, as a tool to understand the broader context within which they occur’ means here.

490. This should be updated to 2001-2024 to be current at time of publication.

Figures A1-A8. The values of ENSO, AMO and PDO associated with the event are missing from the plots.