Answers to the comments made by reviewer Vikki Thompson to the article "*A flexible methodology to evaluate natural variability in ClimaMeter*"

May 2026

First and foremost, we would like to thank the two reviewers for taking the time to carefully read the manuscript. Their comments are the result of careful consideration and are very insightful. We have done our best to address all of their remarks, and we believe that our work has truly benefited from this review and that the manuscript has improved significantly.

In this document, the reviewer's comments are shown in blue and our responses in black. All line numbers refer to the submitted version of the article.

Following on from the current ClimaMeter methodology, this paper presents a new method for attributing changes in extreme weather events using analogues, remaining within the same conceptual framework. The method presented, 'ClimaMeter 2.0', does provide improvements on the original method, and the paper also outlines the remaining limitations of such a method. The figures themselves are clear and nicely presented. The discussion of limitation of the original ClimaMeter method is comprehensive. However, in my view the presentation of the method and particularly the comparison of results using the two methods should be enhanced before publication. Many of the most important findings are not adequately discussed within the text and suitable recommendations for implementing the method are not included.

Major Comments:

1. Abstract fails to state the most significant findings within the paper.
   We have taken the suggestion from the major comment #3 and will clarify that there are differences between the results of the two methodologies, and that we will discuss them in the paper.
   We will add at line 14: "*We perform extensive testing of ClimaMeter 2.0 using pre-industrial climate simulations and observational data, and we compare the results with those of ClimaMeter to highlight any differences in the performance of the two approaches.*"

2. In the assessment using a pre-industrial climate model it appears that ClimaMeter method overestimates the impact of climate change. I find this to be a key conclusion, as it suggests the method currently being used by ClimaMeter is not suitable. This result should be better highlighted in the abstract and discussion. Thank you for this comment, which allows us to highlight a very important aspect of our work. Based on the application of ClimaMeter to the pre-industrial simulation, we conclude that there is indeed an overestimation of the impact of climate change. This conclusion, however, applies exclusively for preindustrial simulations and for this specific test and cannot be generalized to the reanalyses data. The purpose of the pre-industrial test is rather to demonstrate that it is possible to introduce a more flexible methodology, one that adapts to both observed data and pre-industrial conditions and future conditions where greenhouse gases forcing may slow down or overshoot scenarios, and is therefore a more refined and general version. In this sense, the key aspect of ClimaMeter 2.0 is the test of the hypothesis that ACC influences observed changes, performed through the analysis of long-term trends in specific quantiles of the target hazard (detailed in Section 4.2). We emphasize that, while in the pre-industrial experiment this test yields a negative result most of the time, as expected, in the observations it yields a positive result virtually always for temperature extremes, while the result is more variable for precipitation and wind speed. However, overall, in most cases the hypothesis is confirmed, and thus the result aligns with the ClimaMeter version, which systematically assumes this hypothesis to be true. In light of this, this work does not support the claim that ClimaMeter is unsuitable, but rather uses this as a basis for proposing a more general and flexible extension.

   We will better stress this important aspect in the manuscript, adding in the conclusion (line 580): "*The application of the ClimaMeter 2.0 methodology to pre-industrial events yields promising results, as it reduces the proportion of events identified as being amplified by climate change, linking them in most cases to natural variability. Additionally, this result is robust across the three models analyzed, providing strong evidence for the robustness of the new method. It is important to stress that the pre-industrial experiment points to an overestimation of impact of climate change when using the ClimaMeter method, but such conclusion applies only within the scope of this test and cannot be generalized to reanalyses or observational data. With the pre-industrial test, we aim to prove that it is possible to introduce a more flexible methodology, one that adapts to both observed data, pre-industrial simulations, or future climate simulation where greenhouse forcing could be reduced (for example, overshooting) and which therefore measures the role of climate change in a more refined way.*"

3. Although gauges from both methods are shown in Fig 7-9, subplots e and f, there is no comment on these in the text. The gauges appear to show significant differences between the two methods. This should be clearly

highlighted in the abstract, results, and discussion.
We will better highlight the differences in the revised manuscript:
- Regarding the 2023 September European Heatwave, we will add at line 509: "*The difference between the two gauges therefore stems from the calculation of the weights of the variability modes, which makes the indicator continuous. In this case, this calculation results in a lower estimate of the role of climate change compared to the ClimaMeter methodology.*"
- Regarding storm Hans, we will add at line 529: "*In this case, the weight assigned to the PDO nearly doubles the gauge's value. It is therefore interesting to note that, for Storm Hans, ClimaMeter would underestimate the role of climate change.*"
- Regarding storm Ciaran, we will add at line 553: "*In this case, however, the large discrepancy between the values of the two gauges stems from the fact that there is no significant trend in wind speed in the target region. It is important to note that, since the event affected the entire west coast of the country, the target region also includes nearly half of France's inland areas. It may therefore be a too large region to detect a significant trend.*"
- In the conclusions, at line 584, we will add: "*In both models and reanalysis, the greatest discrepancies between the two methods occur when the trend in the hazard variable is not significant. In this case, the ClimaMeter 2.0 gauge often points towards low values, indicating that natural variability is the predominant factor in determining the conditional hazard changes. If, on the other hand, the trend is significant, then the differences stem from the calculation of the weights of the modes of variability, which modulate the value of the 2.0 gauge.* "
- in the abstract, at line 14: "*We perform extensive testing of ClimaMeter 2.0 using pre-industrial climate simulations and observational data, and we compare the results with those of ClimaMeter to highlight any differences in the performance of the two approaches.*"

4. As the assessment of the pre-industrial suggests ClimaMeter 1.0 overestimating the role of ACC and Fig7-9 show differences between the two methods, a systematic comparison of more real-world events between v1.0 and v2.0 should be undertaken, to ascertain if this is the case. A more general comparison of the results from the two versions would undoubtedly be of great interest. It could highlight, for example, the types of events for which the two methods yield the most divergent results. At the same time, this work focuses on the methodological aspect. We believe that the three case studies presented in the paper (analyzed in both the pre-industrial simulation from three climate models and the ERA5 reanalysis) are sufficient to illustrate the reasons that led us to develop ClimaMeter 2.0 and the innovations we introduce with this method.
We will indicate that a more general comparison of the two methods on a large variety of events is left for future work by adding at line 604: "*Future directions for this work will involve an in-depth comparison of the results obtained using the ClimaMeter and ClimaMeter 2.0 methodologies across*

a wide range of events. This would be of great interest, as it would enable us, for example, to identify the types of events for which the two methods produce the most divergent results, as well as the modes of variability that are predominant in different regions and for different types of events."

5. The gauges in Fig 7-9 e do not agree with previously published ClimaMeter reports for the three events, this should be explained or corrected.
   We will clarify it in the revised manuscript. We will add at line 499: "*In the event report already published on the ClimaMeter website (Faranda et al., 2024a), the value of the natural variability gauge is 5%. This difference is due to the fact that in this study we decided to introduce a minimum interval of 7 days between consecutive analogues to avoid the oversampling of the same meteorological event. This may slightly alter the pool of selected analogs, and consequently, the p-values obtained from the CvM test may differ. In this case, it is the one associated with ENSO (here greater than 5%) that accounts for the difference between the present analysis and the one already published by the ClimaMeter Consortium.*"
   Also, we will add at line 543: "*Once again, the difference from the gauge value in the previously published study (35%, Alberti and Faranda (2024)) is due to the minimum interval of three days between consecutive analogues that we chose to impose in the present analysis.*"

6. The paper fails to give a clear recommendation on how the findings should be implemented by ClimaMeter. Should past reports be reassessed given the possibly overestimation of the role of climate change? Should both methods be used together for a period of comparison? Is further research required before analogues are used for attribution?
   The plan is to implement the Natural Variability Gauge 2.0 within the framework once the present submission is accepted or published.
   We will better stress this by adding at line 560: "*This new method and the associated gauge 2.0 (whose novel representation is detailed in section 6.1) will soon be integrated into the protocol and implemented at the operational level in the analyses available on the online platform.*"

   As previously mentioned, we believe that a detailed comparison of the two methodologies would offer significant added value and provide interesting insights into potential differences between ClimaMeter and ClimaMeter 2.0 for specific categories of events. As indicated above, we will leave this comparison for future work. However, we do not believe that reanalyzing all (or various) previous events would be done in a systematic way: ClimaMeter has always published its reports highlighting its own limitations and with a transparent protocol. As in any fields of science, this work represent a scientific evolution and, as such it will be used for new studies. Furthermore, ClimaMeter results of V1 have been always weighted by the expert judgment of the authors and also relied on the published literature so that uncertainties have been acknowledged in each report.

   Moreover, the use of analogues in attribution frameworks is a well-established

and widely used technique that offers an interesting complement to probabilistic techniques. In this work, its relevance is not called into question.

7. Assumption that modes do not change because of ACC: A more substantial lit. review would be useful. Only literature in favour of 'no change' is presented.
   The literature we cited is based on a review of Chapters 2 and 3 of the IPCC AR6. The limited available data and the absence of any sustained change have led us to make this hypothesis for ENSO and PDO. In the case of ENSO, we will add a few more details in the revised manuscript, at the line 219 (section 3.2.2): "*For ENSO, there is evidence showing that its amplitude from 1950 to the present has been higher than during the pre-industrial period (from 1400). This suggests that such an increase in variance may have been caused by enhanced greenhouse gas levels (Hope et al., 2017). In addition, there is evidence that human emissions and changes in aerosols have an impact on the persistence of El Niño events, although these effects have compensated each other over the course of the 20th century (Stevenson et al., 2019). However, no clear evidence indicates a recent and sustained shift beyond the range of decadal to millennial variability (Gulev et al., 2021), and there is high confidence that human influence has not yet modulated its variability beyond its range of internal variability (Maher et al., 2018).*"

   The case of the AMO is slightly different, since, as already reported in the submitted article at line 222, there is robust evidence that anthropogenic and natural external forcings have modulated it over the historical period. At the same time, however, there is low confidence in the estimated magnitude of human influence, and therefore we believe that the hypothesis also applies to the AMO, given the current state of knowledge.

   The two following paragraphs, taken from the IPCC AR6, may help clarify the matter:
   "*Since the late 19th century, major modes of climate variability show no sustained trends but do exhibit fluctuations in frequency and magnitude at inter-decadal time scales, with the notable exception of the Southern Annular Mode, which has become systematically more positive (high confidence). There is high confidence that these modes of variability have existed for millennia or longer, but low confidence in detailed reconstructions of most modes prior to direct instrumental records.*" (IPCC Ch.2 Executive Summary)

   "*Human influence has not affected the principal tropical modes of interannual climate variability or their associated regional teleconnections beyond the range of internal variability (high confidence). There is medium confidence that anthropogenic and volcanic aerosols contributed to observed changes in the Atlantic Multi-decadal Variability (AMV) index and associated regional teleconnections since the 1960s, but there is low confidence in the magnitude of this influence. There is high confidence that internal vari-*

ability is the main driver of Pacific Decadal Variability (PDV) observed since pre-industrial times, despite some modelling evidence for potential human influence. Uncertainties remain in quantification of the human influence on AMV and PDV due to brevity of the observational records, limited model performance in reproducing related sea surface temperature (SST) anomalies despite improvements from CMIP5 to CMIP6 (medium confidence), and limited process understanding of their key drivers." (IPCC Ch.3 Executive Summary)

Minor comments:

1. typo in line 37 (full stop in wrong place) corrected
2. Missing ref in line 39 added
3. Brackets may be wrong in lines 119-121 fixed
4. Line 142: further details of the 'expert knowledge' used to determine the analogue region would be useful.
   We will add at line 142: "*For example, in the case of the Los Angeles wildfires in January 2025 (Cazzaniga and Faranda, 2025), a domain was selected that includes both the high-pressure system over the Great Basin—which generated the strong, dry northeasterly winds blowing toward Southern California—and the area most severely affected by the fires. In the case of a heat wave caused by an atmospheric blocking, instead, both the area associated with the high-pressure system and the regions that recorded the highest temperatures would be included.*"
5. Line 202: it is unclear what is meant, please reword modified
6. The limitation of the length of available data (only since 1950s) should be highlighted. Some modes of variability act on longer timescales (as commented on in line 437).
   We will add at line 148: "*Of course, over that time period, human-induced effects on the climate are far from negligible. This choice is imposed by the length of the available data.*"
7. Typo in line 463 (gauges): corrected
8. Typo in line 484 (missing space): added