Comment by Philip Gooding

https://egusphere.copernicus.org/preprints/2025/egusphere-2025-5264/#discussion

RC2: 'Comment on egusphere-2025-5264', Philip Gooding, 15 Dec 2025 reply

Due to my training and expertise, my comments focus largely on the use of weather records. I note the first reviewer has provided more substantive comments on the interpretation of different ML and reanalysis data - I don't have anything to add here. I do, however, second their view that 'the authors present an impressive amount of information that will be of broad interest to readers of Climate of the Past.'

L525. I wonder if the authors can do more to address the issue of 'suspicious values' in weather records. Their approach - leaving them in to 'see how the approaches deal with potentially erroneous measurements' and because they 'do not have the means to objectively check all observations' - is certainly valid. However, it doesn't necessarily have to be a case of 'inclusion' or 'exclusion' of suspicious values. Have the authors considered developing a confidence scale, thereby indicating in quantitative terms which sets of weather records they a more confident in? They allude to certain datasets as being probably of higher quality (e.g. Barcelona) than others (e.g. Marschlins). Would quantifying such 'quality' (and displaying it in the form of a table?) enhance the analysis?

We were unsure about including our assessment of confidence in the station measurements. We refrained because it was done ex post facto, using semi-quantitative methods without state-of-the-art techniques such as homogenization. However, we recognize that the information could be valuable to some readers and will include a column in Table 1 presenting the outcome of the assessment numerically: 2 means that we have good confidence, 1 means medium, and 0 means poor confidence.

L632. 'There is potential in many dimensions.' This is true. However, could the authors be more specific? One of the strengths of this manuscript is that it incorporates data and perspectives from diverse sources and methods. Many readers (including myself) may only be a specialist in one of these fields. Thus, I wonder if the authors could speak to specific avenues for research that such specialists could undertake (apart from, of course, building more interdisciplinary collaborations)? What could a specialist in old weather records/reanalysis/ML take from your article, and what could they contribute to your interdisciplinary efforts moving forward? What questions for further research does your manuscript bring up for specialists in different fields?

Thank you for this comment. It has made us think more about the roles you mention, and we may change the last paragraph of the manuscript into a slightly longer explanation, such as

"There is potential in many dimensions. Specialists in historical weather observations and measurements can help by providing accurate data. Records from areas with sparse

coverage—such as the North Atlantic in our analyses—are particularly helpful. In turn, this also highlights the value of metadata, quality checks, unit conversions or homogenization. Regional reanalysis experts may contribute by developing frameworks that can include more potentially valuable measurements. While reanalyses often discard values that deviate too strongly from expectations, such outliers sometimes reflect actual weather conditions and thus hold significant value. Machine learning approaches, in turn, must learn to cope with the uncertainties inherent in early measurements, e.g. by designing models that can assess potential errors without discarding useful information. Overall, this calls for collaboration between specialists in historical data recovery, reanalysis, and machine learning—each addressing complementary aspects of data quality, uncertainty, and model representation."

The clarity of figure 10 would be enhanced if it were in colour (as in the other figures).

We had tried a few color schemes. However, we felt that they either visually over-represent one column (e.g. brighter colors, or orange, etc.) or could lead to confusion with similar color assignments to other variables in the manuscript. Finally, we used the grey scale because of 'visual equidistance' of the bars.