

All your comments are found below, with our answers written in **bold**.

1. The most critical concern is the numerous empty parentheses “()” across the manuscript. Notably, these empty “()” occur 16 times throughout the text, including in places where definitions or citations are required. This significantly undermines readability. For example, in the Introduction, “However, for highly localised observable variables, such as precipitation, several studies have shown that ERA5 is not always able to accurately capture local conditions ().” and “, e.g., when target-resolution gridded data products are available for training ().”

We are really sorry about all the empty parentheses. It appears that something went wrong with our references when we submitted the manuscript to arxiv, and many of the references simply disappeared in the compiled document. We did not notice this when checking the final document, because the majority of the references were still available in the compiled pdf file. The empty parentheses are therefore areas in the text where one or more references disappeared during compilation. We will fix this as soon as possible.

2. The manuscript uses ERA5 and GSOD data that, in principle, extend well beyond 2010. However, the analysis is restricted to 1950–2010 without a clear justification. The authors should provide a more explicit explanation for why only data until 2010 were used.

Originally, the experiment was set up this way, due to an associated project where we were interested in different reference periods in the period from 1950 to 2010. However, we fully agree that it would be better to extend the analysis to include recent years. We will redo the analysis in this manner in the revision.

3. The selection of donor stations solely based on Euclidean distance may not accurately reflect climatic similarity, especially in maritime or complex terrain regions. Sensitivity to the choice of K would be better discussed not only through skill scores but also through climatological reasoning.

We agree that the distance measure could be extended to also include other measures of similarity than Euclidean distance, such as climatic similarity. We will provide an improved discussion about this point in the revision.

4. Figures 7, 8 and 12: The hexagonal binning obscures station-level heterogeneity. It would be better if it included small multiples of scatter plots to show skill score vs. elevation difference/coastal distance for key metrics

Thank you for this comment. We agree that it makes sense to examine scatter plots of skill score or absolute scores versus, e.g., elevation difference. We have now created such plots that could be added into the manuscript. We still believe that the plots of mean elevation difference and mean annual precipitation provide value, and are interesting to look at when considering all the skill score maps in the Appendix, but it might make sense to move these maps into the Appendix also, and to replace the figures in the main text with scatter plots.