Response to the reviewers

We are grateful to the reviewers for all of their comments. Based on these comments, we have strengthened the analysis by adding more years of data, other methods of visualising the results, and a more thorough evaluation and discussion regarding the spatial and inter-variable dependence properties of downscaled weather ensembles. All comments are provided below, with our responses written in **bold**.

To make it easier to distinguish all our changes to the manuscript, we have added a new document, created using the `latexdiff` program, that displays all changes since our previous submission. Note, however, that we have been unable to get the program to work well for figures, figure captions and references.

Reviewer 1

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 were told by the associate editor to wait until the revision process was complete before we updated the manuscript on arxiv, but we will fix the arxiv document as soon as we can.

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 have therefore updated the reference period to 1950–2024, and rerun the analysis. All the necessary data for 2025 were not yet available when we started rerunning the analysis, which is why 2025 was not included.

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 have now provided an improved discussion about this point in the discussion section of the manuscript.

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 and added them to the manuscript, while moving the map plots into the Appendix, as we still believe that the map plots provide some information and value.

Reviewer 2

1. Regarding the precipitation occurrence modeling, the authors employ a Bernoulli distribution while acknowledging that temporal dependence structures cannot be adequately incorporated into the precipitation occurrence process due to the infeasibility of transforming binary precipitation occurrence into Gaussian random variables. This represents a notable limitation, as the temporal structure in precipitation occurrence may not be sufficiently preserved under this approach. Given this acknowledged technical constraint, I would encourage the authors to discuss why traditional Markov chain-based methods, which are well-established for modeling temporal dependencies in precipitation occurrence, were not adopted as an alternative.

   Thank you for pointing this out. Based on your suggestion, we now model precipitation occurrences in the full downscaling model using a 1. order Markov chain, with time- and location-dependent transition probabilities that are modelled using the same covariates as in the original occurrence model. This new model strongly outperforms the old full precipitation downscaling model. More details about the model are provided in the revised manuscript.

2. While the authors acknowledge that their proposed downscaling method treats precipitation and temperature independently and thus cannot capture multivariate relationships between these variables, it remains unclear whether the framework is capable of preserving spatial or inter-site dependence structures within the downscaled precipitation and temperature fields themselves. The ability to maintain realistic spatial coherence is critical for many hydrological applications, and I recommend that the authors provide additional results and discussion addressing the performance of their method in reproducing spatial dependence patterns across the downscaled domain.

   Thank you for this comment. We have now performed additional experiments to evaluate the spatial coherence when downscaling temperature and precipitation to multiple nearby locations. The added experiments show that, although the spatiotemporal ensembles struggle to capture certain spatial properties of the

observed data, they overall perform well, and overall outperform the CPRCM simulations and ERA5 with respect to many different spatial properties. We have added a section about these experiments in the Appendix of the revised manuscript.