

“Capturing the Extremes: A Quasi-Comonotonicity-Based Algorithm for Disaggregating Daily to Hourly Rainfall” is an interesting and novel study, which I recommend for publication after revisions to the manuscript.

We thank Referee #4 for the positive assessment of our work and for recommending the manuscript for publication after revisions. We appreciate the recognition of the novelty and relevance of the proposed approach, and we address the specific comments below.

Please see my feedback and comments below:

a) The abstract should include quantifiable values of the main results.

We thank the reviewer for this suggestion. We will revise the abstract to include quantifiable values of the main results, highlighting key performance metrics of Q-CODA compared to the other disaggregation methods.

b) Figure 5 – Remove  $P^{\max}_{1h}$  from subplot title and rescale (h). Recommended to add median boxplots.

We thank the reviewer for the suggestion. We will remove  $P^{\max}_{1h}$  from the subplot titles and rescale subplot (h) accordingly. Regarding the inclusion of a median plot, we believe this could be unnecessary in this context. The current boxplots already capture the mean, variance, and extreme values, providing a comprehensive summary of the data distribution. In addition, the Q-Q plots requested in comment d) will complement the boxplots by directly illustrating the alignment between simulated and observed values across the probability distribution. In our opinion, including a median panel would therefore be redundant, while the current combination of boxplots and Q-Q plots conveys both central tendency and variability in the disaggregated rainfall.

c) Figure 7 – Subplot titles should include only “T = XX years.”

We thank the reviewer for this suggestion. We will update the subplot titles in Figure 7 so that they include only “T = XX years”.

d) Include a Q-Q plot comparing simulated and observed rainfall datasets for both daily hourly maxima and full hourly scales for selected stations. Selection could be based on elevation or regional grouping.

We thank the reviewer for this helpful suggestion. We agree that Q-Q plots are a useful complement to the summary metrics already presented, as they provide a more direct visual assessment of the agreement between observed and simulated rainfall distributions.

In response, we generated Q-Q plots for selected groups of stations following a regional grouping criterion (Mediterranean and southern stations, inland stations, and northern and northwestern stations), and evaluated both daily maximum 1-hour precipitation (Image 1) and the full hourly rainfall distribution (Image 2). These diagnostics were produced for all disaggregation methods considered in the comparison.

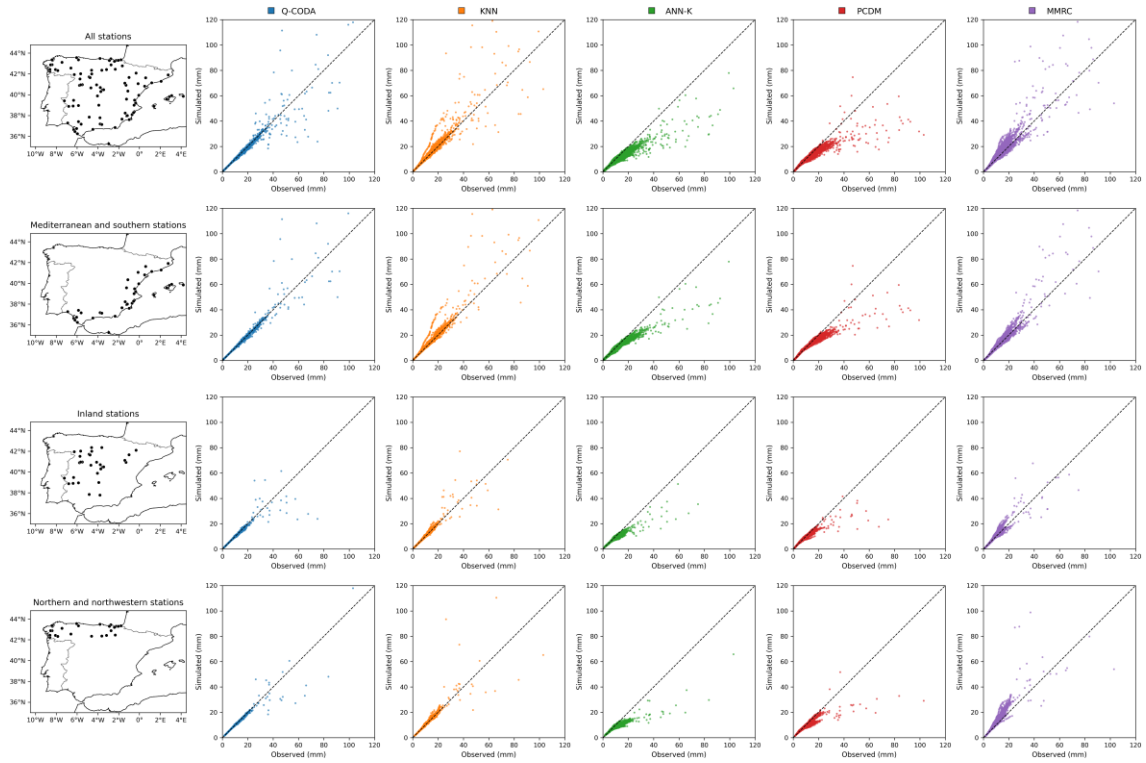


Image 1: Q-Q plots of observed versus simulated daily maximum 1-hour precipitation for the evaluated disaggregation methods across different station groups in Spain (all stations, Mediterranean and southern, inland, and northern and northwestern regions).

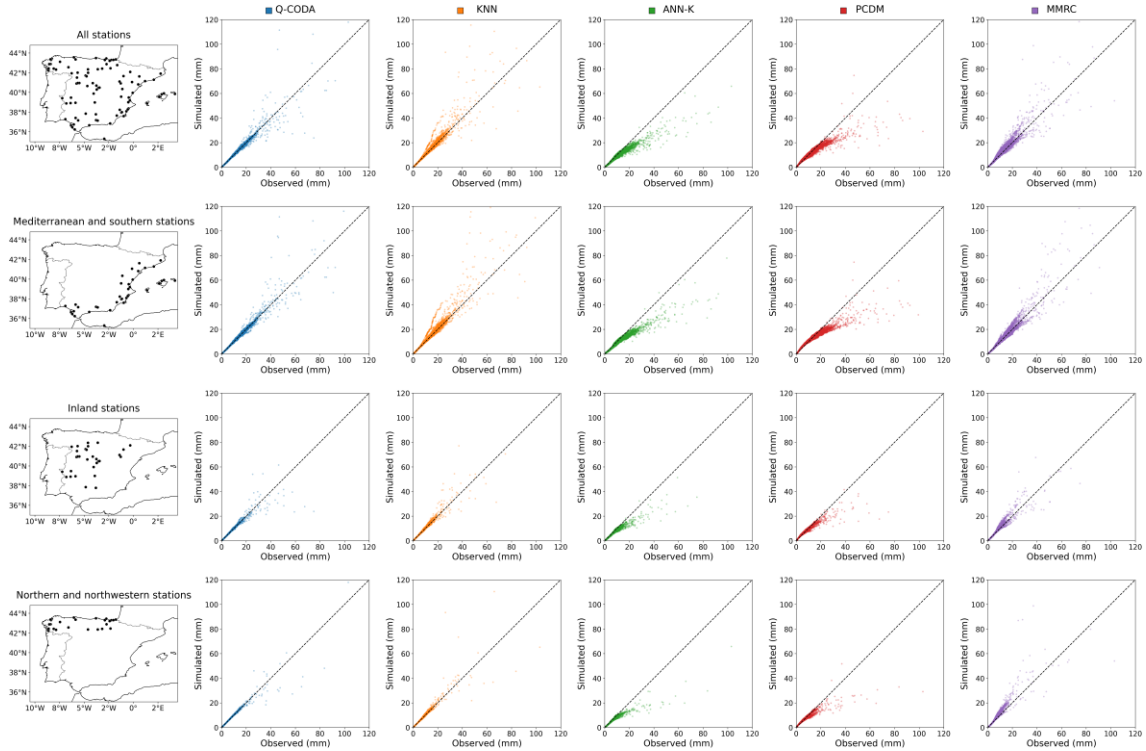


Image 2: Q-Q plots of observed versus simulated full hourly precipitation for the evaluated disaggregation methods across different station groups in Spain (all stations, Mediterranean and southern, inland, and northern and northwestern regions).

After examining the results, we found that both sets of Q-Q plots display very similar comparative patterns across methods, so including both would introduce redundancy while adding limited new insight. Therefore, in the revised manuscript we think it is a better option to include only the Q-Q plots corresponding to daily maximum 1-hour precipitation, as these are more directly related to the main focus of the study on sub-daily rainfall extremes. We will incorporate this figure into the new Discussion section, where it will also help illustrate the behaviour of Q-CODA across different regional and climatic contexts in Spain. This addition will be explicitly linked to the points raised by Referee #1 (<https://doi.org/10.5194/egusphere-2025-4469-AC2>) regarding the applicability of the methodology under different climatic regimes.

We apologize if the panels appear with limited resolution here; high-quality vectorized PDFs of the figures will be provided in the final version, should the manuscript successfully pass the review process.

e) It would be much better if the author(s) could add a separate section for the discussion.

We thank the reviewer for this valuable suggestion. We agree that adding a separate discussion section will improve the clarity and structure of the manuscript. Accordingly, we will include a dedicated section that synthesizes the points already addressed in our responses to Referee #1 (<https://doi.org/10.5194/egusphere-2025-4469-AC2>) and Referee #2 (<https://doi.org/10.5194/egusphere-2025-4469-AC3>), such as the applicability of Q-CODA to other climatic regimes, sensitivity to deviations from quasi-comonotonicity, and the analysis of outliers. In addition, we will address the issues raised by Referee #3 (<https://doi.org/10.5194/egusphere-2025-4469-RC3>) regarding the potential inclusion of parameterization in Q-CODA, the handling of the spatial component, and the associated limitations. This new discussion section will provide a focused reflection on the broader implications, strengths, and constraints of the proposed approach.

More Comments:

1. L115 – How consistent is the data on a yearly basis? Although a station may have >90% overall data availability, some years could still have significant missing data.

We thank the reviewer for raising this important point regarding annual data consistency. To address this, we performed a detailed assessment of annual data completeness across all 91 AEMET stations, using a stripe plot visualization and threshold analysis. Our results (see Image 3) show that while some stations have missing data in early years (e.g., 1996, the first year of the first fold), the overall coverage across all folds is sufficient to reliably evaluate and train the disaggregation methods.

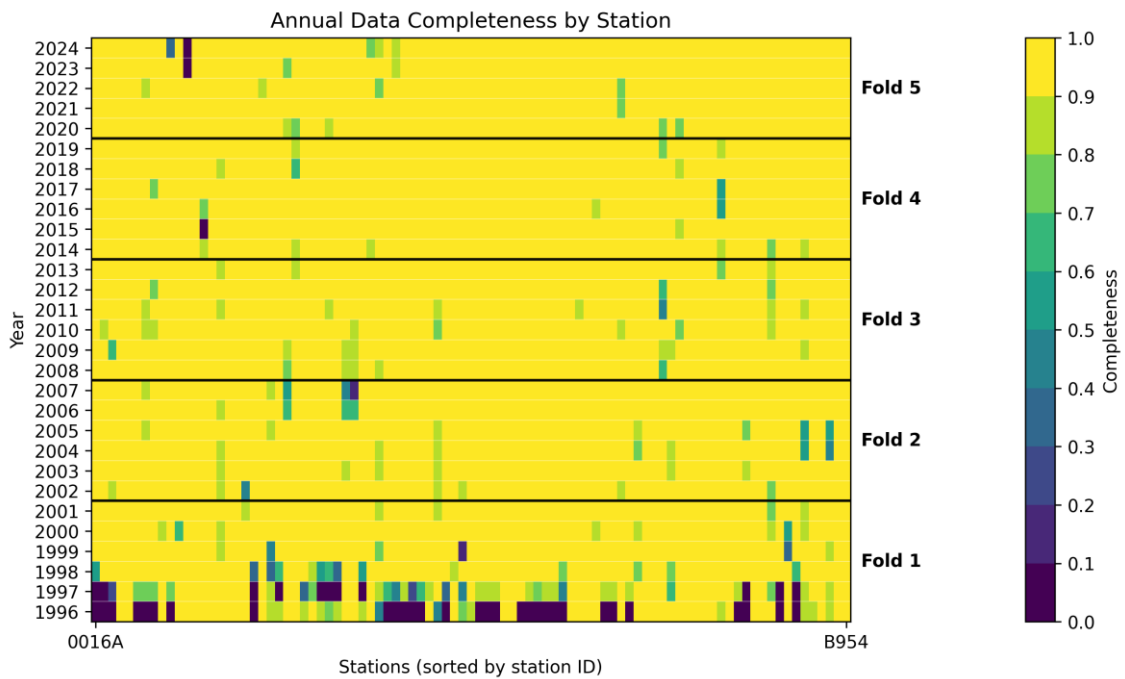


Image 3: Annual data completeness by station (1996-2024). Rows correspond to years, columns to stations (sorted by ID), and color indicates the proportion of available hourly data. White lines separate years; thicker black lines indicate fold boundaries.

2. General – Inconsistent use of “section” and “sect.”

We thank the reviewer for pointing out the inconsistent use of “section” and “Sect.” throughout the manuscript. We will carefully revise the text to ensure uniform and consistent terminology in all references to sections.

3. L91–95 – Remove or move to another section. The same applies to L96–104 and L120–125.

We agree that the mentioned passages could interrupt the flow of the introduction and will consider either removing them or relocating them to more appropriate sections of the manuscript.

4. 158 – Specify which previous methods are being referred to.

We thank the reviewer for this comment. The expression “previous rainfall disaggregation approaches” refers specifically to methods such as KNN, ANN-K, MMRC, Poisson-cluster rainfall generators, and other copula-based disaggregation methods mentioned in the Introduction. We will clarify this point in the revised manuscript to ensure that readers clearly understand which methods are being contrasted with our proposed approach.

5. L206 – Why was a 5-fold used? Also, divide the sentence for clarity.

A 5-fold cross-validation was selected as it provides a well-established compromise between bias (systematic error due to limited training data per fold) and variance (sensitivity of the evaluation to the specific fold partitioning) in the assessment of statistical and machine learning methods, and it has been commonly employed in scientific studies (Gutiérrez et al., 2019; Hernanz et al., 2022; Gong et al., 2026). With this setup, each fold contains enough data for training and testing while keeping computational demands reasonable.

We will divide the sentence to improve clarity.

Gutiérrez, J. M., Maraun, D., Widmann, M., Huth, R., Hertig, E., Benestad, R., ... and Pagé, C.: An intercomparison of a large ensemble of statistical downscaling methods over Europe: Results from the VALUE perfect predictor cross-validation experiment, *Int. J. Climatol.*, 39(9), 3750–3785, doi:10.1002/joc.5462, 2019.

Hernanz, A., Garcia-Valero, J. A., Domínguez, M., Ramos-Calzado, P., Pastor-Saavedra, M. A., and Rodríguez-Camino, E.: Evaluation of statistical downscaling methods for climate change projections over Spain: Present conditions with perfect predictors, *Int. J. Climatol.*, 42(2), 762–776, doi:10.1002/joc.7271, 2022.

Gong, F., Dai, L., Benchawattananon, R., Foong, S. Y., Cheng, Z., & Weng, C. (2026). Modern Pollen–Environment Calibrations in Thailand–Malaysia: Normalised Difference Vegetation Index (NDVI) and Precipitation as Quantitative Constraints for Vegetation Reconstruction. *Journal of Biogeography*, 53(3), e70191.

6. L243 – Explain why 0.04 mm was selected.

We thank the reviewer for this comment. A tolerance of 0.04 mm was selected to keep adjustments below the rounding precision of one decimal place used in the dataset. This ensures that the final daily totals are not affected while maintaining internal consistency in the hourly series.

7. L293 – Check the use of the word “by.”

It will be modified.

8. L331 – Rewrite this sentence for clarity.

We thank the reviewer for this comment. In line with Comment 3 from Referee 3 (see <https://doi.org/10.5194/egusphere-2025-4469-RC3>), which suggested removing the discussion of the Neyman-Scott Rectangular Pulse (NSRP) model as it is a continuous simulation approach rather than a disaggregation method, the paragraph containing line 331 will be removed. This ensures that the section focuses solely on relevant disaggregation methods and improves the clarity and conciseness of the manuscript.

9. 342 – Use the full form of MELODIST in the previous sentence.

We will use the full form of MELODIST in the previous sentence.

10. L375 – Clarify whether this analysis was done for each station individually or for all stations together.

We thank the reviewer for this comment. The analysis was conducted individually for each station. The boxplots presented in the manuscript summarize the results across all 91 stations, providing a concise overview of the station-wise distributions. We will clarify this point in the manuscript at the indicated line.

11. L408, L416 – Explain how these are benchmark methods; if not justified, consider removing the term.

We thank the reviewer for this comment. The term “benchmark disaggregation methods” is used to refer to a set of well-established, state-of-the-art methodologies commonly employed in rainfall disaggregation studies, including KNN, ANN-K, MMRC, and Poisson-cluster generators. These methods were selected because they represent a representative sample of current approaches against which Q-CODA can be meaningfully compared. We will clarify this justification in the manuscript at the indicated lines.

12. L413–415 – Rewrite to avoid redundancy with the figure reference. Check the term “daily maximum”- it appears to refer to a daily maximum 1-hour value.

We thank the reviewer for this observation. The term “daily” in line 415 referring to “daily maximum (Fig. 5h)” was a writing error. The Fig. 5h actually presents the maximum of all sub-daily 1-hour maxima, and we will remove “daily” to correct this. Additionally, we will reconsider the sentence structure to avoid excessive redundancy with the figure reference.

13. L417 – Replace with “high-temporal-resolution.”

It will be replaced with “high-temporal-resolution.”

14. L446, L473 – The use of “robust” needs more justification.

We thank the reviewer for this comment. Following the suggestion, we will revise the text in lines L446 and L473 to replace “robust” with more precise wording. The updated sentences will read:

“Overall, the results in Fig. 5 demonstrate the superior ability of Q-CODA to reproduce both the magnitude and statistical characteristics of peak sub-daily precipitation. Its consistent performance across multiple metrics and spatial locations highlights its value as a disaggregation tool that reliably reproduces observed rainfall features.”

“Similarly, the metrics in Fig. 6 show that Q-CODA achieves the best or near-best performance across all statistical and temporal features of the hourly rainfall series, providing a consistent advantage over state-of-the-art methods.”

These revisions clarify the intended meaning while avoiding unsupported terms.

15. L450–453, L465 – Rewrite these lines as suggested above for clarity and conciseness.

We thank the reviewer for this comment. We have revised lines 450-453 and 465 to enhance clarity and conciseness. In particular, the description of the analyzed metrics (bias in the temporal mean, temporal variance, zero proportion, mean event duration, and autocorrelations at lags 1, 2, 6, and 12) and the corresponding discussion of Q-CODA's performance have been updated as follows:

L450-453:

“The analyzed metrics include bias in the temporal mean (Fig. 6a), temporal variance (Fig. 6b), zero proportion (Fig. 6c), mean event duration (Fig. 6d), and autocorrelations at lags 1, 2, 6, and 12 (Figs. 6e-h).”

L465:

“Across all autocorrelation lags (lag-1, lag-2, lag-6, and lag-12; Figs. 6e-h), Q-CODA consistently exhibits the lowest median bias.”