This study presents the new algorithm Q-CODA for disaggregating precipitation data from daily to sub-daily values. The method leverages the relationship between daily precipitation totals and sub-daily maximum values, which is assumed and shown to be quasi-comonotonic, meaning that high daily precipitation totals are strongly associated with high sub-daily maximum values. The quasi-comonotonicity assumption is incorporated into the methodology via an upper bound copula. The algorithm is applied to 91 observational stations in Spain, and multiple other methodologies are included in the study as benchmarks.

A wide variety of evaluation metrics are applied, shedding light on which aspects of sub-daily precipitation the various methodologies handle well and which they struggle with (e.g., temporal consistency, the fraction of dry days, representation of extreme rainfall events).

The paper is generally well written and, as far as I can tell, methodologically sound. It appears that both Q-CODA and the benchmark methods are implemented according to best practices, which ensures a fair comparison.

I recommend that the paper be accepted with minor revisions, but have some suggestions that could strengthen the existing analysis.

We would like to sincerely thank the reviewer for their careful reading of our manuscript and for their constructive and encouraging feedback. We greatly appreciate the positive overall assessment of the study, particularly the recognition of the methodological soundness of Q-CODA, the fair implementation of benchmark methods, and the breadth of the evaluation framework. We are grateful for the recommendation of acceptance with minor revisions.

We have carefully considered all suggestions and believe they will help improve the clarity, completeness, and broader relevance of the paper. In the revised manuscript, we have addressed formatting issues, added missing references, clarified aspects of the methodology, and expanded the discussion where appropriate. In particular, we have strengthened the presentation of the quasi-comonotonicity assumption and its implications, improved the accessibility of Section 3, and added additional context regarding data quality control, evaluation metrics, performance variability across stations, and computational aspects of the algorithm. Additionally, we have introduced a dedicated Discussion section following the Results section to more thoroughly examine the broader methodological questions raised by the reviewer. In this new section, we expand the analysis of the quasi-comonotonicity assumption and discuss the applicability of Q-CODA beyond Spain by considering other regions of the world with contrasting precipitation regimes, including those where the dependence between daily totals and sub-daily maxima may be weaker.

Below, we respond point-by-point to each of the reviewer’s specific comments and indicate the corresponding changes made in the manuscript.

Specific Comments

Page 4, lines 120-121

There are some references missing from the list at the back. I found two, Lee et al. (2022) and Chodhury et al. (2025), but there may be more throughout the document. I recommend that the authors check all citations and ensure they are included in the reference list.

Thank you for pointing this out. We have carefully reviewed the manuscript and identified references that were cited in the text but unintentionally omitted from the reference list. In particular, the following references have now been added: Lee et al. (2022) and Chowdhury et al. (2025).

Lee, J., Kim, U., Kim, S., and Kim, J.: Development and application of a rainfall temporal disaggregation method to project design rainfalls, Water, 14, 1401, doi:10.3390/w1409140, 2022.

Chowdhury, D. P., Roy, D., and Saha, U.: Incorporation of weather parameters in MMRC-K model for rainfall disaggregation, Stoch. Environ. Res. Risk Assess., 39, 289–308, doi:10.1007/s00477-024-02863-4, 2025.

In addition, we have conducted a thorough check of all in-text citations to ensure that no other references are missing. We have also updated the reference list to include any new citations introduced as part of the revisions made during the review process.

Page 4, lines 126-127

Could you please describe the quality control that was applied to the observational data? It doesn't have to be a full and detailed account, but a brief mention of what type of homogeneity testing was applied or a reference to the dataset would be helpful.

Thank you for this helpful suggestion. We agree that a brief description of the quality control and homogeneity assessment improves transparency.

In the revised manuscript, we will expand this section to clarify that, in addition to the completeness filtering already described (≥90% data availability), we performed a homogeneity assessment of the annual precipitation series derived from the hourly records. Specifically, we applied three widely used breakpoint detection tests: the Pettitt test, the Standard Normal Homogeneity Test (SNHT), and the Buishand range test, using a significance level of 0.05. To further verify these cases, we performed a spatial consistency check by comparing detected change points with those of neighboring stations within a 50 km radius. This analysis indicated that 86 out of the 91 AEMET stations showed no evidence of inhomogeneities. The small subset of stations flagged by the tests were geographically clustered in northern Spain and exhibited coincident change points associated with particularly wet years, rather than documented station relocations or instrumental changes, suggesting false positives linked to natural variability. Importantly, none of these stations correspond to the outliers identified in the evaluation metrics of the disaggregation methods, indicating that these potential inhomogeneities do not affect the interpretation of the performance results.

We will incorporate a concise description of this procedure into the data section referencing the applied homogeneity tests.

Page 5-6, Section 3

The methodology could benefit from clearer presentation in certain sections. For example, on line 150-151, you write about "the upper bound copula C+, representing perfect positive dependence (comonotonicity)..." which is then presented in Equation (2). After the equation, you write: "This copula implies that X and Y increase together almost surely, which aligns with the quasi-comonotonic behaviour observed between daily precipitation totals and sub-daily maxima." The transition from perfect dependence to quasi-comonotonic behaviour is not entirely clear.

Additionally, while the benchmarks are described clearly, the description of the Q-CODA methodology is more difficult to follow, perhaps due to the level of detail. The flowchart certainly helps, and after reading Section 3 a couple of times it does make sense, but it could be more accessible. Could you add a simplified example showing how the comonotonic upper bound is translated into target maxima and used during the iterative adjustment process? This is key to understanding the method.

Thank you for this thoughtful and constructive comment. We appreciate the reviewer's careful reading of Section 3 and agree that the transition from the theoretical concept of perfect comonotonicity to the empirically observed quasi-comonotonic behaviour deserves further clarification, and that the methodological description could be made more accessible.

In the revised manuscript, we will clarify this point by explicitly stating that the Fréchet–Hoeffding upper bound copula is not assumed to represent the exact dependence structure in the data, but rather to define an idealized upper envelope corresponding to perfect positive dependence. We now explain more clearly that the empirical relationship between daily totals and sub-daily maxima is only quasi-comonotonic (i.e., very strong but not perfect), and that the upper bound copula is used as a practical way to translate this near-monotonic relationship into conditional target maxima that act as physically consistent constraints during disaggregation. In Section 3 we will add the following paragraph:

"In this study, the Fréchet–Hoeffding upper bound copula is not assumed to represent the exact dependence structure between daily precipitation totals and sub-daily maxima. Rather, it is used to define an idealized limiting case corresponding to perfect positive dependence (comonotonicity), which provides a useful theoretical reference for constructing physically consistent constraints during the disaggregation process. Empirical analyses (see Fig. 2) show that daily totals and sub-daily maxima exhibit a very strong but not perfectly monotonic relationship. We therefore refer to this behaviour as quasi-comonotonic: large daily totals are consistently associated with large sub-daily maxima, although variability remains due to event structure and timing. Within Q-CODA, the upper bound copula is used as a practical mechanism to translate this near-monotonic relationship into conditional target values for sub-daily extremes. These targets do not impose a deterministic mapping but instead provide an upper-envelope constraint that guides the reconstruction toward solutions that are statistically consistent with the observed dependence structure."

We will delete "*almost*" and change "*aligns*" to "*aligns approximately*" in your cited sentence:

“This copula implies that X and Y increase together ~~almost~~ surely, which aligns approximately with the quasi-comonotonic behaviour observed between daily precipitation totals and sub-daily maxima.”

We will also add the following clarification introducing Sect. 3.4:

“Conceptually, the method operates in percentile space. For each wet day, the target daily precipitation total is first located within the empirical distribution of daily totals derived from the training data. Under the assumption of comonotonic behaviour, the same percentile is then used to infer a corresponding target value for the sub-daily maximum precipitation from its empirical cumulative distribution function. This step is equivalent to applying the Fréchet–Hoeffding upper bound, which links the marginal distributions through a shared rank structure. The resulting target maximum is subsequently used as a constraint within the iterative adjustment procedure to ensure that the reconstructed hourly sequence is consistent with both the daily total and the expected magnitude of sub-daily maxima.”

To improve accessibility, we will also expand the methodological description to better guide the reader through the logic of Q-CODA. In particular, we have added a simplified illustrative example in Appendix A showing how a given daily total is mapped to percentile space, how this percentile is used to derive target sub-daily maxima from the training distribution using the comonotonic assumption, and how these targets are then used within the iterative adjustment process to constrain the hourly sequence while preserving the daily total and the temporal structure. This addition complements the flowchart and helps clarify how the theoretical dependence structure is operationalized in practice.

Overall, these revisions aim to make the connection between the copula framework, the quasi-comonotonic empirical evidence, and the iterative reconstruction procedure more transparent and easier to follow.

“**Appendix A: Detailed step-by-step example of the Q-CODA process for a single day.**

This appendix provides a fully reproducible worked example showing how Q-CODA works.

A.1 Step 1

Suppose a given day has a total precipitation amount of 60 mm. Based on the training dataset, suppose this value corresponds to the 90th percentile of the empirical distribution of daily totals. Under the comonotonic assumption, we assign the same percentile (90th) to the empirical cumulative distributions of sub-daily maxima, using Eq. (4) of the manuscript, yielding the following set of target constraints:

$\boldsymbol{M} = (P_{1h}^{max}, P_{2h}^{max}, P_{6h}^{max}, P_{12h}^{max}) = (10.0, 16.0, 30.0, 45.0)$

A.2 Step 2

Q-CODA uses a seed hourly profile selected from the historical record using a k-nearest neighbours (KNN) approach. In this example, the KNN procedure identifies a past day with similar daily precipitation magnitude and seasonal characteristics. The 24-hour rainfall distribution from this analogue day provides an initial temporal pattern that already contains realistic intra-day variability and persistence:

$\boldsymbol{h} = (h_1, h_2, \ldots, h_{24})$ = [0.0 , 0.0, 0.0, 0.9, 2.7, 5.4, 8.1, 7.2, 5.4, 3.6, 1.8, 0.9, 0.0, 0.0 , 0.0 , 0.0 , 0.9, 2.7, 3.6, 2.7, 1.8, 0.9, 0.0 , 0.0]

This seed profile is first rescaled so that its 24 hourly values sum exactly to the prescribed daily total (60 mm):

$\boldsymbol{h}$ = [0.0, 0.0, 0.0, 1.0, 3.0, 6.0, 9.0, 8.0, 6.0, 4.0, 2.0, 1.0, 0.0, 0.0, 0.0, 0.0, 1.0, 3.0, 4.0, 3.0, 2.0, 1.0, 0.0, 0.0]

However, after rescaling, the derived sub-daily máxima (computed from rolling 1 h, 2 h, 6 h and 12 h accumulations) will generally not match the percentile-based targets $\boldsymbol{M}$ obtained in step 1, $\boldsymbol{S} \neq \boldsymbol{M}$.

$\boldsymbol{S} = (S_{1h}^{max}, S_{2h}^{max}, S_{6h}^{max}, S_{12h}^{max}) = (9.0, 17.0, 35.0, 50.0)$

$\boldsymbol{M} = (P_{1h}^{max}, P_{2h}^{max}, P_{6h}^{max}, P_{12h}^{max}) = (10.0, 16.0, 30.0, 45.0)$

A.3 Step 3

Q-CODA then applies an iterative adjustment procedure to progressively enforce the set of magnitude constraints while preserving physical and temporal consistency. At each iteration, the algorithm locates the window where the accumulated rainfall is currently maximal, then computes the difference with respect to the target value and redistributes the excess or deficit across the τ hours within that window. After that, it enforces non-negativity and rescales slightly to keep the daily total.

A.3.1 Compute current maxima from the series

So compared with targets:

$\tau = 1$ h: $S_{1h}^{max} = 9$ vs $P_{1h}^{max} = 10 \rightarrow S_{1h}^{max}$ too low

$\tau = 2$ h: $S_{2h}^{max} = 17$ vs $P_{2h}^{max} = 16 \rightarrow S_{2h}^{max}$ too high

$\tau = 6$ h: $S_{6h}^{max} = 35$ vs $P_{6h}^{max} = 30 \rightarrow S_{6h}^{max}$ too high

$\tau = 12$ h: $S_{12h}^{max} = 50$ vs $P_{12h}^{max} = 45 \rightarrow S_{12h}^{max}$ too high

A.3.2 Local correction of each duration

The algorithm now loops over each constraint.

(a) Adjust $\tau = 1$ h

Target: $P_{1h}^{max} = 10$ mm

Current: $S_{1h}^{max} = 9$ mm

Difference: $\Delta_\tau = +1$ mm

It finds where the maximum occurs (hour 7) and adds:

$\Delta_\tau / \tau = 1 / 1 = +1$ mm

So hour 7 becomes 9 → 10 mm

(b) Adjust $\tau = 2$ h

Target: $P_{2h}^{max} = 16$ mm

Current: $S_{2h}^{max} = 18$ mm

Difference: $\Delta_\tau = -2$ mm

It finds where the maximum occurs (hours 7-8) and distributes this difference equally across those 2 hours:

$\Delta_\tau / \tau = -2 / 2 = -1$ mm

So hour 7 becomes 10 → 9 mm and hour 8 becomes 8 → 7 mm. New 2-hour sum 16 mm

(c) Adjust $\tau = 6$ h

Target: $P_{6h}^{max} = 30$ mm

Current: $S_{6h}^{max} = 34$ mm

Difference: $\Delta_\tau = -4$ mm

It finds where the maximum occurs (hours 5-10) and it subtracts: $\Delta_\tau / \tau = -4 / 6 = -0.\overline{6}$ mm from each of those 6 hours.

(d) Adjust $\tau = 12$ h

Target: $P_{12h}^{max} = 45$ mm

Current: $S_{12h}^{max} = 48$ mm

Difference: $\Delta_\tau = -3$ mm

It finds where the maximum occurs (hours 4-15) and it subtracts: $\Delta_\tau / \tau = -3 / 12 = -0.25$ mm from each of those 12 hours

A.3.3 Enforce non-negativity

If any value became negative after corrections, it is clipped to 0.

A.3.4 Rescale to preserve daily total

Because the function has added and removed rainfall locally, the total may no longer equal 60 mm. So it rescales the entire vector: $\boldsymbol{h} = \boldsymbol{h} \times (P_d / \Sigma h_i)$. This keeps the shape but restores the exact daily mass.

A.3.5 Recompute maxima and repeat

After the process the maxima is $\boldsymbol{S} = (S_{1h}^{max}, S_{2h}^{max}, S_{6h}^{max}, S_{12h}^{max}) = (9.2, 16.4, 36.1, 45.7)$, still slightly off ($\boldsymbol{S} \neq \boldsymbol{M}$).

So the algorithm repeats the same process: relocate peak windows; apply small corrections, and rescale total. This continues until: |calculated − target| < 0.1 mm for all durations, or until 20 iterations are reached. At the end, a final precise rescaling is applied if the daily total differs from P_d by more than predefined tolerance (0.04 mm) obtaining the following hourly vector (rounded to the first decimal): [2.1, 2.1, 2.1, 2.7, 2.7, 4.3, 10.0, 5.9, 4.3, 2.7, 2.7, 3.2, 0.0, 0.0, 0.0, 0.0, 1.1, 3.2, 4.3, 3.2, 2.1, 1.1, 0.0, 0.0]

A.3.6 Refinement for temporal autocorrelation

The goal is to improve vector lag-1 autocorrelation without deviating from $\boldsymbol{M}$.

Q-CODA identifies small isolated rainfall (< 5 mm) immediately before a non-zero hour and then it moves forward only if it increases autocorrelation and preserves $\boldsymbol{M}$.

In [2.1, 2.1, 2.1, 2.7, 2.7, 4.3, 10.0, 5.9, 4.3, 2.7, 2.7, 3.2, 0.0, 0.0, 0.0, 0.0, 1.1, 3.2, 4.3, 3.2, 2.1, 1.1, 0.0, 0.0] there is no isolated values, so Q-CODA does not indroduces any modification. However, suppose that in the 15th hour there is 0.4 mm instead of 0.0 mm. In that hypothetical case, the isolated 0.4 mm would be shifted to the following hour, as this would increase autocorrelation while keeping the sub-daily maxima within the acceptable tolerance.

Finally, Q-CODA permutes short windows (3–5 hours) by skipping windows containing zeros and trying all permutations within the window, accepting only if values in $\boldsymbol{M}$ remain within tolerance and autocorrelation improves. Permutations converts [2.1, 2.1, 2.1, 2.7, 2.7, 4.3, 10.0, 5.9, 4.3, 2.7, 2.7, 3.2, 0.0, 0.0, 0.0, 0.0, 1.1, 3.2, 4.3, 3.2, 2.1, 1.1, 0.0, 0.0] → [2.1, 2.1, 2.1 2.7, 2.7, 4.3, 0.0, 0.0, 0.0, 0.0, 1.1, 2.1, 3.2, 4.3, 3.2, 1.1, 0.0] which increases autocorrelation."

Page 11, line 290

The parenthesis in the cited paper should be after the name when it is in running text, i.e., "The ELU activation function was chosen by Bhattacharyya et al. (2024)..." rather than "(Bhattacharyya et al. 2024)".

This issue recurs later on lines 302, 303, and 344, but I may have missed other instances. While this is a minor formatting issue, I recommend that the authors carefully check all citations to ensure they follow the correct format.

Thank you for noting this formatting issue. We have corrected the citation style in the indicated instances (lines 290, 302, 303, and 344) so that, when authors are mentioned in running text, the year now appears in parentheses following the name (e.g., "Bhattacharyya et al. (2024)").

In addition, we conducted a thorough review of the entire manuscript to identify and correct any other similar cases, ensuring that all in-text citations consistently follow the appropriate formatting conventions throughout the paper.

General Comments

Evaluation Metrics

The evaluation metrics are described in the main text, but references or equations are not provided. While this makes the main text more readable, it would be useful to include an appendix with a fuller description of the metrics, including equations and references (e.g., for the Nash-Sutcliffe Efficiency and Wasserstein distance). This would improve the reproducibility of the study. I understand the code will also be made available, but it would still be good if the paper included all information needed to interpret and recreate the results.

Thank you for this valuable suggestion. We agree that providing a more complete description of the evaluation metrics improves reproducibility.

In response, we have added a new appendix (Appendix B) to the revised manuscript that includes a detailed description of all evaluation metrics used in the study, together with their mathematical formulations and key references (e.g., for the Nash–Sutcliffe Efficiency and the Wasserstein distance). The appendix clearly defines how each metric is computed for both daily maximum 1-hour precipitation and the hourly time series, including error measures, distribution-based metrics, bias indicators, intermittency diagnostics, event-duration statistics, and autocorrelation-based measures. While the main text retains a concise overview to preserve readability, this additional material ensures that all necessary information is now contained within the paper itself, allowing readers to fully interpret and independently reproduce the evaluation framework without relying solely on the accompanying code. Below is the appendix that will be added at the end of the manuscript.

Appendix B: Evaluation metrics

This appendix provides a detailed description of the evaluation metrics used in this study to assess the performance of the precipitation disaggregation methods. All metrics are computed by comparing simulated series against observations at each meteorological station, and results are summarized spatially using boxplots in Fig. 5 and Fig. 6. While the full implementation is provided with the released code, the definitions below allow the results to be independently interpreted and reproduced.

B.1 Metrics for Daily Maximum 1-hour Precipitation (Fig. 5)

Let O_i and S_i denote the observed and simulated daily maximum 1-hour precipitation values (P_{1h}^{max}) at time step i, respectively, with $i = 1,\ldots, N$.

B.1.1 Mean Absolute Error (MAE)

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|O_i - S_i| , \tag{B1}$$

The MAE measures the average magnitude of the errors without regard to their sign and provides an intuitive measure of absolute deviation.

B.1.2 Nash–Sutcliffe Efficiency (NSE)

$$NSE = 1 - \frac{\sum_{i=1}^{N}(S_i - O_i)^2}{\sum_{i=1}^{N}(O_i - O_{mean})^2} , \tag{B2}$$

where O_{mean} is the mean of the observed series. NSE evaluates the predictive skill of the model relative to the observed mean; values close to 1 indicate high skill, while negative values indicate performance worse than a climatological mean predictor.

B.1.3 1-D Wasserstein Distance

The 1-D Wasserstein distance (also known as the Earth Mover's Distance) quantifies the distance between the empirical distributions of simulated and observed values:

$$W_1(O,S) = \inf_{\gamma \in \Gamma(O,S)} \int |x-y| \, d\gamma(x,y) , \tag{B3}$$

Where $W_1(O, S)$ is the set of all joint distributions with marginals O and S. This metric captures differences in the full distribution, including extremes, and is particularly suitable for precipitation analysis.

B.1.4 Bias metrics

Relative biases (in percent) are computed for several statistical moments and extreme quantile:

$$Bias(X) = 100\frac{X_O - X_S}{X_O}, \quad (B4)$$

Where X represents: the mean, the variance, the 99th percentile, the 99.9th percentile and the maximum. These metrics showed in Fig.5 assess the ability of the disaggregation methods to reproduce not only central tendencies but also high-end extremes relevant for impact studies.

B.2 Metrics for Hourly Precipitation Time Series (Fig. 6)

Hourly precipitation series are constructed by concatenating the 24 hourly values for all available days at each station.

B.2.1 Mean and Variance Bias

Biases in the mean and variance of hourly precipitation are computed using the same relative bias formulation as in Section B.1.4 using complete hourly precipitation series instead of daily maximum 1-hour precipitation values (P_{1h}^{max}).

B.2.2 Zero Precipitation Proportion Bias

$$Z = \frac{1}{M}\sum_{j=1}^{M}\mathbb{I}(p_j = 0), \quad (B5)$$

Where p_j is hourly precipitation and $\mathbb{I}$ is the indicator function. The relative bias in Z quantifies the model's ability to reproduce intermittency.

B.2.3 Mean Precipitation Event Duration Bias

A precipitation event is defined as a sequence of consecutive hours with non-zero precipitation. The mean event duration is computed as the average length (in hours) of all events in the series. Bias is again expressed in relative terms.

B.2.4 Autocorrelation Bias

Temporal dependence is evaluated using lag-k autocorrelation coefficients:

$$p_k = \frac{cov(p_t, p_{t+k})}{\sigma^2}, \quad (B6)$$

Biases are computed for lags k = 1, 2, 6, and 12 hours.

Assumption of Quasi-Comonotonicity

The assumption of quasi-comonotonicity underpins the methodology, and the authors argue that the heterogeneity of precipitation regimes in Spain makes the results broadly applicable. While this argument is pretty convincing, I can think of precipitation regimes where the quasi-comonotonicity assumption might not hold as well. For example:

- In regions like northwestern Europe or the Pacific Northwest of the US and Canada, storm tracks bring sustained moderate precipitation over long periods of time, which may weaken the relationship between daily totals and sub-daily maxima.
- In monsoon regions, short-lived, strong convective systems could contribute to very high sub-daily maxima without corresponding high daily totals. This could weaken or break the dependence between daily totals and sub-daily maxima.

Could the authors expand on the discussion of quasi-comonotonicity and the applicability of Q-CODA in other regions? This might fit best in the Discussion section.

How sensitive is Q-CODA to deviations from the quasi-comonotonicity assumption? For example, could the authors test the algorithm on artificial datasets where the correlation between daily totals and sub-daily maxima is systematically reduced? This would help determine how far the assumption can deviate before the method's performance degrades.

We thank the reviewer for this thoughtful and constructive comment. We agree that empirical quasi-comonotonicity is a central element of the proposed methodology, and that its validity may vary across precipitation regimes. In response, we have expanded the discussion and carried out additional analyses in regions explicitly mentioned by the reviewer, as well as sensitivity experiments using synthetic datasets.

Applicability to other climatic regimes

We explored two contrasting precipitation regimes suggested by the reviewer: (i) mid-latitude maritime climates with long-duration frontal precipitation, and (ii) tropical monsoon environments with intense convective rainfall.

Because publicly available long-term hourly precipitation datasets with consistent coverage are limited, we focused on stations from the NCEI/NOAA archive where sufficiently long records exist. Note that reanalysis products are generally unable to adequately capture precipitation extremes unless they are generated with convection-permitting resolutions, as coarse-grid configurations tend to smooth or underestimate short-lived, high-intensity convective events; for this reason, they were not considered in the present study.

First, we analyzed five stations in the Pacific Northwest of the United States (Köppen Cfb climate), a region characterized by persistent frontal systems and prolonged moderate rainfall. The selected stations include three located on the Olympic Peninsula in Washington State (COOP:450013, COOP:456114, COOP:456624) and two coastal stations in Alaska (COOP:500352, COOP:509941). These stations are shown in Image 1. Before continuing, to avoid any possible confusion with the numbering of the figures in the revised manuscript, we will refer to the newly included figures in this response document as “Images”.

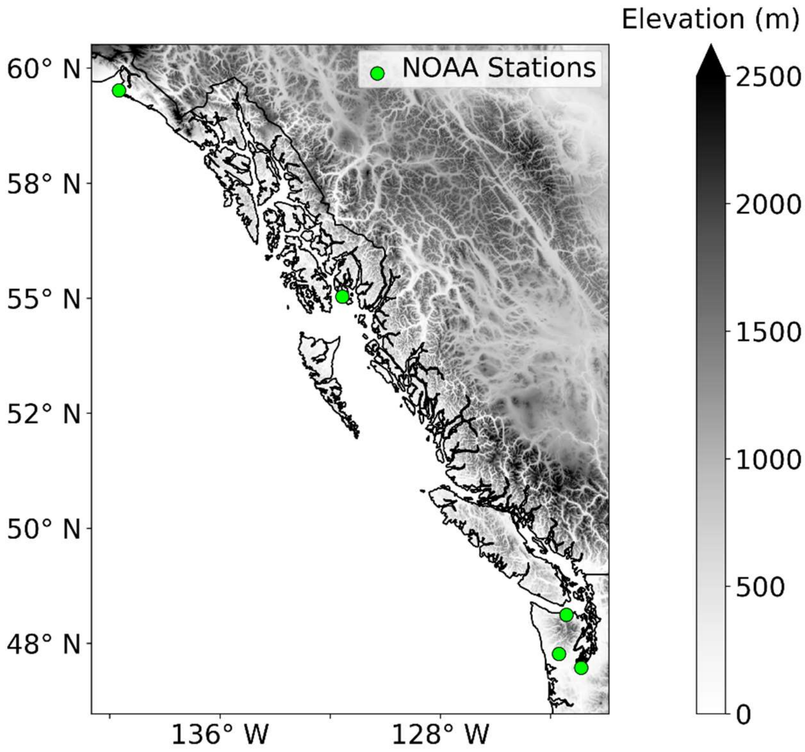

Image 1: Location of the five NCEI/NOAA stations selected in the Pacific Northwest of the United States for testing the quasi-comonotonicity assumption under a maritime mid-latitude precipitation regime (Köppen Cfb). Three stations are located on the Olympic Peninsula (COOP:450013, COOP:456114, COOP:456624) and two along the southern coast of Alaska (COOP:500352, COOP:509941). These stations are characterized by long-duration frontal precipitation systems and persistent moderate rainfall.

For these locations, we computed the Spearman correlation between daily totals and sub-daily máxima $\rho_s(P_{1h}^{max}, P_d)$ and evaluated the same performance metrics used in the main study (Wasserstein distance for the P_{1h}^{max} distribution, bias in the 99.9th percentile, lagged autocorrelation bias, and IDF-based RMSE). For consistency with the Spanish case study while accounting for data availability in the NCEI–NOAA archive, the evaluation period for the international stations was extended to 50 years (1951–2000). The same validation framework was applied using a 5-fold cross-validation scheme, where each fold corresponds to a continuous 10-year block. This configuration ensures temporal robustness in the assessment while maximizing the use of the available long-term hourly precipitation records.

The results (see Image 2) show that the quasi-comonotonic relationship remains strong even in this more persistent rainfall regime, although slightly weaker than in Spain. The lowest observed Spearman correlation decreases to approximately 0.984, whereas in the Spanish dataset even the weakest stations rarely fell below 0.990. Despite this reduction, Q-CODA continues to perform robustly and consistently better than the benchmark methods, particularly for extreme-value reproduction. This is reflected in the low bias in the 99.9th percentile of hourly maxima and the low RMSE in the IDF curves for a 100-year return period. These results suggest that the method remains applicable even when the quasi-comonotonicity assumption is somewhat weaker than in Mediterranean-type regimes.

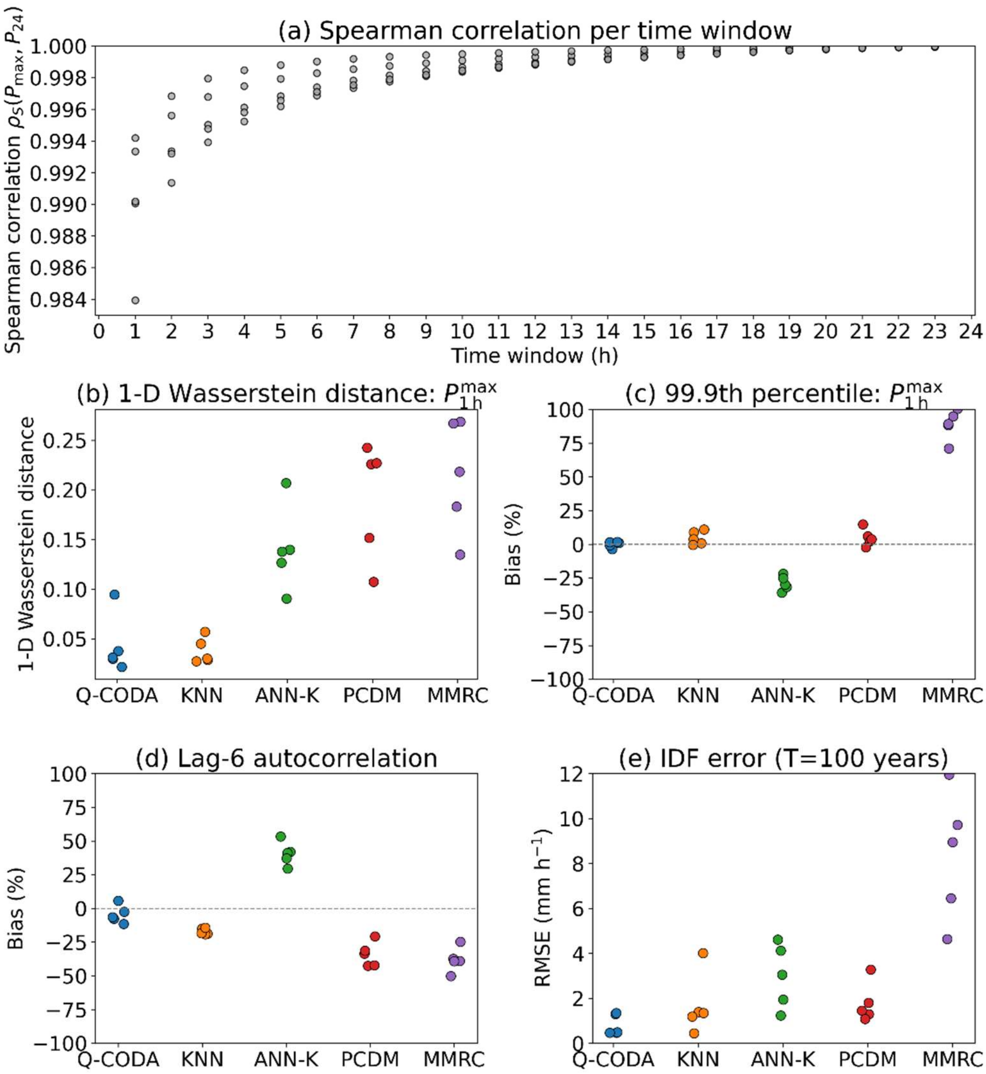

Image 2: Performance of Q-CODA and benchmark disaggregation methods (KNN, ANN-K, PCDM, and MMRC) for the five Pacific Northwest stations (Köppen Cfb). Panel (a) shows the Spearman correlation between sub-daily maxima (for different aggregation windows) and daily totals. Panels (b–e) present the evaluation metrics for the simulated series: (b) 1-D Wasserstein distance for P_{1h}^{max}, (c) bias of the 99.9th percentile of P_{1h}^{max}, (d) bias in lag-6 autocorrelation of the complete hourly series, and (e) RMSE of the IDF curves for a 100-year return period.

We also analyzed three stations located in the Florida Peninsula (COOP:085663, COOP:086323, COOP:088780), representative of a tropical monsoon climate (Köppen Am), where short-lived convective bursts could in principle weaken the dependence between daily totals and sub-daily maxima.

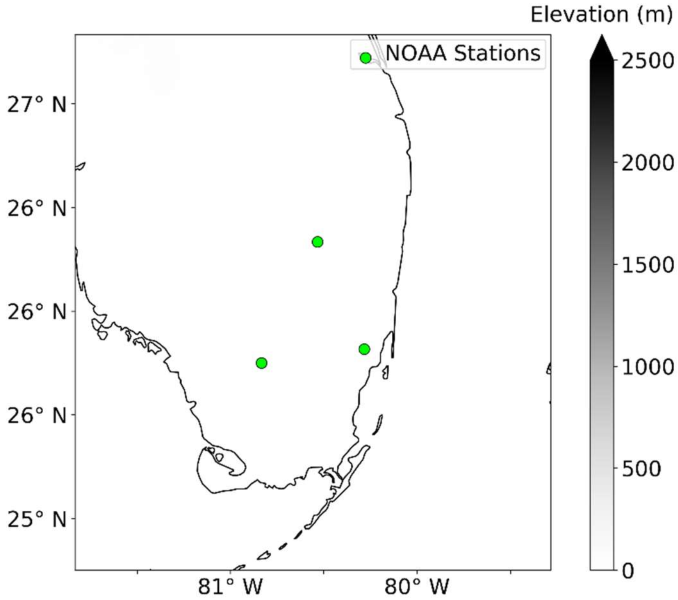

Image 3: Location of the three NCEI/NOAA stations selected in the Florida Peninsula to evaluate Q-CODA under a tropical monsoon climate (Köppen Am). The stations (COOP:085663, COOP:086323, COOP:088780) are representative of environments dominated by short-lived, high-intensity convective precipitation events.

However, in these stations the Spearman correlation between daily totals and 1-hour maxima exceeds 0.997 (see Image 4), which falls well within and even at the upper end of the range observed across AEMET stations in Spain. This indicates that, despite the tropical monsoonal regime and the presence of intense convective rainfall, the empirical dependence between daily accumulation and sub-daily extremes remains strongly quasi-comonotonic. Under these conditions, Q-CODA again exhibits performance fully consistent with the results reported in the main manuscript. In particular, it maintains very low bias in the 99.9th percentile of the 1-hour maxima distribution and low RMSE in the 100-year IDF estimates, confirming its ability to accurately reproduce high-end extremes. At the same time, the method preserves the temporal structure of the hourly series, as reflected by stable autocorrelation. Overall, the Florida case study reinforces that when quasi-comonotonicity is strong, even in highly convective tropical environments, Q-CODA remains robust and provides clear advantages over the benchmark disaggregation methods.

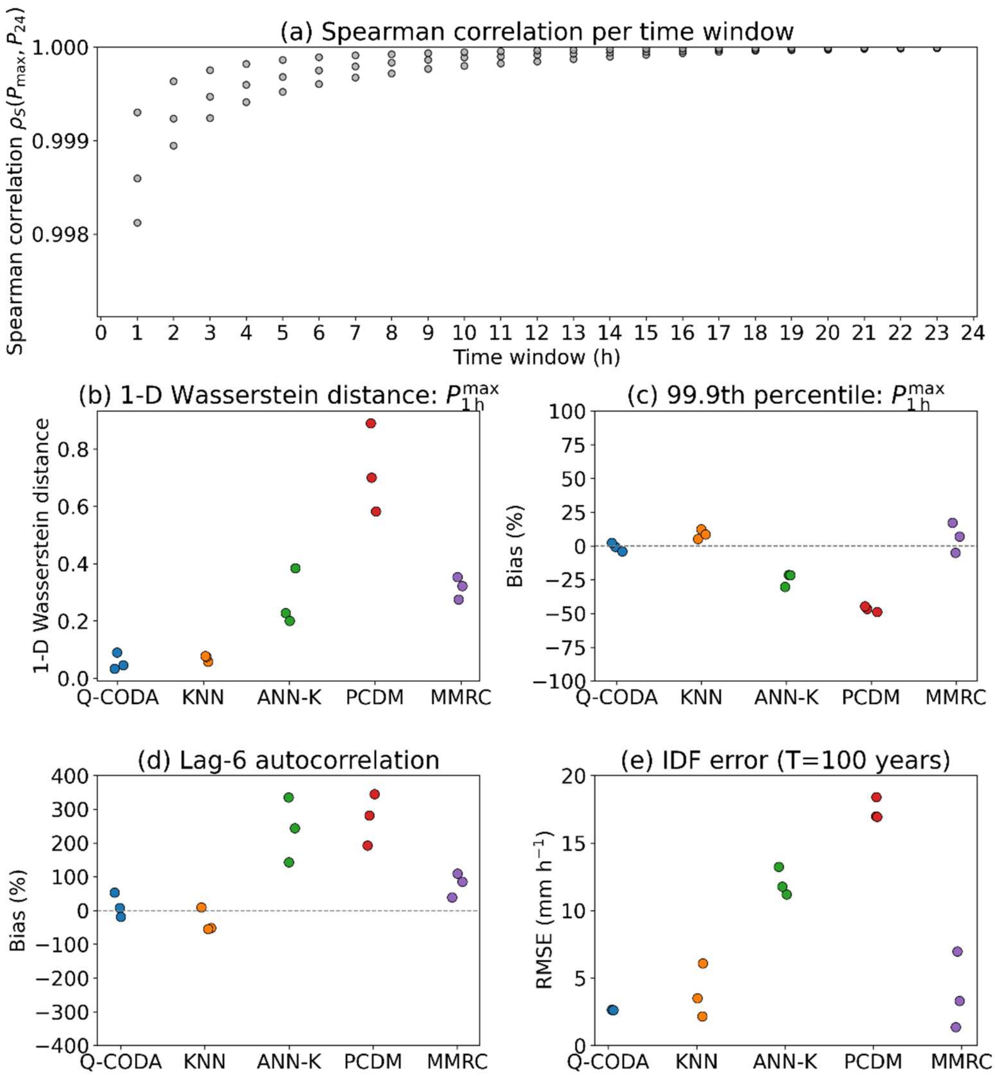

Image 4: Performance of Q-CODA and benchmark disaggregation methods (KNN, ANN-K, PCDM, and MMRC) for the three Florida stations located in a tropical monsoon climate (Köppen Am). Panel (a) shows the Spearman correlation. between sub-daily maxima (for different aggregation windows) and daily totals. Panels (b–e) present the evaluation metrics for the simulated series: (b) 1-D Wasserstein distance for P_{1h}^{max}, (c) bias of the 99.9th percentile of P_{1h}^{max}, (d) bias in lag-6 autocorrelation of the complete hourly series, and (e) RMSE of the IDF curves for a 100-year return period.

Sensitivity to deviations from quasi-comonotonicity

To more directly address the reviewer's question regarding sensitivity to the assumption itself, we performed an additional set of controlled experiments using six artificial hourly precipitation series designed to emulate regimes with progressively weaker dependence between daily totals and hourly maxima. These artifitial series allow us to systematically explore conditions that are rarely observed in real long-term station records.

An analysis analogous to the observational cases was conducted, relating model performance to the Spearman correlation between daily totals and P_{1h}^{max}. The results indicate that Q-CODA remains robust over a wide range of dependence strengths. However, when the Spearman correlation drops below approximately 0.975 (see Image 5), a degradation in performance begins to appear across metrics such as 99.9th percentile bias (Image 5c) and RMSE of the IDF curves (Image 5e). In this regime, the quasi-comonotonicity assumption becomes less reliable, and the added value of the copula-based adjustment diminishes. In such hypothetical cases, the simpler KNN-based disaggregation can become preferable. In any case, Q-CODA continues to outperform ANN-K, PCDM, and MMRC.

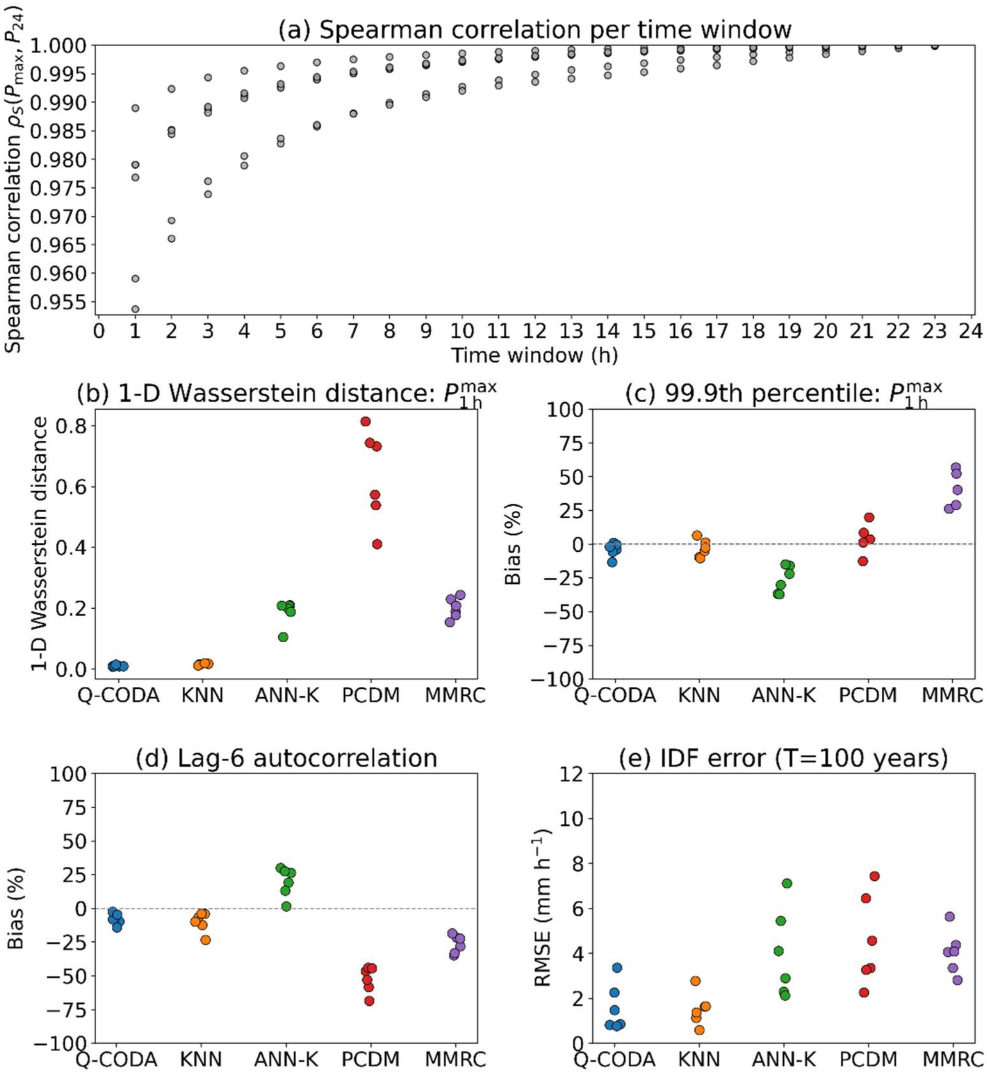

Image 5: Performance of Q-CODA and benchmark methods on 6 artificial hourly precipitation series constructed to progressively weaken the quasi-comonotonic relationship between daily totals and 1-hour maxima (P_{1h}^{max}). Panel (a) shows the Spearman correlation. between sub-daily maxima (for different aggregation windows) and daily totals. Panels (b–e) present the evaluation metrics for the simulated series: (b) 1-D Wasserstein distance for P1hmax, (c) bias of the 99.9th percentile of P_{1h}^{max}, (d) bias

in lag-6 autocorrelation of the complete hourly series, and (e) RMSE of the IDF curves for a 100-year return period.

Results indicate that Q-CODA remains robust for moderately reduced dependence, but a degradation in performance begins to emerge when the Spearman correlation falls below approximately 0.975, providing an empirical threshold for the practical validity of the quasi-comonotonicity assumption. This threshold-based behavior is consistent with the theoretical role of the Fréchet–Hoeffding upper bound in the method: the stronger the rank dependence between daily totals and sub-daily maxima, the more informative the percentile mapping becomes, and the more effectively Q-CODA can constrain the hourly reconstruction.

Summary for inclusion in the Discussion

Overall, these additional analyses support three key conclusions:

1. The quasi-comonotonic relationship between daily totals and sub-daily maxima appears to be remarkably robust across different climatic regimes, including oceanic mid-latitude and tropical monsoon environments.

2. Q-CODA maintains good performance even when this dependence weakens moderately (e.g., Spearman ≈ 0.980-0.990), as observed in the Pacific Northwest.

3. Sensitivity experiments suggest that the method begins to lose its advantage when the dependence drops below approximately 0.975, providing a practical guideline for assessing applicability in new regions.

We will incorporate a discussion of these findings into the revised manuscript in a new Discusion section in order to clarify the climatic generality of the quasi-comonotonicity assumption and to provide a transparent perspective on the range of conditions under which Q-CODA is expected to perform optimally.

Performance Variability Across Stations

The Q-CODA algorithm performs well across most evaluation metrics, but some outliers appear (in Figure 6b and 6f for example), suggesting that the method struggles a bit at certain stations. This is to be expected, but it would be interesting to investigate these cases further. For example:

- Do these outliers correspond to stations with lower correlation values in Figure 2, potentially indicating weaker quasi-comonotonic dependence?
- Are there locations where all methods perform poorly, potentially pointing to data quality issues or precipitation processes that are difficult to capture with disaggregation methods?

We thank the reviewer for this insightful comment. We agree that the presence of some outliers in Figs. 6b and 6f deserves further investigation, and we have conducted an additional analysis to better understand whether these cases are linked to weak quasi-comonotonic dependence, station-specific issues, or structural limitations of disaggregation methods.

First, it is important to clarify that Fig. 6b in the manuscript evaluates the absolute bias in the variance of the full hourly series, while Fig. 6f evaluates the absolute bias in lag-2 autocorrelation of the full hourly series. These metrics characterize the temporal structure of the reconstructed hourly signal rather than the behavior of extreme sub-daily rainfall.

To investigate the origin of the observed outliers, we produced an additional diagnostic panel (see Image 6) relating performance deterioration to the strength of quasi-comonotonic dependence.

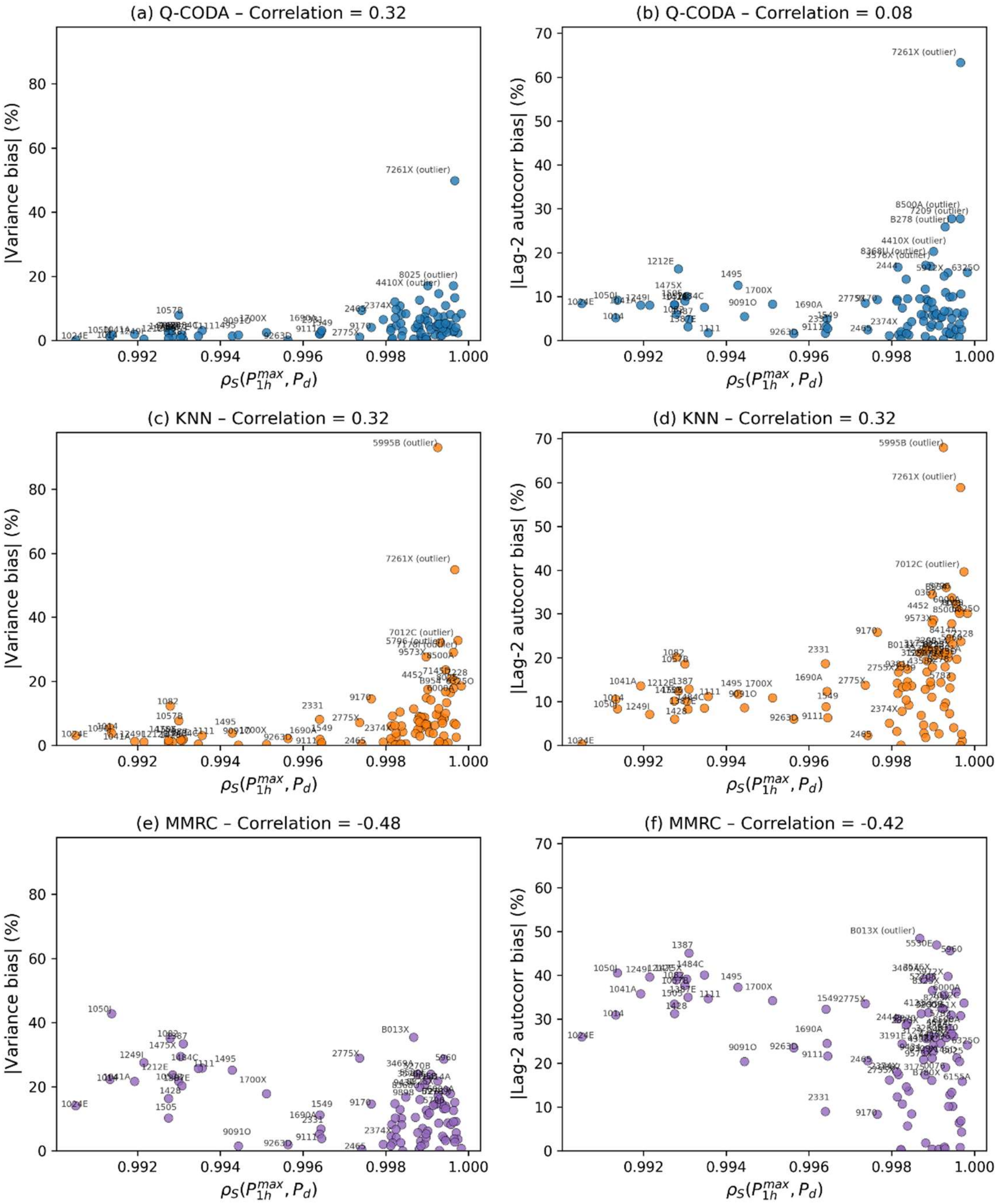

Image 6: Pearson correlation between temporal structure-based metrics and $\rho_s(P_{1h}^{max}, P_d)$ for disaggregation methods Q-CODA, KNN, and MMRC. A label has been added to

each point indicating the corresponding AEMET station ID, and, in parentheses, whether the station is identified as an outlier in Figs. 6b or 6f.

Specifically, we analyzed the relationship between:

- X-axis: Spearman correlation between daily totals and 1-hour maxima (as shown in Fig. 2 in the manuscript),
- Y-axis: absolute bias of the corresponding evaluation metric,

and computed Pearson correlations for each method.

For the metrics associated with Fig. 6 (variance bias and lag-2 autocorrelation bias), the Pearson correlations between performance deterioration and the loss of comonotonicity are generally weak. For Q-CODA, the correlations are 0.32 (variance bias, Image 6a) and 0.08 (lag-2 autocorrelation bias, Image 6b). Comparable or stronger relationships are observed for the benchmark methods. This indicates that the variability seen in Figs. 6b and 6f cannot be primarily attributed to weaker dependence between daily totals and hourly maxima. Additionally, we explicitly tracked the stations identified as outliers in Figs. 6b and 6f and located them in the new diagnostic plots in Image 6. In these plots, labels were added for all stations using their station IDs, with the outlier stations clearly marked by including "(outlier)" in parentheses next to the station ID. The identified outliers do not coincide across methods, which argues against systematic data quality issues at specific stations. Instead, the affected stations vary depending on the method and the metric, suggesting that the observed dispersion is method- and metric-specific rather than station-driven.

The weak relationship with quasi-comonotonicity is also physically consistent. The metrics in Fig. 6 evaluate properties of the complete hourly time series (variance and autocorrelation structure). In Q-CODA, these characteristics are partly inherited from the KNN seed series and subsequently refined, so their behavior is not expected to be strongly controlled by the daily–subdaily dependence structure captured in Fig. 2 of the manuscript. In contrast, the influence of quasi-comonotonicity level becomes much clearer when analyzing the metrics associated with extreme sub-daily rainfall (Fig. 5). We performed an analogous analysis using:

- Wasserstein distance (1-D) for the distribution of 1-hour maxima, and
- Absolute bias of the 99.9th percentile,

again plotted against the Spearman correlation between daily totals and 1-hour maxima.

In this case (see Image 7), stronger relationships emerge. For Q-CODA, the Pearson correlation between Wasserstein distance and the dependence strength is −0.59 (Image 7a), indicating that performance improves as quasi-comonotonicity increases. However, this pattern is weaker than for MMRC (Image 7e), suggesting that classical approaches are more sensitive to low-dependence regimes. For the 99.9th percentile bias, Q-CODA shows substantially lower sensitivity to the loss of dependence compared with KNN and MMRC, reflecting its improved robustness in reproducing extreme hourly values across different regimes (see Images 7b, 7d and 7f).

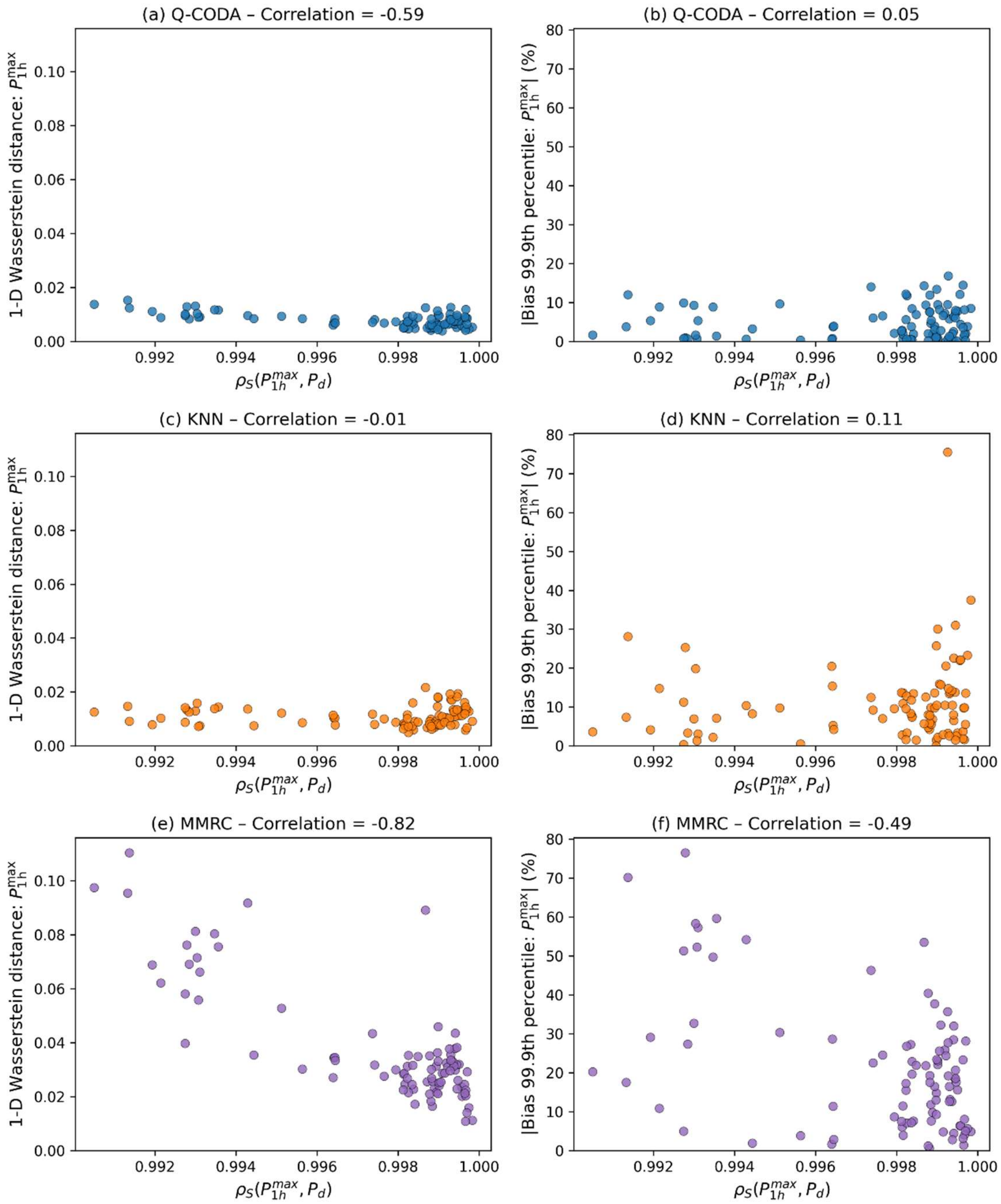

Image 7: Pearson correlation between P_{1h}^{max} –based metrics and $\rho_s(P_{1h}^{max}, P_d)$ for disaggregation methods Q-CODA, KNN, and MMRC.

Importantly, the stations with low Spearman correlation in Fig. 2 tend to correspond to oceanic climates, where rainfall processes are more distributed and less dominated by short, intense convective bursts. In contrast, higher dependence values are more common in temperate, Mediterranean, and drier climates, where intense sub-daily events are more tightly linked to daily totals. This climatological distinction helps explain why quasi-comonotonicity has a stronger effect on the metrics related to hourly extremes than on those describing the overall temporal structure.

In summary:

1. The outliers observed in Figs. 6b and 6f do not correspond to stations with systematically lower daily–subdaily dependence.

2. They do not coincide across methods, which argues against station-specific data quality problems.

3. Weak quasi-comonotonicity mainly affects metrics tied to extreme hourly rainfall (Fig. 5), not those describing the full temporal structure (Fig. 6).

4. Compared with benchmark methods, Q-CODA shows lower sensitivity to low-dependence regimes, particularly for extreme 99.9th percentile.

We will incorporate a brief discussion of these findings in the revised manuscript to clarify the origin of performance variability across stations.

Computational Feasibility

The paper does not discuss the computational costs of Q-CODA relative to the benchmarks. Given the iterative nature of the algorithm, this could be an important consideration, particularly for larger datasets or real-time applications. Please add a discussion of the computational efficiency of Q-CODA as compared to the benchmark methods. Even better, you could perform a comparative test of the computational costs. I am thinking of runtime, where parallelization might be an option for some of these methods, but also computational costs and energy usage (which can be checked with tools like CodeCarbon in python).

Thank you for highlighting this important aspect. We agree that discussing computational efficiency is valuable, particularly given the iterative nature of Q-CODA and its potential application to large datasets.

In response, we have conducted an additional analysis to quantify the computational cost of Q-CODA in comparison with the benchmark methods. Specifically, we used the suggested CodeCarbon tool to estimate runtime and associated computational cost for each method, applying all disaggregation methods to the same reference station used to illustrate the derivation of IDF curves in Fig. 7. All computations were performed on a workstation equipped with an Intel Xeon Gold 5218 processor (32 cores) and 63 GB of RAM. This provides a consistent basis for comparison under identical data and processing conditions. The results of this comparative assessment (see table below) have been incorporated at the end of the Results section, where we briefly discuss the relative runtime and computational footprint of Q-CODA versus the benchmark approaches. This addition helps contextualize the practical feasibility of the method for larger-scale or operational applications.

Method	Q-CODA	KNN	ANN-K	PCDM	MMRC
Duration (s)	266.690	60.967	448.32	82.58	67.59
Energy (kWh)	0.03006	0.00686	0.10236	0,00930	0.00747
Emissions (kgCO_2)	0.00523	0.00119	0.01782	0,00162	0.00130

Table 1: Comparison of computational costs across methods.