

S1: Detailed step-by-step example of the Q-CODA process for a single day.

Supplementary Sect. S1 provides a fully reproducible worked example showing how Q-CODA works.

S1.1 Step 1

- 5 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), yielding the following set of target constraints:

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

S1.2 Step 2

- 10 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:

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

15

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

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

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

20 $\mathbf{S} = (S_{1h}^{\max}, S_{2h}^{\max}, S_{6h}^{\max}, S_{12h}^{\max}) = (9.0, 17.0, 36.0, 40.0)$

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

S1.3 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.

S1.3.1 Compute current maxima from the series

So compared with targets:

30 $\tau = 1$ 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: $S_{6h}^{\max} = 36$ vs $P_{6h}^{\max} = 30 \rightarrow S_{6h}^{\max}$ too high

$\tau = 12$ h: $S_{12h}^{\max} = 40$ vs $P_{12h}^{\max} = 45 \rightarrow S_{12h}^{\max}$ too high

S1.3.2 Local correction of each duration

35 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

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

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

So hour 7 becomes $9 \rightarrow 10$ mm

(b) Adjust $\tau = 2$ h

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

45 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:

$$\Delta_{\tau} / \tau = -2 / 2 = -1 \text{ mm}$$

So hour 7 becomes $10 \rightarrow 9$ mm and hour 8 becomes $8 \rightarrow 7$ mm. New 2-hour sum 16 mm

50 (c) Adjust $\tau = 6$ h

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

Current: $S_{6h}^{\max} = 35$ mm

Difference: $\Delta_{\tau} = -5$ mm

It finds where the maximum occurs (hours 5-10) and it subtracts: $\Delta_{\tau} / \tau = -5 / 6 = -0.8\bar{3}$ mm from each of those 6 hours.

55 (d) Adjust $\tau = 12$ h

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

Current: $S_{12h}^{\max} = 34$ mm

Difference: $\Delta_{\tau} = 11$ mm

It finds where the maximum occurs (hours 4-15) and it adds: $\Delta_{\tau} / \tau = 11 / 12 = 0.91\bar{6}$ mm from each of those 12 hours

60 **S1.3.3 Enforce non-negativity**

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

S1.3.4 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: $\mathbf{h} = \mathbf{h} \times (P_d / \sum h_i)$. This keeps the shape but restores the exact daily mass.

65 S1.3.5 Recompute maxima and repeat

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

So the algorithm repeats the same process: relocate peak windows; apply small corrections, and rescale total. This continues until: $|\text{calculated} - \text{target}| < 0.1$ 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]

S1.3.6 Refinement for temporal autocorrelation

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

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 $\mathbf{M}$.

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 introduces 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.

80 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 $\mathbf{M}$ 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] $\rightarrow$ [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, 2.1, 3.2, 4.3, 3.2, 1.1, 0.0, 0.0] which increases autocorrelation.

85 S2: Evaluation metrics

Supplementary Sect. S2 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.

S2.1 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 $i = 1, \dots, N$.

S2.1.1 Mean Absolute Error (MAE)

$$MAE = \frac{1}{N} \sum_{i=1}^N |O_i - S_i|, \quad (S1)$$

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

S2.1.2 Nash–Sutcliffe Efficiency (NSE)

$$NSE = 1 - \frac{\sum_{i=1}^N (S_i - O_i)^2}{\sum_{i=1}^N (O_i - O_{\text{mean}})^2}, \quad (S2)$$

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.

S2.1.3 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:

$$W_1(O, S) = \inf_{\gamma \in \Gamma(O, S)} \int |x - y| d\gamma(x, y), \quad (S3)$$

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.

S2.1.4 Bias metrics

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

$$Bias(X) = 100 \frac{X_O - X_S}{X_O}, \quad (S4)$$

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.

115 **S2.2 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.

S2.2.1 Mean and Variance Bias

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

120 **S2.2.2 Zero Precipitation Proportion Bias**

$$Z = \frac{1}{M} \sum_{j=1}^M \mathbb{I}(p_j = 0) , \quad (S5)$$

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.

S2.2.3 Mean Precipitation Event Duration Bias

125 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.

S2.2.4 Autocorrelation Bias

Temporal dependence is evaluated using lag- k autocorrelation coefficients:

$$p_k = \frac{\text{cov}(p_t, p_{t+k})}{\sigma^2} , \quad (S6)$$

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

S3: Annual data completeness across 91 AEMET stations with > 90% hourly data availability (1996-2024)

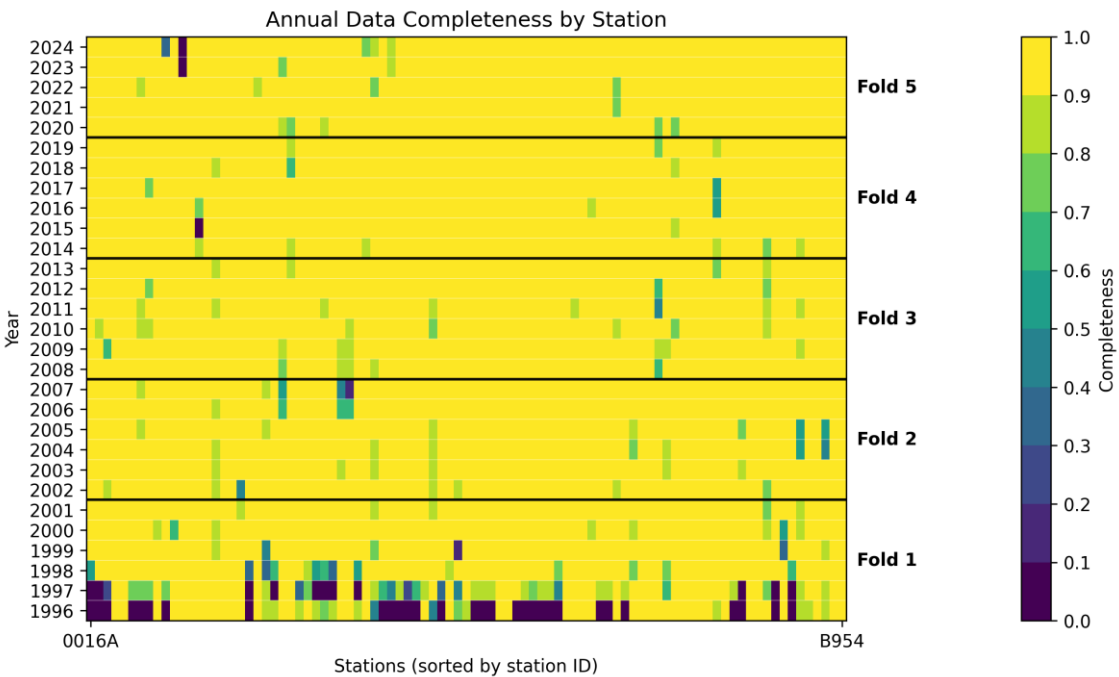


Figure S1: Annual data completeness by station (1996-2024). Rows correspond to years, columns to stations (sorted by ID), and colour indicates the proportion of available hourly data. White lines separate years; thicker black lines indicate fold boundaries.

S4: Performance Variability Across Stations: analysis

Supplementary Sect. S4 provides a detailed analysis of the station-level outliers observed in the performance of Q-CODA and the benchmark disaggregation methods. In particular, it investigates whether the outliers identified in Fig. 6 are associated with weaker quasi-comonotonic dependence, station-specific data issues, or structural limitations of the methods. To this end, we examine the relationship between the strength of daily–subdaily dependence (measured as the Spearman correlation between daily precipitation totals and 1-hour maxima) and the absolute bias of two key evaluation metrics, variance bias (Fig. 6b) and lag-2 autocorrelation bias (Fig. 6f) of the reconstructed full hourly series. These metrics characterize the temporal structure of the reconstructed hourly signal rather than the behaviour of extreme sub-daily rainfall. To investigate the origin of the observed outliers, we produced an additional diagnostic panel (see Fig. S2) relating performance deterioration to the strength of quasi-comonotonic dependence.

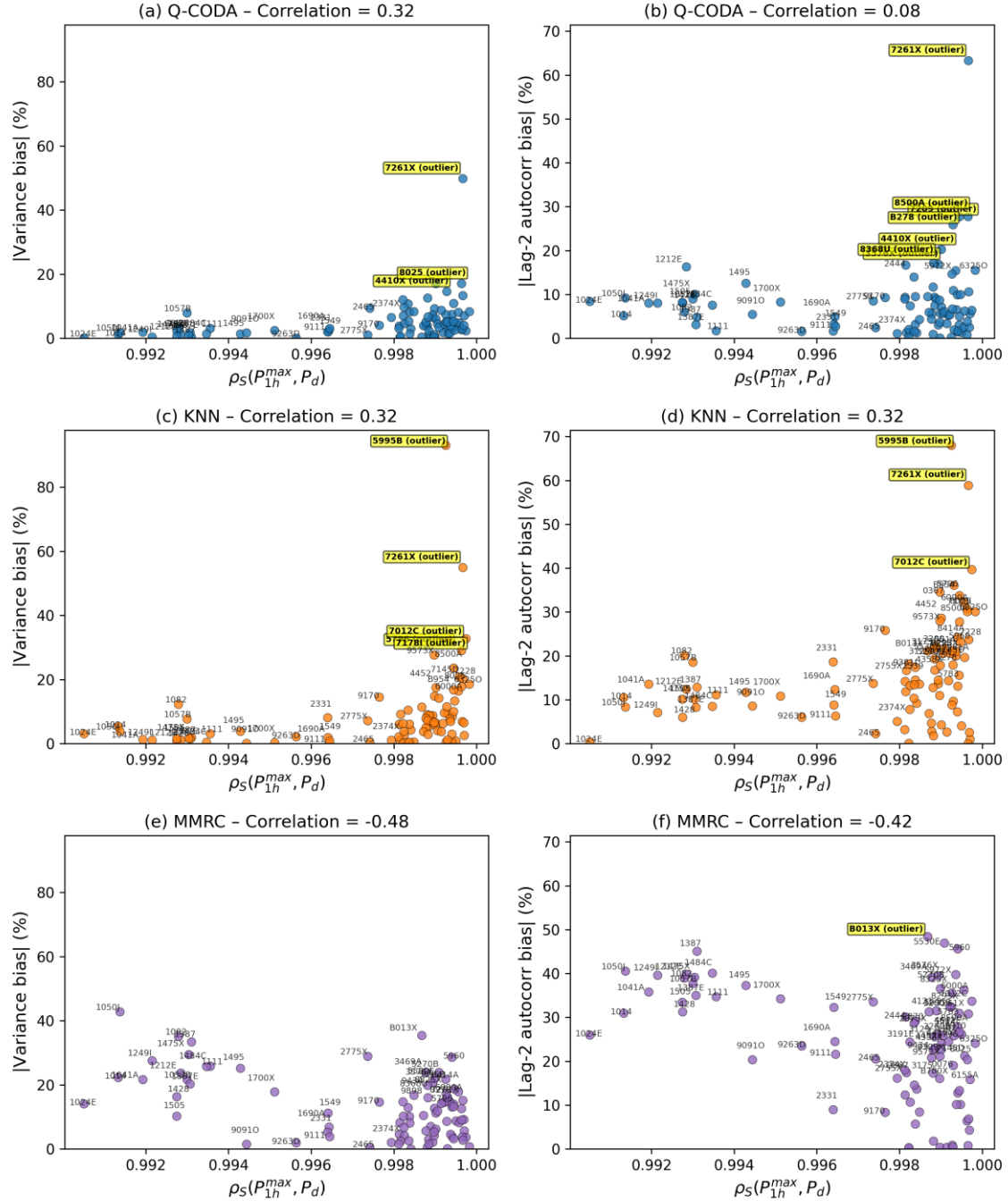
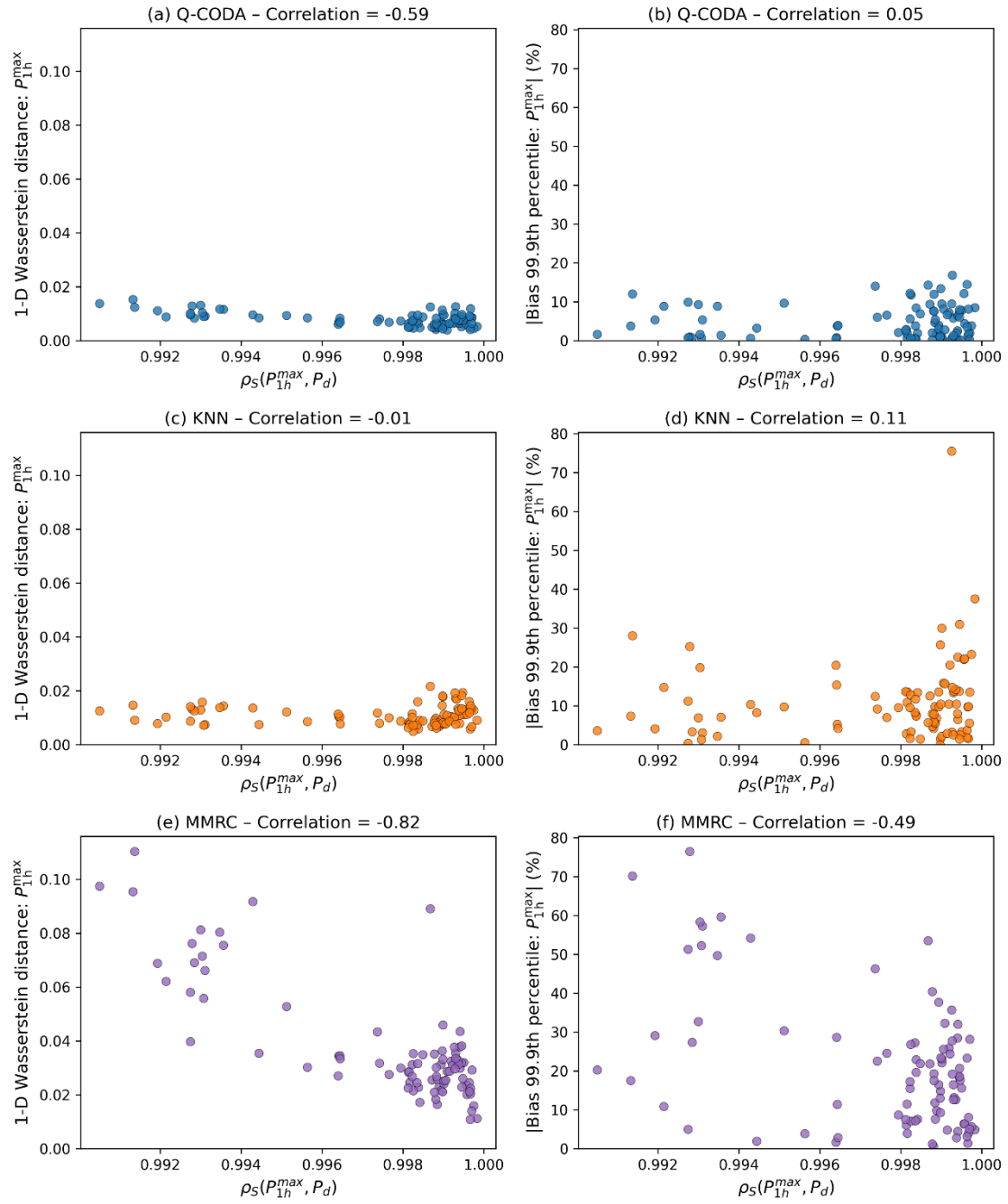


Figure S2: Pearson correlation between temporal structure-based metrics and $\rho_s(P_{1h}^{\max}, P_d)$ for disaggregation methods Q-CODA, KNN, and MMRC. A label has been added to each point indicating the corresponding AEMET station ID, and, in parentheses and yellow, whether the station is identified as an outlier in Figs. 6b or 6f.

Specifically, we analysed the relationship between: (a) X-axis: Spearman correlation between daily totals and 1-hour maxima (as shown in Fig. 2), and (b) 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, Fig. S2a) and 0.08 (lag-2 autocorrelation bias, Fig. S2b). 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 Fig. S2. In these plots, labels were added for all stations using their station IDs, with the outlier stations clearly marked by including “(outlier)” in parentheses and yellow 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 behaviour is not expected to be strongly controlled by the daily–subdaily dependence structure captured in Fig. 2. In contrast, the influence of quasi-comonotonicity level becomes much clearer when analysing 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 Fig. S3), stronger relationships emerge. For Q-CODA, the Pearson correlation between Wasserstein distance and the dependence strength is -0.59 (Fig. S3a), indicating that performance improves as quasi-comonotonicity increases. However, this pattern is weaker than for MMRC (Fig. S3e), 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 ability to reproduce extreme hourly values across different regimes (see Figs. S3b, S3d and S3f).



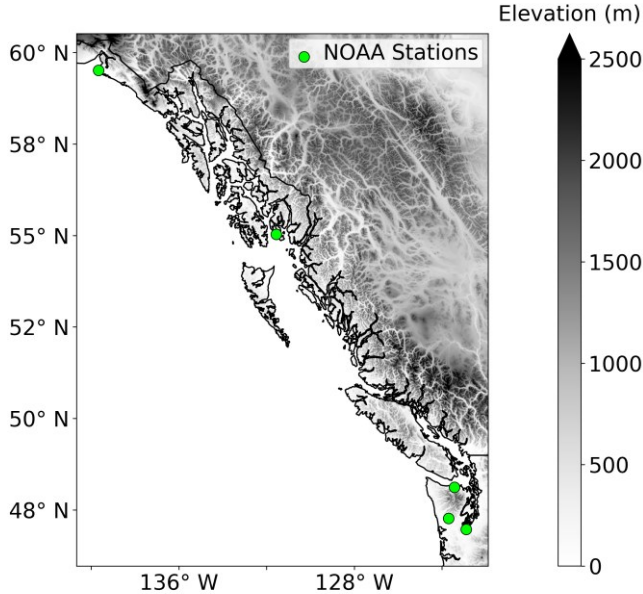
175

Figure S3: Pearson correlation between $P_{1h}^{\max}$ -based metrics and $\rho_S(P_{1h}^{\max}, P_d)$ for disaggregation methods Q-CODA, KNN, and MMRC.

180 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.

185 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 state-of-the-art methods, Q-CODA shows lower sensitivity to low-dependence regimes, particularly for extreme 99.9th percentile.

S5: Maps of Other Climatic Regimes



190 **Figure S4: Location of the five NCEI/NOAA stations selected in the Pacific Northwest of the United States for testing the quasi-comonotonicity assumption under an oceanic 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.**

195

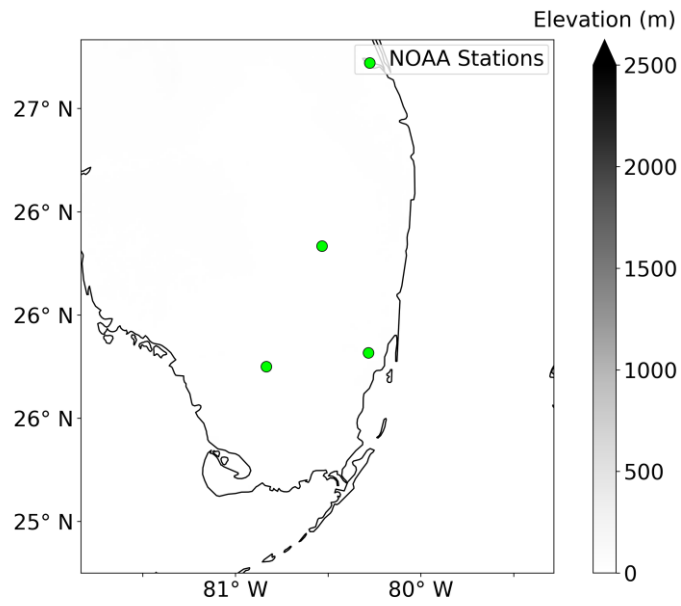


Figure S5: Location of the three NCEI/NOAA stations selected in the Florida Peninsula to evaluate Q-CODA under a tropical monsoon climate (Köppen Am). The stations (COOP:085663, COOP:086323, COOP:088780) are representative of environments dominated by short-lived, high-intensity convective precipitation events.

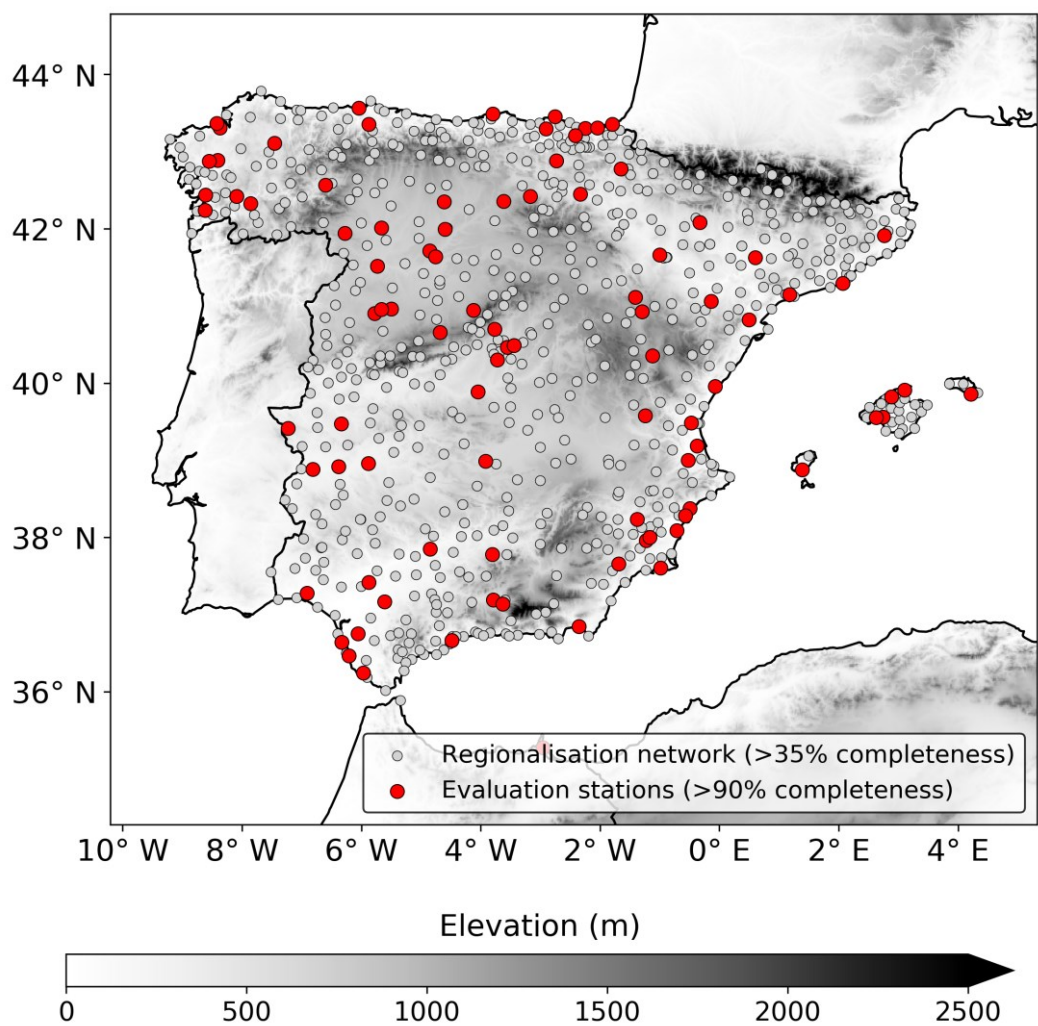


Figure S6: Spatial distribution of the stations used in the regionalised semi-parametric Q-CODA-R framework. The 91 evaluation stations (> 90% hourly data completeness over 1996-2024) were used exclusively for independent leave-one-station-out validation of Q-CODA-R, whereas the regionalisation network also includes additional stations with at least 35% completeness over the same period.

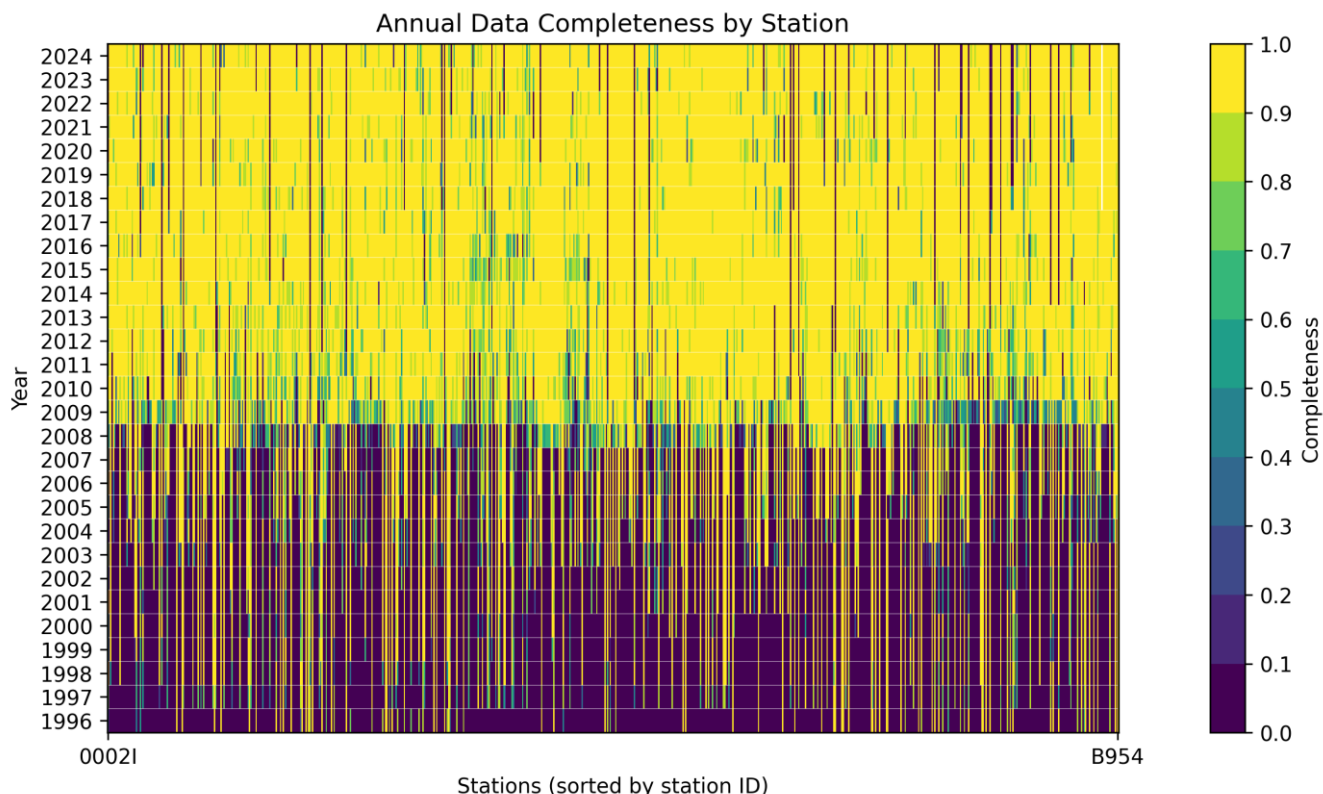


Figure S7: Annual hourly data completeness of the stations used in the Q-CODA-R regionalisation framework over 1996-2024. Each column represents a station and each row a year, with colours showing the proportion of available hourly precipitation data. Stations were included in the regionalisation network if their total completeness exceeded 35% over the full period (approximately ≥ 10 years of hourly observations). The stripe shows that the universal kriging of the seasonal Bernoulli–Gamma parameters was supported by a spatially broader but temporally less complete network than the high-quality evaluation subset, making the regionalisation experiment more demanding. It can also be observed that, from 2009 onwards, most of the network exhibits generally high hourly data completeness.

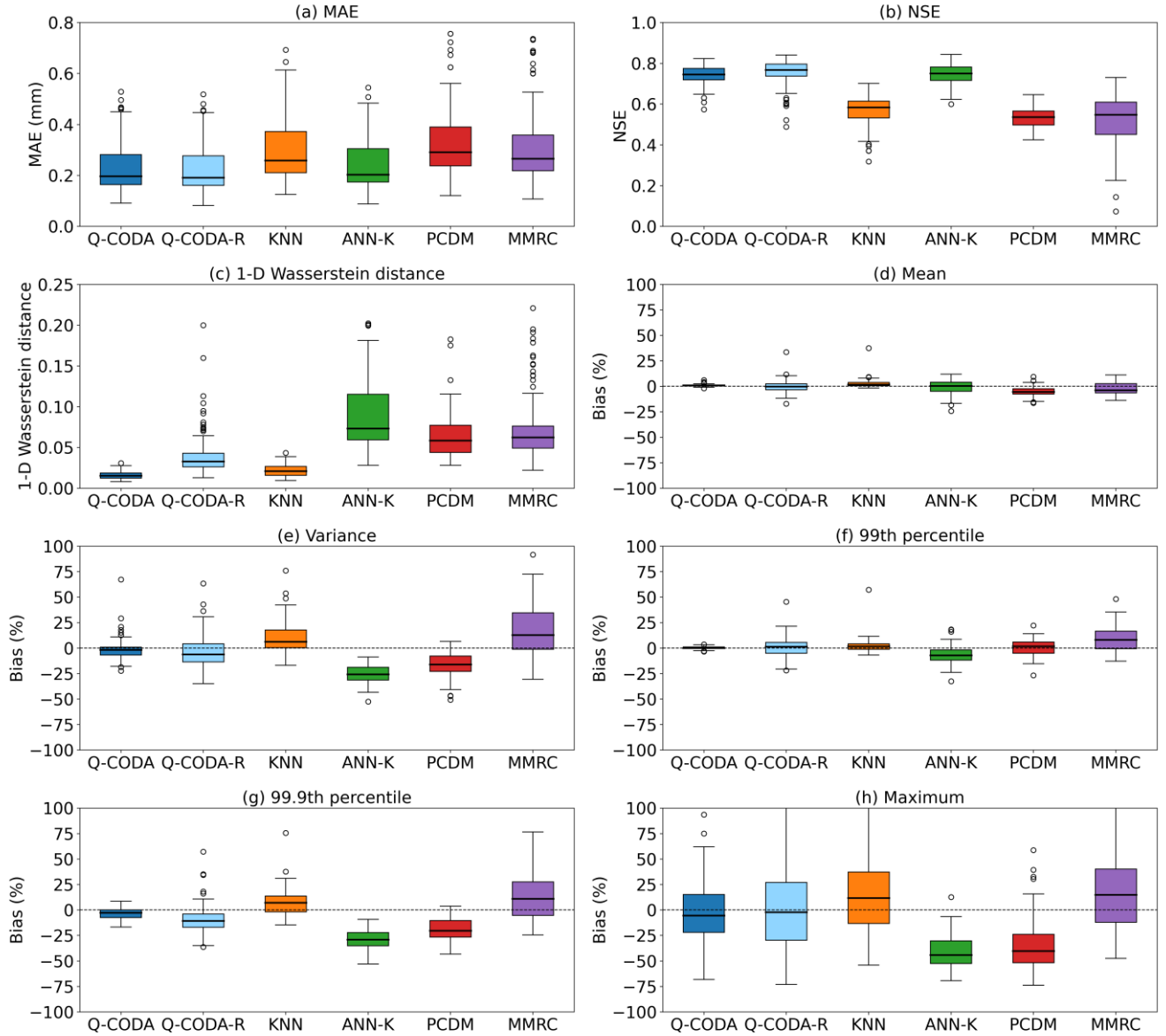


Figure S8: Comparative evaluation of disaggregation methods for reconstructing daily maximum 1-hour precipitation ($P_{1h}^{\max}$) across 91 meteorological stations. Each panel shows boxplots summarizing the spatial distribution of key performance metrics: (a) Mean Absolute Error (MAE), (b) Nash-Sutcliffe Efficiency (NSE), (c) 1-D Wasserstein distance, (d) bias in the mean, (e) bias in the variance, (f) bias in the 99th percentile, (g) bias in the 99.9th percentile, and (h) bias in the maximum. All methods except Q-CODA-R are evaluated using 5-fold cross-validation, whereas Q-CODA-R is evaluated in a leave-one-station-out (LOSO) framework, ensuring that no data from the target station are used in training Q-CODA-R.

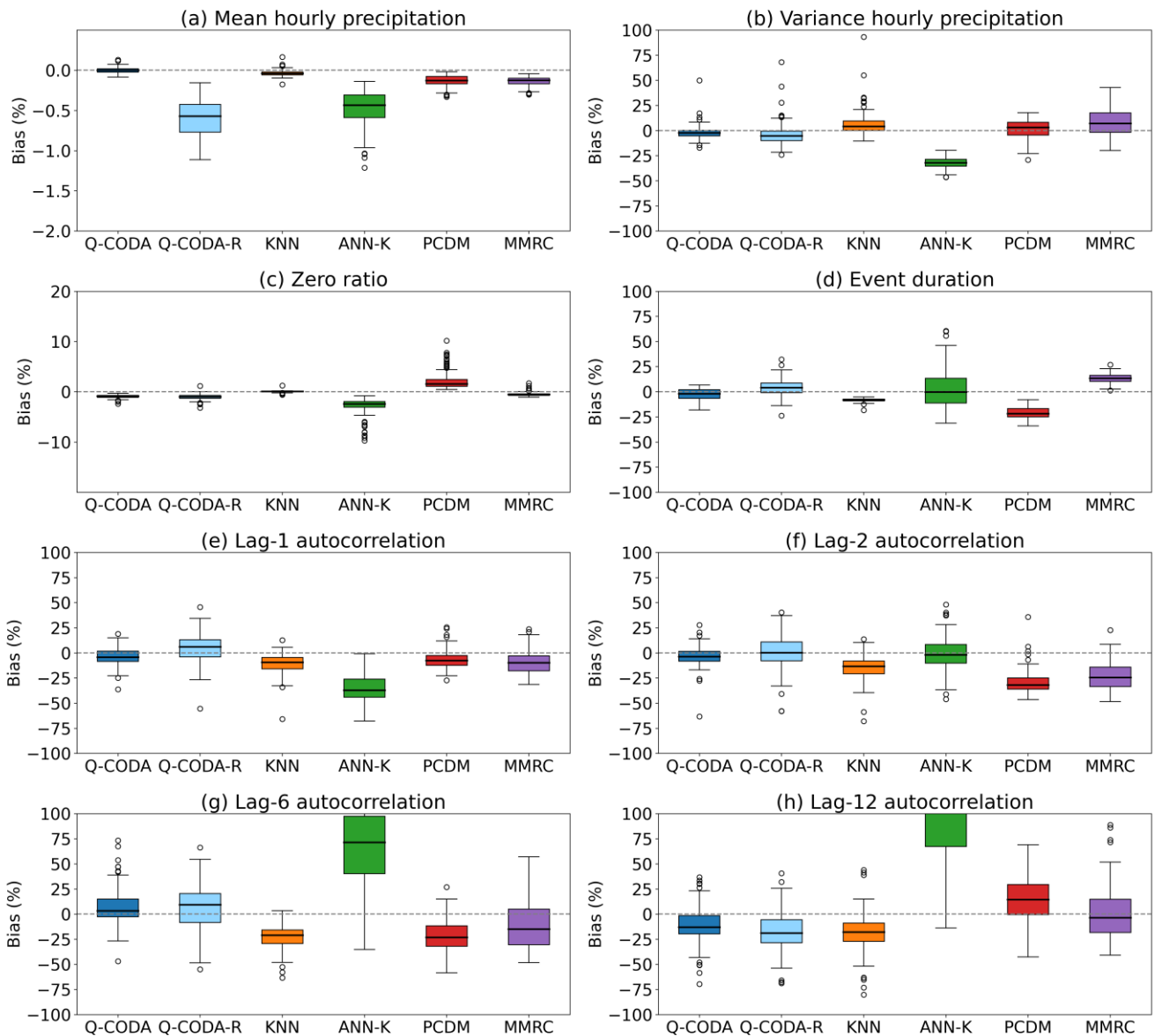


Figure S9: Comparative evaluation of disaggregation methods for reconstructing complete hourly precipitation time series across 91 meteorological stations. Each panel presents boxplots summarizing the spatial distribution of key performance metrics: (a) mean hourly precipitation, (b) hourly precipitation variance, (c) bias in zero proportion, (d) bias in mean precipitation event duration, (e) bias in lag-1 autocorrelation, (f) bias in lag-2 autocorrelation, (g) bias in lag-6 autocorrelation, and (h) bias in lag-12 autocorrelation. All methods except Q-CODA-R are evaluated using 5-fold cross-validation, whereas Q-CODA-R is evaluated in a leave-one-station-out (LOSO) framework, ensuring that no data from the target station are used in training Q-CODA-R.

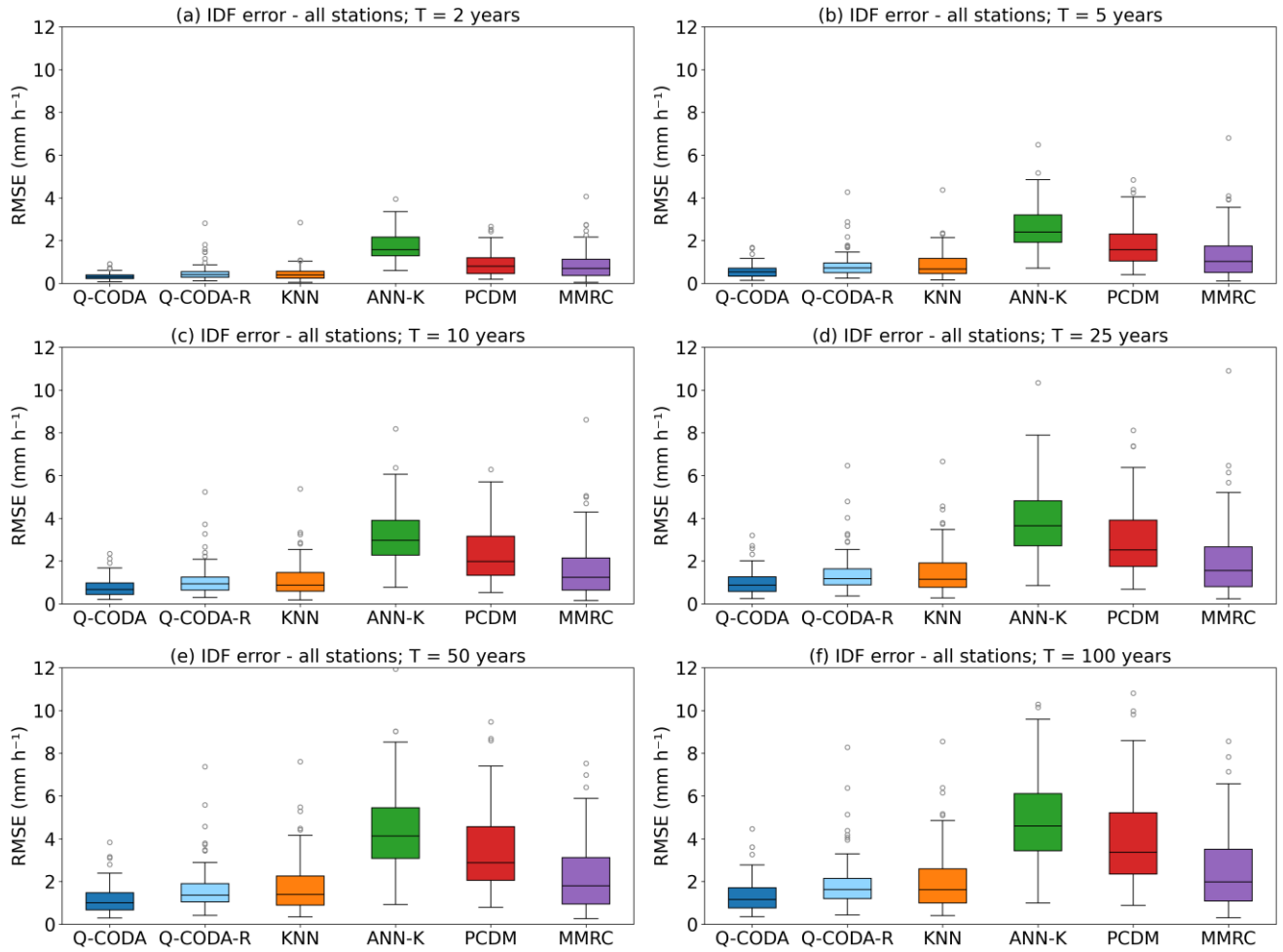


Figure S10: Spatial distribution of root mean squared error (RMSE) between observed and simulated IDF (Intensity-Duration-Frequency) curves across 91 meteorological stations, grouped by disaggregation method. All methods except Q-CODA-R are evaluated using 5-fold cross-validation, whereas Q-CODA-R is evaluated in a leave-one-station-out (LOSO) framework, ensuring that no data from the target station are used in training Q-CODA-R.