https://doi.org/10.5194/egusphere-2025-2710
Preprint. Discussion started: 26 June 2025

A Framework for the Assessment of Rainfall Disaggregation Methods in Representing Extreme Precipitation

Claudio Sandoval[1], Jorge Gironás[1,2,3,4], and Cristián Chadwick [5]

[1]Departamento de Ingeniería Hidráulica y Ambiental, Pontificia Universidad Católica de Chile, Santiago, Chile
[2]Centro de Desarrollo Urbano Sustentable (CEDEUS), Santiago, Chile
[3]Centro de Investigación para la Gestión Integrada del Riesgo de Desastres (CIGIDEN), Santiago, Chile
[4]Centro Interdisciplinario de Cambio Global, Pontificia Universidad Católica de Chile, Santiago, Chile
[5]Facultad de Ingeniería y Ciencias, Universidad Adolfo Ibáñez, Santiago, Chile

Correspondence: Jorge Gironás (jgironas@uc.cl)

Abstract. High-resolution precipitation data are essential to analyze extreme rainfall, critical for hydrological modeling, infrastructure design and climate change assessments. As high-resolution rainfall data are limited, disaggregation methods become an alternative to access such data. Although many studies have evaluated these methods, there is no framework for their selection based on their performance in reproducing extreme attributes. This paper presents a framework for evaluating daily-to-hourly rainfall disaggregation methods, measuring the performance on representing extreme precipitation behavior. The framework assesses this performance using Intensity-Duration-Frequency (IDF) curves and extreme rainfall indices (ERIs). IDF curve disaggregation performance evaluation uses accuracy and precision metrics (i.e., how close and consistent disaggregated values are to observed data, respectively), while ERIs are assessed by comparing the variability and bias of disaggregated annual series to observed data using a modified Kling-Gupta efficiency. The framework was applied to five sites with diverse climates, using three disaggregation methods: (1) a stochastic pulse-type method (SOC), (2) a non-parametric k-nearest neighbor (k-NN), and (3) a method based on Huff curves (HUFF). Results show that k-NN tends to outperform other methods in replicating IDF curves, modeling extreme rainfall percentiles and capturing the occurrence and magnitude of intense precipitation events, as well as most critical dry situations. SOC performs well in precision but has a lower ability in accuracy while HUFF is best at modeling 5-hour maximum rainfall. Nonetheless, these performances are not consistent across all locations, with the best-performing method varying per site, highlighting the importance of context-specific evaluations enabled by the framework.

1 Introduction

High-resolution precipitation data (e.g., hourly) are essential for a wide range of hydrological and engineering applications. These data are critical for analyzing extreme rainfall characteristics (e.g., magnitude, duration, frequency), used in the design and planning of hydraulic infrastructure (Bruni et al., 2015; Ochoa-Rodriguez et al., 2015). They support real-time control systems (Hingray et al., 2002), enable accurate runoff estimation in rapidly responding catchments (Bennett et al., 2016; Reynolds et al., 2017), including urban areas (Schilling, 1991), and allow for reliable assessment of rainfall-induced impacts on various systems (Li et al., 2020). Because extreme rainfall events often produce the highest overland flow and discharge

https://doi.org/10.5194/egusphere-2025-2710
Preprint. Discussion started: 26 June 2025

rates in hydrological models (Back et al., 2012), their misestimation can result in poor infrastructure design and potential system failure (Abreu et al., 2022). Moreover, fine-scale precipitation records are fundamental for compiling extreme rainfall and hydroclimatic indices used in climate trend analysis and for monitoring long-term changes affecting both human and ecological systems (Zhang et al., 2011).

Despite the importance of high-resolution precipitation records, their limited availability has been a persistent challenge, as sub-daily rainfall is generally available at only a small number of locations and often contains a large percentage of missing data (Woldemeskel, 2016). Thus, rainfall disaggregation methods become crucial, as they allow using coarser, comparatively more abundant records when dealing with finer scales. Numerous conceptual approaches and methods for fine-resolution rainfall disaggregation have been proposed (e.g., Hershenhorn and Woolhiser, 1987; Cowpertwait, 1991; Glasbey et al., 1995). These methods allow simulating a variety of feasible scenarios at fine resolutions, considering the complex and scalable nature of the rainfall process and the governing mechanisms (Pui et al., 2012). Some of the most popular methods are (1) the Bartlett–Lewis and Neyman–Scott rectangular pulse models, both grounded in point process theory (Rodriguez-Iturbe et al., 1987; Onof and Wheater, 1993; Khaliq and Cunnane, 1996), and (2) random cascade models, which draw upon scale-invariance theory (Gupta and Waymire, 1993; Cârsteanu and Foufoula-Georgiou, 1996; Molnar and Burlando, 2005).

Disaggregation methods are typically validated using common statistics, such as mean, standard deviation, lag 1-hour autocorrelation, and proportions of dry periods (e.g., Onof and Wheater, 1993; Socolofsky et al., 2001), with the aim of ensuring that the disaggregated series adequately reproduce the statistical properties of the observed data. Beyond the use of common statistics for general validation, a growing number of studies have also focused on assessing the ability of disaggregation methods to reproduce extreme precipitation characteristics (e.g., Bhattacharyya and Saha, 2023; Molnar and Burlando, 2005). These evaluations typically focus on the methods capacity to replicate observed indicators such as annual maximum precipitation (AMP), Intensity-Duration-Frequency (IDF) curves, dry spell metrics, and extreme percentiles of rainfall intensity.

Overall, the literature describing assessment approaches of disaggregation methods might be classified into three types. A first type of studies propose or test specific methods with an emphasis on extremes. For example, Bhattacharyya and Saha (2023) assessed the capacity to replicate dry hours and reproduce IDF (Intensity–Duration–Frequency) curves, across various return periods and durations, of a daily-to-hourly disaggregation method combining artificial neural networks with k-means clustering. Thober et al. (2014) used indicators such as the total precipitation above the 95th percentile (r95p) and the maximum number of consecutive dry days (CDD) to capture the ability of representing both extreme wet and dry conditions of a monthly-to-daily generator based on multiplicative cascades. Molnar and Burlando (2005) examined dry timestep reproduction and the preservation of aggregated mean precipitation across different durations and return periods when assessing a canonical and microcanonical multiplicative cascades to downscale data from 1280 to 10 min resolution. A second type of studies compare existing disaggregation methods under a unified evaluation scheme that includes extreme precipitation analysis. For instance, Poschlod et al. (2018) used metrics such as dry-hour proportions, wet spell characteristics per rainy day, and AMP across multiple durations and return periods to assess the method of fragments and outputs from the Weather Research and Forecasting regional climate model. Pui et al. (2012) used both general statistics and IDF-derived extremes to assess several methods, including canonical and microcanonical random multiplicative cascades (Gupta and Waymire, 1993; Molnar and Burlando,

https://doi.org/10.5194/egusphere-2025-2710
Preprint. Discussion started: 26 June 2025

2005), the Randomized Bartlett-Lewis Model (RBLM) (Koutsoyiannis and Onof, 2001), and a fragment-based resampling approach (Sharma and Srikanthan, 2006). Similarly, Hingray and Ben Haha (2005) used AMP quantiles at 10 min resolution and across different return periods to assess seven hourly-to-sub-hourly disaggregation methods. Finally, a third type of study presents integrative disaggregation frameworks. Qin and Dai (2024) used quantile-based frequency curves across multiple return periods and durations to assess a framework which integrates six point-process models and four disaggregation schemes over flexible timescales.

To the best of our knowledge, there is no clear agreement to guide the selection of a rainfall disaggregation method based on its ability to effectively replicate extreme precipitation. Overall, the previous studies tend to focus on specific locations, timescales, or metrics in isolation. Most commonly, evaluations are centered around the agreement between disaggregated and observed AMP for different durations and return periods. However, other critical aspects, such as the considerable uncertainty inherent in extreme rainfall estimation, which is closely tied to the disaggregation process (Qin and Dai, 2024), are often overlooked. The objective of this paper is to present and assess a framework to evaluate the performance of daily-to-hourly rainfall disaggregation methods based on their ability to represent observed extreme precipitation. This evaluation is based on a holistic comprehension of extreme rainfall and encompasses two key aspects: the capacity of the methods to accurately replicate observed IDF curves across various return periods and their effectiveness in reproducing some critical extreme rainfall indices (ERIs). The framework is tested for five locations with different climates and three rainfall disaggregation methods: (1) a stochastic pulse-type method based on Socolofsky et al. (2001) work (SOC), (2) a non-parametric resampling method based on the k-NN algorithm and the method of fragments (k-NN) (Lall and Sharma, 1996; Sharma and Srikanthan, 2006), and (3) a nondimensional storm pattern-based approach using Huff curves (HUFF) (Huff, 1967). The structure of the paper is as follows: Section 2 describes the proposed framework and its components. Section 3 presents the three disaggregation methods, their structures, functioning, and essential procedures, while Section 4 describes the different locations and available data; Section 5 presents the results and its discussions, and Section 6 is dedicated to the main conclusions.

2 Disaggregation assessment framework

The disaggregation assessment framework analyzes the performance of disaggregation methods in terms of their capability of replicating AMP and their uncertainty, as well as their capability of reproducing ERIs and their uncertainty. The proposed framework considers three components (Figure 1), which are explained in detail in the following subsections: (1) the Rainfall Disaggregation component (Figure 1a, green part) in which daily precipitation series aggregated from observed hourly series are disaggregated into hourly series using the disaggregation methods under assessment; (2) the IDF curves-based Assessment component (Figure 1a, red part) oriented to evaluate the accuracy and precision of the disaggregation method (i.e., how close and consistent disaggregated IDF's curves are to observed IDF's curves, respectively), using a constructed reference IDF curve; and (3), the Extreme Rainfall Indices Assessment component (Figure 1a, blue part) to evaluate the capability of the disaggregated series (i.e., hourly disaggregated data) in replicating four relevant ERIs annual series of observed records.

https://doi.org/10.5194/egusphere-2025-2710
Preprint. Discussion started: 26 June 2025

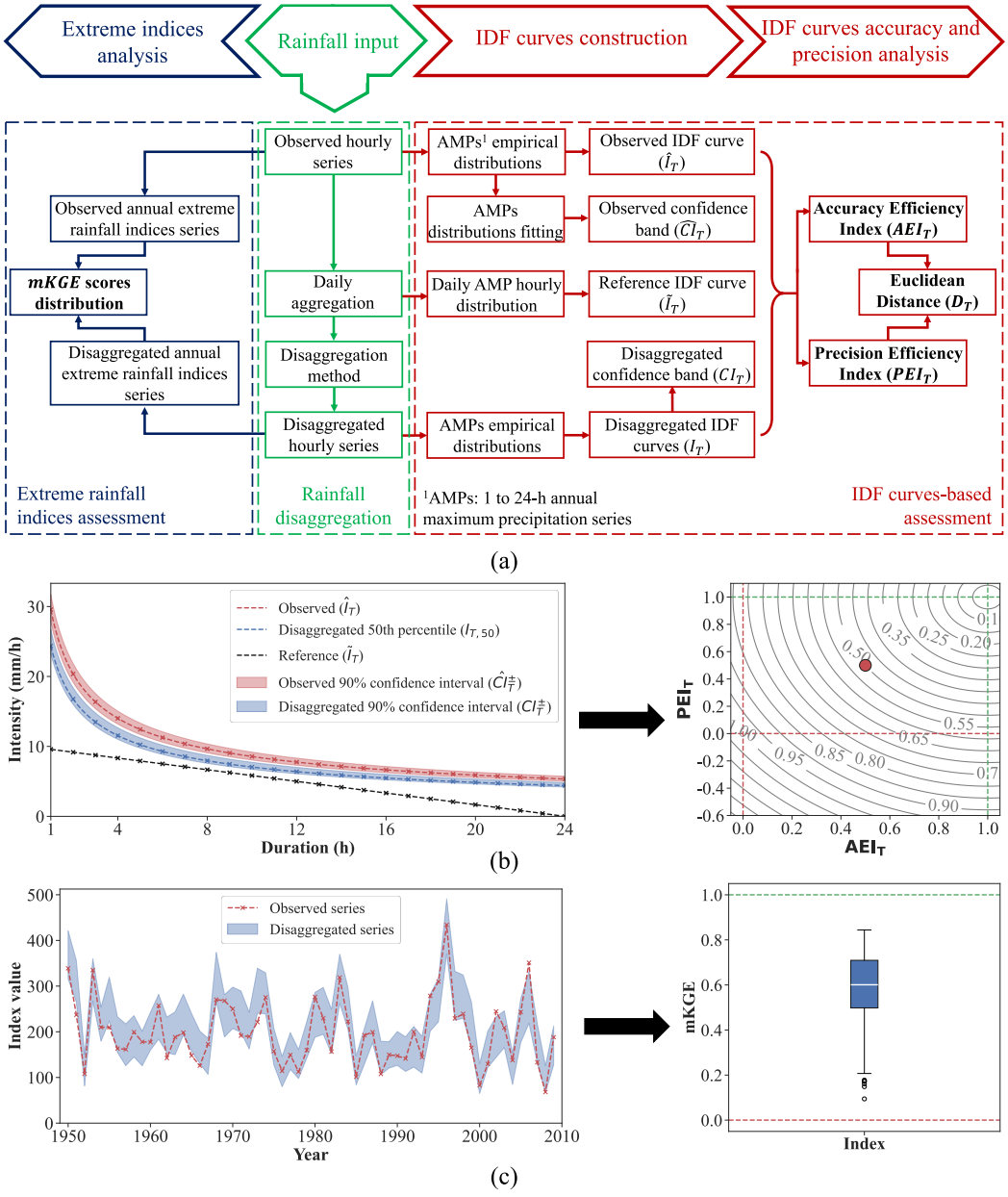

Figure 1. Rainfall disaggregation methods assessment framework. (a) The framework. (b) Example of T-years IDF curves assessment: observed (red), disaggregated (blue), and reference (black). These are used to construct the AEI_T (Accuracy Efficiency Index), PEI_T (Precision Efficiency Index), and D_T (Euclidean Distance, shown as concentric ellipses) plot. (c) Example of ERIs assessment: $mKGE$ values are computed between observed and disaggregated annual series for each ERI; results are shown as boxplots.

https://doi.org/10.5194/egusphere-2025-2710
Preprint. Discussion started: 26 June 2025

2.1 Rainfall disaggregation

The rainfall disaggregation process starts by aggregating the sub-daily precipitation into the daily scale. The second step is to implement the disaggregation methods under assessment to obtain, for each one, many possible disaggregated time series from the daily precipitation, so that the disaggregation uncertainty can be captured.

2.2 IDF curves-based assessment

2.2.1 IDF curves construction

At this stage, the observed and disaggregated IDF curves, along with their corresponding uncertainty bands, are constructed using the multiple disaggregated rainfall series generated with the disaggregation method. These curves are calculated after the empirical frequency analysis using Weibull plotting position (Weibull, 1939) of the average intensities of the AMP for durations d = 1, 2, …, 24 h. The uncertainty of the disaggregated IDF curves is associated with the disaggregation process (i.e., the stochastic uncertainty), while the uncertainty of observed IDF curves is statistical (i.e. uncertainty associated with using a finite sample). For each return period, the disaggregated uncertainty band is calculated as the 90% confidence interval derived from the empirical distribution from all disaggregated IDF curves. Conversely, the observed uncertainty band requires assuming a probability distribution for the annual maximum precipitation (AMP) for durations of 1, 2, …, up to 24 h. For this purpose, the Gumbel distribution (Gumbel, 1935) is fitted using the maximum likelihood method. For each duration, multiple samples of d-hours AMP must be generated from the respective probability distribution. These samples enable the calculation of multiple intensity values for a given return period. Applying this procedure to all the durations, a 90% confidence interval can be constructed to represent the statistical uncertainty of the observed IDF curve. Eventually, an empirical bootstrapping of the historical observed years could also be used, instead of a theoretical distribution, when the number of years is big enough.

For the assessment of the disaggregated IDF curves obtained from the different methods, a reference deterministic IDF curve is generated from the observed daily series. If the disaggregated IDF curve performs worse than this reference curve, the method should be discarded. The reference curve is constructed as follows. Each value in the observed daily AMP (AMP_{24}) series is distributed across 24 hours using a linearly decreasing intensity profile, going from $I(d=0) = \frac{AMP_{24}}{12}$ to $I(d=24) = 0$:

$$I(d) = \frac{AMP_{24}}{12}\left(1 - \frac{d}{24}\right) \tag{1}$$

Intensities are evaluated at d = 0.5, 1.5, …, 23.5, yielding values for durations d = 1, 2, …, 24 h. For each duration, a rolling window of length d is applied to identify the most intense d-hour pulse within the day. This is repeated for all daily AMP values, generating one sample per duration. An empirical frequency analysis is then applied to each sample to estimate the T-year AMP for each duration, which is converted to intensity by dividing by d. The reference IDF curve is obtained by linking these intensities across all durations.

https://doi.org/10.5194/egusphere-2025-2710
Preprint. Discussion started: 26 June 2025

2.2.2 IDF curves accuracy-precision analysis

The final step of this stage involves evaluating the performance of the disaggregated IDF curves for each return period T using three metrics: Accuracy Efficiency Index (AEI_T), Precision Efficiency Index (PEI_T) and Euclidean Distance (D_T). The AEI_T is proposed to quantify the accuracy of the disaggregated IDF curves compared to the observed IDF curve. This metric is inspired by the Modified Coefficient of Efficiency (Legates and McCabe, 1999), a modification of the Nash-Sutcliffe efficiency coefficient (Nash and Sutcliffe, 1970) in which the absolute differences rather than squared differences are calculated to avoid giving more weight to larger differences. The AEI_T is calculated as:

$$AEI_T = 1 - \frac{\sum_{d=1}^{24} \frac{1}{d} |\widehat{I}_d^T - I_{d,50}^T|}{\sum_{d=1}^{24} \frac{1}{d} |\widehat{I}_d^T - \widetilde{I}_d^T|} \tag{2}$$

where $\widehat{I}_d^T$ represents the intensity of duration d from the observed T-year IDF curve, $I_{d,50}^T$ is the intensity of duration d from the disaggregated IDF curve at the 50$^{\text{th}}$ percentile for T years, and $\widetilde{I}_d^T$ is the intensity of duration d from the reference T-year IDF curve. Equation (2) assesses how well the disaggregated 50$^{\text{th}}$ percentile IDF curve represents the observed IDF curve compared to using the reference IDF curve. The weighting by the inverse of the duration (d) is used to better capture higher intensities, as they tend to cause larger floods in quick response basins (Viglione and Blöschl, 2009).

AEI_T ranges from $-\infty$ to 1, with a value of 1 indicating that the disaggregated 50$^{\text{th}}$ percentile curve exactly matches the observed IDF curve values. A value of 0 implies that the disaggregated curve performs as well as the reference IDF curve; lower values indicate that the IDF curve from the disaggregation is worse than the reference.

The PEI_T quantifies if the dispersion of the disaggregated IDF curves is comparable to the statistical uncertainty of the observed IDF curve. PEI_T is calculated as:

$$PEI_T = 1 - \frac{\sum_{d=1}^{24} \frac{1}{d} \left(\frac{CI_{T,d}^{+} - CI_{T,d}^{-}}{I_{d,50}^T} \right)}{\sum_{d=1}^{24} \frac{1}{d} \left(\frac{\widehat{CI}_{T,d}^{+} - \widehat{CI}_{T,d}^{-}}{\widehat{I}_d^T} \right)} \tag{3}$$

where $\widehat{CI}_{T,d}^{\pm}$ represents the upper (+) and lower (-) confidence interval bounds (i.e., 95$^{\text{th}}$ and 5$^{\text{th}}$ percentiles, respectively, of statistical sampling uncertainty) of the observed T-year IDF curve for a duration d. Similarly, $CI_{T,d}^{\pm}$ denotes the upper and lower confidence interval bounds (i.e., 95$^{\text{th}}$ and 5$^{\text{th}}$ percentiles, respectively, of the stochastic disaggregation uncertainty) of the intensities for the same return period T and duration d, but derived from the disaggregated IDF curves. As in the AEI_T calculation, PEI_T incorporates a weighting factor based on the inverse of the duration (d) to emphasize high intensity precipitation.

PEI_T ranges from $-\infty$ to 1. A value of 1 indicates no uncertainty attributable to the disaggregation process, while a value of 0 implies that the disaggregation uncertainty is numerically equivalent to sample statistical uncertainty. Negative values indicate that the uncertainty associated with the disaggregation process exceeds the natural variability, which is undesirable. Therefore, PEI_T values between 0 and 1 are preferred, recognizing that 1 is an asymptotic limit since the stochastic nature of the disaggregation process inherently carries some level of uncertainty.

https://doi.org/10.5194/egusphere-2025-2710
Preprint. Discussion started: 26 June 2025

Table 1. ERIs used in this study, adapted from Zhang et al. (2011).

Index	Description	Units
TDD	Hours with no rain ($P < 0.1$ mm)	h
RX5	Maximum precipitation in 5 consecutive hours	mm
R95%	Number of hours in which $P > P_{95}$ of the observed record	h
P>R95%	Summation of P above P_{95} of the observed record	mm

Finally, a metric that integrates both AEI_T and PEI_T, referred to as D_T, is proposed. It represents the Euclidean distance, for a given IDF curve of T years return period, between the point (AEI_T, PEI_T) and the optimal point (1,1), which symbolizes ideal performance in terms of accuracy and precision.

$$D_T = \sqrt{w_{AEI}\,(AEI_T - 1)^2 + w_{PEI}\,(PEI_T - 1)^2} \tag{4}$$

A value of $D_T = 0$ indicates perfect performance in the disaggregation process, meaning that both AEI_T and PEI_T have reached the maximum possible value of one. Note that the weighting factors w_{AEI} and w_{PEI} allow assigning greater importance to one metric over the other when calculating the distance to the optimal point. Figure 1b provides a generic example for a specific return period. In right side of Figure 1b, the plot synthesizes the three performance indices in a single figure, with AEI_T in the horizontal axis, PEI_T in the vertical axis, and D_T as the distance to the optimal (1,1) point, measured as concentrical ellipses (i.e., the elliptical shape appears from different weighting factors in D_T).

2.3 Extreme rainfall indices assessment

2.3.1 Extreme rainfall indices (ERIs)

Table 1 shows a variety of ERIs previously proposed by Zhang et al. (2011) and widely used in previous studies (e.g., Klein Tank and Können, 2003; Taschetto and England, 2009; Kioutsioukis et al., 2009; Vicuña et al., 2013). Although developed to be used with daily precipitation records, we adapted them to hourly series for the purpose of the present study. The calculation methodology is analogous for hourly data: it estimates an ERI value per year to form an annual time series for each ERI.

The selection of these ERIs is grounded in their ability to represent different characteristics of extreme rainfall behavior across years. For the assessment approach we consider (1) Total Dry Duration (TDD) as a rainfall intermittency index that captures the most critical dry conditions, (2) Maximum precipitation in 5 consecutive hours (RX5) as an index that capture high precipitation pulses for a duration long-enough to impact flow rates in a wide range of basins, (3) Number of hours in which precipitation exceeds the observed 95$^{\text{th}}$ percentile P_{95} (R95%), and (4) Summation of precipitation above percentile P_{95} record (P>R95%). The last two indices assess duration and magnitudes respectively, associated with extreme precipitation pulses. While these specific ERIs are proposed in this study, other indices could be used to perform equivalent analyses.

https://doi.org/10.5194/egusphere-2025-2710
Preprint. Discussion started: 26 June 2025

2.3.2 Modified Kling-Gupta efficiency (mKGE)

For the ERIs assessment, we propose the following modified Kling-Gupta efficiency coefficient ($mKGE$) that is based on the original KGE metric proposed by Gupta et al. (2009):

$$mKGE = 1 - \sqrt{\left(\frac{\sigma_{sim}}{\sigma_{obs}} - 1\right)^2 + \left(\frac{\mu_{sim}}{\mu_{obs}} - 1\right)^2} \quad (5)$$

where σ_{obs} and σ_{sim} are the standard deviation of the observed and disaggregated ERIs series, while μ_{obs} and μ_{sim} are the mean of these series, respectively. Unlike the original KGE, the $mKGE$ excludes the correlation component r between the disaggregated and observed series (Gupta et al., 2009) because the priority is to represent the statistical behavior (mean and variability) of extreme values, and not the temporal alignment of these extremes (a task not expected to be accomplished by a disaggregation method). As in the original formulation, $mKGE$ ranges from $-\infty$ to 1, with 1 being the optimal value (i.e. standard deviations and means of the disaggregated and observed series are equal). To establish a target minimum performance threshold, the mean of the observed hydrological time series is used as a benchmark, as suggested by Knoben et al. (2019). In this case, for each ERI, $\mu_{sim} = \mu_{obs}$ and $\sigma_{sim} = 0$, and the reference $mKGE$ value ($mKGE_{ref}$) (i.e. target minimum performance threshold) becomes:

$$mKGE_{ref} = 1 - \sqrt{\left(\frac{0}{\sigma_{obs}} - 1\right)^2 + \left(\frac{\mu_{obs}}{\mu_{obs}} - 1\right)^2} = 0 \quad (6)$$

The $mKGE$ value is calculated for each ERI time series, comparing the observed and disaggregated series. By considering all possible disaggregation series from each method, given the stochastic nature of disaggregation, there is a dispersion in the calculated $mKGE$ values, which should be as narrow as possible. Thus, high $mKGE$ values and reduced dispersion are sought. Figure 1c provides an example of the calculation of $mKGE$, where the dashed green line represents the optimal $mKGE$ value, while the red one indicates $mKGE_{ref}$.

3 Disaggregation methods

3.1 Stochastic pulse-type method (SOC)

This pulse-type rainfall disaggregation technique, developed by Socolofsky et al. (2001) and further applied in a case study in Texas (Choi et al., 2008), disaggregates daily rainfall depths into individual sub-daily events using rectangular pulses whose intensity, duration, and start time is randomly assigned. The stochastic parameters used in the generation of each rainfall pulse are derived from observed hourly rainfall records at a monthly basis. Furthermore, the method uses a monthly minimum event threshold ε to prevent underestimating the no rain probability (Choi et al., 2008). For each month, ε is calibrated to represent well the variance, lag 1-h autocorrelation coefficient and probability of no rain of the observed series. What follows is a summary of the method; readers are referred to the original work (Socolofsky et al., 2001) for more details:

Step 1: Fitting the cumulative distribution function (CDF) to the rainfall depths of events occurring each month. The rainfall record is first divided into discrete events using a minimum dry time between events. In this study, a time of 6 h is used

https://doi.org/10.5194/egusphere-2025-2710
Preprint. Discussion started: 26 June 2025

to ensure independence between events, according to Huff (1967). Then, for each month a CDF is fitted to the event's depths occurring in that month. While Socolofsky et al. (2001) used the empirical CDF, this study fits both Gamma and Exponential distributions to allow greater flexibility in generating random rainfall depths. The best distribution is selected based on the Kolmogorov-Smirnov (KS) test.

Step 2: Defining the disaggregated depth magnitude from daily precipitation. The daily precipitation amount to be disaggregated, PP_{t1}, is passed through the CDF of its respective month to obtain its corresponding non-exceedance probability a_1. Then a random uniform number $u_1 \sim U(0, a_1)$ is used to obtain the first precipitation depth pulse (PP_1) from the CDF.

Step 3: Updating the remaining daily precipitation and checking for convergence. Once the pulse PP_1 is known, the remaining daily precipitation to be disaggregated is $PP_{t2} = PP_{t1} - PP_1$. If $PP_{t2} > \varepsilon$, then PP_{t2} is passed through the respective CDF to obtain the non-exceedance probability a_2. Similar to the previous step, a new precipitation depth PP_2 from the CDF is obtained after generating a new uniform number $u_2 \sim U(0, a_2)$. The process is repeated iteratively until the condition $PP_{ti} < \varepsilon$ is true and the disaggregation process is finished. In our implementation of the method, the depth of the final pulse will directly correspond to the remaining depth PP_{ti} to ensure mass balance within the disaggregated day.

Step 4: Setting duration and start time of the generated sub-daily pulses. The duration d_{pulse} and starting time must be assigned for each pulse. If the pulse is the last one (i.e. if $PP_t \leq \varepsilon$), then d_{pulse}= 1 h. Else, d_{pulse} is selected directly from the event database of the month, using the duration of the event with the most similar depth. If two events have the same depth, the duration of the first event in chronological order is chosen (Socolofsky et al., 2001). For this study, since the selection of PP_i follows a probability distribution, the duration is determined using a k-NN algorithm (described in detail in Section 3.2), by selecting the duration of a randomly selected, analogous event in terms of depth. For the start time, a uniform random number $U(0, 24 - d_{pulse})$ is generated, ensuring that the pulses are placed within the disaggregated day without overlapping at midnight. Finally, place each pulse in its respective block within the day, considering a constant rainfall intensity $I = \frac{PP_i}{d_{pulse}}$ throughout the duration. If two events overlap, their intensities are added.

Step 5: Applying the disaggregation process to all target days. Repeat steps 1 to 4 independently for each day t to be disaggregated until completing the procedure for all daily rainfall values in the dataset.

3.2 *k*-nearest neighbors method (k-NN)

This non-parametric method follows a resampling logic, utilizing a "fragments" vector that represents the proportions of hourly precipitation derived from the original daily rainfall amount (Pui et al., 2012). This vector is selected from a historical day analogous to the day being disaggregated, for which the hourly distribution is well known, allowing disaggregated series to have similar attributes to those observed by maintaining temporal dependence at a daily scale, and then using a non-parametric approach to assign dependence at hourly timescales (Sharma and Srikanthan, 2006). The analogous day is identified using the *k*-NN algorithm (Lall and Sharma, 1996). In this approach, a pool of historical rainy days is considered as candidates. Each candidate day is compared to the target day by calculating the Euclidean distance between their total daily rainfall values. The *k*-nearest neighbors are the k days with the smallest distances, representing the most similar rainfall conditions. Finally, one of these neighbors is randomly selected, giving a higher probability to the most similar neighbor, introducing stochasticity into

https://doi.org/10.5194/egusphere-2025-2710
Preprint. Discussion started: 26 June 2025

the process. The "fragments" vector from the chosen day is then applied to the target day to distribute its daily precipitation into hourly values. The implementation follows some well-defined steps:

Step 1: Constructing reference daily rainfall and sub-daily fragment time series. The method relies on two datasets: (i) the daily rainfall to be disaggregated, and (ii) a reference dataset containing days for which both daily rainfall totals and their corresponding sub-daily (hourly) time series are available. The reference daily rainfall PP_r and the associated dimensionless sub-daily fragments $f_{r,m}$ are given by:

$$PP_r = \sum_{m=1}^{24} X_{r,m} \tag{7}$$

$$f_{r,m} = \frac{X_{r,m}}{PP_r} \tag{8}$$

where, $X_{r,m}$ refers to the hourly rainfall at hour m on reference day r. The reference datasets serve as the pool of analog days used in the disaggregation procedure.

Step 2: Forming the analog days vector ($\boldsymbol{R}$) to disaggregate daily rainfall ($\boldsymbol{PP_t}$). To disaggregate a given daily rainfall value PP_t, potential analog days PP_r are looked in each one of the N years from the reference dataset, restricted to a window of length l centered on day t to ensure capturing precipitation seasonality. All the PP_r days of the restricted seasonality window are concatenated in a single vector $\boldsymbol{R}$ of potential analogs:

$$\boldsymbol{R} = [R_1, R_2, \ldots, R_j, \ldots R_{l \cdot N}] \tag{9}$$

It is important to note that no rainfall classification scheme (such as conditioning on the wet or dry status of the days before and after) was applied in this process, as proposed by Sharma and Srikanthan (2006). This modification gives the method greater flexibility in selecting analog days, which is particularly useful for stations with sparse or irregular rainfall patterns.

Step 3: Estimating Euclidean distance of precipitation. Compute the absolute differences $|PP_t - R_j|$ between the daily rainfall amount to be disaggregated PP_t and each daily rainfall amount R_j in the potential analog days vector $\boldsymbol{R}$. These differences are then sorted in ascending order to identify the analogs and form the vector $\widetilde{\boldsymbol{R}}$ with the k values from $\boldsymbol{R}$ most similar to PP_t. Here, $k = \sqrt{n}$, as suggested heuristically by Lall and Sharma (1996), where $n = l \cdot N$ is the sample size or length of the potential analog vector $\boldsymbol{R}$.

Step 4: Randomly selecting an analog day. Randomly select a j-th neighbor from the analog days vector $\widetilde{\boldsymbol{R}}$ using the conditional probability of selecting neighbor j, p_j (Lall and Sharma, 1996; Mehrotra and Sharma, 2006):

$$p_j = \frac{\frac{1}{j}}{\sum_{i=1}^{k} \frac{1}{i}} \tag{10}$$

The corresponding neighbor $\widetilde{R}_j$ is selected from $\widetilde{\boldsymbol{R}}$ after randomly generating a uniformly distributed number $U(0,1)$

Step 5: Disaggregating $\boldsymbol{PP_t}$ using the hourly fragments of the selected analog day $\boldsymbol{j}$. Once the analog day j has been selected from vector $\widetilde{\boldsymbol{R}}$, the daily rainfall PP_t is disaggregated using the hourly rainfall fragments $f_{j,m}$ associated with the

https://doi.org/10.5194/egusphere-2025-2710
Preprint. Discussion started: 26 June 2025

selected analog day and defined according to Equation (10). The hourly precipitation values for day t are then computed as:

$$X_{t,m} = f_{j,m} \cdot PP_t \tag{11}$$

This procedure ensures the preservation of both the mass balance and the temporal distribution of the selected analog from the original daily value PP_t.

Step 6: Applying the disaggregation procedure to all target days. Steps 2 to 5 are repeated independently for each daily rainfall value PP_t until all target days in the dataset have been disaggregated.

In this study, the window size l is calibrated based on the approach proposed by Alam and Elshorbagy (2015), which selects an optimal half-window length to effectively capture the memory of the hydrological regime. The optimal half-window size is determined by minimizing the inverse duration-weighted average root mean square error $RMSE_{wei}$, calculated between the observed ($X_{i,d}$) and disaggregated ($\widehat{X}_{i,d}$) normalized annual maximum precipitation $AMP_{i,d}$ (and $\widehat{AMP}_{i,d}$ for various durations d ranging from 1 to 24 h, of year i. Equations (14), (15), (16) and (17) outline the computation of observed ($X_{i,d}$) and disaggregated ($\widehat{X}_{i,d}$) $RMSE_d$ and $RMSE_{wei}$, respectively, with N being the total number of years in the records. As in the accuracy-precision metrics (AEI_T and PEI_T), the purpose of using inverse duration as weighting factors is to better capture the AMP for shorter duration, which are related to larger floods (Viglione and Blöschl, 2009).

$$X_{i,d} = \frac{AMP_{i,d} - \min(AMP_{i,d})}{\max(AMP_{i,d}) - \min(AMP_{i,d})} \tag{12}$$

$$\widehat{X}_{i,d} = \frac{\widehat{AMP}_{i,d} - \min(\widehat{AMP}_{i,d})}{\max(\widehat{AMP}_{i,d}) - \min(\widehat{AMP}_{i,d})} \tag{13}$$

$$RMSE_d = \sqrt{\sum_{i=1}^{N} \frac{\left(X_{i,d} - \widehat{X}_{i,d}\right)^2}{N}} \tag{14}$$

$$RMSE_{wei} = \frac{\sum_{i=1}^{24} \frac{1}{d} \cdot RMSE_d}{\sum_{i=1}^{24} \frac{1}{d}} \tag{15}$$

3.3 Nondimensional storm pattern-based method (HUFF)

As a third approach, a simple daily-to-hourly rainfall disaggregation process proposed by Huff (1967) is used. The method defines the so-called Huff curves, which represent in a dimensionless form the relationship between elapsed time and accumulated precipitation over the duration of historical precipitation events at a specific location (i.e. both variables are scaled within a range of 0 to 100%). For the construction of the Huff curves, precipitation events are classified according to four quartiles, each representing the part of the event where most of the rainfall occurs. Within each quartile, multiple curves are identified, which show the temporal distribution of the accumulated dimensionless event depth for different exceedance probabilities. The implementation as a disaggregation method for a given year is described as follows.

https://doi.org/10.5194/egusphere-2025-2710
Preprint. Discussion started: 26 June 2025

Step 1: Constructing a rainfall event-based record. Using the observed hourly time series, a record of storm events is created. An event will be distinguished from another by having six or more dry hours between them (Huff, 1967).

Step 2: Assigning each event by quartile and constructing the curves. Events are classified into quartiles according to the storm segment contributing the largest portion of rainfall. Both elapsed time and cumulative precipitation are then scaled from 0 to 100%. The dimensionless profiles are discretized in 1% increments.

Step 3: Performing a frequency analysis for each quartile and time interval to construct the Huff curves. For each quartile and time interval, the empirical distribution is used to estimate the cumulative dimensionless precipitation for different exceedance probabilities. Nine curves are obtained associated with exceedance probabilities of 10, 20, ..., up to 90%. Each curve connects time-accumulated volume points with the same exceedance probability.

Step 4: Choosing a Huff curve and disaggregating daily precipitation (PP_t). For each quartile q, the empirical probability p_q of selecting that quartile is calculated as:

$$p_q = \frac{n_q}{N_e} \tag{16}$$

where n_q is the number of events in quartile q, and N_e is the total number of recorded events. These probabilities are arranged sequentially along the unit interval [0,1]. To disaggregate a daily precipitation PP_t, a random uniform number $U(0,1)$ is first drawn and used to select the corresponding quartile. Then, another random number $U(0.05, 0.95)$, discretized to the nearest 0.1, is used to choose the corresponding Huff curve. The duration d and the start time for PP_t are determined following the same approach as in the SOC method: d is selected using a k-NN algorithm that identifies an analogous event based on precipitation depth, and the start time is assigned by drawing a random number $U(0, 24-d)$. Finally, the selected Huff curve is used to determine the hourly distribution of PP_t over its duration d.

Step 5: Applying the disaggregation procedure to all target days. Repeat Step 4 for each daily precipitation value PP_t to be disaggregated, until all days in the dataset have been processed.

4 Case study

4.1 Application sites

The proposed framework was evaluated at five different locations to assess its performance under different precipitation patterns. Hourly rainfall data for stations Altheim (Germany), Hikone (Japan), Pathhead (United Kingdom, UK), and Oregon (United States, US), were obtained from the highly reliable Global Sub-Daily Rainfall Dataset, part of the INTENSE project (Lewis et al., 2019, 2021). Additionally, the Quinta Normal station of the Chilean Meteorological Directorate was also used, providing 10-minute precipitation data that were subsequently aggregated to an hourly scale. The datasets span at least 10 years and have less than 0.8% missing data. Table 2 summarizes the main characteristics of the available records, while Figure 2a shows the location of each station and the precipitation's seasonal variability, and Figure 2b presents the time series of annual precipitation in each site.

https://doi.org/10.5194/egusphere-2025-2710
Preprint. Discussion started: 26 June 2025

Table 2. Selected stations.

Station	Latitude	Longitude	Missing data (%)	Start date	End date
Altheim (Germany)	48.15	9.46	0.001	Jan 2005	Dec 2015
Hikone (Japan)	35.27	136.24	0.006	Jan 1976	Dec 2008
Oregon (US)	44.13	-123.22	0.273	Jan 1950	Dec 2009
Pathhead (UK)	55.85	-2.99	0.156	Jan 1994	Dec 2014
Quinta Normal (Chile)	-33.44	-70.68	0.801	Jan 1917	Dec 1960

4.2 Implementation approach

For the disaggregation process, a leave-one-year-out cross-validation scheme was implemented, following a common practice in rainfall disaggregation studies (e-g., Bárdossy and Pegram, 2016; Poschlod et al., 2018; Guan et al., 2023). This approach maximizes the use of the limited hourly observations available at each site. For each realization, the hourly data of a given year are first aggregated to the daily scale and then disaggregated using the hourly data from the remaining years. This procedure is repeated for all years, and the resulting disaggregated annual series are concatenated to form a complete disaggregated hourly time series.

Applying the previous procedure for each location and method, 300 disaggregated series were obtained. This number of disaggregated series provides a robust estimate of the non-exceedance probability q for a given quantile x_q, based on the Weibull's plotting position formula (Weibull, 1939):

$$P(X \leq x_q | N = n) = q\left(\frac{n}{n+1}\right) \tag{17}$$

where x_q is the q^{th} quantile and n is the number of disaggregations. Note that the empirical rank of x_q is approximately $q \cdot n$, which becomes increasingly accurate as n grows. For instance, with $n = 50$, the empirical estimate underrepresents q by about 1.96%, while for $n = 300$, the error is approximately 0.33%. Increasing n to 1000 further reduces the error to 0.10%, but the marginal gain in accuracy does not justify the significantly higher computational cost.

In addition to the main disaggregation runs, 300 disaggregated series were also generated for each location for the calibration process of the SOC (monthly threshold ε) and the k-NN method (window size l) (see Sections 3.1 and 3.2 for details on the procedures). The calibration results for both methods are presented in Appendix A (Tables A1 and A2, respectively). Additionally, the Huff curves for each location are presented in Appendix B (Figure B1).

For the IDF-based assessment, return periods of 2, 5, 10 and 20 years were considered. Higher T values were excluded to avoid additional uncertainty associated with their estimation from short-size samples. For each T, one IDF curve was constructed per disaggregated series, allowing for the estimation of the stochastic uncertainty associated with the disaggregation process. In parallel, 1000 AMP series were generated for each duration to estimate the sampling uncertainty of the observed T-year IDF curve, following the procedure described in Section 2.2.1. This number of samples enables a robust estimation

https://doi.org/10.5194/egusphere-2025-2710
Preprint. Discussion started: 26 June 2025

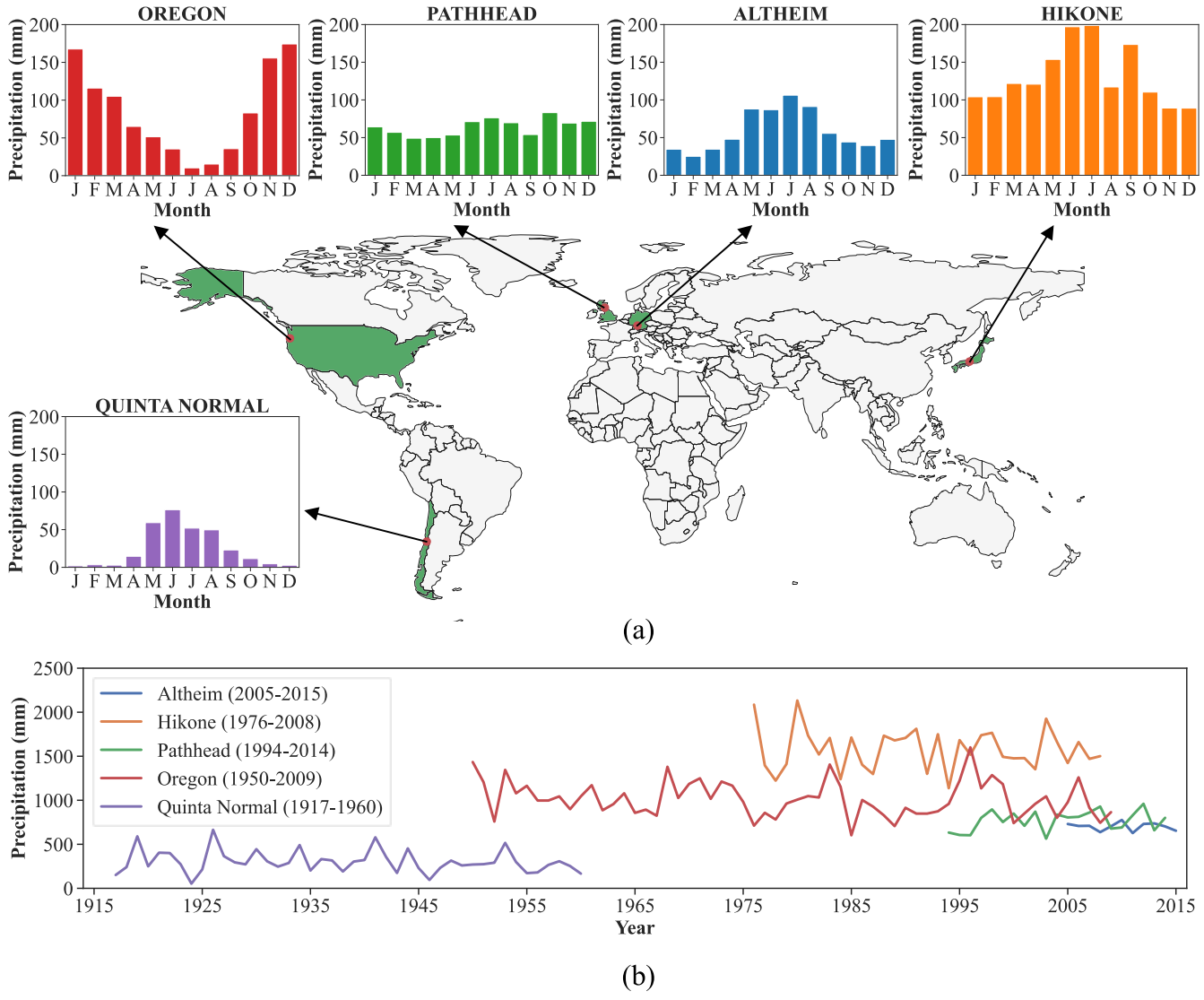

Figure 2. Selected stations for the assessment. (a) Location and monthly mean precipitation. (b) Annual precipitation series.

of sampling variability at a relatively low computational cost. For the estimation of D_T (Eq. 4), weights of $w_{AEI} = \frac{2}{3}$ and $w_{PEI} = \frac{1}{3}$ were applied, emphasizing accuracy over precision in the representation of the observed IDF curves. This avoids overvaluing precision, which reflects the consistency of the estimates rather than their alignment with the observed intensities.

Finally, for the ERIs assessment, one observed annual series and 300 disaggregated annual series were obtained for each index. This yielded 300 $mKGE$ values per index, for which their dispersion was computed as illustrated in Figure 1c.

https://doi.org/10.5194/egusphere-2025-2710
Preprint. Discussion started: 26 June 2025

5 Results and discussion

5.1 IDF curves analysis

Figure 3 shows the results of the accuracy-precision analysis of the IDF curves for each site and disaggregation method. Overall, all methodologies perform consistently well across the evaluated climates and methodologies regarding the AEI_T metric, which never falls below 0. For PEI_T, at least one result below 0 is observed at all five sites evaluated. This outcome reflects the inherent uncertainty introduced by the disaggregation process although it should not necessarily be interpreted as poor model performance. Values of PEI_T close to 0 are expected as they indicate that the uncertainty added through disaggregation remains comparable to the variability associated with statistical sampling. The differences observed in the ability of methods to replicate AEI_T and PEI_T highlight that their performance varies depending on which aspect of the IDF curves is being evaluated, and that this sensitivity is strongly influenced by the disaggregation method and the precipitation regime. These distinctions, made visible through the proposed evaluation framework, underscore the importance of assessing multiple performance dimensions simultaneously.

Although some variation across sites and methodologies is observed, the overall results suggest that the disaggregation methods under evaluation manage to adequately preserve the expected level of uncertainty. On the other hand, the combined D_T metric is more influenced by AEI_T in its calculation, as w_{AEI} is twice the value of w_{PEI}, leading to expected positive results. This weighting structure, explicitly defined within the proposed framework, allows to balance the emphasis between accuracy and precision depending on the specific application needs, with the flexibility to prioritize one dimension over the other.

Another takeaway from Figure 3 is that, although the performance of the disaggregation strongly depends on the local climate and disaggregation method, it does not vary significantly across different return periods (i.e. same shape markers are relatively clustered), except for Quinta Normal. Table 3 synthesizes the performance of each disaggregation method per location, by presenting the average value of each metric across the four return periods for each rainfall disaggregation approach. Overall, k-NN tends to achieve higher average AEI_T values across most sites, which also translates into better D_T values, due to the higher weighting of AEI_T in the combined metric. SOC generally exhibits lower performance regarding AEI_T but achieves better results for PEI_T, reflecting lower internal variability in the disaggregated IDF curves across realizations compared to the other methods.

These contrasting behaviors illustrate how the ability of disaggregation methods to replicate different aspects of rainfall structure varies depending on the method and location, a distinction that the framework is specifically designed to highlight. The superior ability of k-NN to represent AMPs for different durations (and therefore IDF curves) has been previously reported in several studies (e.g., Pui et al., 2012; Poschlod et al., 2018). In contrast, SOC's tendency to underestimate extreme rainfall intensities is consistent with findings on pulse-type models (e.g., Kaczmarska et al., 2014; Cross et al., 2018), as its random assignment of event durations may underrepresent short-duration, high-intensity events critical for IDF curves.

Table 4 summarizes how often each methodology achieves the best result across the three evaluated metrics, using their average values $\overline{AEI_T}$, $\overline{PEI_T}$ and $\overline{D_T}$. Although k-NN shows an overall tendency toward better performance, it is outperformed

https://doi.org/10.5194/egusphere-2025-2710
Preprint. Discussion started: 26 June 2025

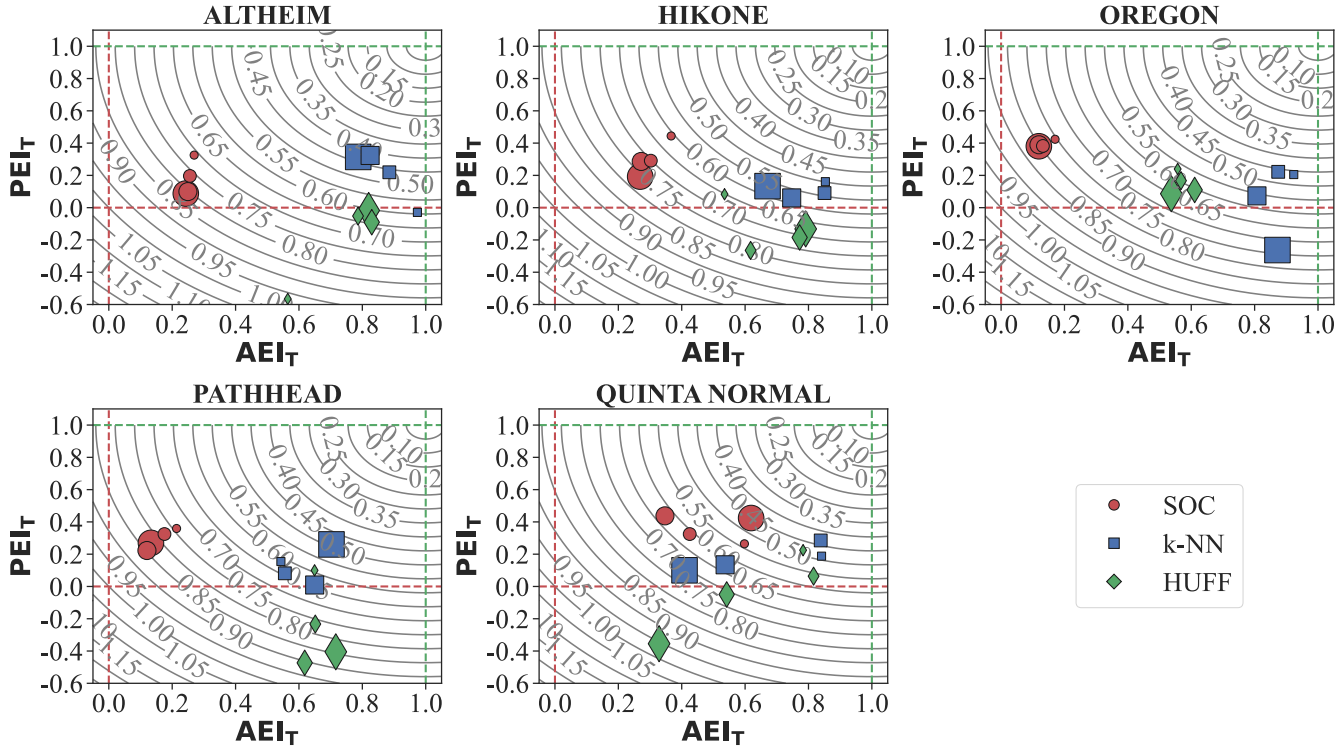

Figure 3. IDF curves accuracy (AEI_T) and precision (PEI_T) analysis. The 2-, 5-, 10- and 20-year return periods are presented with a bigger mark size related to a bigger return period.

Table 3. IDF curves mean accuracy ($\overline{AEI_T}$), mean precision ($\overline{PEI_T}$), and mean Euclidean distance ($\overline{D_T}$) values. $\overline{D_T}$ corresponds to the average D_T across the four return periods.

Location	$\overline{AEI_T}$			$\overline{PEI_T}$			$\overline{D_T}$		
	SOC	k-NN	HUFF	SOC	k-NN	HUFF	SOC	k-NN	HUFF
Altheim	0.254	0.868	0.750	0.177	0.207	-0.181	0.773	0.476	0.713
Hikone	0.367	0.780	0.679	0.304	0.111	-0.125	0.698	0.547	0.707
Oregon	0.135	0.869	0.568	0.394	0.059	0.151	0.788	0.555	0.605
Pathhead	0.161	0.613	0.659	0.294	0.126	-0.253	0.797	0.597	0.777
Quinta Normal	0.497	0.656	0.617	0.363	0.177	-0.029	0.557	0.564	0.677

by other methods for some metrics and sites, reinforcing that no method universally dominates across all conditions. For AEI_T, k-NN achieves the best results in four out of five locations, whereas for PEI_T, SOC leads in four out of five locations. When combining both metrics through D_T, k-NN again performs best in four out of five locations, except for Quinta Normal, where SOC achieves the lowest D_T value. HUFF has the least favorable performance overall, achieving the best AEI_T value only

https://doi.org/10.5194/egusphere-2025-2710
Preprint. Discussion started: 26 June 2025

Table 4. Number of locations in which each method performs best in terms of the mean accuracy ($\overline{AEI_T}$), mean precision ($\overline{PEI_T}$), and mean Euclidean distance ($\overline{D_T}$).

Metric	SOC	k-NN	HUFF	Number of stations
$\overline{AEI_T}$	0	4	1	5
$\overline{PEI_T}$	4	1	0	5
$\overline{D_T}$	1	4	0	5

once, at Pathhead. Finally, note that $\overline{AEI_T}$, $\overline{PEI_T}$ and $\overline{D_T}$ were calculated as simple average values, without assigning different weights to the return periods. One could evaluate the impact of using different weights, for example, giving more significance to larger return periods in a risk analysis.

5.2 ERIs analysis

In addition to evaluating the model's ability to replicate AMPs across different durations, it is also important to assess whether the methods replicates other attributes relevant to hydrologic analysis and modeling. Figure 4 shows the results of the ERIs analysis for each location and disaggregation method. Overall, all methods perform well in terms of mKGE, with little dispersion in the performance, except in Altheim. Altheim's higher dispersion in the results is likely due to its shorter time series (only 10 years of data), leading to greater variability and less statistical representativeness, which is visible in a higher number of outliers compared to the other locations. Despite this, all $mKGE$ values are positive and above the pre-defined minimum performance target threshold of zero. In the other locations, the performance is excellent, reflected by desirable $mKGE$ values. Note that a good performance is observed for the TDD index (i.e., TDD typically larger than 0.7), regardless of the location. Such behavior is expected as dry hours are more common than rainy ones.

The 50th percentiles of the $mKGE$ for each ERI is used to compare disaggregation methodologies (Table 5). In terms of the TDD index, k-NN performs the best, leading in three out of five locations with $mKGE$ values greater than 0.8. While for RX5 index, HUFF provides the highest $mKGE$ values in three out of five locations, all exceeding 0.83. For R95% and P>R95%, the best method is k-NN, leading in four out of five locations, with $mKGE$ values greater than 0.84 for R95% and over 0.79 for P>R95%. However, it performs the worst at Altheim, where it underperforms compared to HUFF, and marginally outperforms SOC. Once again, Altheim shows the lowest results across all ERIs and methodologies, although all results generally indicate good performance.

The results reflects the inherent nature of each disaggregation method. The k-NN method performs best when assessing TDD, likely due to its ability to preserve the natural patterns of rainfall intermittency by using historical records for the disaggregation. A good performance was also expected for SOC in this regard, considering its nature as a pulse-type model, which tends to correctly represent rainfall intermittency when sub-daily records are available (Park et al., 2021). However, this was not observed in practice. Although SOC assigns rainfall pulses with random intensity and duration, the allocation of time blocks and the timing between events do not follow a Poisson process, which is characteristic of point-process models.

https://doi.org/10.5194/egusphere-2025-2710
Preprint. Discussion started: 26 June 2025

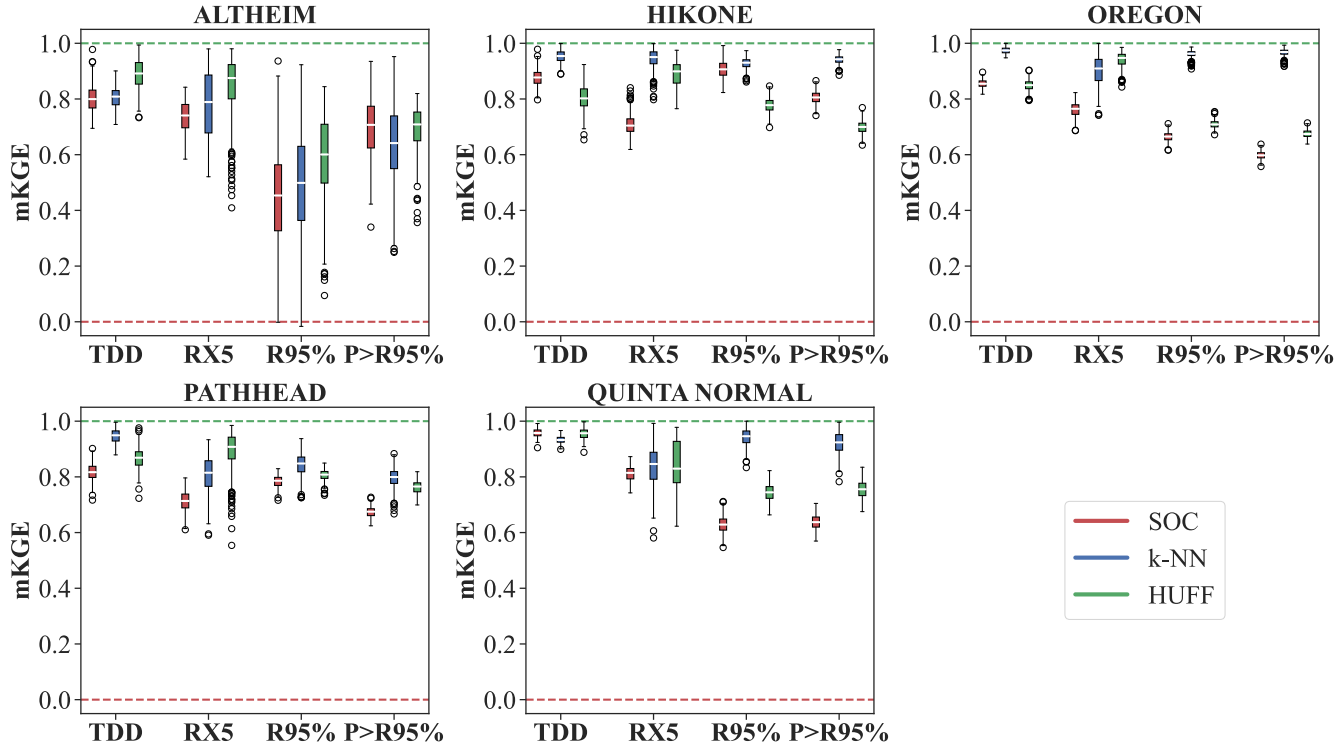

Figure 4. Boxplots with $mKGE$ values obtained for the ERIs series in the different locations under study.

Table 5. 50th percentile $mKGE$ of each ERI series at each location.

Location	TDD			RX5			R95%			P>R95%		
	SOC	k-NN	HUFF	SOC	k-NN	HUFF	SOC	k-NN	HUFF	SOC	k-NN	HUFF
Altheim	0.800	0.809	0.892	0.741	0.789	0.875	0.453	0.499	0.601	0.707	0.642	0.709
Hikone	0.877	0.954	0.802	0.704	0.950	0.899	0.907	0.930	0.778	0.806	0.943	0.700
Oregon	0.855	0.974	0.850	0.765	0.910	0.948	0.665	0.963	0.709	0.598	0.969	0.674
Pathhead	0.817	0.949	0.869	0.713	0.815	0.908	0.786	0.848	0.808	0.675	0.799	0.765
Quinta Normal	0.957	0.932	0.950	0.814	0.846	0.830	0.629	0.946	0.744	0.637	0.924	0.755

Consequently, SOC's pulse structure, generated within discrete intervals rather than over continuous time, limits its ability to fully capture the temporal dryness patterns of rainfall, which are critical for accurate hydrological modeling (Socolofsky et al., 2001). The relatively strong performance of k-NN in representing extreme percentiles is also expected, as it can capture both the magnitude and temporal distribution of extreme events that have occurred historically, unlike SOC and HUFF, which use a stochastic approach to estimate the duration and distribution of precipitation within a day. Finally, as Huff curves encompass the

https://doi.org/10.5194/egusphere-2025-2710
Preprint. Discussion started: 26 June 2025

Table 6. Number of locations where each method performs best for each ERI, based on the 50th percentile of $mKGE$.

Index	SOC	k-NN	HUFF	Number of locations
TDD	1	3	1	5
RX5	0	2	3	5
R95%	0	4	1	5
P>R95%	0	4	1	5

behavior of all historical precipitation events, the HUFF method is especially effective for representing intermediate-duration events, which tend to occur more frequently on average than short or very long events.

Table 6 summarizes the results across locations and methodologies. Overall, k-NN leads in both representing the driest hours of each year (TDD) and capturing the interannual variation in the magnitude (R95%) and duration (P>R95%) of the most extreme rainfall hours. In turn, HUFF provides the best representation of the observed 5-hour maximum precipitation series (RX5). These findings reinforce the value of the proposed framework in highlighting the specific strengths and limitations of each disaggregation method across different dimensions of extreme rainfall representation. Rather than suggesting a universally superior disaggregation method, the framework enables a nuanced understanding of how performance varies depending on the metric, location, and climatic context. Each method tends to excel in a different aspect (be it intensity percentiles, rainfall intermittency, or event structure), demonstrating the need for a comprehensive, multi-metric evaluation when selecting disaggregation approaches.

6 Conclusions

A framework was developed to easily compare the performance of daily-to-hourly rainfall disaggregation methods by testing mainly their ability to represent extreme precipitation. The framework, which could be also extended to other sub-daily timescales, was implemented and validated using three distinct rainfall disaggregation methods (i.e. a pulse-type method (SOC), a k-NN method (k-NN) and a nondimensional storm pattern-based method (HUFF)) and data from five locations across diverse global climate regimes (Altheim, Germany; Hikone, Japan; Oregon, US; Pathhead, UK; Quinta Normal, Chile). Our main conclusions are:

- By jointly analyzing IDF curves and extreme rainfall indices (ERIs), the framework provides a comprehensive and multi-dimensional basis for performance assessment across metrics, durations, and climatic contexts.

- Because multiple performance dimensions are considered simultaneously, the framework is able to reveal method-specific strengths and limitations in replicating different aspects, such as the accuracy and precision in replicating observed extreme precipitation (using IDF curves), percentile-based intensities, and temporal intermittency.

https://doi.org/10.5194/egusphere-2025-2710
Preprint. Discussion started: 26 June 2025

- The results from the case study demonstrate that the performance of disaggregation methods is highly context-dependent, with no single approach consistently outperforming across all sites and metrics. These findings emphasize the necessity of using a structured framework when selecting or developing disaggregation methodologies for hydrological studies.

In addition, the following specific findings emerged from the application of the framework to the case study. These are consistent with the theoretical foundations of the disaggregation methods and support the need for evaluation tools that can capture their behavior across different aspects of extreme precipitation:

- The k-NN method demonstrated strong performance in balancing accuracy and precision in IDF curve replication, while SOC achieved higher precision at the expense of lower accuracy.

- In the ERIs evaluation, k-NN performed best in modeling extreme rainfall percentiles (R95 and P>R95%) and rainfall intermittency (i.e., TDD), while HUFF replicates better intermediate-duration maximum rainfall (i.e., RX5).

The framework developed in this study provides a structured and flexible basis for assessing the ability of rainfall disaggregation methods to represent extreme precipitation. It is designed to accommodate a wide range of disaggregation approaches beyond the ones tested in this paper, and regardless of their underlying assumptions or formulations. By integrating multiple performance metrics and enabling method-to-method comparisons across locations and indicators, the framework facilitates a systematic understanding of the strengths and limitations inherent to each approach. Future work should aim to adapt and expand the framework to address additional challenges within precipitation disaggregation. One potential direction is to extend the methodology toward sub-hourly disaggregation, allowing its application in contexts where finer temporal resolutions are required. Another avenue could involve adapting the framework to evaluate disaggregation methods across spatial scales, facilitating the transfer of precipitation information from larger areas to localized catchments or urban settings.

Code and data availability. The dataset for the Quinta Normal station (Chile) was obtained from the Chilean Meteorological Directorate (DMC) (https://www.meteochile.gob.cl), while the hourly datasets for the remaining four locations were retrieved from the Global Sub-Daily Rainfall Dataset (GSDR) developed by the INTENSE project (Lewis et al., 2019, 2021). Additional data and source code are available from the corresponding author upon reasonable request.

https://doi.org/10.5194/egusphere-2025-2710
Preprint. Discussion started: 26 June 2025

Appendix A: Calibration of minimum event threshold (SOC) and half-window size (k-NN)

Table A1. Calibrated monthly minimum event thresholds (mm) for the SOC method.

Location	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Altheim	1.4	1.3	1.8	2.1	5.7	6.2	8.4	6.7	3.7	3.2	2.2	1.1
Hikone	6.7	2.3	8.9	4.1	5.2	8.2	16.4	14.3	7.9	4.6	3.9	7.8
Oregon	2.7	2.7	2.5	2.4	3.1	3.8	4.4	4.6	4.2	3.4	3.3	3.2
Pathhead	1.7	1.3	3.3	1.8	2.1	2.7	3.3	3.0	2.7	2.3	1.8	1.8
Quinta Normal	9.8	10.8	6.6	13.7	3.9	4.2	4.1	3.0	2.4	3.6	5.9	15.9

https://doi.org/10.5194/egusphere-2025-2710
Preprint. Discussion started: 26 June 2025

Table A2. Calibrated half-window sizes and their respective $RMSE_{wei}$ values for the k-NN method.

Location	Half-window size (days)	$RMSE_{wei}$
Altheim	8	0.304
Hikone	5	0.251
Oregon	14	0.216
Pathhead	1	0.283
Quinta Normal	25	0.161

https://doi.org/10.5194/egusphere-2025-2710
Preprint. Discussion started: 26 June 2025

Appendix B: Huff curves

Figure B1 presents the Huff curves for the five study sites. It is important to note that these curves were derived using the complete set of hourly records, including all available years. Therefore, they serve illustrative purposes only. In the case study application, Huff curves were instead computed for each year individually, using the remaining years by following the leave-one-year-out cross-validation approach.

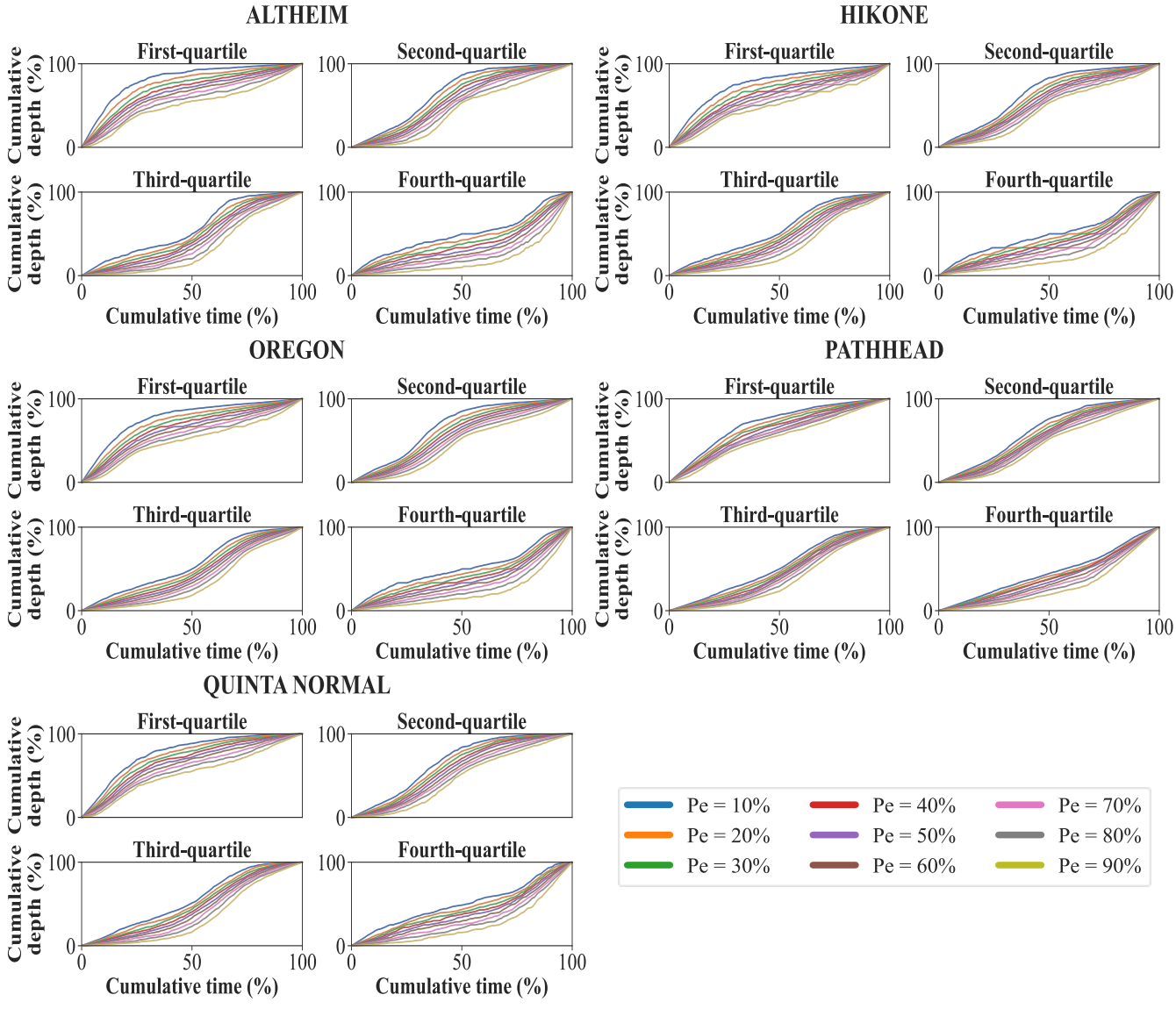

Figure B1. Huff curves, considering the entire observed hourly rainfall record in each case.

https://doi.org/10.5194/egusphere-2025-2710
Preprint. Discussion started: 26 June 2025

Author contributions. CS: study conceptualization, methodology development, data curation and analysis, manuscript preparation. JG and CC: study conceptualization, research supervision, manuscript revision and editing.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. We acknowledge the financial support from FONDECYT Grant No. 1241163. We also thank ANID for the FONDAP Grants No. 1523A0004 and 1523A0009, as well as the support provided by the Research Development Program from University Adolfo Ibáñez through the 2025 FONDECYT Puente Project. We are also grateful for the Arturo Cousiño Lyon Scholarship funded by Fundación San Carlos de Maipo of the Sociedad del Canal de Maipo.

https://doi.org/10.5194/egusphere-2025-2710
Preprint. Discussion started: 26 June 2025

References

Abreu, M. C., Pereira, S. B., Cecílio, R. A., Pruski, F. F., de Almeida, L. T., & da Silva, D. D. (2022). Assessing the application of ratios between daily and sub-daily extreme rainfall as disaggregation coefficients. Physics and Chemistry of the Earth, Parts A/B/C, 128, 103223. https://doi.org/10.1016/j.pce.2022.103223

Alam, M. S., & Elshorbagy, A. (2015). Quantification of the climate change-induced variations in intensity–duration–frequency curves in the Canadian Prairies. Journal of Hydrology, 527, 990-1005. https://doi.org/10.1016/j.jhydrol.2015.05.059

Back, Á. J., Oliveira, J. L. R., & Henn, A. (2012). Relações entre precipitações intensas de diferentes durações para desagregação da chuva diária em Santa Catarina. Revista Brasileira de Engenharia Agrícola e Ambiental, 16(4), 391–398. https://doi.org/10.1590/S1415-43662012000400002

Bárdossy, A., & Pegram, G. G. S. (2016). Space-time conditional disaggregation of precipitation at high resolution via simulation. Water Resources Research, 52(2), 920–937. https://doi.org/10.1002/2015wr018037

Bhattacharyya, D., & Saha, U. (2023). Deep learning application for disaggregation of rainfall with emphasis on preservation of extreme rainfall characteristics for Indian monsoon conditions. Stochastic Environmental Research and Risk Assessment, 37(1021–1038). https://doi.org/10.1007/s00477-022-02331-x

Bennett, J. C., Robertson, D. E., Ward, P. G. D., Hapuarachchi, H. A. P., & Wang, Q. J. (2016). Calibrating hourly rainfall-runoff models with daily forcings for streamflow forecasting applications in meso-scale catchments. Environmental Modelling & Software, 76, 20–36. https://doi.org/10.1016/j.envsoft.2015.11.006

Bruni, G., Reinoso, R., van de Giesen, N. C., Clemens, F. H. L. R., & ten Veldhuis, J. A. E. (2015). On the sensitivity of urban hydrodynamic modelling to rainfall spatial and temporal resolution. Hydrology and Earth System Sciences, 19(2), 691–709. https://doi.org/10.5194/hess-19-691-2015

Cârsteanu, A., & Foufoula-Georgiou, E. (1996). Assessing dependence among weights in a multiplicative cascade model of temporal rainfall. Journal of Geophysical Research, 101(D21), 26363–26370. https://doi.org/10.1029/96JD01657

Choi, J., Socolofsky, S. A., & Olivera, F. (2008). Hourly disaggregation of daily rainfall in Texas using measured hourly precipitation at other locations. Journal of Hydrologic Engineering, 13(6), 476–487. https://doi.org/10.1061/(ASCE)1084-0699(2008)13:6(476)

Cowpertwait, P. S. P. (1991). Further developments of the Neyman–Scott clustered point process for modeling rainfall. Water Resources Research, 27(7), 1431–1438. https://doi.org/10.1029/91WR00479

Cross, D., Onof, C., Winter, H., & Bernardara, P. (2018). Censored rainfall modelling for estimation of fine-scale extremes. Hydrology and Earth System Sciences, 22(1), 727–756. https://doi.org/10.5194/hess-22-727-2018

Glasbey, C. A., Cooper, G., & McGechan, M. B. (1995). Disaggregation of daily rainfall by conditional simulation from a point-process model. Journal of Hydrology, 165(1–4), 1–9. https://doi.org/10.1016/0022-1694(94)02598-6

Guan, X., Nissen, K., Nguyen, V.D., Merz, B., Winter, B., & Vorogushyn, S. (2023). Multisite temporal rainfall disaggregation using methods of fragments conditioned on circulation patterns. Journal of Hydrology, 621, Article 129640. https://doi.org/10.1016/j.jhydrol.2023.129640

Gumbel, E. J. (1935). Les valeurs extrêmes des distributions statistiques. Annales de l'institut Henri Poincaré, 5(2), http://www.numdam.org/item?id=AIHP_1935__5_2_115_0

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. Journal of Hydrology, 377(1–2), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

https://doi.org/10.5194/egusphere-2025-2710
Preprint. Discussion started: 26 June 2025

Gupta, V., & Waymire, E. (1993). A statistical analysis of mesoscale rainfall as a random cascade. Journal of Applied Meteorology, 32, 251–267. https://doi.org/10.1175/1520-0450(1993)032<0251:ASAOMR>2.0.CO;2

Gutierrez-Magness, A. L., & McCuen, R. H. (2004). Accuracy Evaluation of Rainfall Disaggregation Methods. Journal of Hydrologic Engineering, 9(2), 71-78. https://doi.org/10.1061/(ASCE)1084-0699(2004)9:2(71)

Hershenhorn, J., & Woolhiser, D. A. (1987). Disaggregation of daily rainfall. Journal of Hydrology, 95(3–4), 299–322. https://doi.org/10.1016/0022-1694(87)90008-4

Hingray, B., & Ben Haha, M. (2005). Statistical performances of various deterministic and stochastic models for rainfall series disaggregation. Atmospheric Research, 77(1–4), 152–175. https://doi.org/10.1016/j.atmosres.2004.10.023

Hingray, B., Monbaron, E., Jarrar, I., Favre, A. C., Consuegra, D., & Musy, A. (2002). Stochastic generation and disaggregation of hourly rainfall series for continuous hydrological modelling and flood control reservoir design. Water Science and Technology, 45(2), 113–119. https://doi.org/10.2166/wst.2002.0035

Huff, F. A. (1967). Time distribution of rainfall in heavy storms. Water Resources Research, 3(4), 1007–1021. https://doi.org/10.1029/WR003i004p01007

Kaczmarska, J., Isham, V., & Onof, C. (2014). Point process models for fine-resolution rainfall. Hydrological Sciences Journal, 59(11), 1972–1991. https://doi.org/10.1080/02626667.2014.925558

Khaliq, M. N., & Cunnane, C. (1996). Modelling point rainfall occurrences with the modified Bartlett-Lewis rectangular pulses model. Journal of Hydrology, 180(1–4), 109–138. https://doi.org/10.1016/0022-1694(95)02894-3

Kioutsioukis, I., Melas, D., & Zerefos, C. (2009). Statistical assessment of changes in climate extremes over Greece (1955–2002). International Journal of Climatology, 30(11), 1723–1737. https://doi.org/10.1002/joc.2030

Klein Tank, A. M. G., & Können, G. P. (2003). Trends in indices of daily temperature and precipitation extremes in Europe, 1946–99. Journal of Climate, 16(24), 3665–3680. https://doi.org/10.1175/1520-0442(2003)016

Knoben, W. J. M., Freer, J. E., & Woods, R. A. (2019). Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. Hydrology and Earth System Sciences, 23(10), 4323–4331. https://doi.org/10.5194/hess-23-4323-2019

Koutsoyiannis, D., & Onof, C. (2001). Rainfall disaggregation using adjusting procedures on a Poisson cluster model. Journal of Hydrology, 246(1-4), 109-122. https://doi.org/10.1016/S0022-1694(01)00363-8

Lall, U., & Sharma, A. (1996). A nearest neighbor bootstrap for resampling hydrologic time series. Water Resources Research, 32(3), 679-693. https://doi.org/10.1029/95WR02966

Legates, D. R., & G. J. McCabe Jr. (1999). Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. Water Resources Research, 35(1), 233–241. https://doi.org/10.1029/1998WR900018

Lewis, E., Fowler, H. J., Alexander, L. V., Dunn, R., McClean, F., Barbero, R., Guerreiro, S., Li, X.-F., & Blenkinsop, S. (2019). GSDR: A global sub-daily rainfall dataset. Journal of Climate, 32(14), 4715–4729. https://doi.org/10.1175/JCLI-D-18-0143.1

Lewis, E., Pritchard, D., Villalobos-Herrera, R., Blenkinsop, S., McClean, F., Guerreiro, S., Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Rustemeier, E., & Fowler, H. J. (2021). Quality control of a global hourly rainfall dataset. Environmental Modelling & Software, 144, 105169. https://doi.org/10.1016/j.envsoft.2021.105169

Li, Z., Li, J. J., & Shi, X. P. (2020). A Two-Stage Multisite and Multivariate Weather Generator. Journal of Environmental Informatics, 35(2). https://doi.org/10.3808/jei.201900424

Mehrotra, R., & Sharma, A. (2006). Conditional resampling of hydrologic time series using multiple predictor variables: A K-nearest neighbour approach. Advances in Water Resources, 29(7), 987-999. https://doi.org/10.1016/j.advwatres.2005.08.007

https://doi.org/10.5194/egusphere-2025-2710
Preprint. Discussion started: 26 June 2025

Molnar, P., & Burlando, P. (2005). Preservation of rainfall properties in stochastic disaggregation by a simple random cascade model. Atmospheric Research, 77(1-4), 137–151. https://doi.org/10.1016/j.atmosres.2004.10.024

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I — A discussion of principles. Journal of Hydrology, 10(3), 282-290. https://doi.org/10.1016/0022-1694(70)90255-6

Ochoa-Rodriguez, S., Wang, L.-P., Gires, A., Pina, R. D., Reinoso-Rondinel, R., Bruni, G., Ichiba, A., Gaitan, S., Cristiano, E., van Assel, J., Kroll, S., Murlà-Tuyls, D., Tisserand, B., Schertzer, D., Tchiguirinskaia, I., Onof, C., Willems, P., & ten Veldhuis, M.-C. (2015). Impact of spatial and temporal resolution of rainfall inputs on urban hydrodynamic modelling outputs: A multi-catchment investigation. Journal of Hydrology, 531(Part 2), 389–407. https://doi.org/10.1016/j.jhydrol.2015.05.035

Onof, C., & Wheater, H. S. (1993). Modelling of British rainfall using a random parameter Bartlett-Lewis Rectangular Pulse Model. Journal of Hydrology, 149(1–4), 67–95. https://doi.org/10.1016/0022-1694(93)90100-N

Park, J., Cross, D., Onof, C., Chen, Y., & Kim, D. (2021). A simple scheme to adjust Poisson cluster rectangular pulse rainfall models for improved performance at sub-hourly timescales. Journal of Hydrology, 598, 126296. https://doi.org/10.1016/j.jhydrol.2021.126296

Poschlod, B., Hodnebrog, Ø., Wood, R. R., Alterskjær, K., Ludwig, R., Myhre, G., & Sillmann, J. (2018). Comparison and Evaluation of Statistical Rainfall Disaggregation and High-Resolution Dynamical Downscaling over Complex Terrain. Journal of Hydrometeorology, 19(12), 1973–1982. https://doi.org/10.1175/JHM-D-18-0132.1

Pui, A., Sharma, A., Mehrotra, R., Sivakumar, B., & Jeremiah, E. (2012). A comparison of alternatives for daily to sub-daily rainfall disaggregation. Journal of Hydrology, 470–471, 138–157. https://doi.org/10.1016/j.jhydrol.2012.08.041

Qin, X., & Dai, C. (2024). Investigation of rainfall disaggregation with flexible timescales based on point process models. Journal of Hydrology, 634, 131101. https://doi.org/10.1016/j.jhydrol.2024.131101

Reynolds, J. E., Halldin, S., Xu, C. Y., Seibert, J., & Kauffeldt, A. (2017). Sub-daily runoff predictions using parameters calibrated on the basis of data with a daily temporal resolution. Journal of Hydrology, 550, 399–411. https://doi.org/10.1016/j.jhydrol.2017.05.012

Rodriguez-Iturbe, I., Cox, D. R., & Isham, V. (1987). Some models for rainfall based on stochastic point processes. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 410(1839), 269–288. https://doi.org/10.1098/rspa.1987.0039

Sharma, A., & Mehrotra, R. (2010). Rainfall generation. In M. Gebremichael (Ed.), *Rainfall: state of the science* (pp. 215–246). Washington: American Geophysical Union. http://dx.doi.org/10.1029/2010GM000973

Sharma, A., & Srikanthan, R. (2006). Continuous rainfall simulation: a nonparametric alternative. In *30th Hydrology and Water Resources Symposium*. Launceston, Tasmania: Conference Design.

Schilling, W. (1991). Rainfall data for urban hydrology: What do we need? Atmospheric Research, 27(1–3), 5–21. https://doi.org/10.1016/0169-8095(91)90003-F

Socolofsky, S., Adams, E. E., & Entekhabi, D. (2001). Disaggregation of daily rainfall for continuous watershed modeling. Journal of Hydrologic Engineering, 6(1), 300–309. https://doi.org/10.1061/(ASCE)1084-0699(2001)6:4(300)

Taschetto, A. S., & England, M. H. (2009). An analysis of late twentieth century trends in Australian rainfall. International Journal of Climatology, 29(6), 791–807. https://doi.org/10.1002/joc.1736

Thober, S., Mai, J., Zink, M., & Samaniego, L. (2014). Stochastic temporal disaggregation of monthly precipitation for regional gridded data sets. Water Resources Research, 50(11), 8714–8735. https://doi.org/10.1002/2014wr015930

Vicuña, S., Gironás, J., Meza, F. J., Cruzat, M. L., Jelinek, M., Bustos, E., Poblete, D., & Bambach, N. (2013). Exploring possible connections between hydrological extreme events and climate change in central south Chile. Hydrological Sciences Journal, 58(8), 1598–1619. https://doi.org/10.1080/02626667.2013.840380

https://doi.org/10.5194/egusphere-2025-2710
Preprint. Discussion started: 26 June 2025

Viglione, A., & Blöschl, G. (2009). On the role of storm duration in the mapping of rainfall to flood return periods. Hydrology and Earth System Sciences, 13(2), 205–216. https://doi.org/10.5194/hess-13-205-2009

595 Weibull, W. (1939). A statistical theory of the strength of materials. Ing. Vetensk. Akad. Handl., 151, 1–45.

Woldemeskel, F. M., Sharma, A., Mehrotra, R., & Westra, S. (2016). Constraining continuous rainfall simulations for derived design flood estimation. Journal of Hydrology, 542, 581–588. https://doi.org/10.1016/j.jhydrol.2016.09.028

Zhang, X., Alexander, L., Hegerl, G. C., Jones, P., Klein Tank, A., Peterson, T. C., Trewin, B., & Zwiers, F. W. (2011). Indices for monitoring changes in extremes based on daily temperature and precipitation data. WIREs Climate Change, 2(6), 851–870.

600 https://doi.org/10.1002/wcc.147