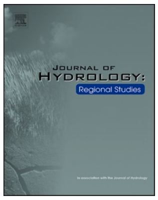
Journal of Hydrology: Regional Studies 66 (2026) 103720

Contents lists available at ScienceDirect

Journal of Hydrology: Regional Studies

journal homepage: www.elsevier.com/locate/ejrh



Uncertainty-informed regionalization of sub-daily rainfall extremes in northwestern Italy using Bilinear Surface Smoothing with credible-interval-based data screening

Nikolaos Malamos [a,*], Paola Mazzoglio [b], Theano Iliopoulou [c,d]

[a] *Department of Natural Resources Development & Agricultural Engineering, Agricultural University of Athens, Iera Odos 75, Athens GR-11855, Greece*
[b] *Department of Environment, Land and Infrastructure Engineering, Politecnico di Torino, Torino 10129, Italy*
[c] *Department of Geography, School of Environment, Geography and Applied Economics, Harokopio University of Athens, Athens GR-17676, Greece*
[d] *Department of Water Resources and Environmental Engineering, School of Civil Engineering, National Technical University of Athens, Athens GR-15780, Greece*

A R T I C L E I N F O

Keywords:
Sub-daily rainfall extremes
Spatial interpolation
Uncertainty quantification
Bilinear Surface Smoothing
Credible intervals
Italy

A B S T R A C T

Study region: This study focuses on the northwestern part of Italy, where accurate spatial characterization of sub-daily rainfall extremes is essential for hydrological design, flood risk assessment, and infrastructure planning. Observational data are drawn from the I^2-RED dataset, comprising approximately 820 rain gauges with records spanning 1927–2020. The region is characterized by complex terrain and substantial spatial variability in rainfall extremes.
Study focus: The Bilinear Surface Smoothing with Explanatory variable (BSSE) framework is applied to regionalize mean annual maximum rainfall depths at 1, 3, 6, 12, and 24 h timescales. Elevation is incorporated as an explanatory variable. The stochastic formulation of BSSE was exploited to identify stations whose observations fall outside the credible intervals of the fitted spatial field. These stations were removed, and the BSSE parameters were re-optimized to assess the effect of data screening on in-sample and leave-one-out cross-validation performance.
New hydrological insights for the region: The identified outliers predominantly correspond to legacy gauges with short or temporally unrepresentative records, whose anomalously low mean annual maxima likely reflect sampling limitations and, in some cases, documented intensification of rainfall extremes in the region. Their removal leads to systematic performance improvements across all timescales. The results show that heterogeneous rain gauge networks can introduce significant uncertainty in regionalization analyses and that BSSE offers a reproducible framework for spatial interpolation and model-based data quality assessment.

1. Introduction

Regionalization of rainfall extremes is a central problem in hydrology because design, risk analysis, and infrastructure assessment often require spatially continuous estimates at locations where direct observations are sparse, short, or unavailable (Casson and Coles, 1999; Buishand et al., 2008; Svensson and Jones, 2010; Claps et al., 2022). The challenge is especially important for sub-daily

* Corresponding author.
E-mail addresses: nmalamos@aua.gr (N. Malamos), paola.mazzoglio@polito.it (P. Mazzoglio), tiliopoulou@hua.gr (T. Iliopoulou).

https://doi.org/10.1016/j.ejrh.2026.103720
Received 29 April 2026; Received in revised form 2 July 2026; Accepted 3 July 2026
Available online 10 July 2026

timescales, where the spatial coherence of extreme events decreases substantially, the influence of small-scale convective processes becomes dominant and both sampling uncertainty and local physiographic controls increase variability (Coles and Tawn, 1996; Dunn, 2001; Renard et al., 2010; Hu et al., 2013).

Various approaches have shown that precipitation estimation over complex terrain benefits substantially from auxiliary variables such as elevation or gridded rainfall products, especially when those variables are available on a denser support than the station network. Daly et al. (1994) first showed how topographic information could be used to improve precipitation mapping in mountainous terrain. Daly et al. (2002) further developed that logic into a knowledge-based mapping framework for climatic variables: topographic and climatic variability is modeled by incorporating terrain facets, coastal proximity, and a two-layer atmospheric structure to account for effects such as aspect, exposure, temperature inversions, and orographic precipitation patterns. Thornton et al. (1997) demonstrated that dense physiographic information can materially improve interpolation when the point network is sparse or irregular, a principle that remains central in more recent terrain-aware gridded estimation systems. Over the Alps, Frei and Schär (1998) demonstrated that a simple and unique precipitation–height relationship does not exist because much of the topographic signal is associated with slope and shielding rather than height effects. In parallel, geostatistical and regression-based approaches such as kriging with external drift, cokriging, and local-regression methods showed that rainfall regionalization becomes more realistic when it uses explanatory information that encodes terrain and spatial structure rather than relying on distances among stations alone (Goovaerts, 1997, 2000; Brunsdon et al., 2001; Diodato and Ceccarelli, 2005; Di Piazza et al., 2011; Crespi et al., 2018; Prudhomme and Reed, 1998, 1999).

In the context of extreme rainfall regionalization, the orographic effect introduces a systematic and scale-dependent signal. Studies have shown that the relationship between elevation and precipitation extremes varies significantly across temporal aggregation scales, with orographic enhancement of precipitation becoming progressively more pronounced at longer timescales (Libertino et al., 2018; Mazzoglio et al., 2022, 2023; Iliopoulou et al., 2024). At short timescales, the so-called reverse orographic effect tends to prevail (Allamano et al., 2009; Avanzi et al., 2015; Formetta et al., 2022; Mazzoglio et al., 2022, 2023; Malamos et al., 2025). In this case, convective processes tend to dominate and exhibit more erratic spatial structures, whereas at longer timescales large-scale atmospheric forcing and orographic lifting become increasingly important, leading to more spatially coherent patterns. At the same time, literature has emphasized that uncertainty quantification is not an optional supplement to spatial estimation, but part of the estimation problem itself. Coles and Tawn (1996) showed, in the context of rainfall extremes, that point estimates alone are inadequate because extreme-event inference is inherently uncertain and often requires extrapolation beyond the observed sample. Pardo-Igúzquiza et al. (2006) argued that uncertainty must be treated as a core component of rainfall-field estimation, not merely as an afterthought. More generally, hydrologic predictive uncertainty is known to reflect multiple components, including input, structural, and parameter uncertainty, and meaningful inference becomes difficult when these sources are poorly separated or when the data themselves contain inconsistencies (Renard et al., 2010).

A further implication is that uncertainty bounds can also serve as diagnostic tools. In rainfall applications, such intervals may help reveal stations that are difficult to reconcile with the fitted spatial model because of local anomalies, short or fragmented records, metadata problems, or measurement inconsistencies. This point is especially relevant for extreme-rainfall datasets, where finite-sample uncertainty is large and where retaining every available record does not necessarily improve the final regionalization. However, existing regionalization approaches rarely exploit model-based uncertainty estimates to diagnose inconsistencies in observational datasets, treating data quality control and spatial interpolation as largely separate problems. This separation limits the ability to fully leverage available information and may lead to suboptimal regionalization results.

Within this context, Bilinear Surface Smoothing (BSS) and its extension with an explanatory variable, Bilinear Surface Smoothing with Explanatory variable (BSSE), provide a suitable methodological framework. The method was introduced as a linear smoother defined on a bilinear grid, with optional incorporation of an explanatory field observed on a denser support than the target point data and parameter selection through generalized cross-validation (Malamos and Koutsoyiannis, 2016a, 2016b). A more recent extension added posterior variance estimation and Bayesian credible intervals, allowing uncertainty to be embedded directly into the interpolation model itself (Malamos et al., 2025). This is especially useful because it enables uncertainty to quantify the reliability of the fitted surface and, at the same time, to provide a model-based criterion for identifying observations that are statistically incompatible with the inferred spatial structure.

The above-presented framework is applied to northwestern Italy as a representative case study of a region with complex terrain and strong sub-daily rainfall variability, where previous studies have documented marked spatial heterogeneity and timescale-dependent orographic effects (Mazzoglio et al., 2022, 2023). We investigate whether credible-interval-based correction can improve the regionalization of sub-daily average annual maximum rainfall by identifying and removing stations that behave as outliers relative to the global spatial explanatory structure. At the same time, we examine whether the resulting reduction in noise leads the BSSE optimization procedure to select a systematically different surface geometry, as expressed through changes in surface resolution and tension parameters.

Ultimately, this study proposes a unified framework for the combined treatment of data uncertainty and regionalization performance, enabling a more robust evaluation of how uncertainty propagates into both the structure and the outputs of regional models. Although geostatistical methods such as kriging with external drift or cokriging can also incorporate auxiliary spatial information and quantify interpolation uncertainty, the proposed BSSE framework integrates credible-interval-based data consistency assessment directly into the stochastic regionalization model.

The paper is organized as follows. Section 2 describes the study area and the observational dataset. Section 3 presents the BSSE methodological framework, including the stochastic model formulation, the credible interval-based outlier correction procedure, and the evaluation criteria adopted. Section 4 reports the results of the study. Section 5 discusses the main findings in a broader context.

Section 6 summarizes the conclusions.

2. Area of interest and dataset

In this work, we selected the northwestern part of Italy as a case study. The area of interest includes the administrative region of Piedmont, enlarged with a 50 km buffer to include nearby areas of Aosta Valley, Liguria, Lombardy and Emilia-Romagna (Fig. 1). The area has been selected for its complex topography including the Alps in the North and West sides, the hilly regions of Langhe and Monferrato in the South of Piedmont, extensive plains that extend toward the East, and its proximity to the Ligurian Sea.

Annual maximum rainfall depths were derived from the Improved Italian – Rainfall Extreme Dataset or I^2-RED (Mazzoglio et al., 2020), a homogenized and quality-controlled database of rain gauge observations developed to analyze extreme precipitation events across Italy. The dataset integrates long-term pluviometric records from a dense national monitoring network and applies rigorous procedures for data validation, consistency checks, and homogenization. In this work, we considered sub-daily (1, 3, 6, 12, 24 h) time scales, which are particularly relevant for hydrological applications, flood risk assessment, and infrastructure design. The available time series span the period 1927–2020. A total of approximately 820 gauged locations were selected for analysis, each characterized by at least 10 years of available data for the timescales considered. This threshold was adopted to ensure a minimum level of statistical robustness in the estimation of the mean value of the annual maxima, while preserving adequate spatial coverage across the study area. Table 1 contains summary statistics of the stations considered in this study.

The digital elevation model used in this work is the Shuttle Radar Topography Mission (SRTM) at 30 m spatial resolution (Farr et al., 2007), resampled at 1-km resolution using a cubic interpolation, as in Daymet (Thornton et al., 2021).

The elevation of each rain gauge was obtained from this resampled DEM using the geographic coordinates of the stations. This procedure follows the methodology proposed in PRISM (Daly et al., 1994, 2002) and later implemented in Daymet (Thornton et al., 1997, 2021) and in previous works conducted in Italy (Mazzoglio et al., 2023). This step is methodologically significant and not trivial, as the way elevation is assigned can substantially influence spatial interpolation and the representation of orographic effects. Indeed, the PRISM developers emphasized that the relationship between precipitation and elevation is better captured when local elevations are derived from a moderate- or low-resolution DEM, rather than from very fine-scale terrain data. A coarser DEM provides a smoothed representation of the terrain, which more accurately reflects the spatial scale of orographic processes that control precipitation patterns. In contrast, high-resolution elevation data can introduce local topographic noise, leading to unrealistic gradients and biases in precipitation-elevation relationships.

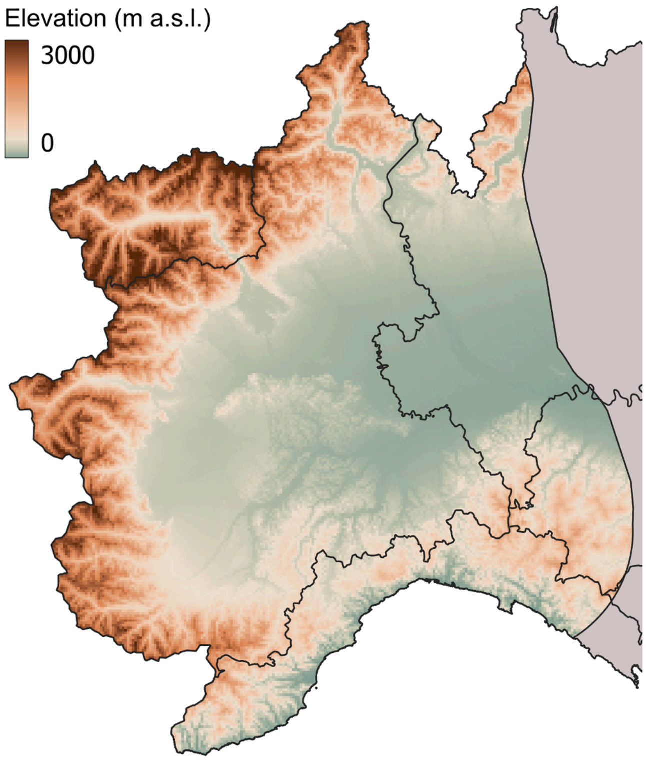


Fig. 1. Area of interest.

Table 1
Number of stations with sub-daily data at each timescale along with the corresponding descriptive statistics values of the dataset.

Timescale	Number	Minimum (mm)	Maximum (mm)	Mean (mm)	Median (mm)	σ_o (mm)
1 h	822	9.9	64.7	30.3	29.8	9.8
3 h	822	17.1	106.1	45.3	43.1	15.4
6 h	821	24.6	150.3	59.6	56.1	20.6
12 h	822	32.5	204.6	79.5	74.2	27.7
24 h	821	43.8	257.8	105.6	100.5	37.3

3. Methodological framework

3.1. Bilinear Surface Smoothing with Explanatory variable (BSSE) framework

Bilinear Surface Smoothing (BSS; Malamos and Koutsoyiannis, 2016a, 2016b) represents an alternative and flexible approach for spatial interpolation. The method involves fitting a bilinear surface within a regression framework, defined by known breakpoints and adjustable smoothing parameters. This formulation allows for a balance between local detail and overall surface continuity. Furthermore, the approach can be extended to objectively incorporate the influence of auxiliary explanatory variables (particularly when such variables are available from denser datasets) thereby enhancing the accuracy and representativeness of the interpolated surface, leading to the Bilinear Surface Smoothing with Explanatory Variable (BSSE).

Let $\boldsymbol{z} = (z_1, \ldots, z_n)^{\mathrm{T}}$ be a set of n observed rainfall values at the points with coordinates (x_i, y_i), where i=1, …, n. We also assume an explanatory variable $t_i(x_i, y_i)$ which is available on a denser grid so:

$$\boldsymbol{T} = \mathrm{diag}(t_1(x_1, y_1), \cdots, t_n(x_n, y_n)) \tag{1}$$

For two integers mx, my, we define a bilinear grid with m nodes as:

$$m = (\mathrm{mx} + 1)(\mathrm{my} + 1) - 1 \tag{2}$$

Let $\boldsymbol{\Pi}$ with size $n \times (m + 1)$ be the bilinear basis matrix. In this case, the BSSE model that represents the fitted field at the given locations is (Malamos and Koutsoyiannis, 2016a):

$$\widehat{\boldsymbol{z}} = \boldsymbol{\Pi d} + \boldsymbol{T\Pi e} \tag{3}$$

where $\boldsymbol{d}$, $\boldsymbol{e}$ are the vectors with size $m + 1$ of the unknown applicates of the bilinear surfaces d and e, respectively.

Roughness along x and y axes is controlled by matrices $\boldsymbol{\Psi}_x$, $\boldsymbol{\Psi}_y$ each with size $(m–1)\times (m+1)$ and four smoothing parameters λ_x, λ_y, μ_x, μ_y which in BSSE are parameterized through the tension parameters $\tau_{\lambda x}$, $\tau_{\lambda y}$, $\tau_{\mu x}$, $\tau_{\mu y}$ with their values in [0, 1) (Malamos and Koutsoyiannis, 2016a).

The BSSE estimate ($\boldsymbol{d}$, $\boldsymbol{e}$) minimizes the penalized least squares functional:

$$f(\boldsymbol{d}, \boldsymbol{e}) = \|\boldsymbol{z} - \widehat{\boldsymbol{z}}\|^2 + \lambda_x\, \boldsymbol{d}^\top \boldsymbol{\Psi}_x^\top \boldsymbol{\Psi}_x \boldsymbol{d} + \lambda_y\, \boldsymbol{d}^\top \boldsymbol{\Psi}_y^\top \boldsymbol{\Psi}_y \boldsymbol{d} + \mu_x\, \boldsymbol{e}^\top \boldsymbol{\Psi}_x^\top \boldsymbol{\Psi}_x \boldsymbol{e} + \mu_y\, \boldsymbol{e}^\top \boldsymbol{\Psi}_y^\top \boldsymbol{\Psi}_y \boldsymbol{e} \tag{4}$$

where all smoothing parameters are included.

Because $f(\boldsymbol{d}, \boldsymbol{e})$ is quadratic, its minimizer is linear, thus the fitted values admit the following representation for any fixed configuration of the involved parameters, $\theta = (\mathrm{mx}, \mathrm{my}, \tau_{\lambda\mathrm{x}}, \tau_{\lambda\mathrm{y}}, \tau_{\mu\mathrm{x}}, \tau_{\mu\mathrm{y}})$:

$$\widehat{\boldsymbol{z}} = \boldsymbol{A}(\theta)\, \boldsymbol{z} \tag{5}$$

where $\boldsymbol{A}(\theta)$ is a $n \times n$ symmetric smoother matrix, given by:

$$\boldsymbol{A}(\theta) = \boldsymbol{\Pi}\ \boldsymbol{T\Pi}\ \boldsymbol{\Omega}^{-1} (\boldsymbol{\Pi}\ \boldsymbol{T\Pi})^{\mathrm{T}} \tag{6}$$

where

$$\boldsymbol{\Omega} := \begin{bmatrix} \boldsymbol{\Pi}^{\mathrm{T}} \boldsymbol{\Pi} + \lambda_x \boldsymbol{\Psi}_x{}^{\mathrm{T}} \boldsymbol{\Psi}_x + \lambda_y \boldsymbol{\Psi}_y{}^{\mathrm{T}} \boldsymbol{\Psi}_y & \boldsymbol{\Pi}^{\mathrm{T}} \boldsymbol{T\Pi} \\ \boldsymbol{\Pi}^{\mathrm{T}} \boldsymbol{T\Pi} & \boldsymbol{\Pi}^{\mathrm{T}} \boldsymbol{T}^{\mathrm{T}} \boldsymbol{T\Pi} + \mu_x \boldsymbol{\Psi}_x{}^{\mathrm{T}} \boldsymbol{\Psi}_x + \mu_y \boldsymbol{\Psi}_y{}^{\mathrm{T}} \boldsymbol{\Psi}_y \end{bmatrix} \tag{7}$$

The above formulation denotes that the geometry of the BSSE surfaces is determined by the parameter set, θ, that includes both the grid resolution mx, my and the four tension parameters. Malamos and Koutsoyiannis (2016b) stated that larger values of mx, my (i.e. finer grid) allow high complexity, but their increase beyond a certain point, along one direction, results in almost constant values of the smoothing parameters that refer to the opposite axis.

On the other hand, smaller tensions permit stronger curvature and greater variance amplification at coarser grid (i.e. small numbers of mx, my). When smoothing (tension) parameters are fixed, the structural complexity of the model is fully determined by mx and my.

3.2. BSSE's stochastic model and prediction risk

We interpret the n observations, z, as a noisy realization of a true field j as:

$$\boldsymbol{z} = \boldsymbol{j} + \varepsilon \tag{8}$$

where $\boldsymbol{j}$ is the n-dimensional vector of the unknown true field and $\boldsymbol{\varepsilon}$ is the corresponding independent Gaussian noise vector with common variance σ^2 at all stations, i.e. $\boldsymbol{\varepsilon} \sim N(0, \sigma^2 \boldsymbol{I}_n)$.

For any smoother matrix $\boldsymbol{A}(\theta)$ the fitted values are:

$$\widehat{\boldsymbol{z}} = \boldsymbol{A}(\theta)\boldsymbol{z} \tag{9}$$

and the prediction error at the stations is:

$$\boldsymbol{j} - \widehat{\boldsymbol{z}} = \boldsymbol{j} - \boldsymbol{A}(\theta)\boldsymbol{z} \tag{10}$$

Thus

$$\boldsymbol{j} - \widehat{\boldsymbol{z}} = (\boldsymbol{I} - \boldsymbol{A}(\theta))\boldsymbol{j} - \boldsymbol{A}(\theta)\boldsymbol{\varepsilon} \tag{11}$$

Taking the expectation of the squared error yields the classic bias-variance decomposition, where $\boldsymbol{R}(\theta)$ is defined as the expected prediction error, i.e. the risk corresponding to the BSSE smoother $\boldsymbol{A}(\theta)$:

$$\boldsymbol{R}(\theta) = \frac{1}{n}\mathbb{E}[\| \boldsymbol{j} - \widehat{\boldsymbol{z}}\|^2]. \tag{12}$$

By inserting the decomposition:

$$\| \boldsymbol{j} - \widehat{\boldsymbol{z}}\|^2 = \| (\boldsymbol{I} - \boldsymbol{A}(\theta))\boldsymbol{j} - \boldsymbol{A}(\theta)\boldsymbol{\varepsilon}\|^2 \tag{13}$$

while expanding and taking expectations by considering that the cross term has expectation zero because $\mathbb{E}(\boldsymbol{\varepsilon}) = 0$, we get:

$$\boldsymbol{R}(\theta) = \frac{1}{n}\| (\boldsymbol{I} - \boldsymbol{A}(\theta))\boldsymbol{j}\|^2 + \frac{1}{n}\mathbb{E}[\| \boldsymbol{A}(\theta)\boldsymbol{\varepsilon}\|^2] \tag{14}$$

Since $\boldsymbol{\varepsilon} \sim N(0, \sigma^2 \boldsymbol{I}_n)$ we have:

$$\mathbb{E}[\| \boldsymbol{A}(\theta)\boldsymbol{\varepsilon}\|^2] = \mathbb{E}[\boldsymbol{\varepsilon}^\top \boldsymbol{A}(\theta)^\top \boldsymbol{A}(\theta)\boldsymbol{\varepsilon}] = \sigma^2 \mathrm{trace}(\boldsymbol{A}(\theta)^\top \boldsymbol{A}(\theta)) \tag{15}$$

And finally:

$$\boldsymbol{R}(\theta) = \frac{1}{n}\| (\boldsymbol{I} - \boldsymbol{A}(\theta))\boldsymbol{j}\|^2 + \frac{1}{n}\sigma^2 \mathrm{trace}(\boldsymbol{A}(\theta)^\top \boldsymbol{A}(\theta)) \tag{16}$$

Eq. (16) includes the bias term $\|(\boldsymbol{I} - \boldsymbol{A}(\theta))\boldsymbol{j}\|^2$, that decreases when the two surfaces (i.e. d and e) become more complex, i.e. larger m, and the variance amplification term σ^2 trace $(\boldsymbol{A}(\theta)^{\mathrm{T}}\boldsymbol{A}(\theta))$. This variance amplification term is not the same as the estimated residual variance of the data. It reflects the smoother's sensitivity to noise and increases as the smoother becomes more complex.

Generalized cross-validation (GCV) is defined as:

$$\mathrm{GCV}(\theta; \boldsymbol{z}) = \frac{\frac{1}{n}\| (\boldsymbol{I} - \boldsymbol{A}(\theta))\boldsymbol{z}\|^2}{\left[\frac{1}{n}\mathrm{trace}(\boldsymbol{I} - \boldsymbol{A}(\theta))\right]^2} \tag{17}$$

Golub et al. (1979) and Wahba (1990) showed that under the model of Eq. (8), the expectation of GCV matches the prediction risk:

$$\mathbb{E}[\mathrm{GCV}(\theta; \boldsymbol{z})] \cong \boldsymbol{R}(\theta) \tag{18}$$

Thus, GCV is an approximately unbiased estimator of the true prediction risk $\boldsymbol{R}(\theta)$, so minimizing GCV approximates minimizing true prediction error. In this light, the previously mentioned tradeoff between bias and variance amplification governs the GCV optimum, thus the optimal grid.

Furthermore, for a given smoother $\boldsymbol{A}$ and data vector $\boldsymbol{z}$, the residuals are (Malamos and Koutsoyiannis, 2018; Malamos et al., 2025):

$$\boldsymbol{r} = (\boldsymbol{I} - \boldsymbol{A})\boldsymbol{z} \tag{19}$$

Thus, the residual sum of squares is:

$$\mathrm{RSS} = \| \boldsymbol{r}\|^2 \tag{20}$$

and the effective degrees of freedom for the error are:

$$\nu = \mathrm{trace}(\boldsymbol{I} - \boldsymbol{A}) \tag{21}$$

3.3. Outliers correction and BSSE geometry

According to Malamos et al. (2025), the estimator of the noise variance is:

$$\widehat{\sigma}^2 = \frac{\mathrm{RSS}}{\nu} \tag{22}$$

and the posterior covariance matrix of the fitted values is:

$$\boldsymbol{\Sigma} = \widehat{\sigma}^2(\boldsymbol{I} - \boldsymbol{A}) \tag{23}$$

Thus, the BSSE credible interval at station i is:

$$\widehat{z}_i \pm t_{\alpha/2,\nu}\sqrt{\widehat{\sigma}^2(1 - a_{ii})} \tag{24}$$

where $a_{ii}$is the i-th diagonal element of $\boldsymbol{A}$.
Points with:

$$|z_i - \widehat{z}_i| > t_{\alpha/2,\nu}\sqrt{\widehat{\sigma}^2(1 - a_{ii})} \tag{25}$$

are credible interval outliers, indicating inconsistency with the structured bilinear surface field.
From Eqs. (16)-(22), it becomes clear that GCV (Eq. 17) is associated to RSS and the noise variance as:

$$\mathrm{GCV} = \frac{n\mathrm{RSS}}{\nu^2} = \frac{n}{\nu}\,\widehat{\sigma}^2 \tag{26}$$

So, when the dataset contains outliers, both RSS and $\widehat{\sigma}^2$are inflated leading to a higher GCV score and a preference for coarser grids. After the outliers are removed, both RSS and the estimated noise variance $\widehat{\sigma}^2$ decrease. Through Eq. (26), this reduction directly lowers the GCV.

At the same time, the weight on the variance term of Eq. (16) becomes smaller, reducing the penalty related to model complexity. As a result, when the GCV is reoptimized on the outlier-corrected dataset, the bias-variance tradeoff shifts towards more flexible surfaces. Thus, the BSSE surfaces geometry expressed through the bilinear grid (mx, my) results in an increase in mx or my, and/or more relaxed tensions along the direction of meaningful variability. In the case of fixed tension parameters, this typically leads to the selection of a finer bilinear grid, i.e. increasing mx or my or both.

3.4. Evaluation criteria

The error metrics used for the evaluation of the methodology's performance are the mean bias error (MBE), the mean absolute error (MAE) and the root mean square error (RMSE). In addition, we computed the Nash-Sutcliffe Efficiency (NSE; Nash and Sutcliffe, 1970), the coefficient of determination (r^2) and the normalized root mean square error (NRMSE) (Willmott, 1981, 1982; Tegos et al., 2017; Malamos and Koutsoyiannis, 2018; Koutsoyiannis, 2025):

$$\mathrm{MBE} = \frac{1}{n}\sum_{i=1}^{n}(p_i - o_i) \tag{27}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|p_i - o_i| \tag{28}$$

$$\mathrm{RMSE} = \left[\frac{1}{n}\sum_{i=1}^{n}(p_i - o_i)^2\right]^{1/2} \tag{29}$$

$$r^2 = \left(\frac{\sum_{i=1}^{n}(o_i - \overline{o})\quad(p_i - \overline{p})}{\sqrt{\sum_{i=1}^{n}(o_i - \overline{o})^2}\ \sqrt{\sum_{i=1}^{n}(p_i - \overline{p})^2}}\right)^2 \tag{30}$$

$$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{n}(p_i - o_i)^2}{\sum_{i=1}^{n}(\overline{o} - o_i)^2} \tag{31}$$

$$\mathrm{NRMSE} = \frac{\left[\frac{1}{n}\sum_{i=1}^{n}(p_i - o_i)^2\right]^{1/2}}{\overline{o}} \tag{32}$$

where n is the number of predictions / observations, o_i is the i^{th} observation, p_i is the i^{th} prediction, $\overline{o}$ is the observations average while $\overline{p}$ is the predictions average.

4. Results

The complete list of lower outliers removed from the analysis is reported in Table 2. It contains station ID, geographic location, time period covered by the time series and number of timescales for which the station is considered an outlier (on a total of 5).

In almost all cases, the detected outliers correspond to stations that are no longer active and are located in close proximity to either recently installed gauges or older stations that are still operating. This spatial configuration suggests that the anomalous behavior in the mean values is unlikely to be driven by local geomorphological factors, but rather by issues related to different time periods considered and thus different extremes.

A representative example is station 12050, which was active from 1989 to 2006 and provides 15 years of data. The mean value of its 1-hour annual maxima is 28.6 mm. However, the nearest station, 12049, active from 2001 to 2020 with 19 years of observations, records a substantially higher mean value of 40.8 mm for the same timescale. Given the short distance between the two gauges, such a large discrepancy is difficult to justify on purely meteorological grounds and likely reflects differences in record length and time periods covered.

A similar situation can be observed, e.g., for station 12153, which operated from 1959 to 1973 and provides 13 years of data, with a mean 1-hour extreme of 27.7 mm. This station is located close to two more recent and longer-operating gauges: station 12113 (1996–2020, 24 years of data), with a mean value of 41.2 mm, and station 12130 (2004–2020, 17 years), with a mean value of 38.3 mm. Again, the marked difference in average extreme values between neighboring stations suggests that the shorter and older record may not be fully representative of the local precipitation regime.

This difference in the mean values is confirmed by Mazzoglio et al. (2025b). In this work, the authors detected positive trends in rainfall extremes of sub-daily timescales, especially in areas close to the reliefs. This result provides a plausible explanation for the systematic discrepancies observed between older, shorter records and more recent stations operating in the same area.

These examples reinforce the idea that many of the identified outliers are associated with legacy stations characterized by limited record length or outdated measurement setups. The BSSE framework thus proves useful not only for spatial interpolation but also for highlighting potential inconsistencies within historical observational networks.

The outlier-correction procedure resulted in changes in the estimated surfaces exceeding 5% relative to the original ones, in the immediate vicinity of the identified outlier with the magnitude of the differences gradually decreasing with distance (Fig. 2). Negative differences are also observed, indicating that the recalibration affected the overall surface geometry and not only the areas directly adjacent to the outliers. The largest positive and negative differences are observed for the 3 h timescale, suggesting that this scale is the most sensitive to the outlier-correction procedure. As discussed below in relation to the optimized BSSE parameters, this stronger response is mainly associated with the 3 h post-correction change in the altitude-effect smoothing structure, particularly the sharp decrease of $\tau_{\mu y}$ from the upper bound to a value close to the lower bound. The spatial patterns of differences are broadly similar across all timescales, comprising localized increases in the vicinity of the removed outliers and more spatially dispersed decreases elsewhere.

The final regionalization surfaces (Fig. 3) exhibit coherent spatial patterns across all timescales, with higher average rainfall maxima in mountainous areas and along the coastal zones. This behavior is consistent with the known climatological characteristics of the region, reflecting the influence of orographic enhancement and coastal atmospheric processes as reported in Libertino et al. (2018) and Mazzoglio et al. (2025a).

The optimal configuration of the BSSE method for the initial dataset is presented in Table 3, while the descriptive statistics of the corresponding fitted surfaces are shown in Table 4. The method identifies an optimal vertical segmentation (mx) equal to 11 for all timescales. In contrast, the horizontal segmentation (my) varies substantially across timescales, increasing from 292 for the 1 h timescale to 485 for the 24 h one. The invariance of the optimal vertical segmentation (mx = 11) across all timescales, and its relatively small value compared to my, can be understood by considering the physiographic structure of the study region. Northwestern Italy is characterized by a marked asymmetry in the spatial organization of terrain and, consequently, of precipitation. Moving from west to

Table 2
List of stations designated as lower outliers. Coordinates are expressed in WGS 84 UTM zone 32 N (EPSG:32632).

Station ID	Longitude (m)	Latitude (m)	Start year	End year	Counts as outlier
8007	462905.1	4916299	1972	1983	5
8016	429481.4	4894356	1943	1973	3
8059	407422.4	4859342	1935	1999	4
8071	514668.7	4916211	1934	1948	3
8133	450183.3	4894127	1972	2010	5
8145	477624	4927540	1937	1986	4
8175	435973.8	4906697	1940	1995	4
8179	549161.5	4918234	1954	1977	5
8187	545321.5	4897835	1959	1988	3
12050	455382	5091683	1989	2006	5
12126	387337	4993290	1951	1996	3
12153	390278	5011770	1959	1973	3
12181	393150	5026530	1955	1996	4
17119	500062.8	4923602	1933	1977	5

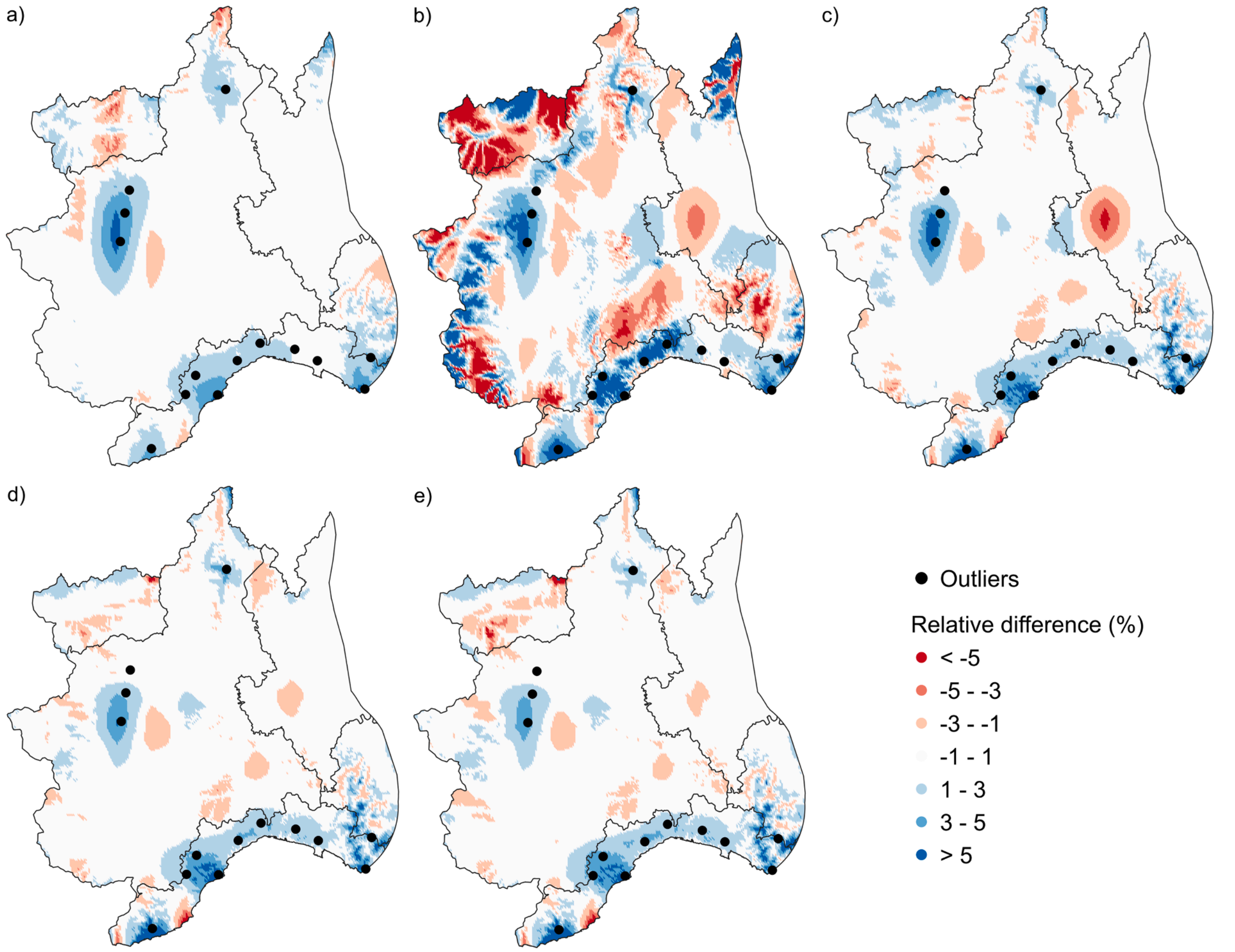


Fig. 2. Relative differences between the outlier-free map and the original one (calculated with respect to the original one). Timescales: 1 (a), 3 (b), 6 (c), 12 (d) and 24 h (e).

east, the landscape transitions gradually from the Alps to a hilly area to the western Po Plain. This west-to-east gradient is relatively smooth and monotonic, with precipitation decreasing progressively as orographic influence diminishes. A moderate number of horizontal segments (mx = 11) is therefore sufficient to capture this gentle spatial trend. The situation along the vertical axis (south to north) is considerably more complex. The southern boundary of the domain is marked by the Ligurian Alps and the proximity to the Mediterranean coast, both of which are associated with high precipitation extremes driven by orographic uplift and maritime moisture availability. Moving northward, precipitation decreases across the lowland and hilly areas, before increasing again toward the Alpine arc in the northern part of the domain. This non-monotonic, multi-peaked structure along the vertical direction requires a larger number of segments (my) to be adequately resolved, explaining why the optimal vertical segmentation is substantially higher than the horizontal one. The fact that mx remains stable at 11 across all timescales further suggests that the west-to-east gradient is a robust and persistent feature of the precipitation climatology of the region, largely independent of the temporal aggregation scale. The greater variability in my across timescales, by contrast, reflects the more complex and scale-sensitive interplay between orographic, convective, and synoptic forcing along the south-to-north transect.

This behavior is also consistent with the bias–variance mechanism underlying the GCV optimization. At shorter timescales, the stronger influence of measurement noise and the erratic nature of fine-scale precipitation increases the variance amplification term in the GCV expression, thereby favoring smoother surfaces. At longer timescales, improved spatial coherence and reduced relative noise allow GCV to select more flexible surfaces and refine the grid, reflected in larger values of my. Interestingly, the 12 h and 24 h timescales yield identical spatial segmentation, suggesting that similar large-scale spatial structures dominate the extreme rainfall generating process at these timescales. Overall, the scale-dependent variation in horizontal segmentation is consistent with the transition from noise-dominated fine timescales to more spatially coherent large-scale rainfall structures as temporal aggregation increases.

The outlier-corrected results are consistent with the previous observations and provide further insight into the impact of the outlier-removal process (Table 5). Following correction, the optimal horizontal segmentation (mx) remains unchanged across all timescales,

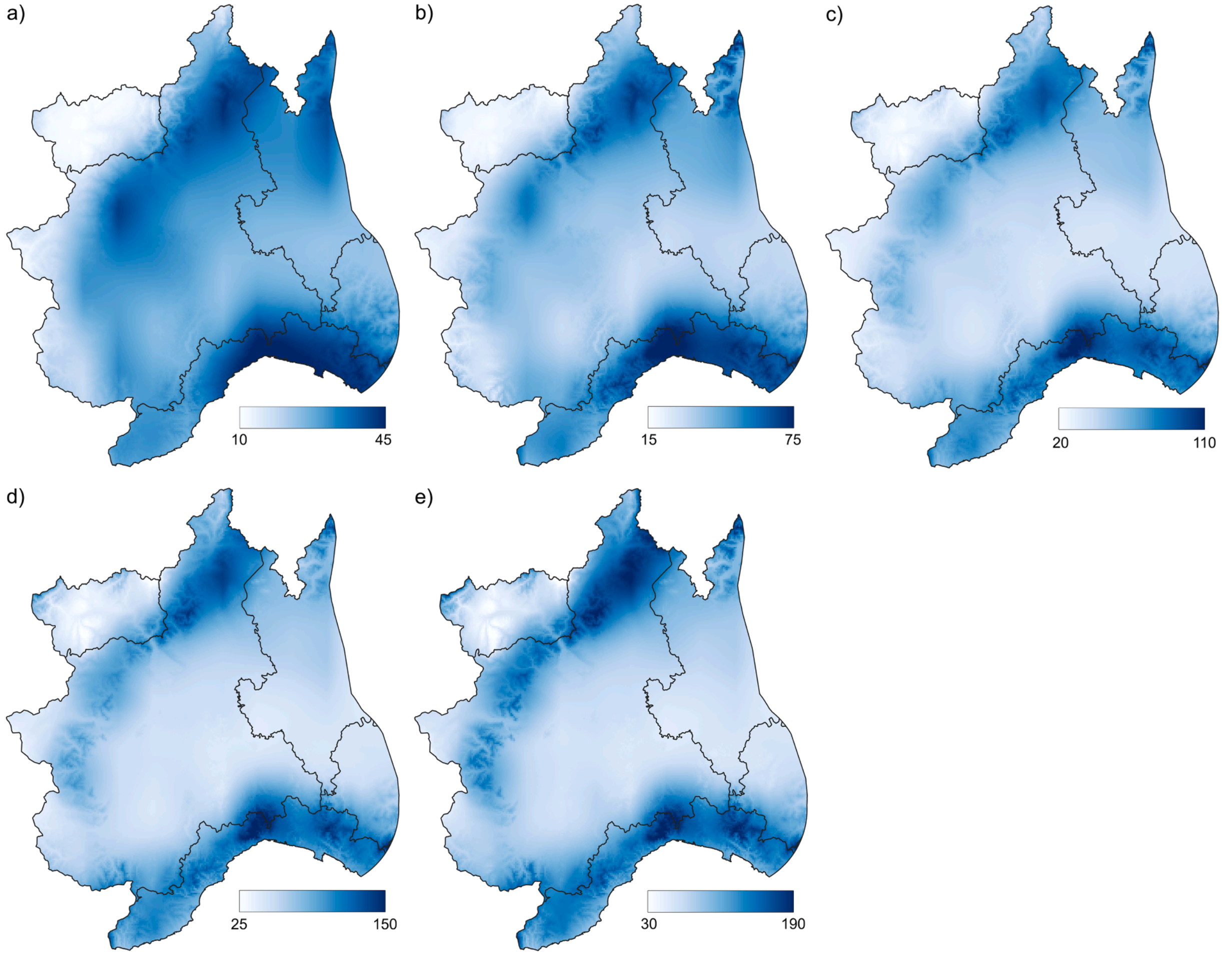


Fig. 3. Outlier-free mean values of the annual maxima (mm). Timescales: 1 (a), 3 (b), 6 (c), 12 (d) and 24 h (e).

Table 3
BSSE optimized parameters for the 1, 3, 6, 12 and 24 h timescales before CI correction.

Timescale (h)	Segments mx	Segments my	$\tau_{\lambda x}$	$\tau_{\lambda y}$	$\tau_{\mu x}$	$\tau_{\mu y}$	GCV	$\hat{\sigma}^2$
1	11	292	0.001	0.001	0.99	0.99	24.3	22.0
3	11	368	0.001	0.001	0.99	0.99	47.0	41.6
6	11	449	0.001	0.001	0.887553	0.005902	77.6	65.4
12	11	485	0.001	0.001	0.310983	0.001	148.1	122.6
24	11	485	0.001061	0.001	0.090691	0.001	292.0	240.8

Table 4
BSSE descriptive statistics values for the 1, 3, 6, 12 and 24 h timescales before CI correction to the station locations.

Timescale (h)	Maximum (mm)	Minimum (mm)	Mean (mm)	Median (mm)	σ_p (mm)
1	49.5	11.0	30.3	30.4	8.5
3	79.1	18.1	45.3	43.9	13.8
6	106.9	26.5	59.6	57.2	18.8
12	147.9	34.7	79.5	76.4	25.3
24	197.5	44.5	105.6	103.0	33.7

whereas the vertical segmentation (my) increases consistently for all timescales except the 1 h case, for which the same configuration is retained.

It should be emphasized that after outlier removal the BSSE parameters were re-optimized using the GCV criterion, allowing the

Table 5
BSSE optimized parameters for the 1, 3, 6, 12 and 24 h timescales after CI correction.

Timescale (h)	Segments mx	Segments my	$\tau_{\lambda x}$	$\tau_{\lambda y}$	$\tau_{\mu x}$	$\tau_{\mu y}$	GCV	$\hat{\sigma}^2$
1	11	292	0.001	0.001	0.99	0.99	22.5	20.3
3	11	401	0.001	0.001	0.968316	0.013795	41.9	35.7
6	11	485	0.001	0.001	0.99	0.004991	68.5	57.0
12	11	521	0.001	0.001	0.332432	0.001	134.1	109.6
24	11	521	0.001	0.001	0.078663	0.001	269.0	218.7

bias-variance tradeoff to adjust to the revised dataset. Relative to the pre-correction optima (Table 3), GCV decreases systematically at all timescales by approximately 7–12%, indicating a lower estimated expected prediction error, consistent with the theoretical bias-variance decomposition presented in Section 2. Importantly, the estimated noise variance, $\hat{\sigma}^2$, also decreases for every timescale after correction. Within the GCV framework, this systematic reduction in the variance term weakens the penalty associated with model complexity, permitting the selection of a finer spatial grid (increased my) and a clearer representation of the underlying spatial structure of extreme rainfall.

The absence of segmentation changes at the 1 h timescale likely reflects the persistence of high fine-scale variability at this aggregation level, where noise-related effects remain sufficiently strong to constrain further grid refinement even after outlier removal. Finally, the stability of the vertical segmentation across timescales suggests that the dominant large-scale spatial structure along this axis is robust and relatively insensitive to temporal aggregation or data correction.

Beyond segmentation, Tables 3 and 5 clarify how smoothing is distributed between the rainfall surface itself and the altitude-related explanatory component. In the adopted parameterization, the tension parameters control smoothness in each direction, with low values implying a rougher surface and values near 0.99 implying a smoother surface. Under this interpretation, the λ tensions ($\tau_{\lambda x}$ and $\tau_{\lambda y}$), which control the curvature of the bilinear surface *d*, are consistently selected at the lower bound ($\approx$0.001) for all timescales both before and after correction. This indicates that, within the explored search space, the GCV optimum persistently favors a comparatively rough representation of the rainfall field, and that the refinement observed after correction is expressed mainly through grid segmentation, notably the systematic increase in my, rather than by increasing the smoothness of the rainfall surface via higher τ_{λ}.

In contrast, the μ tensions ($\tau_{\mu x}$, $\tau_{\mu y}$), which control the curvature of the bilinear surface *e*, i.e. the altitude-effect component, exhibit substantial and timescale dependent changes after outlier removal, implying that the correction step primarily alters how the model regularizes the altitude contribution. At 1 h, both $\tau_{\mu x}$ and $\tau_{\mu y}$ remain near 0.99 before and after correction, indicating a very smooth altitude effect in both directions. At 3 h, $\tau_{\mu y}$ drops sharply after correction while $\tau_{\mu x}$ remains high, implying that the post-correction optimum allows a much rougher altitude effect in the horizontal direction while maintaining a smooth altitude effect in the vertical direction. For 6–24 h, $\tau_{\mu y}$ remains very low, indicating a persistently rough altitude effect along the horizontal axis, whereas $\tau_{\mu x}$ adjusts in a timescale dependent and non-monotonic manner since it increases at 6–12 h and decreases at 24 h. Overall, these patterns suggest that after CI-based screening the model reallocates smoothness between directions primarily within the altitude-effect component, while the rainfall surface itself remains governed by the segmentation choices selected under GCV.

Because the removed outliers predominantly correspond to lower rainfall values, the recalibrated surfaces exhibit moderate increases in the estimated rainfall maxima (Table 6).

The outlier correction leads to systematic improvements in the performance assessment statistics across all timescales (Table 7), although the initial performance was already satisfactory (NSE $>$ 0.8). In relative terms, the largest improvement is observed at the 3 h timescale, with NSE and r^2 increasing by approximately 3% and NRMSE decreasing by nearly 9%, while improvements at the remaining timescales are more moderate, with NSE and r^2 increasing by approximately 2% and NRMSE decreasing by roughly 4–7%.

Comparing the dataset statistics (Table 1) with the fitted statistics at station locations before and after CI correction (Tables 4 and 6) shows that BSSE reproduces the bulk of the observed distribution with fidelity, while applying only limited smoothing that primarily moderates extremes. Across all timescales, the fitted means are essentially identical to the dataset means, while the fitted medians remain very close to the dataset medians. This indicates that the modelling does not induce excessive shrinkage of central tendency.

As expected under a smoothing framework, the fitted maxima at station locations are lower than the raw dataset maxima, and the fitted standard deviations are also reduced relative to the dataset, reflecting moderation of the upper tail rather than distortion of the central mass. Notably, after CI correction the fitted standard deviation increases slightly and moves systematically closer to the observed dataset standard deviation at every timescale, indicating that the CI-based screening does not introduce systematic bias and that the BSSE fit remains strongly anchored to the bulk of the observations, rather than being driven by excessive smoothing toward a

Table 6
BSSE descriptive statistics values for the 1, 3, 6, 12 and 24 h timescales after CI correction to the stations locations.

Timescale (h)	Maximum (mm)	Minimum (mm)	Mean (mm)	Median (mm)	σ_p (mm)
1	50.5	10.9	30.4	30.4	8.6
3	80.9	18.5	45.3	43.7	14.2
6	108.9	26.4	59.6	56.8	19.1
12	150.2	34.6	79.5	75.3	25.7
24	198.8	44.3	105.7	102.3	34.3

Table 7
Performance assessment statistics for the regionalization of the initial and the CI corrected dataset for all timescales.

Surface statistics (All data)	Initial					CI corrected				
	1 h	3 h	6 h	12 h	24 h	1 h	3 h	6 h	12 h	24 h
MBE (mm)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
MAE (mm)	3.3	4.3	5.2	7.2	9.9	3.2	4.0	5.0	6.8	9.4
RMSE (mm)	4.5	6.1	7.4	10.1	14.1	4.3	5.5	6.9	9.5	13.3
NSE	0.793	0.845	0.870	0.868	0.857	0.811	0.873	0.889	0.884	0.873
r^2	0.794	0.845	0.870	0.868	0.858	0.812	0.873	0.889	0.885	0.874
NRMSE (%)	8.1	6.8	5.9	5.9	6.6	7.8	6.2	5.5	5.5	6.2

grand mean.

Table 7 summarizes the in-sample performance at station locations and shows that CI correction improves the regionalization consistently across all timescales. Relative to the initial surfaces, the CI-corrected surfaces yield lower errors for every timescale. Both MAE and RMSE decrease and their reductions correspond to improvements of about 3.0–7.0% in MAE and 4.4–9.8% in RMSE. Importantly, MBE remains essentially zero across timescales, indicating that the correction step does not introduce systematic bias and that the improvement is primarily due to reduced scatter around observations.

Table 8 provides the corresponding LOOCV assessment and confirms that the improvements persist under a predictive setting. As expected, LOOCV errors are larger and skill scores are lower than in-sample values because each station is withheld in turn. Nevertheless, CI correction again improves LOOCV metrics at every timescale. Bias remains negligible as well (MBE $\leq$ 0.2 mm), reinforcing that the CI-based screening primarily reduces variance-related error components and improves generalization rather than merely polishing the in-sample fit.

Another interesting observation is that in both Tables 7 and 8, NSE and r^2 are numerically almost identical. According to Koutsoyiannis (2025), when $\overline{p} \approx \overline{o}$, thus bias is negligible, NSE is maximized for a given r when $\sigma_p \approx r\,\sigma_o$, with its maximum value being equal to r^2. Therefore, the near equality between NSE and r^2, together with the very small MBE values, indicate that the regionalized surfaces operate very close to the maximum efficiency theoretically allowed by the achieved correlation. This means that regionalization incurs essentially no additional NSE penalty due to bias or variance mis-scaling beyond that implied by the correlation structure itself.

5. Discussion

The results presented in this study offer insights into both the methodological capabilities of the BSSE framework and the characteristics of the observational network used for the regionalization of sub-daily rainfall extremes in northwestern Italy. Few interconnected themes emerge from the analysis of the results: the nature and interpretation of the identified outliers, the trade-off between spatial coverage and temporal reliability, the scale-dependent spatial structure of the regionalized surfaces, and the implications of embedding data screening within the interpolation procedure.

5.1. Interpretation of the identified outliers

The stations identified as low outliers correspond to gauges often no longer active, often with short records or early periods of operation. In some cases, older time series are also quite short due to station relocation. Their anomalously low mean annual maxima can be attributed to two concurring factors. First, short time series are inherently more sensitive to sampling variability: a record spanning only 10–15 years may not capture high-intensity episodes that occur at longer recurrence intervals, leading to a systematic underestimation of mean extreme rainfall depths. In contrast, longer and more recent records are more likely to include such events, resulting in higher estimated averages. Second, temporal non-stationarity may amplify this effect. Mazzoglio et al. (2025b) documented increasing trends in sub-daily rainfall extremes across the region, particularly in areas influenced by orographic forcing. Under such evolving climatic conditions, older stations that ceased operation decades ago may represent a precipitation regime characterized by lower intensities, appearing anomalously low not because of measurement errors but because they do not capture the ongoing intensification of short-duration rainfall. Given their spatial proximity to other stations still operating under similar physiographic

Table 8
Performance assessment statistics of LOOCV for the regionalization of the initial and the CI corrected dataset for all timescales.

Surface statistics (LOOCV)	Initial					CI corrected				
	1 h	3 h	6 h	12 h	24 h	1 h	3 h	6 h	12 h	24 h
MBE (mm)	0.0	0.0	0.1	0.1	0.2	0.0	0.1	0.1	0.1	0.2
MAE (mm)	3.6	4.8	6.2	8.7	12.1	3.5	4.6	6.0	8.4	11.7
RMSE (mm)	4.9	6.8	8.7	12.1	17.1	4.7	6.4	8.2	11.7	16.6
NSE	0.751	0.807	0.821	0.809	0.790	0.772	0.828	0.841	0.825	0.804
r^2	0.751	0.807	0.821	0.809	0.791	0.772	0.828	0.841	0.825	0.805
NRMSE (%)	8.9	7.6	6.9	7.0	8.0	8.6	7.2	6.6	6.8	7.7

conditions, genuine local climatic differences alone cannot explain the observed discrepancies.

5.2. The trade-off between spatial coverage and temporal reliability

These considerations raise a broader question about data selection and the trade-off between spatial coverage and temporal reliability when performing spatial interpolation, extreme value analysis, or uncertainty assessment. Restricting the analysis to stations with sufficiently long, homogeneous, and recent records improves the robustness of spatial modeling, as short and obsolete series can introduce noise and bias the estimation of spatial gradients. However, older and shorter records retain value for the historical reconstruction of intense events captured only by temporary gauges. The key issue is therefore not the quantity of data per se, but its representativeness and consistency with the temporal scale of the analysis. Therefore, an informed balance must be achieved between maximizing spatial coverage and ensuring temporal reliability. The identification of low outliers through the BSSE framework provides a useful diagnostic step in this process, enabling the distinction between stations that contribute meaningful climatic information and those whose limited or outdated records may distort the spatial signal. This is particularly relevant for applications such as design rainfall estimation, flood hazard assessment, and infrastructure planning, where underestimation of extremes may have significant safety implications.

5.3. Scale dependence of the regionalized surfaces

These data-related considerations are directly reflected in the behaviour of the regionalization model. The BSSE framework responds not only to the spatial distribution of observations, but also to their representativeness and the underlying physical processes governing rainfall extremes. In this sense, the scale-dependent structure identified by the model should be viewed as the combined outcome of sampling limitations, uncertainty, and the physical mechanisms controlling precipitation across different temporal scales. The optimal spatial configuration identified by BSSE reflects the physical mechanisms controlling extreme rainfall across temporal scales. The BSSE framework favours a more parsimonious segmentation for the 1 h rainfall extremes, whose spatial surface is governed by higher uncertainty, and progressively more refined segmentation as the temporal scale of precipitation increases up to 24 h. These model outputs are consistent with the known climatology of rainfall extremes, with the 1 h scale being largely governed by convective processes that exhibit strong small-scale variability and are typically associated with more erratic spatial structures (Mazzoglio et al., 2022, 2023; Iliopoulou et al., 2024; Malamos et al., 2025). As the aggregation timescale increases, rainfall extremes become progressively more spatially coherent, allowing the model to resolve finer spatial structures in the regionalization surfaces. The convergence of the spatial segmentation for the 12 h and 24 h timescales further suggests that similar large-scale precipitation mechanisms dominate extreme rainfall generation at these timescales across the region. Similarly, the effect of orographic enhancement of precipitation is evident across all timescales, although it becomes more pronounced at longer timescales, particularly at the 24 h scale. The increasing influence of orography on longer duration rainfall extremes has already been pointed out by Libertino et al. (2018) over Piedmont region. This behaviour is consistent with the increasing role of large-scale atmospheric circulation and orographic lifting in longer-duration precipitation events, while short-duration extremes are more strongly controlled by localized convective dynamics (Mazzoglio et al., 2022, 2023). A notable concentration of identified low outliers in coastal areas is also consistent with the recent intensification of rainfall extremes documented in those locations (Mazzoglio et al., 2025b), and the overall spatial patterns of the regionalized surfaces are consistent with climatological analyses reported for the region by Mazzoglio et al. (2023), (2025a), (2025b).

5.4. Integration of interpolation and data quality assessment

The connection between data, process understanding, and model structure is central to the proposed framework, which not only enhances spatial interpolation but also embeds data quality assessment within the regionalization procedure. The BSSE framework features that spatial interpolation and data screening are not treated as separate steps but are linked through the same model structure. In most regionalization studies, potentially problematic observations are identified a priori using external quality-control procedures, and interpolation is then applied to the filtered dataset. Here, by contrast, the identification of inconsistent stations follows directly from the stochastic formulation of BSSE, through the fitted field, the estimated noise variance, and the uncertainty structure summarized by the corresponding credible intervals. The screening criterion is therefore not externally imposed but derived from the same procedure that governs the interpolation itself. In this sense, observations are not flagged simply because they differ from neighbouring values, but because they are not supported by the fitted spatial field within the uncertainty bounds implied by the model. This is particularly relevant for extreme rainfall, where large local differences may reflect either genuine spatial variability or limited temporal representativeness. Under such conditions, purely descriptive or residual-based screening may be difficult to interpret. The present results suggest that the CI-based BSSE framework provides a useful way to distinguish between these cases. Stations identified as inconsistent contribute to higher estimated noise variance and less efficient model configurations, whereas their removal leads to lower GCV, reduced σ^2, and improved predictive performance. The fact that these improvements are observed not only for the full data fit but also under LOOCV indicates that the screening step enhances generalization rather than simply improving calibration. A further implication is that uncertainty plays an operational role in the proposed framework. Rather than being used only for post hoc evaluation, it contributes directly to the assessment of data consistency through the stochastic structure of the model. This provides a reproducible and interpretable basis for identifying observations that warrant further scrutiny. The method therefore has value not only as a regionalization tool, but also as a means of assessing the effective usability of heterogeneous gauge records for spatial analysis. At the same time, such screening should be viewed as complementary to, rather than a substitute for, climatological interpretation and station metadata. Within this scope, the present study shows that BSSE can support both improved interpolation and model-based quality assessment in a unified way.

6. Conclusions

In the context of the present study, the BSSE framework was implemented not only for the spatial interpolation of sub-daily rainfall extremes, but also as a coherent stochastic framework for uncertainty quantification and model-based data screening. By combining a bilinear surface with elevation as a denser explanatory variable and deriving credible intervals directly from the fitted model, the method allows interpolation and quality control to be addressed within the same procedure rather than as separate steps.

Applied to rainfall maxima in northwestern Italy, the framework identified 14 stations behaving as low outliers relative to the regional spatial-explanatory signal (Table 2). These stations predominantly correspond to legacy gauges with short records or early periods of operation, whose anomalously low statistics likely reflect both sampling limitations and the intensification of rainfall extremes documented in the region in recent decades.

Once the effective noise of the dataset was reduced, the optimal BSSE geometry changed. Optimizing GCV again led to a more flexible surface expressed mainly through a systematic increase in the horizontal segmentation my, while mx remained unchanged, together with changes in the tension parameters of the explanatory component. Thus, this removal did not merely affect a few local observations, but it altered the optimal balance between smoothness and spatial detail and, therefore, the spatial structure of the regionalization.

Furthermore, the removal of the outlier stations led to a systematic improvement in model performance. After correction, the BSSE results exhibited lower GCV values, lower estimated noise variance, and improved error metrics, while the corresponding leave-one-out results also indicated better predictive robustness. Thus, the correction step did not simply improve the apparent fit to the calibration data but produced better regionalized rainfall surfaces.

The analysis also highlighted a clear dependence on temporal scale. In this light, shorter timescales remained more irregular and noise-dominated, whereas longer timescales became progressively more spatially coherent and support finer regionalization structures. So, the proposed framework does more than improve interpolation accuracy since it also provides useful means through which the spatial organization of extreme rainfall can be interpreted across timescales.

Finally, the results suggest that in this type of problems, more data do not necessarily imply better regionalization. A larger dataset may include stations that are inconsistent with the spatial process being modeled and may therefore inflate noise, weaken model efficiency, and distort the inferred gradients. Conversely, a smaller but more internally coherent dataset can yield a more reliable regionalized surface. This should not be read as a general argument for discarding data, but as evidence that representativeness and consistency may be more important than sheer data volume when regionalizing rainfall extremes, especially in regions with diverse rainfall regimes subject to long-term trends. In operational use, this screening should be interpreted as a model-based diagnostic rather than as an automatic deletion rule, and should be considered together with station metadata, record length, period of operation and hydrological reasoning.

CRediT authorship contribution statement

Theano Iliopoulou: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Formal analysis. **Paola Mazzoglio:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation. **Nikolaos Malamos:** Writing – review & editing, Writing – original draft, Supervision, Software, Methodology, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

References

Allamano, P., Claps, P., Laio, F., Thea, C., 2009. A data-based assessment of the dependence of short-duration precipitation on elevation. Phys. Chem. Earth 34. https://doi.org/10.1016/j.pce.2009.01.001.

Avanzi, F., De Michele, C., Gabriele, S., Ghezzi, A., Rosso, R., 2015. Orographic signature on extreme precipitation of short durations. J. Hydrometeorol. 16 (1), 278–294. https://doi.org/10.1175/JHM-D-14-0063.1.

Brunsdon, C., McClatchey, J., Unwin, D.J., 2001. Spatial variations in the average rainfall–altitude relationship in Great Britain: an approach using geographically weighted regression. Int. J. Climatol. 21, 455–466. https://doi.org/10.1002/joc.614.

Buishand, T.A., de Haan, L., Zhou, C., 2008. On spatial extremes: with application to a rainfall problem. Ann. Appl. Stat. 2 (2), 624–642. ⟨http://www.jstor.org/stable/30244220⟩.

Casson, E., Coles, S., 1999. Spatial regression models for extremes. Extremes 1, 449–468. https://doi.org/10.1023/A:1009931222386.

Claps, P., Ganora, D., Mazzoglio, P., 2022. Rainfall regionalization techniques. In: Morbidelli, R. (Ed.), Rainfall. Elsevier, pp. 327–350. https://doi.org/10.1016/B978-0-12-822544-8.00013-5.

Coles, S.G., Tawn, J.A., 1996. A Bayesian analysis of extreme rainfall data. Appl. Stat. 45, 463. https://doi.org/10.2307/2986068.

Crespi, A., Brunetti, M., Lentini, G., Maugeri, M., 2018. 1961–1990 high-resolution monthly precipitation climatologies for Italy. Int. J. Climatol. 38, 878–895. https://doi.org/10.1002/joc.5217.

Daly, C., Gibson, W.P., Taylor, G.H., Johnson, G.L., Pasteris, P., 2002. A knowledge-based approach to the statistical mapping of climate. Clim. Res. 22 (2), 99–113. https://doi.org/10.3354/cr022099.

Daly, C., Neilson, R.P., Phillips, D.L., 1994. A statistical-topographic model for mapping climatological precipitation over mountainous terrain. J. Appl. Meteor. 33 (2), 140–158. https://doi.org/10.1175/1520-0450(1994)033%3C0140:ASTMFM%3E2.0.CO;2.

Di Piazza, A., Conti, F.L., Noto, L.V., Viola, F., La Loggia, G., 2011. Comparative analysis of different techniques for spatial interpolation of rainfall data to create a serially complete monthly time series of precipitation for Sicily, Italy. Int. J. Appl. Earth Obs. Geoinf. 13, 396–408. https://doi.org/10.1016/j.jag.2011.01.005.

Diodato, N., Ceccarelli, M., 2005. Interpolation processes using multivariate geostatistics for mapping of climatological precipitation mean in the Sannio Mountains (southern Italy). Earth Surf. Process. Landf. 30. https://doi.org/10.1002/esp.1126.

Dunn, P.K., 2001. Bootstrap confidence intervals for predicted rainfall quantiles. Int. J. Climatol. 21, 89–94. https://doi.org/10.1002/joc.596.

Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., Alsdorf, D., 2007. The Shuttle Radar Topography Mission. Rev. Geophys. 45. https://doi.org/10.1029/2005RG000183.

Formetta, G., Marra, F., Dallan, E., Zaramella, M., Borga, M., 2022. Differential orographic impact on sub-hourly, hourly, and daily extreme precipitation. Adv. Water Resour. 159, 104085. https://doi.org/10.1016/j.advwatres.2021.104085.

Frei, C., Schär, C., 1998. A precipitation climatology of the Alps from high-resolution rain-gauge observations. Int. J. Climatol. 18, 873–900. https://doi.org/10.1002/(SICI)1097-0088(19980630)18:8<873::AID-JOC255>3.0.CO;2-9.

Golub, G.H., Heath, M., Wahba, G., 1979. Generalized Cross-Validation as a method for choosing a good ridge parameter. Technometrics 21 (2), 215–223. https://doi.org/10.1080/00401706.1979.10489751.

Goovaerts, P., 1997. Applied geostatistics series. Geostatistics for Natural Resources Evaluation. Oxford University Press.

Goovaerts, P., 2000. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. J. Hydrol. 228, 113–129. https://doi.org/10.1016/S0022-1694(00)00144-X.

Hu, Y.M., Liang, Z.M., Li, B.Q., Yu, Z.B., 2013. Uncertainty assessment of hydrological frequency analysis using bootstrap method. Math. Probl. Eng., 724632 https://doi.org/10.1155/2013/724632.

Iliopoulou, T., Koutsoyiannis, D., Malamos, N., Koukouvinos, A., Dimitriadis, P., Mamassis, N., Tepetidis, N., Markantonis, D., 2024. A stochastic framework for rainfall intensity–time scale–return period relationships. Part II: point modelling and regionalization over Greece. Hydrol. Sci. J. 69 (8), 1092–1112.

Koutsoyiannis, D., 2025. When Are Models Useful? Revisiting the Quantification of Reality Checks. Water 17, 264. https://doi.org/10.3390/w17020264.

Libertino, A., Allamano, P., Laio, F., Claps, P., 2018. Regional-scale analysis of extreme precipitation from short and fragmented records. Adv. Water Resour. 112, 147–159. https://doi.org/10.1016/j.advwatres.2017.12.015.

Malamos, N., Iliopoulou, T., Dimitriadis, P., Koutsoyiannis, D., 2025. Comparative Analysis of IMERG Satellite Rainfall and Elevation as Covariates for Regionalizing Average and Extreme Rainfall Patterns in Greece by Means of Bilinear Surface Smoothing. Geosciences 15, 212. https://doi.org/10.3390/geosciences15060212.

Malamos, N., Koutsoyiannis, D., 2016b. Bilinear surface smoothing for spatial interpolation with optional incorporation of an explanatory variable. Part 2: Application to synthesized and rainfall data. Hydrol. Sci. J. 61 (3), 527–540. https://doi.org/10.1080/02626667.2015.1080826.

Malamos, N., Koutsoyiannis, D., 2016a. Bilinear surface smoothing for spatial interpolation with optional incorporation of an explanatory variable. Part 1: Theory. Hydrol. Sci. J. 61 (3), 519–526. https://doi.org/10.1080/02626667.2015.1051980.

Malamos, N., Koutsoyiannis, D., 2018. Field survey and modelling of irrigation water quality indices in a Mediterranean island catchment: a comparison between spatial interpolation methods. Hydrol. Sci. J. 63 (10), 1447–1467. https://doi.org/10.1080/02626667.2018.1508874.

Mazzoglio, P., Butera, I., Claps, P., 2020. I^2-RED: a massive update and quality control of the Italian annual extreme rainfall dataset. Water 12, 3308. https://doi.org/10.3390/w12123308.

Mazzoglio, P., Butera, I., Claps, P., 2023. A local regression approach to analyze the orographic effect on the spatial variability of sub-daily rainfall annual maxima. Geomat. Nat. Hazards Risk 14 (1). https://doi.org/10.1080/19475705.2023.2205000.

Mazzoglio, P., Butera, I., Alvioli, M., Claps, P., 2022. The role of morphology in the spatial distribution of short-duration rainfall extremes in Italy. Hydrol. Earth Syst. Sci. 26, 1659–1672. https://doi.org/10.5194/hess-26-1659-2022.

Mazzoglio, P., Lompi, M., Marra, F., Dallan, E., Deidda, R., Claps, P., Manfreda, S., Noto, L.V., Viglione, A., Raffa, M., Mercogliano, P., Marani, M., Caporali, E., Borga, M., 2025a. Orographic and land-sea contrast effects in convection-permitting simulations of extreme sub-daily precipitation. Weather. Clim. Extrem., 100798 https://doi.org/10.1016/j.wace.2025.100798.

Mazzoglio, P., Viglione, A., Ganora, D., Claps, P., 2025b. Mapping the uneven temporal changes in ordinary and extraordinary rainfall extremes in Italy. J. Hydrol. Reg. Stud. 58, 102287. https://doi.org/10.1016/j.ejrh.2025.102287.

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

Pardo-Igúzquiza, E., Grimes, D.I.F., Teo, C., 2006. Assessing the uncertainty associated with intermittent rainfall fields. Water Resour. Res. 42, 2004WR003740. https://doi.org/10.1029/2004WR003740.

Prudhomme, C., Reed, D.W., 1998. Relationships between extreme daily precipitation and topography in a mountainous region: a case study in Scotland. Int. J. Climatol. 18, 1439–1453. https://doi.org/10.1002/(SICI)1097-0088(19981115)18:13%3C1439::AID-JOC320%3E3.0.CO;2-7.

Prudhomme, C., Reed, D.W., 1999. Mapping extreme rainfall in a mountainous region using geostatistical techniques: a case study in Scotland. Int. J. Climatol. 19, 1337–1356. https://doi.org/10.1002/(SICI)1097-0088(199910)19:12%3C1337::AID-JOC421%3E3.0.CO;2-G.

Renard, B., Kavetski, D., Kuczera, G., Thyer, M., Franks, S.W., 2010. Understanding predictive uncertainty in hydrologic modeling: the challenge of identifying input and structural errors. Water Resour. Res. 46, 2009WR008328. https://doi.org/10.1029/2009WR008328.

Svensson, C., Jones, D.A., 2010. Review of rainfall frequency estimation methods. J. Flood Risk Manag. 3 (4), 296–313. https://doi.org/10.1111/j.1753-318X.2010.01079.x.

Tegos, A., Malamos, N., Efstratiadis, A., Tsoukalas, I., Karanasios, A., Koutsoyiannis, D., 2017. Parametric modelling of potential evapotranspiration: a global survey. Water 9, 795. https://doi.org/10.3390/w9100795.

Thornton, P.E., Running, S.W., White, M.A., 1997. Generating surfaces of daily meteorological variables over large regions of complex terrain. J. Hydrol. 190, 214–251.

Thornton, P.E., Shrestha, R., Thornton, M., Kao, S.C., Wei, Y., Wilson, B.E., 2021. Gridded daily weather data for North America with comprehensive uncertainty quantification. Sci. Data 8 (1), 190.

Wahba, G., 1990. Spline Models for Observational Data. Regional Conference Series in Applied Mathematics 59. SIAM, Philadelphia, Pennsylvania.

Willmott, C.J., 1981. On the validation of models. Phys. Geogr. 2 (2), 184–194. https://doi.org/10.1080/02723646.1981.10642213.

Willmott, C.J., 1982. Some comments on the evaluation of model performance. Bull. Am. Meteor. Soc. 63, 1309–1313. https://doi.org/10.1175/1520-0477(1982)063<1309:SCOTEO>2.0.CO;2.