

Article

Interpolation and Machine Learning Methods for Sub-Hourly Missing Rainfall Data Imputation in a Data-Scarce Environment: One- and Two-Step Approaches

Mohamed Boukdire ¹, Çağrı Alperen İnan ², Giada Varra ³ , Renata Della Morte ^{3,*}  and Luca Cozzolino ³ 

¹ Department of Civil Engineering and Architecture, University of Catania, 95123 Catania, Italy; mohamed.boukdire@phd.unict.it

² Euro-Mediterranean Center on Climate Change (CMCC), Soil and Water Systems (SOWAS) Division, Via Thomas Alva Edison, 81100 Caserta, Italy; cagri.inan@cmcc.it

³ Department of Engineering, “Parthenope” University, 80143 Naples, Italy; giada.varra@uniparthenope.it (G.V.); luca.cozzolino@uniparthenope.it (L.C.)

* Correspondence: renata.dellamorte@uniparthenope.it

Abstract

Complete sub-hourly rainfall datasets are critical for accurate flood modeling, real-time forecasting, and understanding of short-duration rainfall extremes. However, these datasets often contain missing values due to sensor or transmission failures. Recovering missing values (or filling these data gaps) at high temporal resolution is challenging due to the imbalance between rain and no-rain periods. In this study, we developed and tested two approaches for the imputation of missing 10-min rainfall data by means of machine learning (Multilayer Perceptron and Random Forest) and interpolation methods (Inverse Distance Weighting and Ordinary Kriging). The (a) direct approach operates on raw data to directly feed the imputation models, while the (b) two-step approach first classifies time steps as rain or no-rain with a Random Forest classifier and subsequently applies an imputation model to predicted rainfall depth instances classified as rain. Each approach was tested under three spatial scenarios: using all nearby stations, using stations within the same cluster, and using the three most highly correlated stations. An additional test involved the comparison of the results obtained using data from the imputed time interval only and data from a time window containing several time intervals before and after the imputed time interval. The methods were evaluated with reference to two different environments, mountainous and coastal, in Campania region (Southern Italy), under data-scarce conditions where rainfall depth is the only available variable. With reference to the application of the two-step approach, the Random Forest classifier shows a good performance both in the mountainous and in the coastal area, with an average weighted *F1* score of 0.961 and 0.957, and an average *Accuracy* of 0.928 and 0.946, respectively. The highest performance in the regression step is obtained by the Random Forest in the mountainous area with an R^2 of 0.541 and an *RMSE* of 0.109 mm, considering a spatial configuration including all stations. The comparison with the direct approach results shows that the two-step approach consistently improves accuracy across all scenarios, highlighting the benefits gained from breaking the data imputation process in stages where different physical conditions (in this case, rain and no-rain) are separately managed. Another important finding is that the use of time windows containing data lagged with respect to the imputed time interval allows capturing the atmospheric dynamics by connecting rainfall instances at different time levels and distant stations. Finally, the study confirms that machine learning models outperform spatial interpolation methods, thanks to their ability to manage data with complicated internal structure.



Academic Editors: Jiangjiang Zhang and Junliang Jin

Received: 29 September 2025

Revised: 3 November 2025

Accepted: 5 November 2025

Published: 10 November 2025

Citation: Boukdire, M.; İnan, Ç.; Varra, G.; Della Morte, R.; Cozzolino, L. Interpolation and Machine Learning Methods for Sub-Hourly Missing Rainfall Data Imputation in a Data-Scarce Environment: One- and Two-Step Approaches. *Hydrology* **2025**, *12*, 297. <https://doi.org/10.3390/hydrology12110297>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: rainfall data gap filling; machine learning; multilayer perceptron; random forest; spatial interpolation methods; regression; data imputation

1. Introduction

Reliable high-resolution rainfall data are essential for hydrological and climatic applications. For example, Berne et al. [1] showed that high-resolution rainfall data are required for urban catchment simulation due to the short response time, while Huang et al. [2] demonstrated that the results' quality for hydrologic simulation in a natural catchment improved with the increase in rainfall data resolution. Low-resolution data cannot capture the evolution of short convective rainfall events that occur during the summer season [3], especially those leading to flash flooding in rapid-response catchments [4]. Frequency and intensity of short-duration rainfall events driven by climate change are increasing, further demanding accurate sub-hourly rainfall measurements [4,5]. Collecting uninterrupted rainfall time series at a fine temporal resolution remains an operational challenge because equipment malfunctions, sensor calibration errors, network communication failures, and maintenance downtimes may result in incomplete datasets, affecting the reliability of subsequent analyses [6,7]. Even quality control operations on rainfall data may produce missing data due to the need to exclude outliers and too frequent or too low data [8]. The need for long and continuous high-resolution rainfall data records, as well as the possible presence of missing data in these records, makes the imputation of these values a necessary preprocessing step for obtaining high-quality outcomes produced by hydrologic models.

While studies have largely addressed the imputation of monthly [9,10] or daily precipitation [11,12], the sub-hourly scale introduces new challenges. The probability of finding no-rainfall data in records decreases with the time scale due to rainfall intermittency [13], resulting in zero-inflated time series for high temporal resolution rainfall data. This can reduce the performance of models for the imputation of missing data due to the bias introduced by the excess of zero values [14]. On the other hand, spatial correlations of high-resolution rainfall data decrease more rapidly with distance than coarser resolution data [15], reducing the capacity of neighboring rainfall gauges to aid the imputation of missing data at the target station for high-resolution time series.

The simplest techniques for the imputation of missing data in time-series rely on basic statistical indicators such as mean, median, or mode [16], but these approaches may fall short due to autocorrelation and persistence in the rainfall records [17]. The availability of neighboring rain gauges allows the use of spatial interpolation techniques such as Inverse Distance Weighting (IDW) [9,18–20], even with optimized parameters [21], Multiple Linear Regression [9,18], and Ordinary Kriging (OK) [19,20,22–24], among the others. Geostatistical techniques like OK work better on flat terrain but may perform relatively poorly in mountainous areas [25]. For this reason, Detrended Universal Kriging and Co-Kriging have been introduced to consider data trends and additional explanatory variables such as gauge elevations [24].

Recently, machine learning (ML) approaches have emerged as a promising family of tools for missing environmental data imputation [16], obtaining improved results over classic simple regression models [26] and offering greater flexibility. Among the most widely used machine learning models are Multilayer Perceptrons (MLPs), Support Vector Machines (SVMs), and Random Forests (RFs), which have been successfully applied to various environmental and hydrological problems. Several studies have reported that ML models effectively capture nonlinear relationships in rainfall data. Sattari et al. (2020) [27] demonstrated the potential of kernel and tree-based machine-learning models, such as

Support Vector Regression, Gaussian Process Regression, and RF, for estimating missing rainfall data. Bellido-Jiménez et al. (2021) [28] compared different ML approaches for filling gaps in daily rainfall series in a semiarid area (Andalusia, Spain), finding that MLP obtained the best results in most stations. Lupi et al. (2023) [29] evaluated various ML models, including linear regression, long short-term memory, and convolutional neural networks, for the imputation of missing daily rainfall data. Their study found that these models are effective in handling incomplete datasets, with the quality of the input data playing a critical role in model performance. Praveen Kumar and Dwarakish (2025) [30] considered different classic regression and ML approaches for filling gaps in daily rainfall series in the Kali River basin (India), finding that the K-nearest Neighbor approach behaved best among the models compared. From this very short review, it emerges that ML approaches seem suitable for the gap filling task in rainfall time series under different climates and in various regions.

Of course, ML approaches obtain the best results when they are built to incorporate the structure of the missing data to be imputed [31,32], allowing for an improved representation of the spatial and temporal dynamics that characterize the process [16]. In the case of missing rainfall data, this could be carried out in different ways. The first strategy integrates additional data such as wind velocity and direction, elevation, humidity, air pressure and temperature, allowing for a more robust imputation of the missing data by exploiting the availability of covariates [14,33]. The second strategy, which serves to cope with zero-inflated time series, consists of breaking the imputation process in two steps: in the first step, time intervals with missing data are classified to determine if they are rainy or non-rainy; in the second step, missing rainfall data imputation is carried out only in the time intervals classified as rainy [14,33–36]. In the third strategy, lagged data from different time steps are used to incorporate the time dynamics of the process to be imputed [16].

Especially relevant for the present discussion are the works by Chivers et al. (2020) [14] and Han et al. (2023) [33], where both the strategies of incorporating additional data and using a two-step approach were exploited. Chivers et al. (2020) [14] integrated sub-hourly (30-min) data collected from 37 weather stations (including precipitation, air pressure and temperature, relative humidity, wind speed and direction, volumetric water content) with rain gauge records in the UK (United Kingdom). Several ML models (gradient tree boosting, K-nearest neighbors, Random Forest, Support Vector Machines, and neural networks) were tested, adopting a two-step approach. The best performance was achieved when using the most comprehensive dataset, i.e., the one that included all the available soil moisture and hydro-meteorological features for each monitoring site and all the rain gauge stations surrounding the target site within a 30 km radius. Han et al. (2023) [33] used daily meteorological data (precipitation, humidity, air pressure and temperature, and wind speed) from 30 weather stations across South Korea to reconstruct missing daily precipitation values. They tested the performance of two ML models, Artificial Neural Networks (ANNs) and RF for both classification and regression steps.

Despite these advances, few studies have investigated the application of ML-based imputation to sub-hourly rainfall, particularly under data-scarce conditions. In operational networks, sensors are sparse and additional meteorological variables, including wind speed and direction, and air temperature and pressure, may be limited or absent [28,37]. Under such constraints, imputation at fine temporal resolutions requires techniques that can effectively handle heterogeneous, feature-limited environments.

To address this gap, the present study conducted a comparative analysis of RF, MLP, OK, and IDW to impute missing 10-min rainfall records by applying two different approaches for each model. In the direct approach, raw data are directly used to feed the regression models. In the two-step approach, the RF model is used as a classifier to sepa-

rate rain and no rain instances, and then a regression model is subsequently fed with the instances classified as rainy. This serves the aim of clarifying the relative model strengths and limitations and provides practical guidance for researchers and practitioners working with high-resolution hydrometeorological datasets, specifically in regions where there is limited access to additional meteorological data.

2. Materials and Methods

2.1. Study Area

The Campania region (Figure 1a,b) lies in Southern Italy between the Tyrrhenian Sea to the west and the Apennine Mountain chain to the east [38]. It includes a wide range of topographic elevations, from coastal lowlands and volcanic landforms, such as Mount Vesuvius and Campi Flegrei, to the inland hills and mountains of the Southern Apennines [39]. Covering an area of approximately 14,000 km², the region is characterized by a temperate Mediterranean climate with hot, dry summers and mild, wet winters. Annual precipitation generally ranges from 800 to 1000 mm, largely driven by moist air masses originating from the Tyrrhenian Sea [40]. These weather systems typically follow a southwest–northeast trajectory and are significantly influenced by the region’s complex orography, which plays a key role in precipitation patterns and distribution [41,42]. Due to the short time scale of these processes, the imputation of gaps in the high-resolution rainfall data series of the Campania region is particularly relevant.

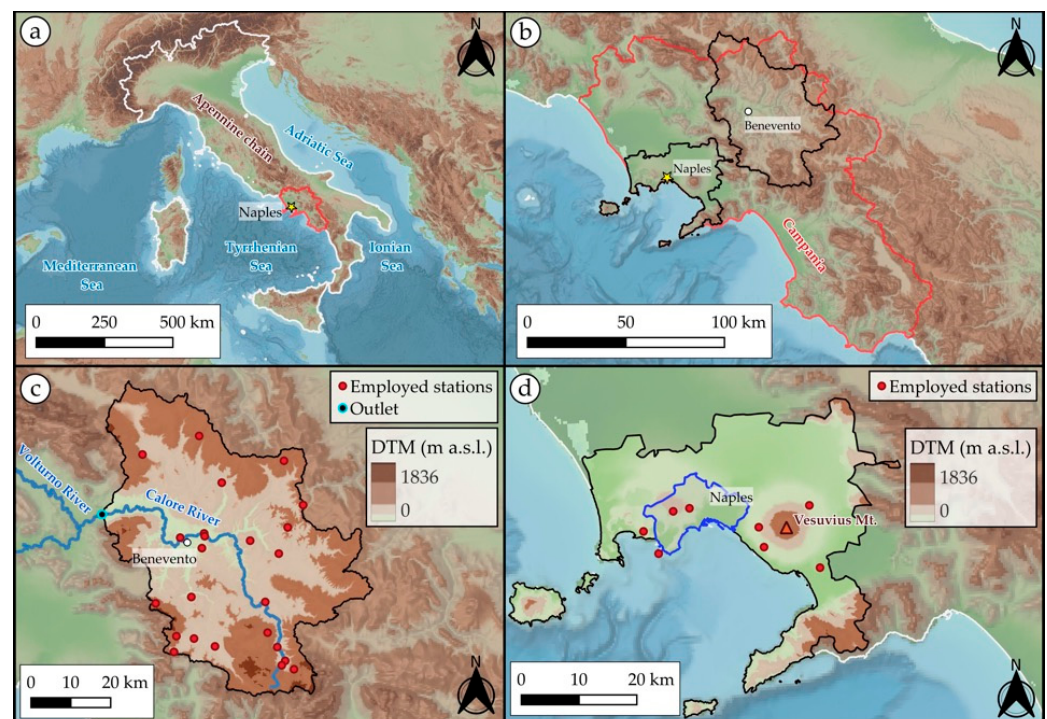


Figure 1. Location of the study areas within (a) Italy and (b) the Campania region (ETOPO 2022 global relief model, 15 arc-second resolution [43]). Distribution of rain gauges within (c) the Calore Irpino River basin and (d) the Metropolitan City of Naples (TINITALY DTM, 10 m resolution, INGV [44]).

Within this larger area, the present study focused on two specific target areas (Figure 1c,d) that exemplify the region’s geographical and environmental diversity: the Calore Irpino River basin (hereafter Calore River basin for the sake of simplicity), located in the inland eastern part of Campania, and the Metropolitan City of Naples, situated along the coast.

Topographic information for the maps in Figure 1 was derived from the ETOPO 2022 [43] global relief model (15 arc-second resolution) for the larger-scale views and from the TINITALY Digital Terrain Model (DTM; 10 m resolution) produced by the Istituto Nazionale di Geofisica e Vulcanologia (INGV) for the detailed panels [44].

2.1.1. Calore River Basin

The portion of the Calore River basin considered (Figure 1c) has an outlet immediately downstream of the area studied in [45,46]. This area was heavily flooded during the rain-storm of 14–15 October 2015, making the imputation of the corresponding high-resolution rainfall data series particularly significant. The basin covers a surface area of approximately 3050 km², with elevations ranging from 55 m to 1810 m above sea level (a.s.l.) [47]. It is primarily a rural, fluvially shaped area influenced by mountainous terrain, characterized by a Mediterranean climate, with an average annual rainfall of 1150 mm (mostly in autumn and winter), and a mean annual temperature of 15.1 °C. Hydrologically, the main river is the Calore Irpino River (hereafter Calore River for simplicity), which flows for 110 km and has a mean annual discharge of approximately 30 m³/s [48]. The area is characterized by a complicated orography that contributes to the spatial variability of precipitation patterns.

Figure 2a shows the average monthly precipitation recorded at a representative rain gauge (Benevento station) in the Calore River basin, demonstrating a marked seasonality with maximum monthly precipitation in November and minimum rainfall in August.

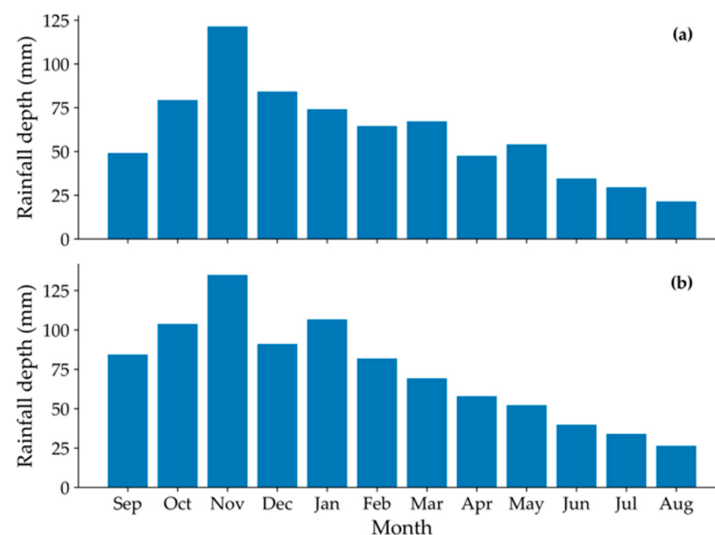


Figure 2. Average monthly precipitation from (a) Benevento station in the Calore River basin, and (b) Napoli Camaldoli station in the Metropolitan City of Naples.

2.1.2. Metropolitan City of Naples

The Metropolitan City of Naples (Figure 1d, black contour line) is a densely urbanized area stretching along the Bay of Naples that includes the City of Naples (Figure 1d, blue contour line), 91 additional municipalities, and three main islands (Ischia, Capri, and Procida), with a total surface of 1.171 km² and elevations ranging from 0 to 1444 m a.s.l. The area exhibits three main orographic groups consisting of the Phlegraean Fields caldera volcano in the northwest, the Monti Lattari chain along the Sorrento peninsula in the southeast, and the stratovolcano Mount Vesuvius in the east. Mount Vesuvius, with a height of 1281 m a.s.l., splits the area along a divide aligned from southwest to northeast.

The City of Naples, which is home to about one million inhabitants and is served by an urban drainage system whose modern history dates to the third quarter of the 19th century [49], is frequently hit by localized pluvial flooding, and its elevation spans between

0 and 470 m a.s.l. The inspection of Figure 2b, where the average monthly precipitation for a representative station (Napoli Camaldoli) is plotted, confirms a marked seasonality with a maximum monthly rainfall in November and a minimum in August.

2.2. Data Description and Preprocessing

The study used sub-hourly rainfall data from the Campania Region Civil Protection agency with measurements recorded every 10 min. Since the rain gauges in the study areas began operating at different times, the dataset includes measurements starting from the year when all stations became operational.

The sensors in both study areas are tipping-bucket rain gauges with a resolution of 0.2 mm per tip, featuring a rain collection vessel of 1000 cm² and operating within a measurement range of 0–300 mm/h and a temperature range of 0–60 °C. The instrument failure and maintenance issues resulted in different amounts of missing data in the time series, varying from 0.07% to 18.45%.

2.2.1. Calore River Basin

The Calore River basin includes 23 rain gauges, distributed over mountainous and hilly terrain, with data covering the period between 2006 and 2024, offering a nearly two-decade record of rainfall. For these rain gauges, Table 1 provides summary information (latitude, longitude, elevation above sea level, and the percentage of missing data). One rain gauge (Pietrastornina, not present in Table 1) was excluded from the analysis due to the presence of artifacts in the data that could have compromised model performance.

Table 1. Characteristics of the rain gauges in the Calore River basin.

Rain Gauge	Longitude (°)	Latitude (°)	Elevation (m a.s.l.)	Missing Data (%)
Bagnoli Irpino	15.07444	40.82731	730	0.82
Montella	15.05020	40.84770	489	4.08
Montella METEO 1	15.03910	40.83830	515	13.52
Cassano Irpino	15.02940	40.88000	460	11.35
Montemarano	15.00360	40.91410	865	0.43
Serino	14.84440	40.89110	327	0.19
Monteforte Irpino	14.72130	40.88470	600	3.33
Avellino-Genio Civile	14.78358	40.91197	360	0.10
Mercogliano	14.73160	40.92000	625	0.19
Altavilla Irpina	14.78386	41.00628	352	0.07
Luogosano	15.00300	40.98380	450	0.17
S. Martino Valle Caudina	14.67630	40.99720	757	0.348
Melito Irpino	15.05330	41.09080	319	13.43
Rocchetta	14.97050	41.12440	670	12.26
Ariano Irpino	15.08542	41.14853	678	0.80
Benevento	14.76242	41.14200	119	0.51
Benevento METEO 1	14.82503	41.11506	236	10.15
Paduli	14.83470	41.14800	130	1.12
Ponte Valentino	14.83580	41.14160	139	0.20
Ariano Irpino METEO 1	15.13608	41.19661	631	10.56
Pago Veiano	14.89761	41.25983	259	11.46
Castelfranco in Miscano	15.08822	41.29953	751	0.48
Morcone	14.66603	41.33536	630	18.45
Colle Sannita	14.83936	41.36883	777	0.13

Pearson Correlation Coefficient and Clustering

For the imputation of missing data, the rain gauges were grouped in different spatial configurations. Aiming at this, the Pearson correlation coefficient (PCC),

$$r_{x,y} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \quad (1)$$

where x_i and y_i are the individual rainfall observations from the two gauges being compared and $\bar{x}$ and $\bar{y}$ are the corresponding mean values, was computed for the time series corresponding to all pairs of stations, considering time lags up to 24 h, accounting for potential delays in rainfall patterns between locations. The optimal lag τ_{opt} is defined here as the value of the lag time that maximizes the PCC.

In Figure 3a, the value $r_{\tau_{opt}}$ of the PCC corresponding to τ_{opt} is plotted against the distance l between stations. The inspection of the figure shows that $r_{\tau_{opt}}$ decreases with l and becomes statistically insignificant beyond a certain distance, which is not clearly defined. For this reason, the PCCs were recalculated using new time series consisting of the difference in rainfall depth between two consecutive time intervals. These new time series are intended as a proxy of the rainfall intensity derivative, which is positive when the rainfall intensifies and negative when it decreases. When these differential time series are used, the PCC clearly drops to zero beyond $l = 25$ km (Figure 3b). Based on this result and the shape of Figure 3a, distance and correlation coefficient thresholds were set at 25 km and 0.4, respectively.

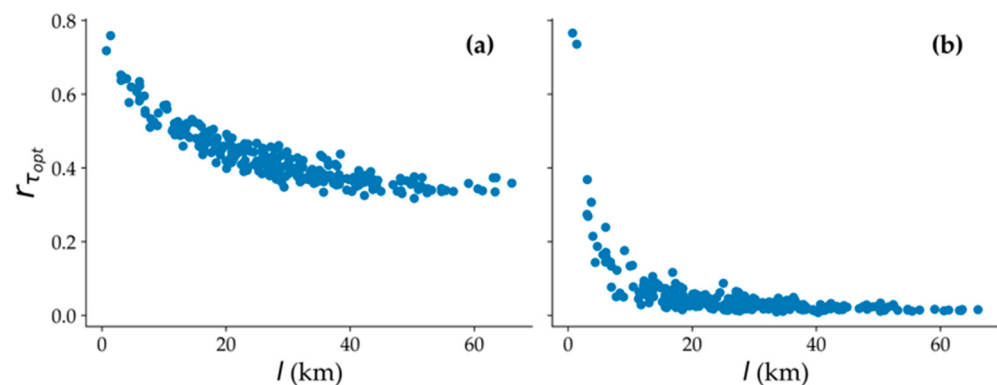


Figure 3. Optimal Pearson correlation coefficients as a function of distance in the Calore River basin: original data series (a), and differential data series (b).

Station clusters were finally formed by applying the above-defined thresholds on the correlation coefficient and spatial distance, ensuring that stations within the same cluster exhibited strong correlation while maintaining geographic proximity. For the Calore River basin, this resulted in five rain gauge clusters. The sensor clusters for Rocchetta and Altavilla Irpina stations are presented in Figure 4 as an example.

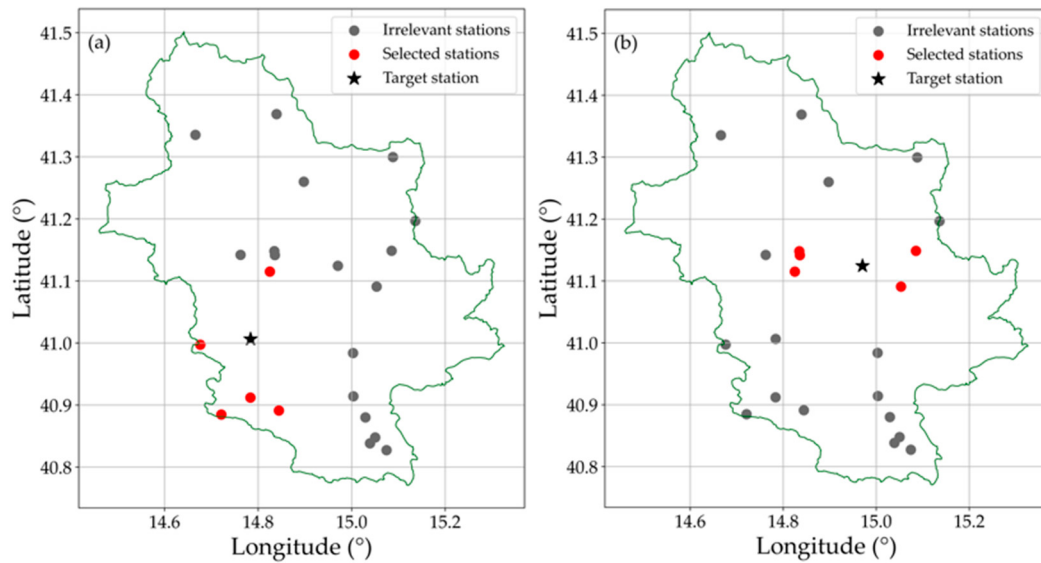


Figure 4. Sensor clusters for (a) Altavilla Irpina and (b) Rocchetta stations in the Calore River basin.

Optimum Time Lag

Figure 5 shows the absolute value of τ_{opt} (in minutes) that maximizes the PCC for each pair of stations as a function of the distance l (in kilometers). Inspection of the figure shows that the optimal time lag between stations increases with distance, suggesting that rain gauges are connected by a precipitative signal moving with finite celerity. The points of Figure 5 are enveloped by a red line with a slope of about 9.39 m/s. This value, which is suggestively reminiscent of the wind velocity in the Campania region mid-troposphere at the elevation of precipitative clusters [40], can be interpreted as the lower bound of the precipitative signal celerity.

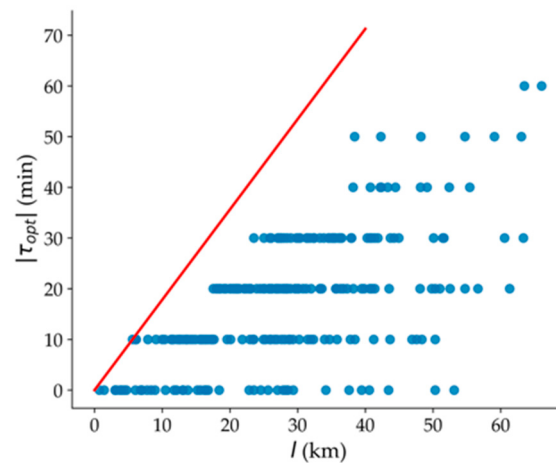


Figure 5. Calore River basin. Optimal absolute lag time τ_{opt} as a function of distance l .

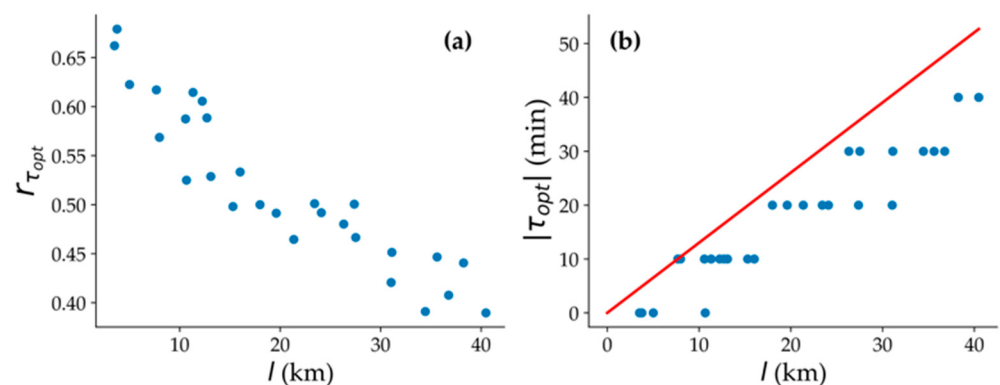
2.2.2. Metropolitan City of Naples

The Metropolitan City of Naples is not enclosed by natural orographic divides (see Figure 1d), and its shape is merely determined by administrative boundaries. For this reason, eight stations, covering the period from 2009 to 2024, were selected based on their proximity to the center of the City of Naples and the sea (see Table 2).

Table 2. Characteristics of the rain gauges in the Metropolitan City of Naples.

Rain Gauge	Longitude (°)	Latitude (°)	Elevation (m a.s.l)	Missing Data (%)
Pompei	14.4925	40.7566	17	0.447
Torre del Greco	14.3805	40.7941	50	3.465
Ottaviano	14.4786	40.8547	180	0.202
Ercolano	14.3722	40.8255	209	0.348
Napoli Camaldoli	14.1988	40.8583	390	0.117
Napoli Capodimonte	14.2325	40.8616	176	0.148
Nisida METEO 1	14.1638	40.7937	88	15.422
Pozzuoli	14.1355	40.8298	127	0.123

In Figure 6a, the values of $r_{\tau_{opt}}$ are reported as a function of the distance l for the Metropolitan City of Naples, again showing that the correlation between two stations decreases with the distance. However, the clustering operation was not carried out for this area due to the small number of sensors. In Figure 6b, the absolute value of τ_{opt} is plotted as a function of l , demonstrating also in this case that information transfer between stations is carried with a minimum celerity equal to 12.8 m/s (slope of the envelope red line in Figure 6b).

**Figure 6.** Metropolitan City of Naples. Optimal Pearson correlation coefficients $r_{\tau_{opt}}$ as a function of distance l (a) and corresponding τ_{opt} in minutes (b).

2.3. Direct vs. Two-Step Approach

Two distinct strategies were investigated for the imputation of missing rainfall data: a Direct Approach (DA) and a Two-Step Approach (TSA). Both methods incorporate spatiotemporal features but differ in their treatment of rain and no-rain events before regression modeling.

TSA is based on the idea that rainy and non-rainy time intervals are characterized by fundamentally different physical phenomena that are irreducible and should not be considered mutually consistent for missing data imputation. The separation between rainy and non-rainy time intervals allowed for the training of the regression algorithms on rainy time intervals only, avoiding a detrimental specialization caused by the abundance of non-rainy time intervals where a null rainfall depth is collected.

Following this idea, an RF binary classifier was trained to distinguish between rain (1) and no-rain (0) time intervals during the first stage of TSA (Figure 7a). In the subsequent stage, which employed one of four regression methods, including MLP, RF, OK, and IDW, only intervals classified as rainy were subjected to regression-based imputation, while no-rain time steps were assigned a value of zero without further modeling (Figure 7a).

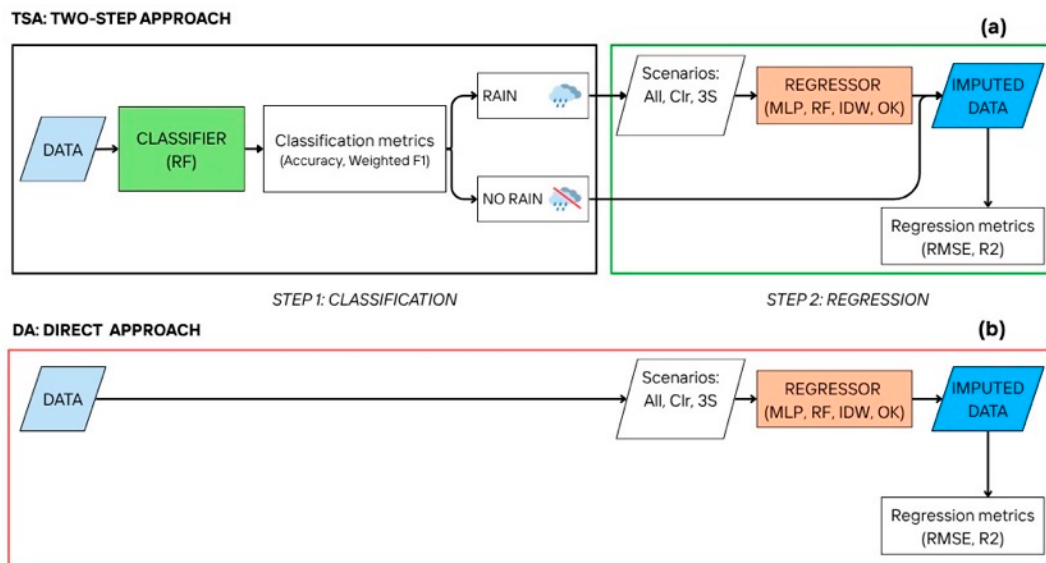


Figure 7. Flowchart of two-step (a) and direct (b) approaches for the imputation of missing data.

In contrast, DA applied the regression models to the full dataset without any prior classification (Figure 7b). All the observations, regardless of the precipitation status, were treated as continuous values and directly imputed.

For each approach, three spatial configurations of input features were considered to impute the missing data in a target station. The first configuration included all rainfall stations (*All* spatial configuration) within the study area, while the second spatial configuration made use of the three best correlated stations only (*3S* spatial configuration). Finally, the third spatial configuration, which was applied to the Calore River basin only, was based on the use of data from gauges falling in the same cluster of the target station (*Clr* spatial configuration). Temporal features (hour, month, and season, with December–January–February as winter season, March–April–May as spring season, June–July–August as summer, and September–October–November as fall season), encoded as additional features for the ML models using sine and cosine transformations [14], were also included:

$$R_{sin} = \sin\left(\frac{2 * \pi * R}{\max(R)}\right), \quad (2)$$

$$R_{cos} = \cos\left(\frac{2 * \pi * R}{\max(R)}\right). \quad (3)$$

In Equations (2) and (3), R represents the original temporal feature value, either hour (ranging from 1 to 24), month (1 to 12), or season (1 for winter, 2 for spring, 3 for summer, 4 for fall). This transformation ensures that cyclical boundaries (e.g., hour 24 and hour 1, or month 12 and month 1) are close in the transformed space, helping ML models to better learn patterns related to time by preserving the continuity and periodicity in temporal data [14].

All models were developed in Python 3.12. NumPy 1.26.4 and Pandas 2.2.2 libraries were used for data handling and preprocessing. The ML models (MLP and RF) were built using Scikit-learn 1.5.2 and Keras 3.6.0 libraries. For spatial interpolation, Kriging was performed with the PyKriging library, while Inverse Distance Weighting (IDW) was implemented through a custom NumPy-based approach. In the following, the classification and regression models are succinctly described.

2.3.1. Random Forest

The concept of random decision forests was initially introduced by Ho (1995) [50] and later formalized and popularized by Breiman (2001) [51]. It is an ensemble learning technique widely used for classification, regression, and various other predictive tasks that operates by generating a large number of decision trees during the training phase and aggregating their outputs to produce a final prediction.

In the RF framework, the aggregation process differs slightly between classification and regression tasks. For classification, each tree in the ensemble provides a class prediction, and the final prediction is made by taking the majority vote from the individual trees. For the regression task, the final prediction $\hat{y}(\mathbf{R})$ is the average of the predictions made by each tree:

$$\hat{y}(\mathbf{R}) = \frac{1}{T} \sum_{t=1}^T y_t(\mathbf{R}) \quad (4)$$

where $y_t(\mathbf{R})$ is the prediction from tree t , T is the total number of trees, and $\mathbf{R}$ is the vector of input data.

Preliminary data preprocessing of Section 2.2 shows that the precipitative signals travel between stations with finite celerity. This suggests that information at preceding or subsequent time intervals in surrounding stations could be relevant for imputing missing data at the targeted station during the targeted time interval. To capture the local temporal context, a sliding time window of 21 steps centered on the targeted time interval was incorporated into the analysis. This window was designed to provide a comprehensive temporal frame that could enhance the predictive capacity of the models across both imputation strategies.

In TSA, the RF classifier was first applied to classify rainfall data into rain and no-rain time intervals before the regression-based imputation stage. To select the most informative stations for the classifier, a progressive strategy was applied by adding the stations one at a time based on their correlation with the target station. Computational experience showed that the best performance was achieved when all available stations were included. In the subsequent step, a RF regressor was used to estimate missing rainfall values using one of the three spatial gauge configurations. In DA, only the RF regressor was used to directly cope with data without prior classification.

To optimize the model performance, a grid search procedure was conducted to fine-tune the hyperparameters of both the classifier and the regressor, for both direct and two-step approaches.

2.3.2. Multilayer Perceptron

A Multilayer Perceptron (MLP) is a type of feedforward artificial neural network (ANN) commonly used in supervised learning tasks. It consists of an input layer, one or more hidden layers, and an output layer, with interconnected nodes trained using the backpropagation algorithm [52]. This approach enables the network to learn internal representations of complex patterns, even when these patterns are not explicitly defined in the input data, making MLPs highly suitable for applications such as hydrological and environmental modeling [53,54]. According to [12], the mathematical representation of MLP for estimation is

$$\hat{y} = \sum_{j=1}^K w_j^{(2)} \sigma \left(\sum_{i=1}^p w_{ji}^{(1)} R_i + b_j^{(1)} \right) + b^{(2)}, \quad (5)$$

where the meaning of the symbols is as follows: p is the number of input features R_i , K is the number of hidden units (neurons) in the hidden layer, $w_{ji}^{(1)}$ is the weight from input R_i to the hidden unit j , and $b_j^{(1)}$ is the bias term associated with the hidden unit j . The activation function $\sigma(\cdot)$, can be any non-linear function such as the sigmoid, hyperbolic tangent (tanh),

or ReLU (rectified linear unit). After the hidden layer, each hidden unit j connects to the output through a weight $w_j^{(2)}$, and a final bias term at the output is represented as $b^{(2)}$. The predicted output, which in this case represents the estimated rainfall, is denoted by $\hat{y}$. Figure 8 represents the multilayer perceptron model with two hidden layers used in the present application.

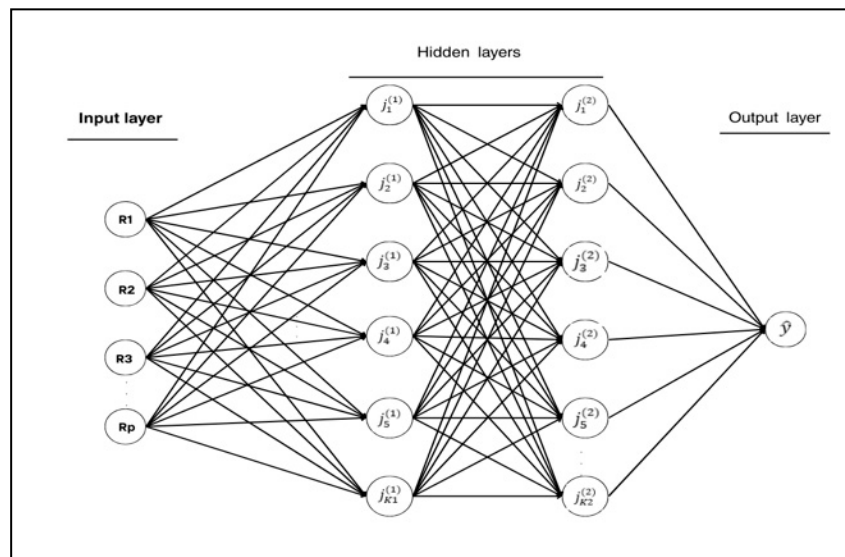


Figure 8. Schematic representation of a Multilayer Perceptron with two hidden layers.

In TSA, MLP is used during rainy time intervals only after classification with RF. In DA, MLP is used on the entire set of data. In both approaches and for each gauge spatial configuration, input data are supplied by a 21-time-interval window centered on the targeted time interval. Before training, a grid search procedure was conducted to fine-tune the model's hyperparameters, ensuring optimal performance.

2.3.3. Inverse Distance Weighting

Inverse Distance Weighting (IDW) is a deterministic interpolation method that estimates unknown values based on a weighted average of nearby known values, where weights are inversely proportional to distance [55]. The general formula is

$$\hat{y}(x_0) = \frac{\sum_{i=1}^N \omega_i(x_0) R(x_i)}{\sum_{i=1}^N \omega_i(x_0)} \quad (6)$$

$$\omega_i(x_0) = \frac{1}{d(x_0, x_i)^p} \quad (7)$$

where $\hat{y}(x_0)$ is the interpolated value at the target location with the coordinate vector x_0 , $R(x_i)$ is the observed value at the station with the coordinate vector x_i , $d(x_0, x_i)$ is the distance between x_0 and x_i , and p is a parameter. For each spatial gauge configuration, only data from the targeted time interval were used.

2.3.4. Ordinary Kriging

Ordinary Kriging (OK) is a geostatistical method that estimates values based on the spatial autocorrelation of observed data, assuming a constant unknown mean across the field. The estimator is a weighted linear combination of neighboring observations

$$\hat{y}(x_0) = \sum_{i=1}^N \lambda_i R(x_i) \quad (8)$$

$$\sum_{i=1}^N \lambda_i = 1 \quad (9)$$

where the weights λ_i depend on the spatial autocorrelation through a semivariogram.

2.4. Performance Analysis

To evaluate the predictive capability of the models, different performance metrics were used for classification and regression tasks. For machine learning regressors, specifically the Multilayer Perceptron (MLP) and Random Forest (RF) models, a fivefold cross-validation procedure was applied. In this approach, the dataset is divided into five equal subsets (folds), and each fold is used as a validation set while the remaining four folds are used for training. This process is repeated five times, ensuring that each sample is used for both training and validation. This technique provides a robust assessment of model performance and helps mitigate overfitting.

For the classification step, which involves the use of the Random Forest classifier to distinguish between rain and no-rain events, *Accuracy*, *Precision*, *Recall*, and *F1 score* metrics are employed. The *Accuracy*

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \quad (10)$$

where *TP* is the number of true positives (rainfall is correctly predicted), *TN* is the number of true negatives (where no-rain conditions are correctly predicted), *FP* is the number of false positives (rainfall incorrectly predicted), and *FN* is the number of false negatives (rainfall occurred but was incorrectly predicted as no rain), represents the ratio between correctly predicted time intervals and the total number of time intervals.

The *Precision* measures the proportion of correctly predicted rain events among all predicted rain events:

$$Precision = \frac{TP}{TP + FP}. \quad (11)$$

The *Recall* measures the proportion of actual rain events that were correctly predicted:

$$Recall = \frac{TP}{TP + FN}. \quad (12)$$

The *F1 score* is the harmonic mean of *Precision* and *Recall*, providing a balanced measure when there is class imbalance:

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall}. \quad (13)$$

In addition, a weighted *F1 score* was calculated to account for class imbalance by averaging the *F1 scores* of each class weighted by their support (the number of true instances for each class).

For the regression step, the model performance was evaluated using the root mean square error (*RMSE*) and the coefficient of determination R^2 . The *RMSE*

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad (14)$$

where n is the number of time interval, y_i is the observed rainfall value for the i -th time interval, and $\hat{y}_i$ is the predicted rainfall value for the corresponding time interval, measures the average difference between the predicted and observed values and is expressed in the same units as the considered variable. Clearly, the smaller the *RMSE*, the better the model performance.

The coefficient of determination

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2}, \quad (15)$$

where $\bar{y}_i$ is the mean of observed values, indicates the proportion of the variance in the observed data that is predictable by the model. The closer R^2 is to 1, the better the model captures the variability of the observed data. However, R^2 can also be negative, which indicates that the model performs worse than simply using the mean of the data as a predictor.

3. Results

The results presented in this section were obtained as the average of fivefold cross-validation, applied consistently across all models and methods. This applied to both the classification step (for the TSA) and the regression step. For a fair and consistent comparison, the same folds were used to evaluate both the machine learning models (MLP, RF) and the interpolation methods (IDW, OK), ensuring that the metrics reflect comparable data splits and conditions.

3.1. Performance of the Classification Step

The performance of the classification step is summarized in Table 3 for the Calore River basin and in Table 4 for the Metropolitan City of Naples.

Table 3. Binary classification metrics for the 23 stations in the Calore River basin.

Station	Accuracy	Precision	Recall	F1	Weighted F1
Bagnoli Irpino	0.924	0.873	0.399	0.548	0.958
Montella	0.945	0.921	0.488	0.638	0.970
Montella METEO 1	0.942	0.907	0.480	0.628	0.968
Cassano Irpino	0.935	0.904	0.401	0.556	0.965
Montemarano	0.922	0.900	0.333	0.486	0.957
Serino	0.927	0.901	0.366	0.520	0.960
Monteforte Irpino	0.922	0.885	0.349	0.500	0.957
Avellino-Genio Civile	0.933	0.918	0.364	0.522	0.963
Altavilla Irpina	0.922	0.911	0.328	0.483	0.957
Luogosano	0.925	0.911	0.281	0.430	0.959
S. Martino Valle Caudina	0.912	0.820	0.383	0.522	0.950
Melito Irpino	0.928	0.865	0.270	0.411	0.961
Rocchetta	0.935	0.917	0.275	0.424	0.965
Ariano Irpino	0.930	0.898	0.293	0.441	0.962
Benevento	0.935	0.908	0.295	0.446	0.965
Benevento METEO 1	0.941	0.894	0.333	0.486	0.968
Paduli	0.951	0.926	0.361	0.520	0.974
Ponte Valentino	0.951	0.900	0.358	0.512	0.974
Ariano Irpino METEO 1	0.924	0.872	0.274	0.417	0.959
Pago Veiano	0.915	0.878	0.239	0.375	0.954
Castelfranco in Miscano	0.910	0.841	0.258	0.395	0.951
Morcone	0.901	0.826	0.291	0.430	0.945
Colle Sannita	0.913	0.860	0.268	0.408	0.953
All stations average	0.928	0.889	0.334	0.483	0.961

Table 4. Binary classification metrics for the 8 stations in the Metropolitan City of Naples.

Station	Accuracy	Precision	Recall	F1	Weighted F1
Ercolano	0.946	0.902	0.346	0.501	0.954
Napoli Camaldoli	0.949	0.898	0.353	0.507	0.957
Napoli Capodimonte	0.949	0.893	0.376	0.529	0.958
Ottaviano	0.950	0.885	0.396	0.547	0.957
Nisida	0.944	0.827	0.337	0.479	0.959
Pompei	0.942	0.835	0.310	0.452	0.959
Pozzuoli	0.943	0.828	0.308	0.449	0.957
Torre del Greco	0.945	0.910	0.333	0.487	0.957
All stations average	0.946	0.872	0.345	0.494	0.957

In the Calore River basin (Table 3), *Accuracy* values across stations are generally high, ranging from 0.901 to 0.951. Most stations exhibit *Precision* values above 0.85, with several stations achieving *Precision* close to or exceeding 0.90. *Recall* values are lower in comparison, typically between 0.24 and 0.49, indicating the challenge in correctly identifying rainy ten-minute time intervals, due to the greater number of no-rain time intervals. Nonetheless, *F1* scores and weighted *F1* scores remain acceptable, with weighted *F1* values exceeding 0.95 for nearly all stations, reflecting the models' overall balanced performance.

For the Metropolitan City of Naples, the results are similar in the eight monitored stations (Table 4). *Accuracy* values range from 0.942 to 0.950, with *Precision* consistently above 0.82. *Recall* values vary between 0.308 and 0.396, while *F1* scores lie within the 0.45–0.55 range. Weighted *F1* scores are uniformly high, between 0.954 and 0.959. The classification models thus demonstrate reliable predictive capacity across both regions, providing a solid foundation for the subsequent regression step.

3.2. Performance of the Regression Step

Tables 5–8 summarize the average *RMSE* and R^2 values related to the regression results across all stations within each study area, covering all methods and both imputation strategies. These average metrics provide a general overview of model accuracy and allow for direct comparison between the two approaches. The R^2 and *RMSE* values for different imputation methods and both study areas are shown in Figures 9–12. In general, across all models, TSA tends to yield higher R^2 and lower *RMSE* values than DA.

Table 5. MLP model overall regression performance for direct (DS) and two-step (TSA) approaches and different station configurations: all stations (*All*), cluster (*Clr*), three most correlated stations (*3S*).

Study Area	Approach	Scenario	RMSE (mm)	R^2
Calore River basin	DA	<i>All</i>	0.128	0.381
		<i>Clr</i>	0.128	0.373
		<i>3S</i>	0.127	0.379
	TSA	<i>All</i>	0.110	0.526
		<i>Clr</i>	0.110	0.519
		<i>3S</i>	0.110	0.521
Metropolitan City of Naples	DA	<i>All</i>	0.158	0.384
		<i>3S</i>	0.158	0.384
	TSA	<i>All</i>	0.146	0.474
		<i>3S</i>	0.146	0.475

Table 6. RF model regression performance for direct (DA) and two-step (TSA) approaches and different station configurations: all stations (*All*), cluster (*Clr*), three most correlated stations (*3S*).

Study Area	Approach	Scenario	RMSE (mm)	R ²
Calore River basin	DA	<i>All</i>	0.129	0.368
		<i>Clr</i>	0.129	0.364
		<i>3S</i>	0.128	0.372
	TSA	<i>All</i>	0.109	0.541
		<i>Clr</i>	0.110	0.528
		<i>3S</i>	0.110	0.529
Metropolitan City of Naples	DA	<i>All</i>	0.160	0.374
		<i>3S</i>	0.159	0.379
	TSA	<i>All</i>	0.144	0.488
		<i>3S</i>	0.145	0.483

Table 7. IDW method regression performance for direct (DA) and two-step (TSA) approaches and different station configurations: all stations (*All*), cluster (*Clr*), three most correlated stations (*3S*).

Study Area	Approach	Scenario	RMSE (mm)	R ²
Calore River basin	DA	<i>All</i>	0.128	0.348
		<i>Clr</i>	0.137	0.238
		<i>3S</i>	0.137	0.239
	TSA	<i>All</i>	0.119	0.442
		<i>Clr</i>	0.123	0.399
		<i>3S</i>	0.123	0.396
Metropolitan City of Naples	DA	<i>All</i>	0.170	0.273
		<i>3S</i>	0.174	0.235
	TSA	<i>All</i>	0.155	0.397
		<i>3S</i>	0.158	0.378

Table 8. OK method overall regression performance for direct (DA) and two-step (TSA) approaches and different station configurations: all stations (*All*), cluster (*Clr*), three most correlated stations (*3S*).

Study Area	Approach	Scenario	RMSE (mm)	R ²
Calore River basin	DA	<i>All</i>	0.135	0.278
		<i>Clr</i>	0.138	0.232
		<i>3S</i>	0.136	0.264
	TSA	<i>All</i>	0.126	0.383
		<i>Clr</i>	0.123	0.396
		<i>3S</i>	0.122	0.411
Metropolitan City of Naples	DA	<i>All</i>	0.179	0.203
		<i>3S</i>	0.175	0.223
	TSA	<i>All</i>	0.163	0.340
		<i>3S</i>	0.158	0.378

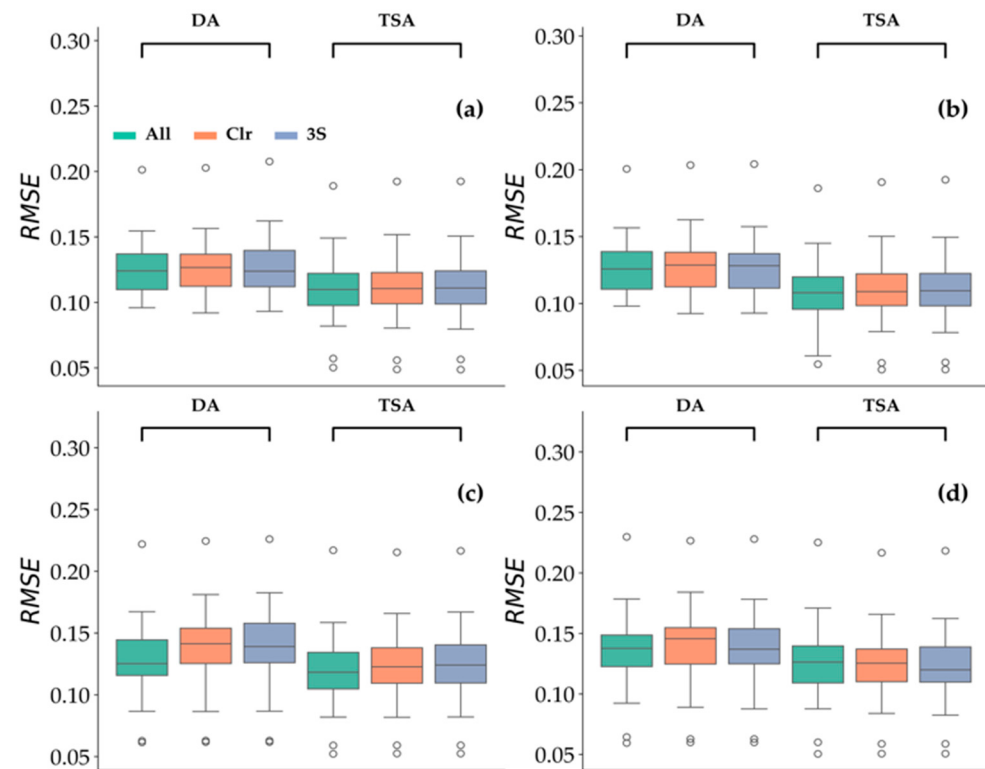


Figure 9. Boxplots of $RMSE$ for different models in the Calore River basin: MLP (a), RF (b), IDW (c), and OK (d) for the direct (DA) and two-step (TSA) approaches under the *All*, *Clr*, and *3S* gauge configurations.

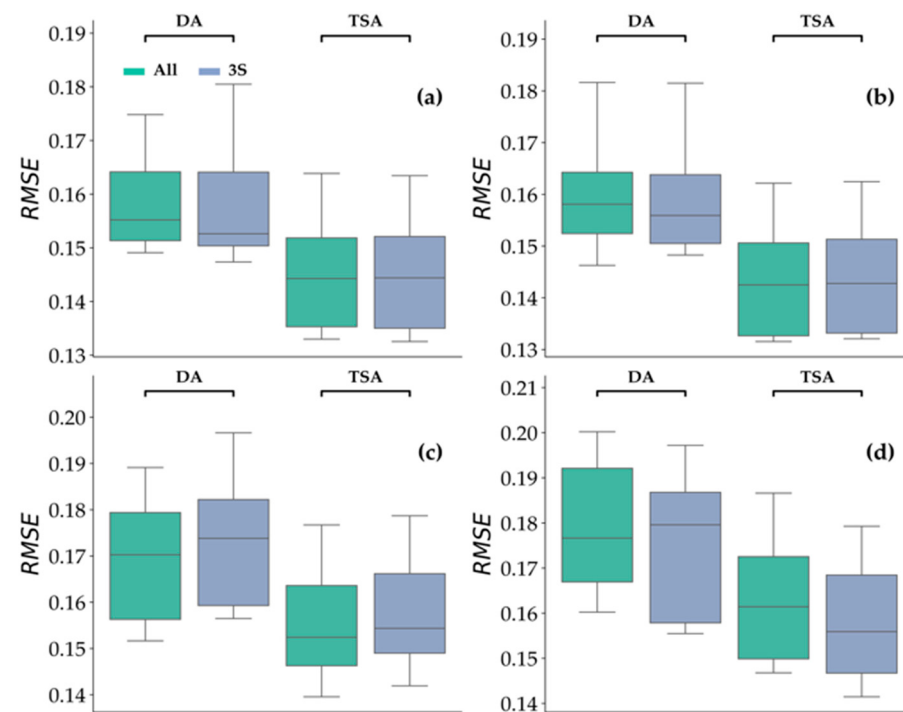


Figure 10. Boxplots of $RMSE$ for different models in the Metropolitan City of Naples: MLP (a), RF (b), IDW (c), and OK (d) for the direct (DA) and two-step (TSA) approaches under the *All* and *3S* gauge configurations.

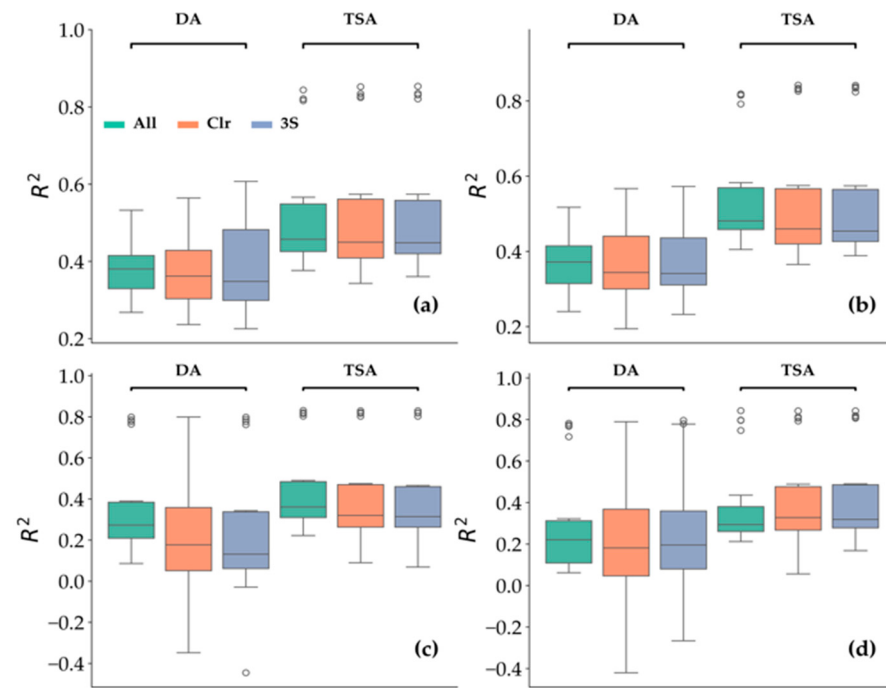


Figure 11. Boxplots of R^2 for different models in the Calore River basin: MLP (a), RF (b), IDW (c), and OK (d) for the direct (DA) and two-step (TSA) approaches under the *All*, *Clr*, and *3S* gauge configurations.

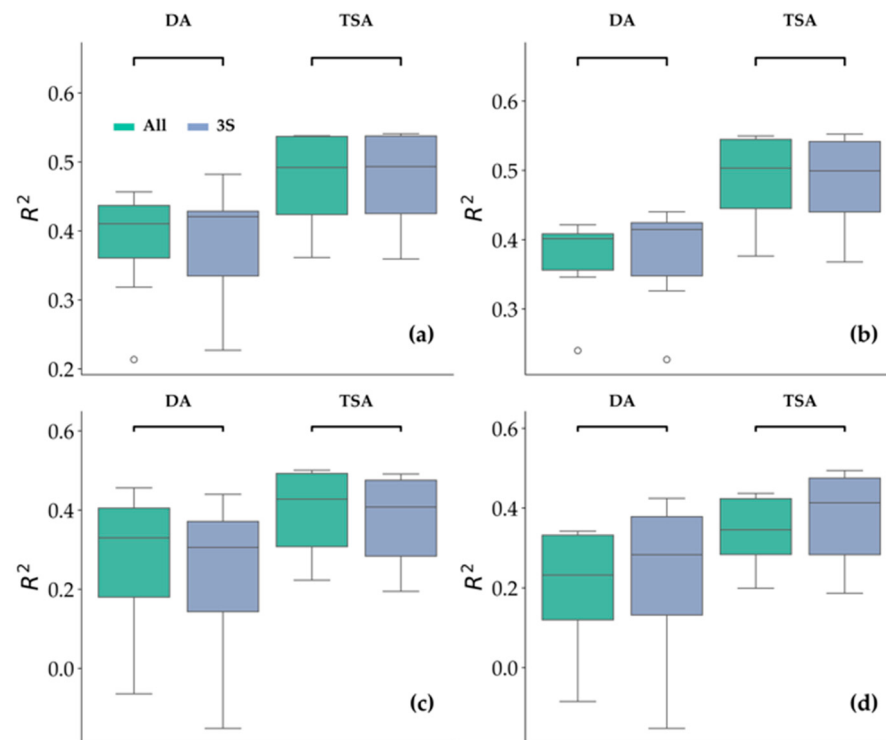


Figure 12. Boxplots of R^2 for different models in the Metropolitan City of Naples: MLP (a), RF (b), IDW (c), and OK (d) for the direct (DA) and two-step (TSA) approaches under the *All* and *3S* gauge configurations.

3.2.1. Calore River Basin

In the case of the MLP model (Table 5, Figures 9a and 11a), TSA consistently results in improved performance across all station selection strategies. The *All* strategy achieves the highest median R^2 (0.521), with a corresponding RMSE of 0.110 mm, indicating more

accurate and consistent predictions. Under DA, performance is lower and more dispersed, particularly for the 3S configuration ($R^2 = 0.379$; $RMSE = 0.127$ mm), which also shows greater variability.

For the RF model (Table 6, Figures 9b and 11b), a similar pattern is observed. TSA produces more compact distributions with higher medians across all station selection methods. For instance, the *All* configuration achieves an R^2 of 0.541 and an $RMSE$ of 0.109 mm under TA, while the same setup yields a lower R^2 of 0.368 and a higher $RMSE$ of 0.129 mm under DA. The performance differences among *All*, *Clr*, and 3S are relatively minor under TSA, with all three configurations achieving nearly identical $RMSE$ values (~0.110 mm), which supports the consistency of the model once no-rain events are filtered out.

The IDW results (Table 7, Figures 9c and 11c) show the lowest overall R^2 values among the tested methods. Under DA, the distributions are highly variable, and for scenarios like *Clr* and 3S, many values of R^2 fall below zero with average values of 0.238 and 0.239, respectively. Correspondingly, the $RMSE$ values are relatively high and equal to 0.137 mm in both cases. TSA leads to modest improvements in R^2 and reductions in $RMSE$. For example, using all stations under TSA, R^2 increases to 0.442 with $RMSE$ dropping to 0.119 mm. Overall, the IDW remains substantially below the machine learning models' performance in both metrics.

For the OK model (Table 8, Figures 9d and 11d), TSA again results in higher and more stable R^2 values compared to DA. Under TSA, the highest R^2 of 0.411 is observed for the 3S strategy, with an $RMSE$ of 0.122 mm. In contrast, under DA, this same configuration only achieves an R^2 of 0.264 with a higher $RMSE$ of 0.136 mm. Although variability is still present, especially in the DA setting, negative R^2 values are less frequent under TSA. The 3S strategies perform similarly, but the *All* configuration generally achieves slightly higher medians and lower $RMSE$ values ($R^2 = 0.383$, $RMSE = 0.126$ mm under TSA).

3.2.2. Metropolitan City of Naples

A similar behavior is observed in the Metropolitan City of Naples, as shown in Figures 10 and 12. Although the *Clr* scenario was not included in this case, both the *All* and 3S spatial configurations show improved performance under TSA across all four regression models. For the MLP model (Table 5), R^2 increases from 0.384 to 0.474 and $RMSE$ decreases from 0.158 mm to 0.146 mm. Similarly, the RF model under TSA yields R^2 values of 0.488 (*All*) and 0.483 (3S), with $RMSE$ values of 0.144 mm and 0.145 mm, respectively (Table 6). These results again confirm the benefit of TSA in improving both predictive accuracy and consistency. Although the differences between *All* and 3S are less pronounced in the Metropolitan City of Naples than in the Calore River basin, TSA remains consistently associated with higher R^2 values, lower $RMSE$ s, and reduced variability across all methods.

3.2.3. Performance Comparison

An inspection of Tables 5–7 and Figures 9–12, clearly shows that independent of the studied area, the spatial configuration of the station and the use of DA or TSA, the ML regressors MLP and RF behave better than the spatial interpolation models ISW and OK. For the Calore River Basin, the ML regressors with TSA supply values of $RMSE$ not higher than 0.110 mm and average values of R^2 not smaller than 0.519. For the same area, the spatial interpolation regressors with TSA supply values of $RMSE$ not smaller than 0.119 mm and average values of R^2 not higher than 0.442. For the Metropolitan City of Naples, the ML regressors with TSA supply values of $RMSE$ not higher than 0.146 mm and average values of R^2 not smaller than 0.474, while the spatial interpolation regressors supply values of $RMSE$ not smaller than 0.155 mm and average values of R^2 not higher than 0.397. Similar observations can be made for the DA strategy. This aligns well with the

existing literature [26–30], confirming that ML models better capture the spatial and time dynamics of precipitation, allowing for improved imputation of missing rainfall data.

Compared to the DA, the TSA consistently enhances model performance in terms of the coefficient of determination R^2 across all models and scenarios. The gain in R^2 varies depending on the regressor model and the study area, but it remains positive in all cases. The greatest improvements are achieved by the interpolation-based methods, particularly the OK, which exhibits an R^2 gain of approximately 71% in the *Clr* scenario within the Calore River basin (Figure 13a). Similarly, in the Metropolitan City of Naples (Figure 13b), both interpolation methods show significant increases across all spatial configurations, confirming the advantage of applying the classification step before regression. The smallest relative improvement is observed for the MLP model in the Metropolitan City of Naples, with an increase of approximately 25% under both the *All* and *3S* configurations (Figure 13b). These results reflect the regression models' improved ability to capture complex rainfall patterns once trained on a noise-free, rain-focused dataset. Although IDW and OK lack the flexibility of machine learning algorithms, the removal of non-rain instances helped reduce variability and improve their overall predictive reliability.

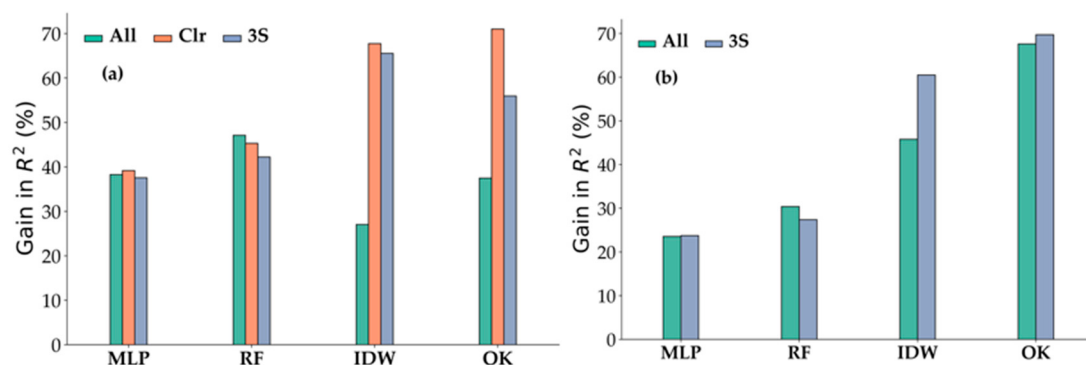


Figure 13. Relative gain in R^2 (%) of the two-step over direct approach. Calore River basin showing all tested scenarios (*All*, *Clr*, *3S*) (a); Metropolitan City of Naples showing the *All* and *3S* scenarios (b).

Under the DA, the IDW and OK methods produced negative R^2 scores for some stations (Luogosano and Pago Veiano in the Calore River basin, and Pompei in the Metropolitan City of Naples), under the *Clr* and *3S* scenarios. This fact, which indicates that the model's predictions are worse than the mean of the observed data [56], happens because spatial interpolation models struggle to accurately capture the dynamic nature of rainfall patterns. Nonetheless, the results for these stations were significantly improved when using the two-step approach, confirming the usefulness of the initial classification step in stabilizing interpolation performance even in challenging conditions.

Among all models and both approaches, the *All*-spatial configuration generally yielded improved or equivalent performance compared to the *3S* and *Clr* scenarios, especially while using TSA. In the Calore River basin, the *All* scenario consistently achieved the highest or near-highest R^2 and lowest RMSE values across models. This outcome highlights the value of leveraging a larger and spatially diverse network of stations, particularly in an orographically complex region where localized rainfall patterns may be missed by a small subset of stations. While the *3S* strategy sometimes approached the performance of the *All* configuration with the TSA, the inclusion of the entire available spatial information generally gave a slight advantage.

4. Discussion

The results presented clearly demonstrate the value of incorporating a classification step before regression for the imputation of missing rainfall data in high-resolution rainfall time series (10-min step). The numerical idea behind this approach is to alleviate the regressors' task by only forcing the imputation of rainy time intervals and avoiding the noise introduced by numerous no-rain intervals, which are present at short time scales due to the intermittency of rainfall phenomena. This is further supported by considering the fundamental physical differences between rainy and no-rain time intervals. During rainy periods, atmospheric processes such as atmospheric lifting, cloud formation, moisture condensation, and precipitation dynamics, dominate. In contrast, dry intervals are governed by entirely different dynamics, including evaporation, atmospheric stability, and the absence of significant convective activity. The differences between these physical regimes make the sets of rainy and dry intervals inherently distinct and non-interchangeable.

4.1. The Classification Step

In this context, the random forest model exhibited consistently high performance in classifying rainy and dry periods across all stations for both the Calore River basin and the Metropolitan City of Naples, even without resorting to the use of additional covariates. A short comparison with the existing literature is instructive. Average *Accuracy* scores for the classification of rain and no-rain time intervals were 0.928 for the Calore River basin and 0.946 for the Metropolitan City of Naples. Average *Precision* scores also remained high, 0.889 for the Calore River basin, and 0.872 for the Metropolitan City of Naples, indicating that the classification model was effective at minimizing false positives. Although the average *Recall* score was lower, 0.334 for the Calore River basin and 0.345 for the Metropolitan City of Naples, the classifier could, however, succeed in isolating a representative subset of rainy time intervals. Finally, the average *F1* score and weighted *F1* score were 0.483 and 0.961, respectively, in the Calore River basin, while their values were 0.494 and 0.957, respectively, in the Metropolitan City of Naples.

In contrast, information about precipitation, air pressure and temperature, relative humidity, wind speed and direction, volumetric water content, with a 30 min step, were integrated in [14] to obtain a weighted *F1* score of 0.938, an *Accuracy* of 0.943, a *Recall* of 0.648, and a *Precision* of 0.847 (averaged on all the stations). Similarly, daily meteorological data (precipitation, humidity, air pressure and temperature, and wind speed) were integrated in [33] to obtain 0.68 for *F1* score, 0.83 for *Accuracy*, 0.64 for *Recall*, and 0.77 for *Precision*. From the comparison with the existing literature, it is evident that at least under a typical Mediterranean climate, it is possible to obtain good performance in the classification step even within rainfall series with a 10-min time step, without resorting to covariates.

4.2. The Regression Step

Overall, the combination of a high-performing classifier and regression models led to more accurate and stable rainfall predictions across both regions and all modeling approaches. The improvements were especially significant in the Calore River basin due to its denser station network, but they also extended to the coastal environment of Naples. The highest performance in the regression step was obtained by the RF in the Calore River basin with an R^2 of 0.541 and an *RMSE* of 0.109 mm, considering a spatial configuration including all stations. These results align well with those presented in [14], where the overall neural network performance was characterized by an R^2 of 0.66 and *RMSE* of 0.141 mm. Similarly, an R^2 value of 0.53 was obtained in [33].

The consistent superiority of the *All*-station configuration in the present work further underscores the importance of maximizing spatial information when it is available, particularly when paired with a method that filters the training data.

4.3. Rainstorm Space-Time Dynamics

Independent of the spatial configuration strategy and the presence of the initial classification step about rain/no-rain intervals, the machine learning regressors (RF and MLP) had sufficient elasticity to consider input data from time intervals that differed from the targeted time level, allowing the internal temporal structure of the rainfall event to be exploited. Again, the numerical approach has an underlying physical justification. In fact, it is expected that data from a targeted station hit by a precipitative cluster during the targeted time interval are well correlated with rainfall data from the positions along the cluster trajectory at the instants of the cluster passage. In other words, targeted data may be better correlated with data from distant positions at delayed time intervals, if these data are all connected by the passage of the same precipitative cluster, and may be less correlated with data at the same targeted time interval from surrounding positions not lying along the cluster trajectory.

Figure 14 contrasts the use of a 21-time interval window with the use of data from the targeted time interval only, within the RF two-step approach for both the *All* and *3S* scenarios (area of Calore River basin). Incorporating this temporal context only slightly increased the median R^2 values but considerably reduced the dispersion of results across stations, indicating more accurate and consistent predictions. By considering a time window centered on the target time step, the model leverages the temporal structure of the data, improving its ability to estimate missing values.

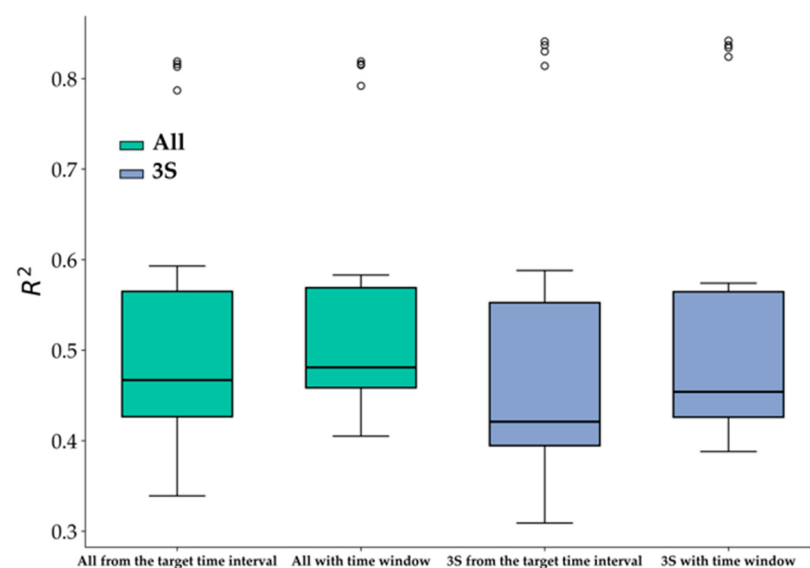


Figure 14. Comparison of R^2 values for two-step Random Forest models, with a 21-time interval window and with only the targeted time interval, across the *All* and *3S* scenarios for the Calore River basin.

4.4. Limitations

Despite the encouraging results obtained in a data-scarce environment, it is important to acknowledge that the regression metrics obtained, particularly the R^2 values, leave room for further enhancement. The predictive capacity of the models could likely be improved by incorporating additional meteorological variables, such as temperature, relative humidity, wind speed, wind direction, and atmospheric pressure, as input features [14,33]. These covariates could provide additional contextual information that influences rainfall

formation and intensity, allowing models to better distinguish spatial and temporal rainfall variability beyond what can be captured by precipitation data alone. The inclusion of covariates would be particularly beneficial in machine learning frameworks such as RF and MLP, which are well-suited to handling multivariate and nonlinear relationships.

Finally, the analysis involved two different areas, a coastal one and a mountainous internal one, of a single region (Campania, Italy). Although this fact does not undermine the results of the present study, future analysis could benefit from procedure verification in additional areas, characterized by different topographies and climates.

5. Conclusions

Complete sub-hourly rainfall datasets are required for practical and theoretical purposes, among which are flood modeling, real-time forecasting, and the understanding of short-duration rainfall extremes. Nonetheless, these datasets often contain missing values due to sensor or transmission failures, implying that filling data gaps in rainfall data series is an important preprocessing activity. In this study, two distinct missing data imputation approaches were compared:

- (a) The direct approach (DA) operates on raw data to directly feed one of the selected imputation models (Multilayer Perceptron, MP, Random Forest, RF, Inverse Distance Weighting, IDW, or Ordinary Kriging, OK), while.
- (b) The two-step approach (TSA) first classifies time steps as rain or no-rain with a RF classifier and subsequently applies one of the selected imputation models to the predicted rainfall depth instances classified as rain.

The two approaches were separately applied to three different rainfall gauge spatial configurations, namely all the available stations, the three best correlated stations, and a cluster of well-correlated stations. The results show that:

- (i) The two-step approach exhibited improved accuracy (with a minimum relative gain in R^2 of 25% over the direct approach), showing that missing rainfall data imputation benefits from breaking the process in stages where different physical conditions such as rain and no-rain are separately managed.
- (ii) Machine learning models (MLP and RF) exhibited improved results with respect to interpolation methods (IDW and OK), thanks to their ability to approximate any continuous function and to learn complicated internal data structures. The highest performance in the regression step was achieved by the RF with an R^2 of 0.541 and an RMSE of 0.109 mm.
- (iii) The spatial configuration that considered all the available stations yielded improved or equivalent performance with respect to the other scenarios, underscoring the importance of maximizing spatial information when it is available and providing practical guidance for similar applications.
- (iv) The use of time intervals lagged with respect to the targeted time intervals, which was easily carried out within machine learning models, allowed for the reconstruction of the dynamics occurring between different rainfall gauges and the tracking of precipitative clusters.

The methodology was tested in a region representative of the Mediterranean climate, characterized by heterogeneous rainfall formation phenomena and strong seasonality. In principle, the approach can be applied to different temporal resolutions (e.g., 30 min, hourly, or daily) and replicated in other regions characterized by diverse climatic and topographic conditions. Future research could further explore the performance and adaptability of the proposed approach under different climatic conditions and precipitation regimes to better assess its transferability to diverse contexts.

In conclusion, the findings of the present study highlight the benefits derived by adapting the imputation approach to the intermittent nature of data (two-step approach) and its space–time dynamics (sliding time window). Although further refinement is still required to improve the accuracy of precipitation predictions, this confirms the value of machine learning methods for recovering missing sub-hourly rainfall data, even in data-scarce contexts where rain gauges may be the only source of high-resolution rainfall data.

Author Contributions: Conceptualization, M.B. and L.C.; methodology, M.B., Ç.A.İ., and L.C.; software, M.B.; validation, M.B.; formal analysis, M.B.; investigation, M.B.; data curation, M.B.; writing—original draft preparation, M.B., Ç.A.İ., G.V., L.C., and R.D.M.; writing—review and editing, M.B., Ç.A.İ., G.V., L.C., and R.D.M.; visualization, M.B. and G.V.; supervision, L.C.; project administration, L.C. and R.D.M.; funding acquisition, L.C. and R.D.M. All authors have read and agreed to the published version of the manuscript.

Funding: The work of Giada Varra is co-funded by the “Southern Apennines River Basin District Authority” through the technical-scientific collaboration agreement with the Engineering Department of the University of Naples “Parthenope” [Supporto tecnico-scientifico in merito alla “Attività di pianificazione tendente al riequilibrio dei processi naturali”, di cui alla Linea 5 del Fondo per lo Sviluppo e la Coesione 2014–2020—PED Acque (CUP-F52G16000010001)]. The work of Luca Cozzolino and Renata Della Morte is co-funded by the project PARADIGM in the framework of “Multi-Risk science for resilient communities under a changing climate (RETURN)—Spoke 5 TS1 “Insediamenti Urbani e Metropolitan” (CUP E63C22002000002), which is part of the NextGeneration EU program financed by the European Union.

Data Availability Statement: The sub-hourly rainfall dataset used in this study is available upon request through the Campania Region Civil Protection agency website (<https://centrofunzionale.regione.campania.it/#/pages/dashboard>, accessed on 1 May 2024). Topographic information was obtained from the ETOPO 2022 global relief model (15 arc-second resolution) for the larger-scale views (<https://www.ncei.noaa.gov/products/etopo-global-relief-model>, accessed on 31 July 2025) and from the TINITALY Digital Terrain Model (DTM; 10 m resolution) produced by the Istituto Nazionale di Geofisica e Vulcanologia (INGV) for the detailed panels (<https://tinitaly.pi.ingv.it/>, accessed on 1 July 2024).

Acknowledgments: Mohamed Boukdire gratefully acknowledges the PhD program in Defense against natural risks and ecological transition of built environment at the University of Catania (Department of Civil Engineering and Architecture—DICAR) for funding his work, and the Department of Engineering at the University of Naples Parthenope for hosting the research activities. All authors are grateful to Vincenzo Capozzi for fruitful discussions. All authors want to acknowledge the Editor and the Referees for their constructive comments that improved the manuscript’s original version.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

MLP	Multilayer perceptron
RF	Random forest
IDW	Inverse distance weighting
OK	Ordinary kriging
PCC	Pearson correlation coefficient
DTM	Digital terrain model
INGV	National Institute of Geophysics and Volcanology
DA	Direct approach
TSA	Two-step approach

All	All stations
Clr	Cluster of stations
3S	The three most correlated stations with the target station

References

- Berne, A.; Delrieu, G.; Creutin, J.D.; Obled, C. Temporal and spatial resolution of rainfall measurements required for urban hydrology. *J. Hydrol.* **2004**, *299*, 166–179. [\[CrossRef\]](#)
- Huang, Y.; Bárdossy, A.; Zhang, K. Sensitivity of hydrological models to temporal and spatial resolutions of rainfall data. *Hydrol. Earth Syst. Sci.* **2019**, *23*, 2647–2663. [\[CrossRef\]](#)
- Singleton, A.; Toumi, R. Super-Clausius–Clapeyron scaling of rainfall in a model squall line. *Quart. J. R. Meteor. Soc.* **2012**, *139*, 334–339. [\[CrossRef\]](#)
- Hosseinzahtalei, P.; Tabari, H.; Willems, P. Climate change impact on short-duration extreme precipitation and intensity–duration–frequency curves over Europe. *J. Hydrol.* **2020**, *590*, 125249. [\[CrossRef\]](#)
- Prein, A.F.; Rasmussen, R.M.; Ikeda, K.; Liu, C.; Clark, M.P.; Holland, G.J. The future intensification of hourly precipitation extremes. *Nat. Clim. Change* **2017**, *7*, 48–52. [\[CrossRef\]](#)
- Meira, M.A.; Freitas, E.S.; Coelho, V.-H.R.; Tomasella, J.; Fowler, H.J.; Ramos Filho, G.M.; Silva, A.L.; Almeida, C.d.N. Quality control procedures for sub-hourly rainfall data: An investigation in different spatio-temporal scales in Brazil. *J. Hydrol.* **2022**, *613A*, 128358. [\[CrossRef\]](#)
- Blenkinsop, S.; Lewis, E.; Chan, S.C.; Fowler, H.J. Quality-control of an hourly rainfall dataset and climatology of extremes for the UK. *Int. J. Climatol.* **2017**, *37*, 722–740. [\[CrossRef\]](#)
- Westerberg, I.; Walther, A.; Guerrero, J.L.; Coello, Z.; Halldin, S.; Xu, C.-Y.; Chen, D.; Lundin, L.-C. Precipitation data in a mountainous catchment in Honduras: Quality assessment and spatiotemporal characteristics. *Theor. Appl. Climatol.* **2010**, *101*, 381–396. [\[CrossRef\]](#)
- Sattari, M.T.; Rezazadeh-Joudi, A.; Kusiak, A. Assessment of different methods for estimation of missing data in precipitation studies. *Hydrol. Res.* **2017**, *48*, 1032–1044. [\[CrossRef\]](#)
- Pinthong, S.; Ditthakit, P.; Salaeh, N.; Hasan, M.A.; Son, C.T.; Linh, N.T.T.; Islam, S.; Yadav, K.K. Imputation of missing monthly rainfall data using machine learning and spatial interpolation approaches in Thale Sap Songkhla River Basin, Thailand. *Environ. Sci. Pollut. Res.* **2024**, *31*, 54044–54060. [\[CrossRef\]](#)
- Famarzadeh, M.; Ehsani, M.R.; Akbari, M.; Rahimi, R.; Moghaddam, M.; Behrangi, A.; Klöve, B.; Haghighi, A.T.; Oussalah, M. Application of Machine Learning and Remote Sensing for Gap-filling Daily Precipitation Data of a Sparsely Gauged Basin in East Africa. *Environ. Process.* **2023**, *10*, 8. [\[CrossRef\]](#)
- Wangwongchai, A.; Waqas, M.; Dechpichai, P.; Hlaing, P.T.; Ahmad, S.; Humphries, U.W. Imputation of missing daily rainfall data; A comparison between artificial intelligence and statistical techniques. *MethodsX* **2023**, *11*, 102459. [\[CrossRef\]](#)
- Schleiss, M.; Jaffrain, J.; Berne, A. Statistical analysis of rainfall intermittency at small spatial and temporal scales. *Geophys. Res. Lett.* **2011**, *38*, L18403. [\[CrossRef\]](#)
- Chivers, B.D.; Wallbank, J.; Cole, S.J.; Sebek, O.; Stanley, S.; Fry, M.; Leontidis, G. Imputation of missing sub-hourly precipitation data in a large sensor network: A machine learning approach. *J. Hydrol.* **2020**, *588*, 125126. [\[CrossRef\]](#)
- van Leth, T.C.; Leijnse, H.; Overeem, A.; Uijlenhoet, R. Rainfall Spatiotemporal Correlation and Intermittency Structure from Micro- γ to Meso- β Scale in the Netherlands. *J. Hydrometeorol.* **2021**, *22*, 2227–2240. [\[CrossRef\]](#)
- Singh, A.; Singh, V.; Gaurav, K. Leveraging neural operators and sliding window technique for enhanced subsurface soil moisture imputation under diverse precipitation scenarios. *J. Geophys. Res. Mach. Learn. Comp.* **2025**, *2*, e2025JH000730. [\[CrossRef\]](#)
- Pappas, C.; Papalexiou, S.M.; Koutsoyiannis, D. A quick gap filling of missing hydrometeorological data. *J. Geophys. Res. Atmos.* **2014**, *119*, 9290–9300. [\[CrossRef\]](#)
- Chen, T.; Ren, L.; Yuan, F.; Yang, X.; Jiang, S.; Tang, T.; Liu, Y.; Zhao, C.; Zhang, L. Comparison of spatial interpolation schemes for rainfall data and application in hydrological modelling. *Water* **2017**, *9*, 342. [\[CrossRef\]](#)
- Yang, R.; Xing, B. A comparison of the performance of different interpolation methods in replicating rainfall magnitudes under different climatic conditions in Chongqing province (China). *Atmosphere* **2021**, *12*, 1318. [\[CrossRef\]](#)
- Workneh, H.T.; Chen, X.; Ma, Y.; Bayable, E.; Dash, A. Comparison of IDW, Kriging and orographic based linear interpolations of rainfall in six rainfall regimes of Ethiopia. *J. Hydrol. Reg. Stud.* **2024**, *52*, 101696. [\[CrossRef\]](#)
- Teegavarapu, R.S.V.; Tufail, M.; Ormsbee, L. Optimal functional forms for estimation of missing precipitation data. *J. Hydrol.* **2009**, *374*, 106–115. [\[CrossRef\]](#)
- Fagandini, F.; Todaro, V.; Tanda, M.G.; Pereira, J.L.; Azevedo, L.; Zanini, A. Missing Rainfall Daily Data: A Comparison Among Gap-Filling Approaches. *Math. Geosci.* **2024**, *56*, 191–217. [\[CrossRef\]](#)

23. Helmi, A.M.; Elgamal, M.; Farouk, M.I.; Abdelhamed, M.S.; Essawy, B.T. Evaluation of Geospatial Interpolation Techniques for Enhancing Spatiotemporal Rainfall Distribution and Filling Data Gaps in Asir Region, Saudi Arabia. *Sustainability* **2023**, *15*, 14028. [\[CrossRef\]](#)
24. Borges, P.d.A.; Franke, J.; da Anunciação, Y.M.T.; Weiss, H.; Bernhofer, C. Comparison of spatial interpolation methods for the estimation of precipitation distribution in Distrito Federal, Brazil. *Theor. Appl. Climatol.* **2016**, *123*, 335–348. [\[CrossRef\]](#)
25. Oriani, F.; Stisen, S.; Demirel, M.C.; Mariethoz, G. Missing Data Imputation for Multisite Rainfall Networks: A Comparison between Geostatistical Interpolation and Pattern-Based Estimation on Different Terrain Types. *J. Hydromet.* **2020**, *21*, 2325–2341. [\[CrossRef\]](#)
26. Richman, M.B.; Trafalis, T.B.; Adrianto, I. Missing Data Imputation Through Machine Learning Algorithms. In *Artificial Intelligence Methods in the Environmental Sciences*; Haupt, S.E., Pasini, A., Marzban, C., Eds.; Springer: Dordrecht, The Netherlands, 2009. [\[CrossRef\]](#)
27. Sattari, M.T.; Falsafian, K.; Irvem, A.; Shahab, S.; Qasem, S.N. Potential of kernel and tree-based machine-learning models for estimating missing data of rainfall. *Eng. Appl. Comput. Fluid Mech.* **2020**, *14*, 1078–1094. [\[CrossRef\]](#)
28. Bellido-Jiménez, J.A.; Gualda, J.E.; García-Marín, A.P. Assessing Machine Learning models for gap filling daily rainfall series in a semiarid region of Spain. *Atmosphere* **2021**, *12*, 1158. [\[CrossRef\]](#)
29. Lupi, A.; Luppichini, M.; Barsanti, M.; Bini, M.; Giannecchini, R. Machine learning models to complete rainfall time series databases affected by missing or anomalous data. *Earth. Sci. Inform.* **2023**, *16*, 3717–3728. [\[CrossRef\]](#)
30. Praveen Kumar, G.; Dwarakish, G.S. Comparison of the multiple imputation approaches for imputing rainfall data: A humid tropical river basin case study. *Water Conserv. Sci. Eng.* **2025**, *10*, 87. [\[CrossRef\]](#)
31. Reichstein, M.; Camps-Valls, G.; Stevens, B.; Jung, M.; Denzler, J.; Carvalhais, N.; Prabhat. Deep learning and process understanding for data-driven Earth system science. *Nature* **2019**, *566*, 195–204. [\[CrossRef\]](#)
32. Pastorini, M.; Rodríguez, R.; Etcheverry, L.; Castro, A.; Gorgoglione, A. Enhancing environmental data imputation: A physically-constrained machine learning framework. *Sci. Total Environ.* **2024**, *926*, 171773. [\[CrossRef\]](#)
33. Han, H.; Kim, B.; Kim, K.; Kim, D.; Kim, H.S. Machine learning approach for the estimation of missing precipitation data: A case study of South Korea. *Water Sci. Technol.* **2023**, *88*, 556–571. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Abraham, Z.; Tan, P.N. A semi-supervised framework for simultaneous classification and regression of zero-inflated time series data with application to precipitation prediction. In Proceedings of the ICDM Workshops 2009—IEEE International Conference on Data Mining, Miami, FL, USA, 6 December 2009; pp. 644–649. [\[CrossRef\]](#)
35. Simolo, C.; Brunetti, M.; Maugeri, M.; Nanni, T. Improving estimation of missing values in daily precipitation series by a probability density function-preserving approach. *Int. J. Climatol.* **2010**, *30*, 1564–1576. [\[CrossRef\]](#)
36. Teegavarapu, R.S.V.; Aly, A.; Pathak, C.S.; Ahlquist, J.; Fuelberg, H.; Hood, J. Infilling missing precipitation records using variants of spatial interpolation and data-driven methods: Use of optimal weighting parameters and nearest neighbour-based corrections. *Int. J. Climatol.* **2018**, *38*, 776–793. [\[CrossRef\]](#)
37. Vidal-Paz, J.; Rodríguez-Gómez, B.A.; Orosa, J.A. A Comparison of Different Methods for Rainfall Imputation: A Galician Case Study. *Appl. Sci.* **2023**, *13*, 12260. [\[CrossRef\]](#)
38. Vitale, S.; Ciarcia, S. Tectono-stratigraphic setting of the Campania region (Southern Italy). *J. Maps* **2018**, *14*, 9–21. [\[CrossRef\]](#)
39. Cuomo, A.; Guida, D.; Palmieri, V. Digital orographic map of peninsular and insular Italy. *J. Maps* **2011**, *7*, 447–463. [\[CrossRef\]](#)
40. Capozzi, V.; Annella, C.; Budillon, G. Classification of daily heavy precipitation patterns and associated synoptic types in the Campania Region (southern Italy). *Atmos. Res.* **2023**, *289*, 106781. [\[CrossRef\]](#)
41. Pelosi, A.; Furcolo, P. An amplification model for the regional estimation of extreme rainfall within orographic areas in Campania region (Italy). *Water* **2015**, *7*, 6877–6891. [\[CrossRef\]](#)
42. Avino, A.; Manfreda, S.; Cimorelli, L.; Pianese, D. Trend of annual maximum rainfall in Campania region (Southern Italy). *Hydrol. Process.* **2021**, *35*, e14447. [\[CrossRef\]](#)
43. NOAA. *ETOPO 2022 15 Arc-Second Global Relief Model*; NOAA National Centers for Environmental Information: Asheville, NC, USA, 2022. [\[CrossRef\]](#)
44. Tarquini, S.; Isola, I.; Favalli, M.; Battistini, A.; Dotta, G. *TINITALY, a Digital Elevation Model of Italy with a 10 Meters Cell Size (Version 1.1)*; Istituto Nazionale di Geofisica e Vulcanologia (INGV): Roma, Italy, 2023.
45. Varra, G.; Della Morte, R.; Tartaglia, M.; Fiduccia, A.; Zammuto, A.; Agostino, I.; Booth, C.A.; Quinn, N.; Lamond, J.E.; Cozzolino, L. Flood Susceptibility Assessment for Improving the Resilience Capacity of Railway Infrastructure Networks. *Water* **2024**, *16*, 2592. [\[CrossRef\]](#)
46. Varra, G.; İnan, Ç.A.; Della Morte, R.; Tartaglia, M.; Fiduccia, A.; Zammuto, A.; Agostino, I.; Cozzolino, L. Assessment of direct rainfall and flood-induced damage to land transport infrastructure using two-dimensional HEC-RAS 6.6 rain-on-grid simulations. *Nat. Hazards* **2025**, *121*, 17615–17645. [\[CrossRef\]](#)
47. Magliulo, P.; Cusano, A. Geomorphology of the Lower Calore River alluvial plain (Southern Italy). *J. Maps* **2016**, *12*, 1119–1127. [\[CrossRef\]](#)

48. Magliulo, P.; Sessa, S.; Cusano, A.; Beatrice, M.; Giannini, A.; Russo, F. Assessing the Morphological Quality of the Calore River (Southern Italy). *Geographies* **2022**, *2*, 354–378. [CrossRef]
49. CUGRI. *Il Sistema Fognario Della Città di Napoli alle Soglie del 2000. Parte Prima*; CUEN: Napoli, Italy, 2000.
50. Ho, T.K. Random Decision Forests. In Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, Canada, 14–16 August 1995; pp. 278–282. Available online: <https://web.archive.org/web/20160417030218/http://ect.bell-labs.com/who/tkh/publications/papers/odt.pdf> (accessed on 1 December 2024).
51. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
52. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533–536. [CrossRef]
53. Campozano, L.; Tenelanda, D.; Sanchez, E.; Samaniego, E.; Feyen, J. Comparison of Statistical Downscaling Methods for Monthly Total Precipitation: Case Study for the Paute River Basin in Southern Ecuador. *Adv. Meteorol.* **2016**, *2016*, 6526341. [CrossRef]
54. İnan, Ç.A.; Artigue, G.; Kurtulus, B.; Pistre, S.; Johannet, A. A Hydrological Digital Twin by Artificial Neural Networks for Flood Simulation in Gardon de Sainte-Croix Basin, France. *IOP Conf. Ser. Earth Environ. Sci.* **2021**, *906*, 012112. [CrossRef]
55. Tan, Y.X.; Ng, J.L.; Huang, Y.F. Estimation of missing daily rainfall during monsoon seasons for tropical region: A comparison between ann and conventional methods. *Carpathian J. Earth Environ. Sci.* **2020**, *15*, 103–112. [CrossRef]
56. Chicco, D.; Warrens, M.J.; Jurman, G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Comput. Sci.* **2021**, *7*, e623. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.