

Automated novel real-time framework for rainfall data imputation in flood early warning systems

Farzad Piadeh^{a,*}, Vahid Bakhtiari^b, Farshad Piadeh^c

^a Centre for Engineering Research, School of Physics, Engineering, Computer Science, University of Hertfordshire, Hatfield, AL10 9AB, UK

^b School of Architecture & Built Environment, Faculty of Science Engineering & Built Environment, Deakin University, Geelong, VIC, 3220, Australia

^c School of Computer Engineering, Islamic Azad University of Mashhad, Ostad Yousefi Blvd., Mashhad, 91871-47578, Iran

ARTICLE INFO

Keywords:

Early warning systems
Ensemble data mining modelling
Event identification
External Benchmark
Missing data
Real-time data imputation

ABSTRACT

Real-time flood warning systems play a crucial role in mitigating impacts of flooding. However, their performance is highly dependent on input data, which can often contain missing values. While data imputation techniques have been widely applied in pre-processing stages, their integration into real-time operations remains underexplored. This study presents a real-time automated decision support system that integrates a soft-voting stacked data mining ensemble model comprising decision tree, K-nearest neighbour, Naive Bayes, Neural Network, Support Vector Machine, Discriminant Analysis, and Gaussian Regression. The system also incorporates hydrological-hydraulic event identification, external benchmarking, and a multi-data fuzzy weighted spatial imputation framework. The effectiveness of the proposed method was evaluated through a real-world case study involving a flood early warning system in an urban drainage network in London, UK. Comparative analyses were conducted against well-established artificial intelligence model, and a sensitivity analysis was performed for further assessment. Results showed that all types of missing data were correctly identified with a precision exceeding 90 % and were accurately imputed - particularly in situations where other models failed to recognise current rainfall values during the onset, peak, and falling limb of events (with no reduction in accuracy compared to the best-performing benchmark models). For the 3-h-ahead flood forecasting, the proposed method reduced the normalised root mean square error by up to 30 % compared to alternative approaches. To ensure the generalisability of the approach, additional locations across the UK were used for validation, which demonstrates the stability and robustness of the system, with only minor error variations.

1. Introduction

Urban areas worldwide increasingly face the challenge of managing excess rainfall and mitigating flood risks (Giroto et al., 2024). Efficient urban drainage systems (UDS) are critical for preventing water accumulation that can cause property damage, traffic disruptions, and threats to public safety (Piadeh et al., 2023a). As the impacts of climate change intensify and urban development expands, these systems become more vulnerable to inundation during extreme rainfall events (Piadeh et al., 2023b). In this context, early warning systems (EWS), particularly those focused on real-time urban flood forecasting (RTUFF), play a vital role in strengthening the resilience of UDS and mitigating the adverse effects of flooding (Li and Burian, 2023). RTUFF models offer a dynamic, data-driven solution by continuously integrating and analysing information from weather stations, river gauges, and other monitoring

sources (Piadeh et al., 2022). These systems deliver real-time assessments of changing rainfall patterns and water levels which equips urban planners, emergency responders, and residents with timely, actionable insights to support informed decision-making during potential flood events (Bakhtiari et al., 2024).

The performance of EWS heavily relies on data availability, as missing or incomplete data can significantly impact their accuracy (Fang et al., 2020). RTUFF typically depend on hydrometeorological variables, particularly rainfall data, collected from weather and gauging stations (Duarte et al., 2022). In contemporary applications, these stations are increasingly adopting Internet of Things (IoT)-based devices to enhance spatial coverage, data frequency, and accessibility. While this transition offers substantial benefits, it also introduces several operational challenges (Bakhtiari et al., 2025). Data transmission from IoT sensors and open-access APIs can be affected by latency, temporary outages, or

* Corresponding author.

E-mail address: f.piadeh@herts.ac.uk (F. Piadeh).

<https://doi.org/10.1016/j.engappai.2025.113348>

Received 18 August 2025; Received in revised form 18 November 2025; Accepted 24 November 2025

Available online 30 November 2025

0952-1976/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

incomplete uploads caused by network interruptions or maintenance issues (Brunner et al., 2021). Additionally, temporal mismatches between rainfall and water-level records may occur during real-time acquisition, necessitating post-processing to ensure synchronisation and continuity (Kamwaga et al., 2018). These constraints highlight the importance of implementing robust data quality control, redundancy protocols, and automated error-handling mechanisms within the real-time framework to maintain the reliability and predictive performance of the EWS (Anbarasan et al., 2020). Furthermore, real-time data infilling is especially important in areas that are prone to urban flash flooding, where flood events can occur suddenly with little time lead for warning. In these situations, accurate forecasts can be the difference between life and death (Bakhtiari et al., 2023). Therefore, research and development of effective real-time data infilling techniques is critical to improve the performance of RTUFF systems and reducing the impact of flood events on communities and the environment (Umar and Gray, 2023).

Addressing these challenges highlights the role of automated missing data imputation methods and their vital contributions in maintaining the reliability and effectiveness of real-time EWS as well as delivering accurate warnings to mitigate potential hazards (Ben Aissia et al., 2017). While there is no clear consensus on the best approach to infill missing data of rainfall for flood EWS systems, several imputation methods have been tested. The initial solutions primarily involved relatively simple techniques, including linear regressions, inverse distance weighting, moving-median substitution, and nonlinear interpolation methods such as Kriging (Bárdossy and Pegram, 2014; Wangwongchai et al., 2023).

Although These methods can perform well for short-term or isolated data gaps of flood EWS systems, their accuracy diminishes for long-term imputation or dynamically evolving rainfall conditions (Bárdossy and Pegram, 2014; Ding et al., 2020; Lv et al., 2020). Consequently, as shown in Table 1, data-driven models have emerged as promising alternatives. Simpler models such as support vector machine (SVM), self-organising map, decision tree (DT), Gaussian process regression (GPR), K-Nearest Neighbourhood (KNN), multi-linear regression, and principal component analysis have been employed effectively for multistep data infilling, especially in historical datasets (Sattari et al., 2020; Boulin et al., 2022; Tavares et al., 2025). However, most of these techniques have been validated only on retrospective or post-event datasets, limiting their real-time applicability.

Recent works have expanded this field using advanced data-driven frameworks. More robust and adaptive techniques, such as feed-forward neural networks, convolutional neural networks, adaptive neuro-fuzzy inference systems, ensemble learning, and hybrid deep learning frameworks, have proven superior for extended imputation scenarios involving temporal anomalies and rainfall fluctuations (Papailiou et al., 2022; Lupi et al., 2023; Kumar and Dwarakish, 2025). For example, Sriwahyuni et al. (2025) integrated spatiotemporal attention networks with radar rainfall products to improve urban flood forecasting accuracy under missing-data conditions. Similarly, Li et al. (2024) developed a hybrid CNN–LSTM architecture for dynamic rainfall infilling, showing significant accuracy improvements but at high computational costs. Golkhatmi and Farzandi (2024) proposed an adaptive ensemble approach combining random forest and Bayesian

Table 1
Recent advances on rainfall data imputation methods in the context of flood EWS*.

Application**	Applied method***	Benchmark methods***	Case study	Reference
HDP	ANFIS	ARX	South Africa	Nawaz et al. (2016)
RTO	PCA	SOM, LR, KNN	Malaysia	Miró et al. (2017)
HDP	MLR	LR	Bangladesh	Jahan et al. (2019)
RTO	KNN + Hot deck + LR	–	Nigeria	Aieb et al. (2019)
HDP	PCA-based FFNN	–	Colombia	Canchala-Nastar et al. (2019)
HDP	Ensemble DT	Data augmentation	China	Chen et al. (2019)
HDP	PCA + EMA	LR, PCA	Malaysia	Chuan et al. (2019)
HDP	Bayesian PCA	KNN	Malaysia	Lai et al. (2019a)
HDP	KNN	–	Malaysia	Lai and Kuok (2019b)
RTO	Ensemble DT + LR	KNN, RF, SVM, FFNN	UK	Chivers et al. (2020)
HDP	Vector sampling	IDW, Kriging, KNN	Denmark, Australia, Swiss, Switzerland	Oriani et al. (2020)
HDP	PCA-based FFNN	–	Colombia	Ocampo-Marulanda et al. (2021)
HDP	SVM	GPR, RF	Turkey	Sattari et al. (2020)
HDP	GPR	–	Canada	Boulin et al. (2022)
HDP	ANN + MLR	–	Greece	Papailiou et al. (2022)
HDP	FFNN	FFNNs	India	Kumawat et al. (2023)
HDP	LSTM + CNN	–	Italy	Lupi et al. (2023)
HDP	MLR	MLRs	South Korea	Narimani et al. (2023)
HDP	LR, OK, SCK	LR	Portugal	Fagandini et al. (2024)
HDP	Bayesian LR	Non-Bayesian LR, MICE, PMM, UMI, RS	Iran	Golkhatmi and Farzandi (2024)
HDP	KNN, EMA	HI, LR	Turkey	Hırca and Eryılmaz Türkkan (2024)
HDP	EMA	PCA, RF, MICE	India	Kannegowda et al. (2024)
HDP	GP	MLR, RF, SVR, MLP	Thailand	Pinthong et al. (2024)
HDP	ANFIS	LR, MLR, OK, SVM ANN	Iran	Darlane and Borhan (2025)
HDP	CART	KNN, PMM, RF, LLR, Bayesian LR	India	Kumar and Dwarakish (2025)
HDP	CNN	MLR, SVM	Indonesia	Sriwahyuni et al. (2025)
HDP	SMA	SMA, NRM, IDW, BE	Brazil	Tavares et al. (2025)
HDP	MTGP	UMI, MICE, KNN	Burkina Faso	Zio et al. (2025)

*:The search in the Scopus database covering the last decade (2015–2025) was based on keywords “imputation”, “infilling”, “missing data”, “rainfall”, “rain”, “storm”, “Harkins”, “thunder”, and “flood early warning system” “flooding”. Research works that applied conventional data imputation methods as a pre-processing step of flood forecasting are excluded. Papers were reviewed then to select appropriated journal papers focusing on AI-based data infilling methods.

:HDP: Historical Data Preparation, RTO: Real-Time Operation, ANN: Artificial Neural Network, ANFIS: Adaptive neuro fuzzy inference system, ARX: Autoregressive with eXtra Input, *: BE: Best Estimator, CART: Classification And Regression Trees, CNN: Convolutional Neural Network, DT: Decision Tree, EMA: Expectation Maximisation algorithm, FFNN: Feed Foreword Neural Network, GP: Genetic Programming, GPR: Gaussian Process Regression, HI: Hot-deck Imputation, IDW: Inverse distance weighting, LLR: Lasso Linear Regression, LR: Linear Regression, KNN: K-Nearest Neighbourhood, MICE: Multiple Imputation by Chained Equations, MLR: Multi Linear Regression, MTGP: Multi-Task Gaussian Process, NRM: Normal Ratio Method, OK: Ordinary Kriging, PCA: Principal Component Analysis, PMM: Predictive Mean Matching, RF: Random Forest, RS: Random Sampling, SCK: Simple Cokriging, SMA: Simple Arithmetic Mean, SOM: Self Organising Map, SVM: Support Vector Machine, UMI: Unconditional Mean Imputation.

updating to handle streaming IoT sensor data with latency issues. More recently, [Murakami \(2024\)](#) demonstrated that transformer-based hybrid models can enhance anomaly detection and real-time imputation but require high data redundancy and cloud-based computation. Despite these advances, few studies have explicitly addressed the operational trade-off between accuracy, computational efforts, and scalability under real-time conditions - highlighting the need for more practical frameworks that integrate automated anomaly detection with benchmark-driven imputation strategies.

However, a significant research gap persists most of these methods in the context of data imputation of flood EWS focus on historical data infilling, where both past and future data are available to estimate missing values ([Chuan et al., 2019](#); [Oriani et al., 2020](#)). In contrast, real-time operation imputation, which requires decisions based solely on past and current data, remains under-explored ([Aieb et al., 2019](#); [Chivers et al., 2020](#)). This limitation is particularly critical during the early stages of rainfall events, where delayed or incorrect imputation (e. g., treating actual rainfall as dry weather) can drastically impair flood forecasting performance ([Miró et al., 2017](#); [Aieb et al., 2019](#)). Therefore, developing accurate, context-sensitive, and responsive real-time data imputation strategies is vital for flood EWS, especially in urban areas prone to flash floods.

Moreover, recent studies emphasise the importance of detecting and handling diverse types of data issues, not just NaN gaps. Outliers ([Dunkerle, 2023](#)), anomalies ([Lupi et al., 2023](#)), incorrect zero values ([Sa'adi et al., 2023](#)), and false wet/dry classifications are among the critical challenges impacting EWS accuracy. For example, a scenario where dry weather and no rainfall are erroneously identified as wet conditions due to the absence of explicit data can lead to false alarms ([Sa'adi et al., 2023](#); [Balcha et al., 2023](#)). Conversely, failing to recognise actual wet conditions can result in underestimating flood risk, creating a mismatch between the recorded values and the true state of the environment ([Golkhatmi and Farzandi, 2024](#)). These issues - often caused by sensor lag, hardware faults, or data transmission errors - can severely compromise flood forecasting performance. [Golkhatmi and Farzandi \(2024\)](#) further demonstrated that zero rainfall values recorded during ongoing rainfall events can mislead the model into assuming dry conditions, leading to significant underestimations. To address these complexities, recent advances have introduced hybrid imputation models that combine external benchmarks with regression techniques, including ordinary kriging, simple cokriging, and Bayesian linear regression. These hybrid approaches have demonstrated significant improvements in real-time data accuracy ([Fagandini et al., 2024](#); [Dariane and Borhan, 2025](#)). It should be mentioned that when referring to an external benchmark, its definition can vary depending on the scope of the study. However, there is a general agreement that it involves using rainfall data obtained from a broader range of collection sources. For instance, the external benchmark may include data from neighbouring rain gauges or radar systems for a single specific rain gauge with missing data, that record rainfall over adjacent or overlapping catchment areas ([Dariane and Borhan, 2025](#); [Sriwahyuni et al. 2025](#)).

Additionally, computational efficiency has become a growing concern, especially in large-scale monitoring networks. [Tavares et al. \(2025\)](#) and [Zio et al. \(2025\)](#) highlighted the importance of adaptive, scalable imputation techniques such as multiple imputation by chained equations and multi-task Gaussian process, which offer robust performance under increasing data loss and system complexity. Together, these advancements underscore the critical need for real-time, anomaly-aware, and computationally sustainable imputation frameworks tailored for flood EWSs.

Finally, while some research has explored distinct states of rainfall, such as dry and wet weather or considering nature of rainfall such as monsoon and non-monsoon, there remains a gap in addressing differences between dry and wet weather conditions when applying imputation methods. This oversight has significant implications for both computational efforts and predictive accuracy ([Aieb et al., 2019](#);

[Narimani et al., 2023](#)). In smaller-scale studies with limited rainfall monitoring points, the impact of computational efforts may be negligible. However, when dealing with a larger geographical area, the continuous operation of models in response to zero rainfall values (indicating dry weather conditions) can lead to excessive computational efforts. In such scenarios, more straightforward methods could be initially employed to provide the necessary level of intervention and prediction, optimising computational resources and ensuring efficiency.

Hence, the present study aims to develop a comprehensive decision-making framework that integrates a multi-choice infilling method with the advantages of external benchmark data sources. To address the challenge of missing data in real-time flood EWS, the proposed approach incorporates external benchmark resources to enhance infilling accuracy by leveraging additional knowledge beyond the local dataset. Furthermore, two automated Decision Support Systems (DSS) were developed: one designed to identify all types of missing data—not limited to NaN values—thereby enabling a more robust and comprehensive data quality assessment; and another designed to apply fuzzy weighted spatial interpolation for imputing the missing data. In addition, a data mining-based method is introduced to accurately distinguish between dry and wet weather conditions, a capability that is particularly critical during the early stages of rainfall events when timely and accurate detection can significantly improve forecasting outcomes. Finally, unlike previous models focused solely on accuracy or algorithmic sophistication, this framework prioritises real-time applicability, anomaly awareness, and computational efficiency, advancing the operational reliability of flood EWS for urban resilience and risk management.

2. Methodology

This methodology is exemplified through IoT-based rainfall monitoring systems and application programming interface (API), which are now widely applicable globally. The scope of this approach, initially designed for rainfall data, can be expanded to encompass other real-time stations such as gauge monitoring in weather stations and radar systems. While originally tailored for rainfall data, the concept of handling missing data can also be extended to include other input sources such as temperature, wind speed and direction, or soil moisture. This method specifically addresses numerical/temporal data and does not cover spatial-based image data, such as those obtained from satellite production. The exclusion of spatial-based image data is justified by the fact that, despite the increasing popularity and global application of such images, the primary source for rainfall data collection worldwide remains temporal-based data collection. Therefore, this paper aims to introduce a concept applicable to general scenarios, focusing on the predominant method of rainfall data collection globally.

This study proposes a real-time hybrid imputation method integrated into a real-time platform linked with an EWS. This platform seamlessly incorporates decision support systems, data mining techniques, statistical analysis, and leverages IoT-based and API-based external benchmark rainfall stations as additional resources. Ideally, the benchmark rainfall stations should form a geographical network surrounding the main rainfall station i.e. at least three stations. This enhances the system's understanding by incorporating physical-based information from various locations, contributing to a more comprehensive understanding of the meteorological conditions. Illustrated in [Fig. 1](#), the methodology comprises three key steps: (1) The DSS continuously monitors incoming data streams from IoT-based rain gauge networks and applies multi-criteria anomaly detection rules. These include range checks, temporal consistency analysis, and dynamic quantile-based thresholds derived from local climatological statistics, (2) Ensemble data mining platform classifies rainfall events into dry or wet conditions to guide the imputation strategy. It employs a soft-voting stacked ensemble model integrating several base models. Each base learner contributes probabilistic outputs, which are aggregated to improve classification robustness

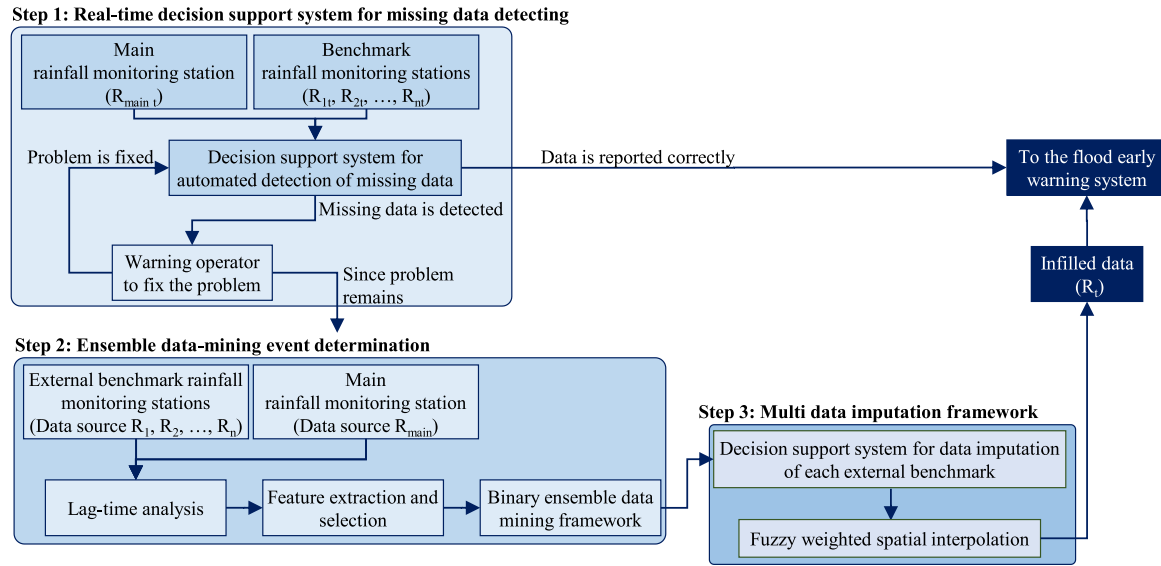


Fig. 1. Real-time data imputation platform for infilling rainfall missing data used for real-time flood early warning systems.

under diverse meteorological scenarios. This classification step is critical for selecting context-appropriate imputation techniques and maintaining forecast accuracy, and (3) a multi-strategy-based framework for data imputation that adapts dynamically to event type and data availability. For dry conditions, linear regression and temporal smoothing are prioritized, while wet conditions invoke advanced probabilistic techniques such as t-Copula and move-median strategies. A key innovation is the Fuzzy Weighted Spatial Interpolation, which integrates fuzzy logic with spatial weighting to handle uncertainty and heterogeneity in rainfall patterns. This assigns membership degrees to candidate stations based on geographic proximity, wind drift, and historical correlation, and then computes adaptive weights to interpolate missing values. This mechanism ensures that spatial variability and physical plausibility are preserved, outperforming traditional deterministic interpolation methods in complex hydrometeorological settings.

This platform initiates by assessing the availability and reliability of the main monitoring rainfall data for each real-time interval, depending on the specifications of the data collection device. This evaluation is conducted through a proposed DSS that uses data obtained from benchmark external rainfall data resources as well as a range of data mining techniques. If missing data is identified, an alarm, including error detection and error type, is sent to the operator for immediate resolution.

Meanwhile, platform will provide missing data required for input of EWS prediction system. To achieve this, the level of required accuracy for imputation is first determined based on predicting the event state, i. e., whether the weather conditions are dry or wet. To accomplish this, a proposed binary ensemble data mining framework, already trained on historical data from both main and external benchmark stations, specifies this state by considering the correlation of each benchmark station with the main station. Ultimately, based on the recognised states, the hybrid method combines several efficient infilling techniques to provide a range of available data. This data is then interpolated using the novel developed kriging with external drafts. Finally, the infilled data is sent to the EWS to be used as part of the input data required for RTUFF.

2.1. Step 1: Real-time decision support system for missing data detection

The proposed DSS framework addresses strategies applied for missing data detection as illustrated in Fig. 2a. The concept of missing data is inspired by Thudumu et al. (2020), encompassing four types of data deficiencies: (1) lack of data or NaN data (Type 1 in Fig. 2b), (2)

zero reported for a non-zero value (indicating dry weather for wet weather - Types 2.1, 2.2, and 2.4 in Fig. 2b), (3) non-zero reported for a zero value (indicating wet weather for dry weather - Type 3.1 Fig. 2b), and (4) abnormal large values for wet weather situations i.e. anomaly (Type 3.4 in Fig. 2b). For each real-time data point (i.e., $R_{main,t}$ in Fig. 2a), the framework defines an appropriate procedure to handle cases where data is not reported or transmitted from the device source.

These processes are shown in detailed in the Algorithm 1, but generally, when the main rainfall station reports a zero value, the proposed DSS verifies its validity by examining both temporal and spatial rainfall patterns. The previous rainfall observation ($R_{main,t-1}$) is first assessed to determine whether the zero represents a genuine dry period or a potential recording error. If the previous value is zero, the system calculates the rainfall network value (RNV) using data from surrounding benchmark stations. A zero value is considered faulty if all benchmarks report rainfall (Type 2.1), while partial rainfall across benchmarks triggers a KNN-based spatial path recognition (inspired by Teegavarapu, 2014; Chiu et al., 2019) to check whether rainfall could have passed over the main station - identifying missing data if a viable path exists (Type 2.2) or confirming correct data otherwise (Type 2.3). When the previous observation is nonzero, the system applies a normalised Continuous Wavelet Transform (CWT) to detect continuity in the rainfall pattern, classifying decreasing wavelet energy as rainfall cessation (Type 2.4) and increasing energy as a false zero during ongoing rainfall (Type 2.5). For positive rainfall readings, the same RNV metric is used to confirm whether wet conditions are consistent across the network. If all benchmark stations indicate dry weather, the value is flagged as missing (Type 3.1). Otherwise, the reported value is validated using historical benchmark correlations (Type 3.2), Isolation Forest analysis (inspired by Gao et al., 2025; Zhang et al., 2024) for joint data plausibility, and Angle-Based Outlier Detection (inspired by De Luca et al., 2024; Liu and Wang, 2024) for novelty evaluation. These models distinguish between overestimated or anomalous rainfall (Type 3.3) and new but credible rainfall conditions (Type 3.4).

To enhance robustness during extreme rainfall conditions, the anomaly detection module incorporates dynamic quantile-based adaptive limits derived from local climatological statistics. Instead of relying on fixed upper thresholds, each observation is compared against a time-varying percentile envelope (typically the 95–99 % percentile), computed from long-term rainfall records using a moving historical window. These adaptive limits allow the system to distinguish physically plausible extreme events from sensor faults, thereby reducing false

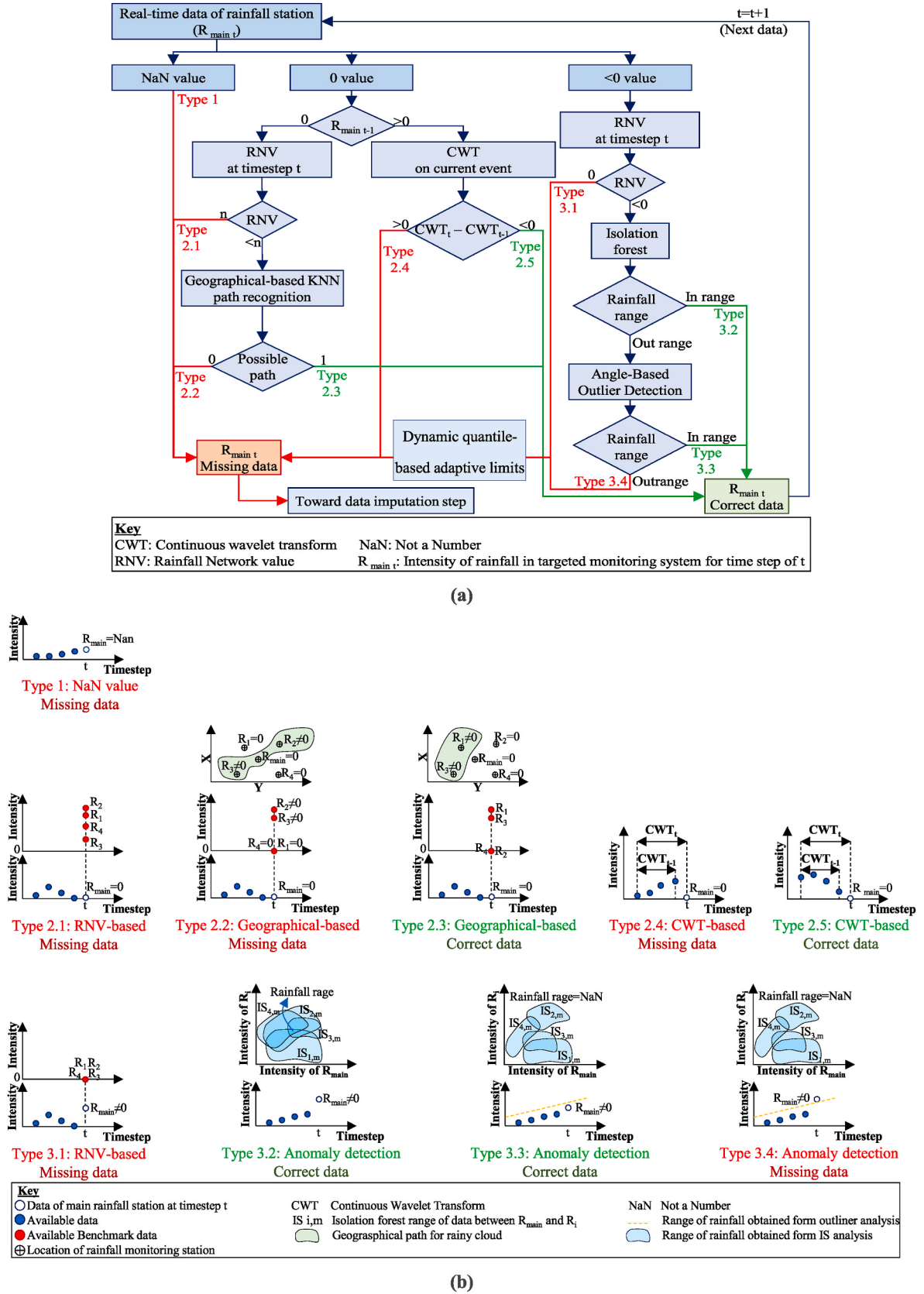


Fig. 2. DSS framework for automated detection of missing data in flood early warning system: (a) workflow of proposed framework, (b) Schematic illustration of various missing and corrected data detection scenarios based on total four external benchmark stations.

positives during high-intensity rainfall and improving the reliability of real-time flood early warning.

Algorithm 1

Pseudocode of DSS framework for automated detection of missing data in flood early warning system

```

% Inputs:
Initiate R_main_t % main station at t (may be NaN)
Initiate R_main_t-1 % main station at t-1 (may be NaN)
Initiate {R_i_t} i = 1..n % benchmark stations at t

% Pretrained/derived (referenced, not implemented here):
Compute RNV({R_i_t}) = % RNV is network rainfall value, Ri_t is the
    ∑_{i=1}^n if (Ri_t ≠ 0) % value of rainfall intensity for ith external
    % benchmark station at timestep t.

Count (R_i_t > 0)
Compute CWT_index: % CWT is continuous wavelet transform for
    ∑_{i=2}^t (R_main_i - R_main_{i-1})^2 % range of data finished by timestep t, R_main_i is
    R_main % the rainfall intensity of main station at
    % timestep i, and R_main is the average intensity of
    % rainfall event.

Path viable: Compute KNN({R_i_t}) % 0/1 feasibility for rain cloud over main
In common rainfall range: Compute Linear regression (R_main_t, {R_i_t})
Isolation forest: Compute inlier ([R_main_t, {R_i_t}])
Angle based outlier: Compute inlier (R_main_t)

% Outputs:
Status = {Correct data, Missing data}
Type ∈ {1, 2.1, 2.2, 2.3, 2.4, 2.5, 3.1, 3.2, 3.3, 3.4}
(optional) Correction → send to imputation step if missing

% Begin
if isNaN(R_main_t):
    Status = Missing data; Type = 1; Goto Imputation
else if R_main_t == 0:
    Compute RNV({R_i_t}) % Two parallel checks: previous step and CWT
    % on current event

    if R_main_t-1 == 0:
        if RNV == n:
            Status = Missing data; Type = 2.1; Goto Imputation
        else if 0 < RNV < n:
            If Path viable = true:
                Status = Missing data; Type = 2.2; Goto Imputation
            else
                Status = Correct data; Type = 2.3; End
        else: %RNV == 0
            Status = Correct data; Type = 2.3; End
    else if R_main_t-1 > 0:
        Compute CWT_t % branch B: CWT change (current event consistency)
        Compute CWT_t-1
        if (CWT_t - CWT_t-1) < 0:
            Status = Correct data; Type = 2.4; End %Rain ended
        else
            Status = Missing data; Type = 2.5; Goto Imputation % Erroneous zero mid-
            event
        Else % unknown R_main_t-1 → fall back to RNV + path
            if RNV == n:
                Status = Missing data; Type = 2.1; Goto Imputation
            else if 0 < RNV < n:
                if path viable == true:
                    Status = Missing data; Type = 2.2; Goto Imputation
                else
                    Status = Correct data; Type = 2.3; End
            else
                Status = Correct data; Type = 2.3; End
    else if R_main_t ≠ 0: %treat as suspect wet reading; validate against network and
    models
        Compute_RNV({R_i_t})
        if RNV == 0:
            Status = Missing data; Type = 3.1; DQBL = True
        Else % validate magnitude
            Isolation forest: Compute inlier ([R_main_t, {R_i_t}])
            if in common rainfall range == true:
                Status = Correct data; Type = 3.2; End
            else
                Angle based outlier: Compute inlier (R_main_t)

```

(continued on next column)

Algorithm 1 (continued)

```

% Inputs:
if angle_based_outlier_is_inlier(R_main_t) == false:
    Status = Missing data; Type = 3.3; DQBL = True
else
    Status = Correct data; Type = 3.4; End
If DQBL == True
    Hist_window ← select H_i = { r_i(τ) | τ ∈ [t-K, t-1] and season(τ) ≈ season(t) }
    if size(Hist_window) < Min_samples_required
        % Fallback to global or longer-term historical distribution if window is too
        small
        Hist_window ← extend_to_longer_history(Hist_main_i)
    Q_i_α_t ← compute_quantile(Hist_window, α) % e.g., 95th–99th percentile
    if R_main_t ≤ Q_i_α_t then
        Status_extreme = Climatologically_plausible
        Flag_anomaly = False
        Return
    else
        Status_extreme = Potential_extreme_or_fault

```

2.2. Ensemble data-mining event determination

The objective of this step is to identify potential states leading to missing data. For instance, missing data may occur during dry weather conditions, allowing for straightforward infilling using simple methods such as linear or non-linear regression. On the other hand, it might occur during rainfall, necessitating more advanced techniques. To address this, a data mining model is developed inspired by Piadeh et al. (2023). For each set of data containing main station and one benchmark station, one stacked ensemble data mining model was developed, helping determine the prevailing weather conditions based on these correlations. To do this, seven weak learner data mining models (detailed provided in Table 2) were selected for their demonstrated potential and widespread use in previous hydrological classification studies (Piadeh et al., 2023b): DT, KNN, Naive Bayes (NB), Neural Network Pattern Recognition (NPR), SVM with Error-Correcting Output, Discriminant Analysis (DA), and GPR. All models were developed and optimised in MATLAB 2025a. Model parameters were tuned using automated hyperparameter optimisation, minimising the 5-fold cross-validation loss over 30 iterations (demonstrated for the one station as an example in Fig. A1 in the Appendix). To ensure equal representation, all identified events were randomly distributed across the training, validation, and testing databases.

A distinct dataset needs to be initially prepared using historical data and transformed into the required features. The set of required data, specifically the necessary timesteps for rainfall intensity, is not uniform across all models. This variability arises because the correlation of each station may demand a specific set of data. To address this, lag time analysis was implemented to ascertain the effective data sets between each external station and the main station that can significantly impact predictions. Here, cross-correlation analysis, as one of mutual information-based methods used here (through the XCORR function in MATLAB, 2025b), is considered an effective method for mutual analysis especially because it accommodates non-linear dependencies and patterns of rainfall occurrence. It is also well-suited for time series analysis, capturing both spatial and temporal dependencies (Wei et al., 2020; Wang et al., 2023). To achieve this, the specific lag time corresponding to the highest coefficient was identified for each external benchmark and the main rainfall station, and this lag time was selected as the optimal data window for the subsequent step.

This step involves extracting distinct rainfall events from the historical records of the main rainfall station, following the methodology proposed by Piadeh et al. (2023b). Subsequently, each rainfall event is further decomposed into extended rainfalls and for every extended rainfall event, a range of required data, equivalent to the best-found lag time, is extracted from the historical data of the benchmark station. All the extracted data will then be systematically organised and stored in the data cube. Similarly, the rainfall events of each benchmark station

will undergo extraction and be stored into the same data cube.

The entirety of the extracted data is then transformed into the features outlined in Table 3. These features encapsulate various characteristics of the benchmark stations' rainfall, such as intensity, duration, normalised absolute energy, normalised continuous wavelet transform, total entropy, antecedent conditions, season occurrence, and weather class. These features serve as the foundation for constructing the necessary data mining framework. The primary output of these models is the predicted weather class. Finally, a voting-based stacked ensemble model was employed to blend the output of each developed base model, thereby enhancing the overall accuracy of the model. All trained models are stored in the model library for use in real-time missing data infilling.

2.3. Multi-data imputation framework

When missing rainfall data are detected, the proposed DSS first determines whether the condition represents a dry or wet state as shown in Algorithm 2. For dry states, the missing values are imputed directly using linear regression based on historical data from the main station, without requiring benchmark inputs. For wet states, the process becomes adaptive, and data driven. The system computes the cross-covariance between each benchmark and the main station to evaluate their dependency using identified lag times. If strong dependency is detected (zero-matrix condition), the missing value is reconstructed using a t-copula approach, which flexibly models correlated rainfall behaviour without strict assumptions on data distribution and is robust to outliers and heavy tails. When dependency is weak, an external-based move-median method is used, which estimates the missing value from benchmark data within a lag-based sliding window. All preliminary imputed values are stored in a data repository for further spatial refinement.

To enhance the spatial and temporal accuracy of rainfall reconstruction, the proposed framework incorporates a fuzzy weighted spatial interpolation model that explicitly integrates topographical, hydrometeorological, and dynamic environmental factors to enhance spatial and temporal accuracy. This approach extends conventional spatial interpolation by embedding external drifts into a fuzzy inference system, enabling more adaptive weighting in regions with sparse monitoring and reducing estimation bias (Bi et al., 2023; Henriksen et al., 2024). Three drift components are calculated for each benchmark station: (1) a geographical drift D_{vi} that quantifies spatial separation using the station coordinates (Equation (1)), (2) a wind drift D_{wi} that captures the directional alignment and relative influence of local wind fields (Equation (2)), and (3) a CWT-based drift representing the similarity in rainfall signal structure. These three inputs form the basis of a Mamdani-type fuzzy inference system (Komsiyah et al., 2023) with

Gaussian membership functions defined for low, medium, and high levels of each drift component. The fuzzy architecture uses a complete rule base of 27 rules ($3 \times 3 \times 3$), through which stations that are eographically close, wind-aligned, and spectrally similar receive higher linguistic weights (high or very high), while stations with weak alignment in any dimension receive proportionally lower weights. The aggregated fuzzy outputs are defuzzied using the centroid method to generate the adaptive station weight λ_i for each benchmark. These weights are subsequently combined with the drift terms in the spatial interpolation formulation, where the final imputed rainfall value R_T at the main station is computed as a normalised drift-weighted sum of benchmark values (Equation (3)). The adaptive weight λ_i , shown in Equation (4), ensures that stations with stronger physical, meteorological, and spectral relevance contribute more heavily to the infilled value. By integrating fuzzy reasoning with externally derived drifts, this method enhances replicability and provides a transparent mathematical structure for the spatial decision-making process in real-time imputation.

$$D_{vi} = \sqrt{(x_i^2 - x_r^2)^2 + (y_i^2 - y_r^2)^2} \quad \text{Equation (1)}$$

where D_{vi} is geography drift for external i th benchmark station, x_i and y_i are the latitude and longitude of i th benchmark station respectively, and x_r and y_r are the latitude and longitude of the main rainfall station with missing data respectively.

$$D_{wi} = W_i \times \frac{\theta_i - \theta_{ri}}{180} \quad \text{Equation (2)}$$

where D_{wi} is wind drift for external i th benchmark station, W_i is the wind speed, θ_i is the wind direction relative to north, θ_{ri} relative difference in angle between the main station and i th external benchmark station. Finally, the missing data are estimated using Equation (9) and integrated into the database supporting the EWS, which is employed for real-time flood forecasting.

$$R_T = \frac{\sum_i R_i \times e^{-D_{vi}^2} \times e^{-D_{wi}^2} \times e^{-CWT_i^2} \times \lambda_i}{\sum_i \lambda_i} \quad \text{Equation (3)}$$

$$\lambda_i = \frac{e^{-D_{vi}^2} \times e^{-D_{wi}^2} \times e^{-CWT_i^2}}{\sum_i e^{-D_{vi}^2} \times e^{-D_{wi}^2} \times e^{-CWT_i^2}} \quad \text{Equation (4)}$$

Where R_T is the infilled missing data and λ_i is the adaptive weight computed as Equation (4).

Table 2
Selected weak learner data mining models to develop Ensemble data-mining event determination.

Selected methods	Description	Optimised hyperparameters
Discriminant Analysis	Multiple linear regression expressing one dependent variable as a combination of other features or measurements	• Delta: Linear coefficient threshold
Decision Tree	The regression tree utilised a top-down recursive tree of an inner node. The decision tree model is divided into smaller subgroups until ultimately separated into an exclusive mutual subset.	• Gamma: Amount of regularisation • Minimum leaf size: Minimum number of leaf node observations
Gaussian Process Regression	The kriging method providing the best linear unbiased prediction at unsampled locations	• Sigma: Initial value for the noise standard deviation
K-Nearest Neighbourhood	Non-parametric method finding the closest neighbourhoods based on similarity	• Distance: Neighbour search method • Neighbours number: Number of nearest neighbours in observant data to find for classifying each point when predicting
Naive Bayes	Supervised learning method applying the theory of Bayes with strong independence assumptions between the different features	• Kernel distribution: Approach of data distribution and data smoothing • Width: Regulating width of Kernel smoothing window
Neural network Pattern Recognition	Finding data regularities and similarities translated by different nodes, biases and weights	• Nodes: Number of small individual units • Layer: Number of hidden nodes group
Supervised Vector Machine with Error-Correcting Output	Linear classification by splitting the data into subsets, e.g. pattern recognition and data classification based on the statistical learning theory and structural risk minimisation principle	• Kernel scale: Approach of data distribution and data smoothing • Box constraint: controller of the maximum penalty aiding to prevent overfitting

Algorithm 2

Pseudocode of DSS framework for automated infilling of missing data by external benchmark data sources

```

%Inputs:
R_main_t is missing
Hist_main % past main-station series for ith rain gauge
{R_i_t} for i = 1:n % benchmark readings at t (maybe missing or zero)
{L_i} for i = 1:n % lag times
{x_i, y_i} % benchmark coordinates
{x_R, y_R} % main-station coordinates
{W_i, θ_i} % wind speed and direction at time t

%Pretrained/derived (referenced, not implemented here):
Compute CWT_i =  $\frac{\sum_{i=2}^t (R_{main\ i} - R_{main\ i-1})^2}{R_{main}}$ 
Copute linear_regression_impute (Hist_main) % leveraging the preceding data from the ith station calculated by “fitlm” code in MATABL
Compute_cross_covariance(CR) = % CR is the cross-correlation at dth lag time, Rmain j is rainfall intensity of main station at timestep j, Ri j is rainfall intensity of jth benchmark station at timestep of j by XCOV code in MATABL

$$\frac{\sum_{j=1}^{t-1} [R_{main\ j} - \bar{R}_{main}] \times [R_{i\ j-d} - \bar{R}_j]}{\sqrt{\sum_1^n [R_{main\ j} - \bar{R}_{main}]^2} \sqrt{\sum_1^n [R_{i\ j-d} - \bar{R}_j]^2}}$$

is_high_dependency(CV_matrix) ∈ {true, false} %per your criterion (text: “zero-matrix” ⇒ high)
Compute_t_copula_impute(R_i_window, R_main_window) % kernel CDF → copula scale → MLE fit → infill
Compute_move_median_impute (R_i_series, window_len) % sliding window median based on lag L_i
Compute_geography_drift (Dvi) = % xi and yi are the latitude and longitude of ith benchmark station respectively, and xr and yr are the latitude and longitude of the main rainfall station with missing data respectively

$$\sqrt{(x_i^2 - x_R^2)^2 + (y_i^2 - y_R^2)^2}$$

Compute_wind_drift (Dwi) =  $\frac{\theta_i - \theta_{Ri}}{180}$  % Wi is the wind speed, Θi is the wind direction relative to north, ΘRi relative difference in angle between the main station and ith external benchmark station
Compute_lambda_i (λi) =  $\frac{e^{-DV_i^2} \times e^{-Dw_i^2} \times e^{-CWT_i^2}}{\sum_i e^{-DV_i^2} \times e^{-Dw_i^2} \times e^{-CWT_i^2}}$ 
Compute_fuzzy_weighted_spatial_interpolation (R_i_pt) =  $\frac{\sum_i R_i \times e^{-DV_i^2} \times e^{-Dw_i^2} \times e^{-CWT_i^2} \times \lambda_i}{\sum_i \lambda_i}$ 

% Outputs:
Determine_state (value) ∈ {Dry, Wet}
R_i_pt_repo % data resource to store intermediate infills (R_i_pt)
Store_repo(repo, station_id, t, value)

Begin
determine global state E_t for the missing main reading E_t ← determine_state(context around t) % Dry or Wet based on classifier
for i in 1:n: %i is the ith external benchmark
    E_t_i = Determine_state(R_i_t) %dry or wet
If E_t_i = Dry
    R_T_i = Linear_regression_impute (Hist_main)
    Store_repo(R_i_pt_repo, i, t, R_i_pt)
    Update database; Stop; i = i+1
Else %E_t_i = Wet % for each benchmark, decide correlated vs not, and create candidate infills R_i_pt
    Initialize set S_candidates = ∅
    R_i_window ← R_i(t-L_i-1: t-1)
    R_main_shifted ← R_main(t-L_i-1: t-1) shifted by L_i
    CV_i ← compute_cross_covariance(R_i_window, R_main_shifted, L_i)
    dep_flag ← is_high_dependency(CV_i)
    if dep_flag = true then %correlated → t-copula infill
        R_i_pt ← t_copula_impute(R_i_window, R_main_shifted)
    Else % weaker dependency → external-based move-median
        window_len ← f(L_i)
        R_i_pt ← move_median_impute(R_i_window, window_len)
    Store_repo(R_i_pt_repo, i, t, R_i_pt)

```

(continued on next page)

Algorithm 2 (continued)

```

%Inputs:
Add (i, R_i.pt) to S_candidates
Update database; Stop; i = i+1
for each (i, R_i.pt) in S_candidates:
    DV_i ← compute_geography_drift(x_i, y_i, x_R, y_R)
    Dw_i ← compute_wind_drift(W_i, θ_i, 0_R)
    λ_i ← compute_lambda_i(DV_i, Dw_i, CWT_i)
    R_T ← fuzzy_weighted_spatial_interpolation((R_i.pt), {λ_i})
write R_T to the EWS database and return R_T

```

2.4. Performance assessment and comparison

To illustrate the application of the proposed model, a selection of benchmark models from the literature is used for comparison. These include: (B1) An external-based LR model tested by Duarte et al. (2022), representing a well-established statistical approach using external data; (B2) FFNN coupled with MLR proposed by Papailiou et al. (2022), considered a robust AI-based method; (B3) A vector sampling method developed by Oriani et al. (2020), which effectively applies statistical modelling to similar datasets; (B4) An event identification approach based on XGBoost followed by LR, introduced by Chivers et al. (2020), representing a leading data mining technique; and (B5) A hybrid external benchmark-based model incorporating hot deck imputation, KNN, MLP, and LR, as demonstrated by Aieb et al. (2019).

The performance of the proposed model and benchmark models are evaluated under a range of scenarios designed to reflect realistic operational conditions. These scenarios include: (1) weather conditions – categorised into dry and wet periods, (2) missing data types – comprising univariate gaps and four-timestep multivariate gaps – the average time required to repair station errors, (3) rainfall characteristics – including intensity and duration, and (4) temporal placement of missing data within a rainfall event – including occurrence at the start, rising limb, steady stage before peak, peak time, steady stage after peak, at the end of the event, and immediately after rainfall.

To evaluate the impact of imputation model accuracy on the performance of the flood EWS and to compare real-time flood forecasting results with and without the proposed framework, an event-based nonlinear autoregressive model with exogenous input (E-NARX) developed by Piadeh et al. (2023b) was employed. Several missing data scenarios, including both unimodal and multimodal gaps, were considered to evaluate the system's robustness under different conditions. The imputation preprocessing for this model was conducted five times—four times using the benchmark models and once using the proposed framework. The results were then compared across different prediction time steps for water level forecasting to examine how the proposed framework enhances model performance.

3. Result & discussion

The proposed methodology is validated through its application to real-time flood forecasting in a real-world UDS pilot study in the UK. A comprehensive evaluation of the results is presented focusing on three key aspects: (1) the standalone accuracy of the rainfall data imputation model, (2) the influence of the imputation quality on the overall performance of the EWS, and (3) generalising the outcomes, using this model is pretrained and run for the same period of time in different location of the UK.

3.1. Study area and time-series data acquisition

Fig. 3a presents the spatial layout of the study area, including the urban catchment, the Ruislip gauging station, and the RAF Northolt rain gauge, which support the RTUFF of the EWS deployed at Ruislip. The open-channel UDS in Ruislip, situated in the Borough of Hillingdon in northwest London, conveys surface runoff from a 9.3 km² catchment via the River Pinn to a tributary of the River Thames. Rainfall events in this area occur regularly throughout the year, with the majority of storms characterised by durations under 600 min (10 h) and rainfall depths typically less than 10 mm (Piadeh et al., 2023a). These events have triggered several fluvial floods across Ruislip's urban neighbourhoods, resulting in traffic disruptions, extensive surface water pooling, and significant damage to infrastructure and private property.

The Ruislip gauging station is equipped with an ultrasonic IoT-based depth sensor that records water levels within the UDS at 15-min intervals. The RAF Northolt rain gauge was selected as the primary meteorological input based on prevailing south-westerly wind patterns

Table 3
Introduced rainfall features extracted for developing data mining models^a.

Extracted rainfall feature	Code	Description/Equation	Transformation key	Unit/Class
Intensity	RI	The ratio of total rainfall depth to the duration	Numerical	mm per timestep interval
Duration	RD	Number of timesteps between the start and end of the rainfall	Numerical	Timestep interval
Normalised absolute energy	RA	Total kinetic energy, here intensity, associated with rainfall calculated as $\frac{\sum_i R_{ji}^2}{R_j}$	Numerical	mm
Normalised continuous wavelet transforms	RC	where R_{ji} is intensity of station j at timestep i , and $\bar{R}$ is the average intensity Time-frequency intensity of rainfall signals calculated as $\frac{\sum_{i=2}^t (R_{main\ i} - R_{main\ i-1})^2}{\bar{R}_{main}}$	Numerical	mm
Total entropy	RH	Uncertainty or disorder in a rainfall distribution calculated as $\sum_i P(R_{ji}) \times \log_2 R_{ji}$	Numerical	–
Antecedent conditions	RP	where P is probability of i th intensity Detecting impact of evapotranspiration. The range of data is equalled to time of concentration of catchment for which EWS applied.	Binary class	0: No 1: Yes
Season occurrence	RS	A different class of humid temperate climate based on recommendation of Peel et al. (2007).	Multi class	1: Dry 2: Mild 3: Rainy
Weather class (predictor)	WC	Transforming rainfall value of main station to predicted state of current data i.e. dry or wet weather condition	Binary class	0: Dry 1: Wet

^a The search in the Scopus database covering the last decade (2015–2025) was based on keywords “rainfall”, “prediction”, “forecasting”, “data driven”, “data mining”, “artificial intelligence”, and “machine learning”. Research works were reviewed to select introduced parameters made by rainfall intensity. All parameters were analysed for the case study through sequential sensitivity analysis and particle component analysis recommended by Miró et al. (2017) and Ocampo-Marulanda et al. (2021), and ineffective parameters were excluded from the data mining framework. Remained selected parameters, reported here, are mainly inspired by, Fan et al. (2023) Liu et al. (2023), Piadeh et al. (2023), Piadeh et al. (2023b), and Zhang et al. (2023).

(See Windrose in Fig. 3a) and cross-correlation analysis between available rainfall stations and water level data (Fig. 3b). It captures rainfall using a high-resolution IoT-enabled tipping bucket system and is supported by five nearby rain gauge stations, which serve as external benchmarks for model validation.

All hydrometeorological data - both rainfall and water levels - are retrieved via open-access application programming interfaces provided by the UK Environment Agency. Fig. 3c depicts the data availability and continuity for both the Ruislip and RAF Northolt stations (DEFRA, 2025). Although data availability varied across stations, all records were standardised to a consistent 12-year monitoring period from 2013 to 2025. The final dataset comprises 365,233 time-stamped entries, of which 70 % were used for training and validating the data mining platform (2013–2022), while the remaining 30 % were reserved for performance evaluation of the model (2022–2025).

3.2. Performance of the proposed methodology

3.2.1. Missing data component performance

Fig. 4 provides an overview of the standalone performance of the pre-data imputation components. This focus on the (1) precision of the DSS in detecting different types of missing and correct rainfall data, (2) the effectiveness of the ensemble data mining framework in weather state recognition, and (3) the contribution of each rainfall feature to model performance. Results show that DSS achieved consistently high detection accuracy across most data types (See Fig. 4a), with precision values generally exceeding 0.90. The system performed particularly well for Type 1 (NaN value detection) and Type 2 series (incorrect zero values during wet weather), reflecting the relative simplicity of these scenarios where the absence of a value or inconsistencies in zero reporting are more readily identified through RNV checks and geographical path validation.

However, performance declined in certain challenging categories. Specifically, for missing data, the lowest scores occurred in Type 3.1 and Type 3.4. According to the scenarios defined in Fig. 2, these correspond to cases where non-zero rainfall is reported during dry conditions (Type 3.1) or where reported rainfall values are anomalously high (Type 3.4). In both situations, the data physically exists, and the system must detect

whether the value is an outlier using isolation forest and angle-based anomaly detection methods. This inherently increases complexity, as the distinction relies on subtle deviations from historical patterns rather than on absolute absence or binary inconsistencies.

For correct data detection, the DSS generally maintained high accuracy, but Type 3.3 showed a comparatively lower precision (still more than 95 % for Type 3.1 and 3.3). This type involves identifying outlier-like values that are, in fact, valid observations. As with Types 3.1 and 3.4, the difficulty raises from distinguishing true extreme rainfall events - potentially intensified by climate variability - from erroneous anomalies. Such cases require a balance between avoiding false alarms and ensuring that the permissible limit values in the dataset are maintained, which explains the slightly lower performance in this category.

Fig. 4 b and c evaluate the role of individual rainfall features in the performance of the ensemble data mining framework, which is responsible for determining the dry or wet state of each event before data imputation. In the baseline configuration (original model), the system was tested against an equal number of events - 2400 dry and 2400 wet - and, according to Table A1, correctly classified 97 % of wet events (true positives) and 98 % of dry events (true negatives), with only 3 % false positives and 2 % false negatives. The sensitivity analysis in Fig. 4b systematically removes each rainfall feature to assess its impact on classification performance. Across all cases, feature removal caused a marked decline in accuracy, with reductions exceeding 20 % in most instances. This is consistent with the changes in the confusion matrix values in Table A1, where removal of any feature led to increases in both false positives and false negatives, highlighting that all features contribute meaningfully to the discrimination between dry and wet states.

Fig. 4c complements this by applying PCA and partial least squares (PLS) to quantify the explanatory power of each feature. The explained variance plots and cumulative variance curves reveal that all seven features contribute non-negligibly to the total variance, reinforcing that none of them can be considered redundant in the classification process. Both PCA and PLS indicate that the information contained in these features is well distributed and removing anyone would impair the model's ability to capture the underlying rainfall dynamics needed for accurate weather state detection.

As illustrated in Fig. 4d, the uncertainty analysis on the size of data feed follows a clear three-regime pattern. First, trimming the dataset from 100 % down to roughly ~ 75 % leads to only marginal change (resistance region) - performance stays close to the full-data baseline (retaining >90 % of baseline skill, with RMSE nearly flat). However, afterward until to ~ 35 %, the relationship is approximately linear in which each incremental reduction produces a roughly proportional deterioration in accuracy (steady RMSE increase). Finally, once the training fraction falls below ~ 35 %, performance drops sharply which results in error escalates rapidly.

3.2.2. Framework performance

3.2.2.1. Error performance. Fig. 5 presents a comprehensive comparison of the proposed method with five benchmark data imputation techniques under varying operational scenarios, using RMSE and R^2 as the performance metric. The results are grouped by rainfall intensity, rainfall duration, and the type of missing data - either univariate (Fig. 5a–f) or multivariate (Fig. 5g–l). In this study, univariate gaps represent single-timestep missing data, while multivariate gaps involve four consecutive timesteps, simulating the average time needed for an operator to restore faulty measurements. This distinction is critical because multivariate scenarios are inherently more challenging: the loss of multiple successive observations removes valuable temporal

continuity, making it more difficult for any imputation model to reconstruct the true signal without distortion.

The first clear observation is that dry-weather cases (Table A2) consistently yielded RMSE values below 0.1 mm for both univariate and multivariate gaps, regardless of method. This is expected because in the absence of rainfall, imputing missing values is trivial - most methods can correctly fill zeros without introducing error. Under wet-weather conditions, however, performance varied considerably with rainfall characteristics and missing data type. For univariate cases with low-intensity, short-duration rainfall (Fig. 5a), all models, including the benchmarks, achieved relatively low RMSE values, with the proposed method performing slightly better due to its accurate detection of event onset and reliance on simple but reliable interpolation when variability is low (0.19 mm for the proposed method versus 0.24–0.41 mm for benchmarks). In low-intensity, long-duration events (Fig. 5b), the gap length remained short, but sustained low rainfall created a more uniform signal, leading to consistently low RMSE across all models (~ 0.23 – 0.31 mm, with the proposed method at 0.23 mm). The difficulty increased sharply with medium- and high-intensity rainfall (Fig. 5c–f).

For univariate medium-intensity, short-duration events (Fig. 5c), RMSE rose across all methods because the short bursts of moderate rainfall offer limited temporal context for imputation, and small errors in peak estimation can significantly affect RMSE (0.46 mm for the proposed method versus 0.77–1.04 mm for benchmarks). In medium-intensity, long-duration events (Fig. 5d), the proposed method

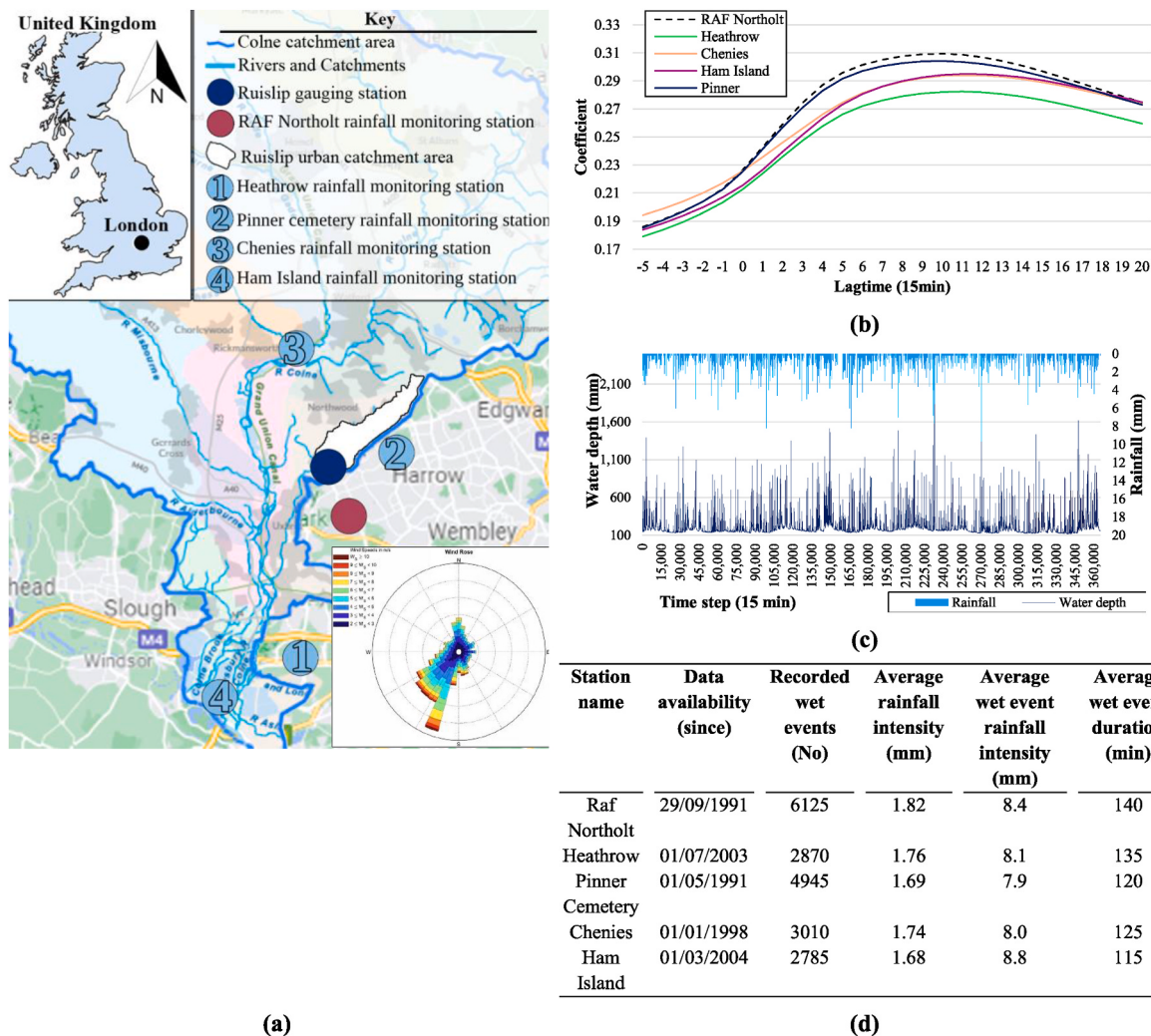


Fig. 3. Geographical map and hydrological data of the pilot study: (a) location of stations, (b) Cross-correlation between selected rainfall and water level data, (c) characteristics of rainfall and water level data, (d) characteristics of the rainfall data source.

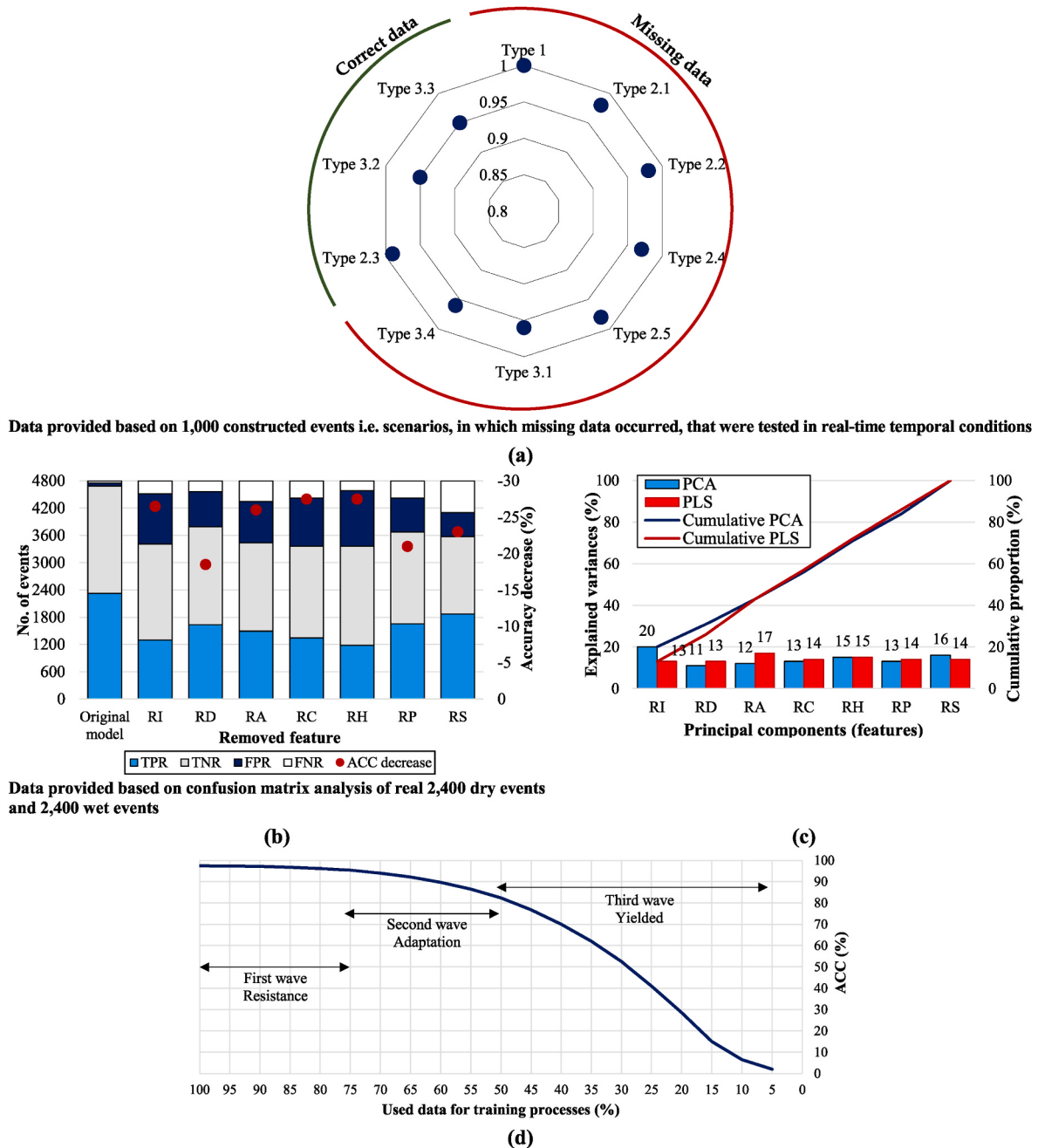


Fig. 4. Performance of pre-data imputation components: (a) precision of DSS framework in detecting missing and correct data, (b) accuracy of data mining framework for weather recognition and impact of removing features, (c) PCA and PLS analysis of feature selection used for data mining, (d) uncertainty analysis of model performance based on size of training data.

retained a marked advantage over benchmarks by using extended temporal correlation patterns from benchmark stations, though RMSE remained higher than in low-intensity cases (0.43 mm versus 0.71–0.94 mm). The greatest challenges arose in high-intensity events (Fig. 5e and f), especially short-duration storms, where rapid changes in intensity and narrow peaks left little margin for interpolation error. Here, the proposed method's hybrid approach helped, but all methods saw substantial error increases.

For multivariate scenarios (Fig. 5g–i), RMSE values were consistently higher than their univariate counterparts under similar rainfall conditions, reflecting the added difficulty of reconstructing longer missing intervals. Low-intensity rainfall (Fig. 5g and h) still produced relatively low RMSE values, and in these categories the proposed model outperformed all benchmark methods by a clear margin, benefiting from its

integration of external benchmark data and applied imputation strategies (0.24–0.34 mm for the proposed method versus 0.57–0.94 mm for benchmarks). In medium- and high-intensity events (Fig. 5i–l), reconstruction errors were amplified due to the loss of fine-scale temporal variability and peak structure across multiple timesteps; nonetheless, this model maintained the lowest RMSE in each case, showing greater resilience to data gaps compared to other techniques (e.g., 0.45–0.65 mm for the proposed method versus 1.10–2.70 mm for benchmarks). The highest RMSE values occurred in high-intensity, long-duration multivariate cases (Fig. 5l), where the combination of sustained heavy rainfall and extended data gaps severely limited the accuracy of all approaches; even in this most challenging scenario, the proposed method still achieved the best performance (0.68 mm versus 1.92–2.89 mm), reducing error more effectively than all benchmarks.

In the multivariate cases in Figs. 5g–l, the R^2 pattern is clear and strongly supports the proposed method. It remains close to the ceiling in easy conditions and decreases pleasantly as events become more severe. In low-intensity rain, R^2 remains very high despite four-step gaps - 98 (short) and 94 (long) - while the best benchmark (B5) trails at 96/70 and the others sit much lower (44–80 short; 40–65 long). Moving to medium intensity, the explanatory power decreases but the advantage remains: the proposed method holds 90 (short) and 88 (long), compared with 69/69 for B5 and roughly 24–44 (short) and 14–21 (long) for the others. This shows that longer gaps and stronger dynamics degrade the benchmarks' fit much more than they do the proposed approach. The most demanding setting is high-intensity rain, where rapid peaks plus four-step gaps suppress R^2 the most. Even so, the proposed method remains strong at 86 (short) and 83 (long), versus 63/61 for B5 and only 12–20 (short) and 4–9 (long) for the other benchmarks. Overall, R^2 falls steadily as conditions toughen - from low to medium and medium to high intensity, and from short to long duration. The proposed method still explains the most variance (98–83). The best competitor drops from 96 to 61, and the others often stay below 50. This widening gap highlights where robust imputation matters most.

3.2.2.2. Bias performance. Fig. 6 and Table A4 indicate a comparison of the proposed method with five benchmark imputation methods under different rainfall characteristics, using bias as the performance metric. For univariate cases (Fig. 6a–f), comparing the proposed method (Fig. 6a) with the strongest benchmark (Fig. 6f), directional biases are consistently smaller for Fig. 6a across intensities. In low-intensity rain, Fig. 6a's over/under are 0.163/0.146 (total -0.014) versus Fig. 6f's

0.194/0.161 (total -0.025) - Fig. 6a reduces directional error by 0.031 (over) and 0.015 (under). At medium intensity, Fig. 6a holds 0.360/0.363 (total $+0.004$) against Fig. 6f's 0.567/0.556 (total 0.000), trimming directional error by ~ 0.20 – 0.21 . Under high intensity, the gap widens: Fig. 6a's 0.326/0.386 (total $+0.043$) versus Fig. 6f's 0.939/0.980 (total -0.007), a reduction of ~ 0.61 (over) and ~ 0.59 (under). Overall, in the univariate case, the proposed method remains almost unbiased, with a slight tendency towards underestimation at the highest intensity ($0.386 > 0.326$), while Fig. 6f's directional errors are much larger even when its total bias appears small due to cancellation.

For multivariate cases (Fig. 6g–l), in low intensity, Fig. 6g records a value of 0.196/0.178 (total -0.021) versus 0.485/0.285 (total -0.149) in Fig. 6l as the best benchmark method, which is accompanied by a reduction of 0.289 (more) and 0.107 (less) of the directional bias. At medium intensity, Fig. 6g is 0.327/0.356 (total $+0.027$) versus Fig. 6l's 0.835/0.664 (total -0.037), reducing ~ 0.51 (over) and ~ 0.31 (under). At high intensity, Fig. 6g remains 0.431/0.492 (total $+0.008$) against Fig. 6l's 1.210/1.179 (total -0.096), reducing directional error by ~ 0.78 (over) and ~ 0.69 (under). Quantitatively, moving from univariate to multivariate increases the proposed method's directional bias by about 0.02–0.11 at low/medium intensity and ~ 0.10 – 0.11 at high intensity; for the best benchmark, the jump is much larger (~ 0.29 – 0.51 at low/medium and ~ 0.24 – 0.20 at high for over/under). Overall, multivariate gaps are harder than univariate. Even in the hardest case (high-intensity, multivariate), the proposed method's directional biases stay low and lean slightly toward underestimation: under = 0.492, over = 0.431. In contrast, the best benchmark has much larger directional errors and leans toward overestimation: over = 1.210, under = 1.179.

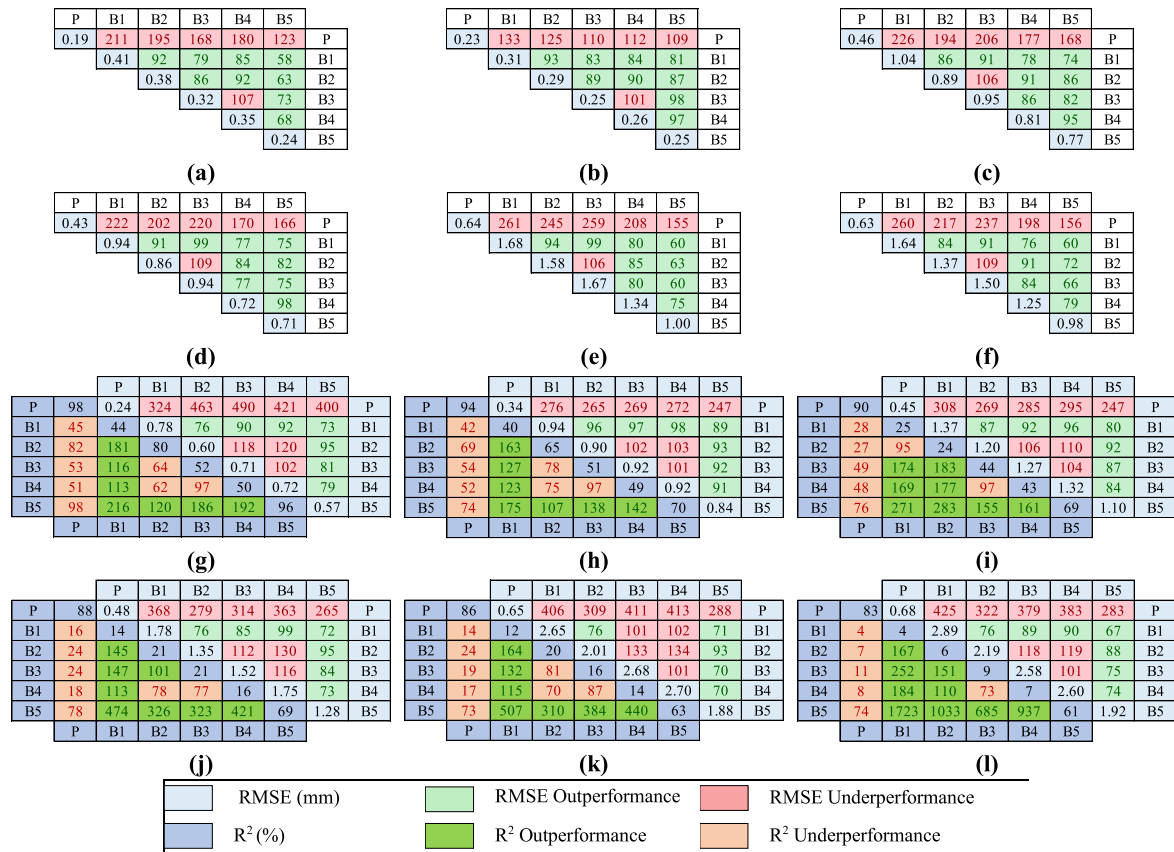


Fig. 5. RMSE and R^2 performance of data imputation methods based on type of missing data and characteristics of rainfall*: (a) UV/LI/SD; (b) UV/LI/LD; (c) UV/MI/SD; (d) UV/MI/LD; (e) UV/HI/SD; (f) UV/HI/LD; (g) MV/LI/SD; (h) MV/LI/LD; (i) MV/MI/SD; (j) MV/MI/LD; (k) MV/HI/SD; (l) MV/HI/LD. Note: Type: UV (Univariate) = one missing data, MV (Multivariate) = 4 missing data (average required time for fixing the problem by operator); Rainfall classification is based on K-means clustering technique recommendations by He et al. (2022): For I (Intensity): H(High) = greater than 12 mm/h, M (Medium) = 8–12 mm/h, L (Low) = less than 8 mm/h; For D(Duration): H(High) = longer than 3hrs, M(Medium) = between 1.5 and 3 h, L(Low) = less than 1.5hr.

3.2.2.3. Sensitivity analysis. Fig. 7 provides a detailed sensitivity analysis of RMSE performance for the proposed and benchmark imputation methods across different temporal positions within wet-weather events (more details in Table A5 in the Appendix). These temporal points track the evolution of rainfall from its initiation (point 1), through its rise (point 2), stable phases before and after the peak (points 3 and 5), the

peak itself (point 4), and the falling limb (point 6), to the event's conclusion (point 7) and the immediate post-event dry period (point 8). The results are separated into univariate gaps (Fig. 7a–c) and multivariate gaps (Fig. 7d–f) for low-, medium-, and high-intensity rainfall events, respectively. In the multivariate case, points 1, 7, and 8 are not applicable because a four-timestep missing-data block cannot be

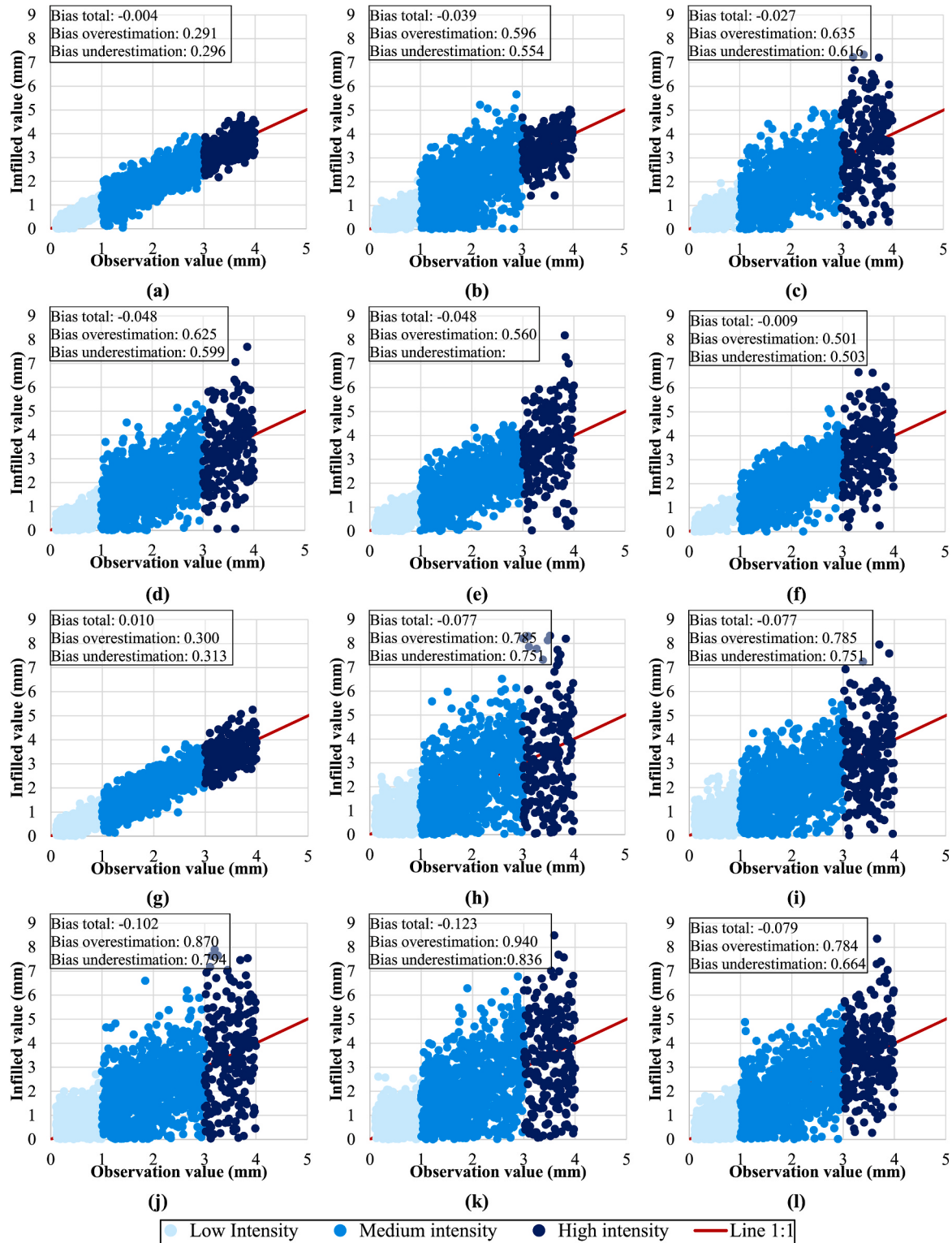


Fig. 6. Bias performance of data-imputation methods by rainfall characteristics. Univariate: (a) Proposed model; (b) B1; (c) B2; (d) B3; (e) B4; (f) B5. Multivariate: (g) Proposed model; (h) B1; (i) B2; (j) B3; (k) B4; (l) B5. Points are coloured by intensity (low/medium/high); the red line is 1:1. Insets report total bias and directional over- and underestimation. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

positioned entirely at the very start or very end of an event without extending into a non-event period, which would distort the evaluation of wet-weather imputation. In practice, all multivariate gaps occur in the central or “active” portion of the event, where rainfall is sustained.

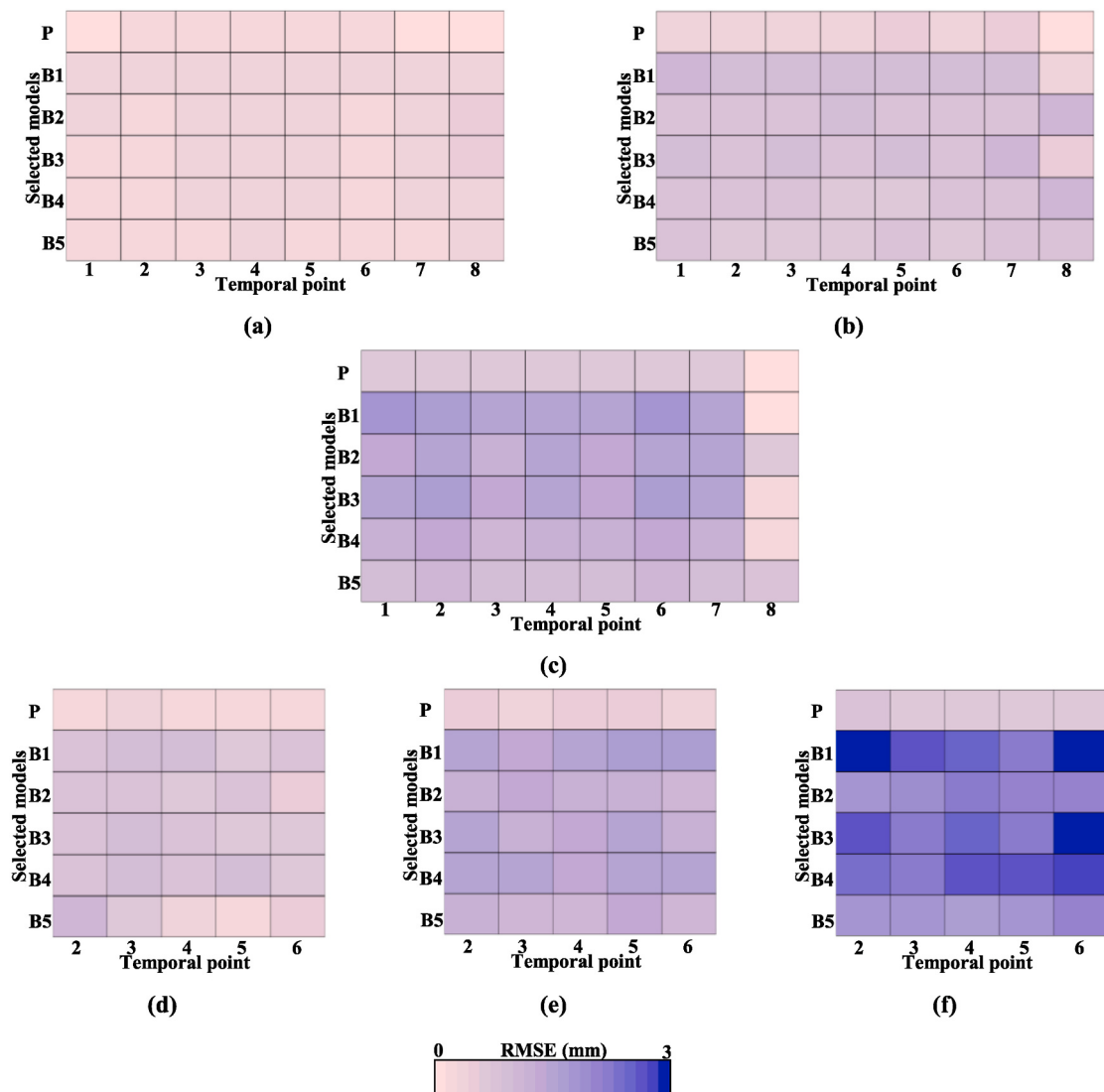
For univariate scenarios, the proposed model achieves consistently low RMSE values across all temporal points in low-intensity events (Fig. 7a), reflecting the relative simplicity of infilling when rainfall changes gradually and errors in peak reconstruction are minimal. As intensity increases to medium (Fig. 7b), RMSE values rise across all points for all methods due to sharper gradients in rainfall, but the proposed approach retains a clear advantage, especially at critical points 4 (peak) and 6 (falling limb), where accurate representation of both the maximum and the decline of rainfall is vital for downstream flood forecasting accuracy. In high-intensity events (Fig. 7c), the challenge intensifies: all benchmarks show significant performance degradation, particularly at point 4, whereas the proposed model maintains comparatively low error, highlighting its ability to handle rapid, high-magnitude changes in intensity.

In multivariate cases, the difficulty is greater because missing-data

blocks span longer intervals, erasing more temporal information. Even so, in low-intensity rainfall (Fig. 7d), the proposed model still delivers the best performance at all applicable points, benefiting from its event-driven external-benchmark integration. In medium-intensity events (Fig. 7e), RMSE grows for all models, but the proposed method continues to outperform others, especially at point 4 (peak), where precise reconstruction prevents critical underestimation of flood potential. High-intensity multivariate scenarios (Fig. 7f) are the most demanding, with benchmarks often exhibiting severe underestimation at the peak and overestimation in the falling limb (point 6). The proposed approach demonstrates resilience here, achieving the lowest RMSE at both critical points and avoiding the dangerous underestimation at the peak-which could delay flood warnings-and the overestimation during decline-which could unnecessarily keep emergency resources on standby.

3.2.3. Performance of the model on extreme unseen events

Fig. 8 and Table A6 (in the Appendix) evaluates the robustness of the proposed and benchmark imputation methods when confronted with an extreme, unseen rainfall event - defined here as a storm with intensity



Missing data temporal position: (1) start of the rainfall event, (2) within the rising limb, (3) during the steady stage before the peak, (4) at the peak of rainfall, (5) during the steady stage after the peak, (6) within the falling limb, (7) at the end of the rainfall event, and (8) during dry weather immediately following the rainfall

Fig. 7. RMSE performance of data imputation methods based on type of missing data, characteristics of rainfall, and location of missing data in the wet weather event: (a) UV/LI; (b) UV/MI; (c) UV/HI; (d) MV/LI; (e) MV/MI; (f) MV/HI.

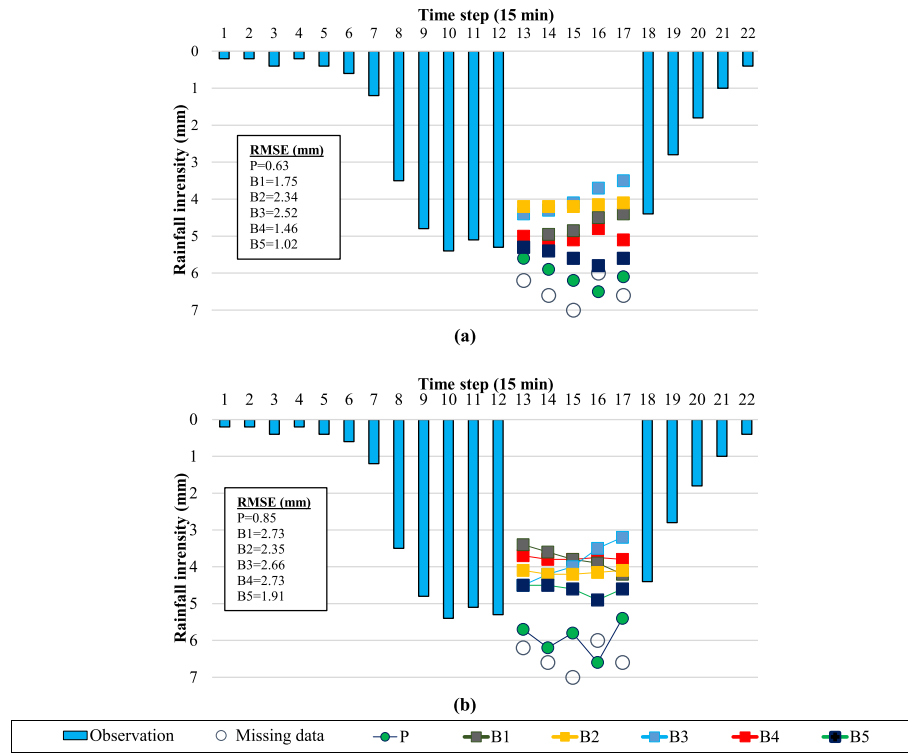


Fig. 8. Demonstration of Performance of data imputation methods for generating rainfall data in extreme unseen events (intensity >5): (a) univariate missing data; (b) multivariate missing data.

greater than 5 mm per 15-min interval. The key distinction is that these events were not included in the training datasets for any model, meaning the results reflect the models' ability to generalise beyond the conditions they were optimised for. Performance is assessed for both univariate (Fig. 8a) and multivariate (Fig. 8b) missing-data scenarios, with RMSE values summarised alongside the rainfall intensity time series and imputed values at the missing points.

In the univariate case (Fig. 8a), where only single timesteps are missing, the proposed model achieved an RMSE of 0.63 mm, substantially outperforming all benchmarks (RMSE ranging from 1.02 to 3.52 mm). This superior accuracy is evident across all five missing points, where the proposed method's estimates are consistently closer to the actual observed intensities. Its hybrid decision framework - combining event detection, cross-covariance analysis, and adaptive method selection - appears to generalise well even in high-intensity, unfamiliar conditions. In contrast, several benchmarks either systematically underestimated or overestimated the peaks, reflecting overfitting to patterns present in their training datasets.

The multivariate case (Fig. 8b) posed a greater challenge, as each missing-data block spanned four consecutive timesteps, erasing more temporal structure. Even here, the proposed model maintained clear superiority, with an RMSE of 0.85 mm, compared to 1.91–2.73 mm for the benchmarks. The improvement is especially critical at peak rainfall points (points 3 and 4), where benchmark methods frequently underestimated intensity by more than 1–2 mm. Such underestimation during extreme events can directly impair flood forecasting accuracy, delaying warnings and reducing lead time for emergency response. Similarly, in the falling limb of the event, the proposed method avoided the tendency of some benchmarks to overestimate intensities, which would unnecessarily prolong emergency stand-by measures.

3.3. Impact of the proposed model on early warning system

Fig. 9 demonstrates the influence of different data imputation strategies on the performance of the one recurrent time-series ML model (E-

NARX) used in an EWS (Details listed in Tables A7–8 in the Appendix). The comparison is made under varying proportions of missing rainfall data, ranging from 5 % to 30 %. The black line in each panel represents the baseline case in which the E-NARX model is run without missing data - this acts as the "ideal" or reference scenario. The green line corresponds to the proposed imputation method, which consistently tracks closest to the black baseline across all missing data percentages, while other benchmark methods (B1–B5) deviate more substantially. The NRMSE for lead times from 15 min (1 timestep ahead) up to 3 h (12 timesteps ahead). As illustrated in these figures, the proposed imputation is superior at every lead time and missing-data level, staying closest to the no-missing baseline. Errors rise roughly linearly with lead time (15 min–3 h) and increase as missingness grows (5 %–30 %). At 1 step ahead the baseline is ~4–6 % NRMSE; the proposed method is typically only 0.5–1.0 absolute percentage higher, while the best benchmark is 1–2.5 absolute percentage higher. By 3 h the gaps widen: with 5 % missing, baseline ~24 %, proposed ~25–26 %, best benchmark ~27–28 %; with 30 % missing, baseline ~27–28 %, proposed ~29–30 %, best benchmark ~33–34 %. Thus, the proposed method maintains the lowest NRMSE among all filled-data runs, and its advantage over the best competitor grows from about 1 to 3 absolute percentage at low missingness to 3–6 absolute percentage under the most severe missing rates and longest lead times.

Across the middle panels of Fig. 9(b–e, h, k, n, q), which chart NSE (%) by lead time, the proposed imputation stays closest to the no-missing baseline at every horizon and for every missing-data level. NSE decreases as forecasts extend from 15 min to 3 h and as missingness rises from 5 % to 30 %. At one step ahead the baseline is ~99–100 %, the proposed method is only ~1–2 absolute percentage lower (~97–99 %), and the best benchmark lags by ~2–4 (~95–97 %). By 3 h the separation is clearer: with 5 % missing, baseline ~85–86 %, proposed ~82–84 %, best benchmark ~79–81 %; with 30 % missing, baseline ~84–85 %, proposed ~72–74 %, best benchmark ~69–71 % (others lower). Overall, while NSE degrades with both longer horizons and higher missingness, the proposed imputation retains the most skill, keeping a small gap

to the baseline at easy settings and opening an advantage over the best competitor of roughly 2–4 absolute percentage at low missingness and 3–5 under the harshest (30 %) case.

The panels on the right of Fig. 9 show the false alarm ratio (FAR), a critical operational metric measuring the share of false flood alarms among all predicted flood events. Here, the performance gap between the proposed method and the benchmarks is clearly visible - and it widens as missing data increase. FAR is more sensitive than NRMSE to errors in peak magnitude and timing: because alarms trigger at thresholds, even small peak underestimates can cause missed alarms, while overestimates inflate false alarms. This sensitivity is already evident in the 5 % missing case (panel f): the no-missing baseline drops from about 95 % to ~60 % over 15 min to 3 h; the proposed method tracks closely (~90 %~55 %), whereas the best benchmark falls much lower (~85 %~45 %), indicating many more false alarms for the benchmark at longer horizons. As missingness rises, the separations grow: at 15 %

(panel i), the proposed method is roughly 8–12 absolute percentage higher than the best benchmark across horizons; by 30 % (panel r), the baseline is ~90 %~55 %, the proposed method ~75–80 %~40 %, and the best benchmark only ~65–70 % to ~30–32 %, leaving a ~8–12 point advantage for the proposed method at short leads and ~10–15 points by 3 h. In short, while NRMSE increases and NSE decreases as data loss and lead time grow, the FAR results show that the proposed imputation consistently keeps alarm reliability closest to the no-missing reference, limiting both missed and spurious triggers even under the harsh 30 % scenario.

3.3.1. Multi catchment training, validation and testing

To provide a more generalised evaluation of the proposed framework and to compare the results from the initial case study (RAF Northolt station) with other regions exhibiting different characteristics, four additional case studies, as shown in Fig. 10, were selected based on

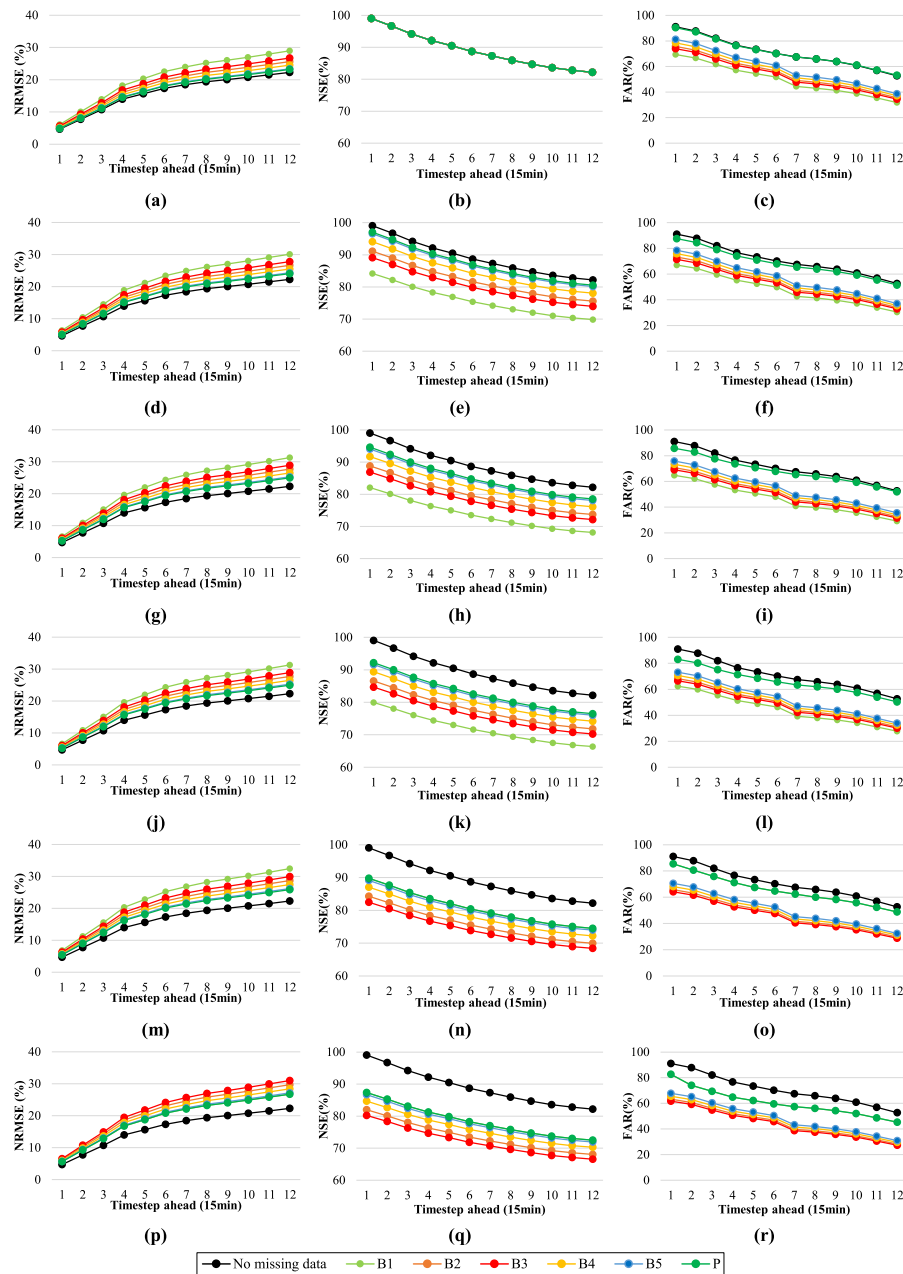


Fig. 9. Performance of proposed data imputation and selected benchmark methods in data infilling of E-NARX model based on percentage of missing data: (a–c) 5 %, (d–f) 10 %, (g–i) 15 %, (j–l) 20 %, (m–o) 25 %, (p–r) 30 %, (left) NRMSE, (middle) NSE, (Right) FAR 3.4. Generalisation of the framework in the context of the UK.

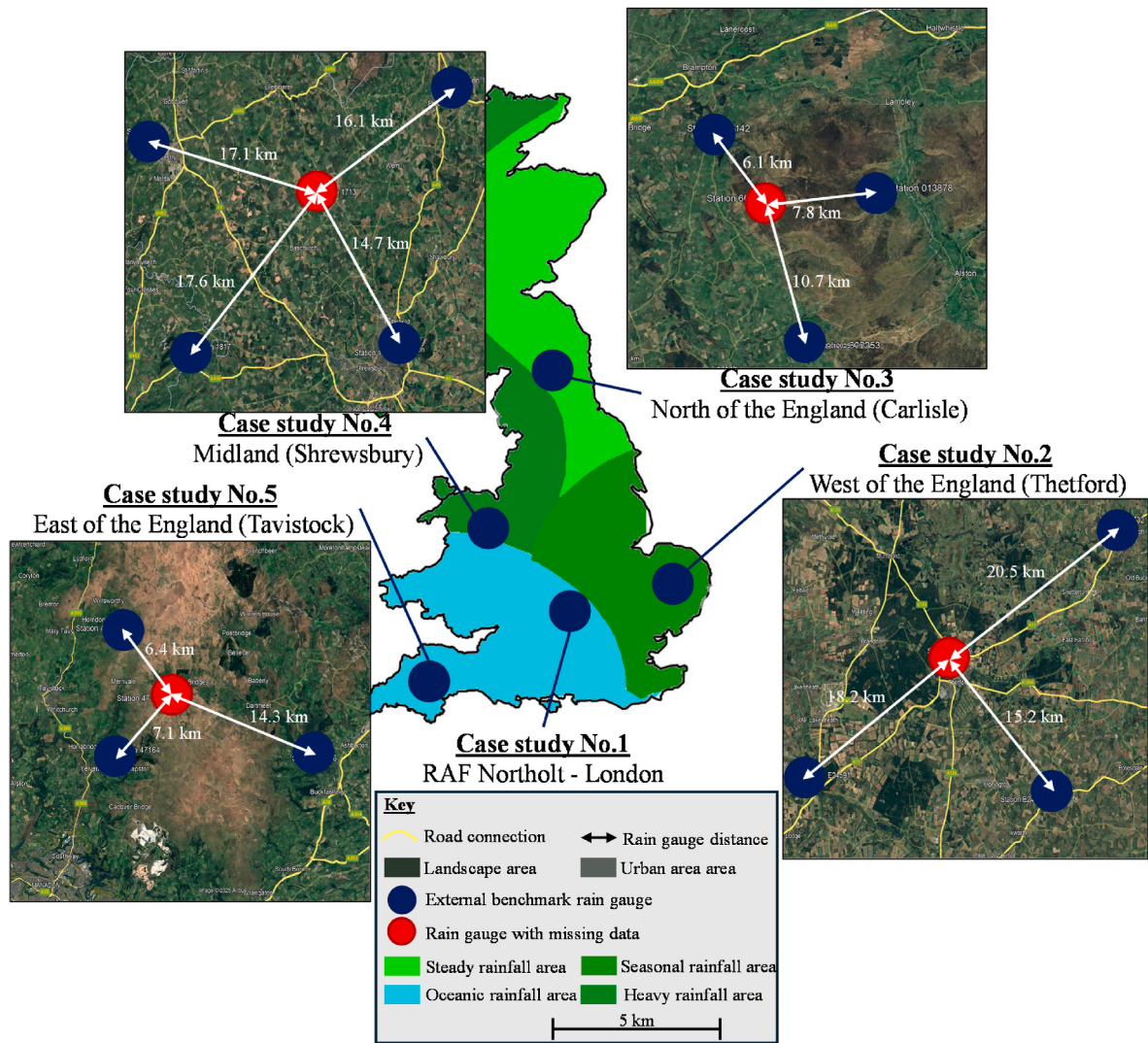


Fig. 10. Geographical distribution of selected case studies.

different rainfall characteristics (classification is inspired from DEFRA,2025). The initial case study, RAF Northolt, is located in an oceanic rainfall region. To examine similar climatic conditions in a less urbanised and more natural setting, the Tavistock station in the east of England was chosen. This comparison allows assessment of the framework's performance under comparable rainfall regimes but differing land-use characteristics. The Thetford station, located in the west of England, was selected to represent a region with more pronounced seasonal rainfall. In this case, the benchmark rainfall stations are

situated farther from the gauge containing missing data, providing an opportunity to test the model's robustness under sparse spatial coverage. For the north of England, where rainfall is generally steady and consistent, the Carlisle station was selected. This area features closely spaced rainfall stations within predominantly natural landscapes, allowing evaluation of the model under dense monitoring conditions. Finally, Shrewsbury in the Midlands was chosen to represent a region with heavy rainfall and relatively urbanised conditions, where rainfall stations are widely spaced. This case provides insights into the

Table 4
Performance comparison of proposed framework with selected case studies.

KPI	Case study				
	RAF Northolt	Thetford	Carlisle	Shrewsbury	Tavistock
Missing data detection					
TPR (%)	97	94	98	91	96
TNR (%)	98	96	98	93	97
Missing data imputation					
RMSE (mm)	0.45	0.6	0.43	0.58	0.46
R ^{2a} (%)	90	87	92	85	91
Bias total (mm)	0.009	0.016	0.033	0.009	-0.052
Bias overestimation (mm)	0.305	0.331	0.246	0.394	0.235
Bias underestimation (mm)	0.296	0.315	0.213	0.385	0.287

^a Data is provided only for multivariate missing data infilling.

framework's performance under high-intensity rainfall and limited data availability.

Table 4 shows that performance stays broadly stable across the five UK catchments. For missing-data detection, rates are uniformly high - TPR 91–98 % and TNR 93–98 % - indicating reliable identification of gaps everywhere. For multivariate infilling, accuracy varies only modestly: RMSE spans 0.43–0.60 mm and R^2 85–92 %. Relative to RAF Northolt (RMSE 0.45; R^2 90), Carlisle is slightly better (0.43; 92), Tavistock is comparable (0.46; 91), while Shrewsbury and Thetford are a little weaker (0.58; 85 and 0.60; 87, respectively). Total bias is near-zero at most sites (RAF 0.009 mm; Thetford 0.016; Carlisle 0.033; Shrewsbury 0.009), with Tavistock showing a small negative total (−0.052 mm, mild underestimation). Directional components reinforce this picture: Tavistock's over/under biases (0.235/0.287 mm) are noticeably smaller than Shrewsbury's (0.394/0.385 mm). Overall, the framework's accuracy does not change noticeably across regions; where gauge networks are denser/closer (e.g., Carlisle, Tavistock) results are a little stronger than RAF Northolt, whereas sparser or more urbanised settings (Shrewsbury, Thetford) show slightly higher errors but remain within a tight performance band.

3.3.2. Evaluation of real-time performance

The RTUFF framework was deployed on a standard workstation equipped with an Intel i7 processor (3.7 GHz, 64 GB RAM) and implemented in MATLAB. The end-to-end processing time, including data acquisition from the Environment Agency API, pre-processing, and model inference, averaged 2.4 s per update cycle for the 15-min data stream. This latency represents less than 3 % of the data refresh interval, indicating that the system is capable of operating in near-real-time mode without compromising responsiveness. Model inference alone accounted for approximately 0.9 s of this total, demonstrating efficient prediction generation suitable for operational decision-making environments.

To evaluate scalability, the framework was tested using simulated parallel data streams from multiple catchments. Results showed that the modular architecture, based on asynchronous data retrieval and batch inference, allows for horizontal scaling across additional stations with minimal computational overhead (less than 8 % increase in latency when expanded to five concurrent streams). Overall, the combination of rapid inference, low communication latency, and modular scalability confirms that the RTUFF framework can operate as a practical real-time component within an early warning system. However, future extensions could integrate GPU-based processing or cloud-computing deployment to further enhance throughput and redundancy under high-demand conditions.

4. Conclusions

This study addressed a critical operational challenge in real-time flood EWS which is the presence of missing rainfall data that can significantly degrade the performance of forecasting models. While most existing data imputation approaches are applied only during pre-processing and are not applied for real-time operations, this research proposes an innovative event-driven decision framework that dynamically selects the most appropriate imputation strategy based on the temporal position, duration, and nature of missing data. The framework integrates data mining AI models, hydrological-hydraulic event identification, and external benchmark data to enhance imputation accuracy in operational settings. A real-world case study in an UDS in London, UK, was used to evaluate the framework against established AI-based imputation methods. The approach not only assesses imputation accuracy but also examines its effect on downstream flood forecasting performance, ensuring that the reconstructed rainfall data enables timely and reliable flood warnings. The main research findings are summarised as follows:

- The proposed framework demonstrated a strong capability in detecting both missing data and confirming the correctness of rainfall records (precision more than 95 %). While detection accuracy was slightly reduced in certain cases, the overall performance remained outstanding (still above 90 %). This is particularly important for ensuring that EWS operate consistently with robustness and reliability.
- Under dry and wet conditions, and in both univariate and multivariate gaps, the proposed method consistently achieves about half the RMSE of the best benchmark; it also keeps total bias near zero with smaller directional over/underestimation (slight underestimation only in the harshest cases) and maintains higher R^2 (retaining markedly more explainable variance across all scenarios). The main difficulties arise in wet weather where sharp peaks are harder to reconstruct - an important direction for future work.
- In event-based sensitivity analysis across the rainfall timeline, the model maintained robust performance from the start of rainfall through peak and recession phases, with particularly strong results in multivariate missing data cases - critical for flood warning reliability where underestimation at the peak or overestimation in the falling limb can have severe operational consequences.
- When tested on unseen events not included in model training, the proposed approach significantly outperformed all benchmark methods, demonstrating its generalisation capability for operational real-time applications. However, still adapting this system with future impact of climate change is recommended for future studies.
- To evaluate the impact of the proposed framework on the flood EWS, the results showed that the proposed method maintained predictive accuracy comparable to the no-missing-data baseline and significantly outperformed alternative approaches in terms of the FAR, particularly at low missing data percentages where threshold sensitivity is most critical.
- With minimal configuration, the framework transferred well across five UK basins, keeping gap-detection accuracy high and imputation accuracy tight. Performance varied predictably with network geometry and setting: Carlisle (denser, more natural) slightly outperformed RAF Northolt, Tavistock was comparable with a small negative total bias, while Shrewsbury and Thetford (wider spacing/more urban) showed modestly higher errors yet stayed within a narrow band. This evidences portability and highlights benchmark-gauge distance and station density as practical levers for further gains.

This study was tested on a single case study and tested again for more four other catchments in the UK for the generalisation. However, it worth it to be tested for other regions rather than the UK. On the other hand, it did not examine the impact of wider network errors, where multiple external benchmark stations may also contain erroneous data. Although such scenarios appear extreme and rare, they should be investigated and are considered a limitation of this work. Furthermore, this study evaluated only one well-established method for flood forecasting. The impact of data imputation on other models, such as LSTM or CNN-layered recurrent models, should also be explored in future research.

Finally, While the proposed framework demonstrates strong performance in real-time anomaly detection and multi-strategy imputation, one limitation is the absence of direct benchmarking against industry-grade operational systems. Although we incorporated widely recognised academic models for comparative evaluation, access to proprietary platforms and datasets was not feasible within the scope of this study. This constraint is acknowledged and is highlighted it as an important avenue for future research. Integrating the framework with commercial early warning systems and validating its performance using large-scale, industry-standard datasets would not only strengthen the generalisability of the approach but also accelerate its adoption in operational environments. This step represents a critical pathway

toward bridging the gap between research innovation and practical deployment in flood risk management.

CRedit authorship contribution statement

Farzad Piadeh: Writing – original draft, Supervision, Resources, Methodology, Investigation. **Vahid Bakhtiari:** Writing – original draft, Conceptualization. **Farshad Piadeh:** Writing – original draft, Methodology.

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.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.engappai.2025.113348>.

Data availability

No data was used for the research described in the article.

References

- Aieb, A., Madani, K., Scarpa, M., Bonaccorso, B., Lefsih, K., 2019. A new approach for processing climate missing databases applied to daily rainfall data in Soummam watershed, Algeria. *Heliyon* 5 (2), e01247.
- Anbarasan, M., Muthu, B., Sivaparthipan, C.B., Sundarasekar, R., Kadry, S., Krishnamoorthy, S., Samuel, D., Dasel, A., 2020. Detection of flood disaster system based on IoT, big data and convolutional deep neural network. *Comput. Commun.* 150, 150–157.
- Bakhtiari, V., Piadeh, F., Behzadian, K., Kapelan, Z., 2023. A critical review for the application of cutting-edge digital visualisation technologies for effective urban flood risk management. *Sustain. Cities Soc.* 99, 104958.
- Bakhtiari, V., Piadeh, F., Chen, A.S., Behzadian, K., 2024. Stakeholder analysis in the application of cutting-edge digital visualisation technologies for urban flood risk management: a critical review. *Expert Syst. Appl.* 236, 121426.
- Bakhtiari, V., Kerchi, H., Piadeh, F., Behzadian, K., Nasirzadeh, F., 2025. Role of the internet of things in flood risk management: a critical review on current practices and future directions. *Nat. Hazards* 7589.
- Balcha, S.K., Hulluka, T.A., Awass, A.A., Bantider, A., Ayele, G.T., 2023. Comparison and selection criterion of missing imputation methods and quality assessment of monthly rainfall in the central rift valley Lakes basin of Ethiopia. *Theor. Appl. Climatol.* 154, 483–503.
- Bárdossy, A., Pegram, G., 2014. Infilling missing precipitation records – a comparison of a new copula-based method with other techniques. *J. Hydrol.* 519 (A), 1162–1170.
- Ben Aissia, M., Chebana, F., Ouara, T.B.M.J., 2017. Multivariate missing data in hydrology – review and applications. *Adv. Water Resour.* 110, 299–309.
- Bi, W., Tian, L., Li, C., Ma, Z., 2023. A Kriging-based probabilistic framework for multi-hazard performance assessment of transmission tower-line systems under coupled wind and rain loads. *Reliab. Eng. Syst. Saf.* 240, 109615.
- Boulin, A., Di Bernardino, E., Laloë, T., Toulemonde, G., 2022. Non-parametric estimator of a multivariate madogram for missing-data and extreme value framework. *J. Multivariate Anal.* 192, 105059.
- Brunner, M., Slater, L., Tallaksen, L., Clark, M., 2021. Challenges in modeling and predicting floods and droughts: a review. *WIREs Water* 8, e1520.
- Canchala-Nastar, T., Carvajal-Escobar, Y., Alfonso-Morales, W., Cerón, W.L., Caicedo, E., 2019. Estimation of missing data of monthly rainfall in southwestern Colombia using artificial neural networks. *Data Brief* 26, 104517.
- Chen, L., Xu, J., Wang, G., Shen, Z., 2019. Comparison of the multiple imputation approaches for imputing rainfall data series and their applications to watershed models. *J. Hydrol.* 572, 449–460.
- Chiu, P.C., Selamat, A., Krejcar, O., 2019. Infilling missing rainfall and runoff data for Sarawak, Malaysia using Gaussian mixture model based K-Nearest neighbor imputation. In: Wotawa, F., Friedrich, G., Pill, I., Koitz-Hristov, R., Ali, M. (Eds.), *Advances and Trends in Artificial Intelligence. from Theory to Practice. IEA/AIE 2019. Lecture Notes in Computer Science*, 11606. Springer, Cham.
- Chivers, B.D., Wallbank, J., Cole, S.J., Sebek, O., Stanley, S., Fry, M., Leontidis, G., 2020. Imputation of missing sub-hourly precipitation data in a large sensor network: a machine learning approach. *J. Hydrol.* 588, 125126.
- Chuan, Z.L., Deni, S.M., Fam, S.F., Ismail, N., 2019. The effectiveness of a probabilistic principal component analysis model and expectation maximisation algorithm in treating missing daily rainfall data. *Asia-Pacific J. Atmosph. Sci.* 56, 119–129.
- Darlane, A.B., Borhan, M.I., 2025. Comparative assessment of classical and machine learning approaches for rainfall data restoration. *Environ. Earth Sci.* 84 (10), 272.
- De Luca, D.L., Ridolfi, E., Russo, F., Moccia, B., Napolitano, F., 2024. Climate change effects on rainfall extreme value distribution: the role of skewness. *J. Hydrol.* 130958.
- Department for Environment, Food and Rural Affairs (DEFRA), 2025. DEFRA Official Website [Online]. environment.data.gov.uk. (Accessed 16 June 2025).
- Ding, Y., Zhu, Y., Feng, J., Zhang, P., Cheng, Z., 2020. Interpretable spatio-temporal attention LSTM model for flood forecasting. *Neurocomputing* 403, 348–359.
- Duarte, L.V., Formiga, K.T.M., Costa, V.A.F., 2022. Comparison of methods for filling daily and monthly rainfall missing data: statistical models or imputation of satellite retrievals? *Water* 14 (19), 3144.
- Dunkerle, D., 2023. Sub-daily rainfall intermittency: is it really stochastic? A test at two Australian locations. *Int. J. Climatol.* 44, 520–570.
- Fagandini, C., Todaro, V., Tanda, M.G., Pereira, J.L., Azevedo, L., Zanini, A., 2024. Missing rainfall daily data: a comparison among gap-filling approaches. *Math. Geosci.* 56 (2), 191–217.
- Fan, D., Wang, H., Qiu, H., Zhi, P., Zhu, Z., 2023. An automatic correction method of marine radar rainfall image based on continuous wavelet transform. *Energy Rep.* 9 (8), 745–753.
- Fang, A.H.S., Lim, W.T., Balakrishnan, T., 2020. Early warning score validation methodologies and performance metrics: a systematic review. *BMC Med. Inf. Decis. Making* 20 (1), 111.
- Gao, J., Ozbay, K., Hu, Y., 2025. Real-time anomaly detection of short-term traffic disruptions in urban areas through adaptive isolation forest. *J. Intell. Transport. Syst.* 29 (3), 269–286.
- Giroto, C.D., Piadeh, F., Bakhtiari, V., Behzadian, K., Chen, A.S., Campos, L.C., Zolgharni, M., 2024. A critical review of digital technology innovations for early warning of water-related disease outbreaks associated with climatic hazards. *Int. J. Disaster Risk Reduct.* 100, 104151.
- Golkhatmi, N.S.N., Farzandi, M., 2024. Enhancing rainfall data consistency and completeness: a spatiotemporal quality control approach and missing data reconstruction using MICE on large precipitation datasets. *Water Resour. Manag.* 38, 815–833.
- Henriksen, R., Hubrechts, J., Möller, J., Knudsen, P., Pedersen, J., 2024. Large-scale rain gauge network optimization using a kriging emulator. *J. Hydrol.* 637, 131360.
- Hırca, T., Eryılmaz Türkkan, G., 2024. Assessment of different methods for estimation of missing rainfall data. *Water Resour. Manag.* 38 (15), 5945–5972.
- Jahan, F., Sinha, N.C., Rahman, M.M., Rahman, M.M., Mondal, M.S.H., Islam, M.A., 2019. Comparison of missing value estimation techniques in rainfall data of Bangladesh. *Theor. Appl. Climatol.* 136, 1115–1131.
- Kamwaga, S., Mulungu, D.M.M., Valimba, P., 2018. Assessment of empirical and regression methods for infilling missing streamflow data in Little Ruaha catchment Tanzania. *Phys. Chem. Earth* 106, 17–28.
- Kannegowda, N., Udayar Pillai, S., Kommireddi, C.V.N.K., Fousiya, 2024. Comparative assessment of univariate and multivariate imputation models for varying lengths of missing rainfall data in a humid tropical region: a case study of Kozhikode, Kerala, India. *Acta Geophys.* 72 (4), 2663–2678.
- Komsiyah, S., Ardyanti, M., Iswanto, I., 2023. Flood-Prone susceptibility analysis in garut using fuzzy inference system Mamdani method. *Procedia Comput. Sci.* 227, 912–921.
- Kumar, G.P., Dwarakish, G.S., 2025. Comparison of the multiple imputation approaches for imputing rainfall data: a humid Tropical River Basin case study. *Water Cons. Sci. Eng.* 10 (2), 87.
- Kumawat, G.R., Munoth, P., Goyal, R., 2023. Improving prediction of missing rainfall data by identifying best artificial neural network model. *J. Earth Syst. Sci.* 132, 192.
- Lai, W.Y., Kuok, K.K., 2019. A study on bayesian principal component analysis for addressing missing rainfall data. *Water Resour. Manag.* 33, 2615–2628.
- Lai, W.Y., Kuok, K.K., Gato-Trinidad, S., Ling, K.X., 2019. A study on sequential K-Nearest neighbor (SKNN) imputation for treating missing rainfall data. *Int. J. Adv. Trends Comput. Sci. Eng.* 8 (3), 363–368.
- Li, J., Burian, S.J., 2023. Evaluating real-time control of stormwater drainage network and green stormwater infrastructure for enhancing flooding resilience under future rainfall projections. *Resour. Conserv. Recycl.* 198, 107123.
- Li, J., Liu, J., Xia, Z., Liu, C., Li, Y., 2024. Accurate simulation of extreme rainfall–flood events via an improved distributed hydrological model. *J. Hydrol.* 645 (A), 132190.
- Liu, X., Wang, Y., 2024. Quantifying impacts of precipitation scenarios projected under climate change on annual probability of rainfall-induced landslides at a specific slope. *Comput. Geotech.* 167, 106063.
- Liu, Z., Wen, Y., Zhang, X., Wang, M., Xiao, S., Chen, Y., He, L., 2023. A novel rainfall forecast model using GNSS observations and CAPE. *J. Atmos. Sol. Terr. Phys.* 253, 106158.
- Lupi, A., Luppichini, M., Barsanti, M., Bini, M., Giannecchini, R., 2023. Machine learning models to complete rainfall time series databases affected by missing or anomalous data. *Earth Sci. Inform.* 16, 3717–3728.
- Lv, N., Liang, X., Chen, C., Zhou, Y., Li, J., Wei, H., Wang, H., 2020. A long short-term memory cyclic model with mutual information for hydrology forecasting: a case study in the xixian basin. *Adv. Water Resour.* 141, 103622.
- Miró, J.J., Caselles, V., Estrela, M.J., 2017. Multiple imputation of rainfall missing data in the Iberian Mediterranean context. *Atmos. Res.* 197, 313–330.
- Murakami, S., 2024. Reduction in the ratio of stemflow to rainfall during heavy rain in two Japanese cedar stands and the influence on rainfall partitioning. *J. Hydrol.* 634, 131100.
- Narimani, R., Jun, C., Michele, C., Gan, T., Nezhad, S., Byun, J., 2023. Multilayer perceptron-based predictive model using wavelet transform for the reconstruction of missing rainfall data. *Stoch. Environ. Res. Risk Assess.* 37, 2791–2802.

- Nawaz, N., Harun, S.B., Othman, R.N., Heryansyah, A., 2016. Neuro-fuzzy systems approach to infill missing rainfall data for Klang river catchment, Malaysia. *Environ. Sci. Eng.* 78 (6–12), 15–21.
- Ocampo-Marulanda, C., Cerón, W.L., Avila-Diaz, A., Canchala, T., Alfonso-Morales, W., Kayano, M.T., Torres, R.R., 2021. Missing data estimation in extreme rainfall indices for the Metropolitan area of Cali - Colombia: an approach based on artificial neural networks. *Data Brief* 39, 107592.
- Oriani, F., Stisen, S., Demirel, M.C., Mariethoz, G., 2020. Missing data imputation for multisite rainfall networks: a comparison between geostatistical interpolation and pattern-based estimation on different terrain types. *J. Hydrometeorol.* 21, 2325–2341.
- Papailiou, I., Spyropoulos, F., Trichakis, I., Karatzas, G.P., 2022. Artificial neural networks and multiple linear regression for filling in missing daily rainfall data. *Water* 14, 2892.
- Piadeh, F., Behzadian, K., Alani, A.M., 2022. A critical review of real-time modelling of flood forecasting in urban drainage systems. *J. Hydrol.* 607, 127476.
- Piadeh, F., Behzadian, K., Chen, A.S., Kapelan, Z., Rizzuto, J.P., Campos, L.C., 2023. (2023a) enhancing urban flood forecasting in drainage systems using dynamic ensemble-based data mining. *Water Res.* 247, 120791.
- Piadeh, F., Behzadian, K., Chen, A.S., Campos, L.C., Rizzuto, J.P., Kapelan, Z., 2023b. Event-based decision support algorithm for real-time flood forecasting in urban drainage systems using machine learning modelling. *Environ. Model. Software* 167, 105772.
- Pinthong, S., Ditthakit, P., Salaeh, N., Hasan, M.A., Son, C.T., Linh, N.T.T., Islam, S., Yadav, K.K., 2024. Imputation of missing monthly rainfall data using machine learning and spatial interpolation approaches in Thale Sap Songkhla River Basin, Thailand. *Environ. Sci. Pollut. Control Ser.* 31 (41), 54044–54060.
- Sattari, M.T., Falsafian, K., Irvem, A., Shahab, S., Qasem, S.N., 2020. Potential of kernel and tree-based machine-learning models for estimating missing data of rainfall. *Eng. Appl. Comput. Fluid Mech.* 14 (1), 1078–1094.
- Sa'adi, Z., Yusop, Z., Alias, N.E., Chow, M.F., Muhammad, M.K.I., Ramli, M.W.A., Iqbal, Z., Shiru, M.S., Rohmat, F.I.W., Mohamad, N.A., Ahmad, M.F., 2023. Evaluating Imputation methods for rainfall data under high variability in Johor River Basin, Malaysia. *Appl. Comput. Geosci.* 20, 100145.
- Sriwahyuni, L., Nurdianti, S., Nugrahani, E.H., Najib, M.K., 2025. Performance of machine learning for imputing missing daily rainfall data in East Java under multiple satellite datamodels. *Geogr. Tech.* 20 (1).
- Tavares, C.D.M.G., Armond, N.B., Ferreira, C.D.C.M., Guerra, A.J.T., 2025. Filling gaps of daily precipitation data at rain gauges in a tropical mountainous region: a case study of the municipality of petropolis, rj, Brazil. *Theor. Appl. Climatol.* 156 (3), 1–14.
- Teegavarapu, R.S.V., 2014. Missing precipitation data estimation using optimal proximity metric-based imputation, nearest-neighbour classification and cluster-based interpolation methods. *Hydrol. Sci. J.* 59 (11), 2009–2026.
- Thudumu, S., Branch, P., Jin, J., Singh, J.J., 2020. A comprehensive survey of anomaly detection techniques for high dimensional big data. *J. Big Data* 7 (42).
- Umar, N., Gray, A., 2023. Comparing single and multiple imputation approaches for missing values in univariate and multivariate water level data. *Water* 15, 1519.
- Wang, T., Deng, Z., Xie, Y., Wang, B., Wu, S., Li, F., Wang, W., Zou, Y., Li, X., Hou, Z., Zeng, J., Wang, M., Peng, C., 2023. Time-lag effects of flood stimulation on methane emissions in the Dongting Lake floodplain, China. *Agric. For. Meteorol.* 341, 109677.
- Wangwongchai, A., Waqas, M., Dechpichai, P., Hlaing, P.T., Ahmad, S., Humphries, U. W., 2023. Imputation of missing daily rainfall data; A comparison between artificial intelligence and statistical techniques. *MethodsX* 11, 102459.
- Wei, Q., Zhang, X., Yi, W., Pan, R., Gao, J., Duan, J., Xu, Z., Cheng, Q., Bai, L., Zhang, Y., Su, H., 2020. Association between floods and hospital admissions for schizophrenia in Hefei, China: the lag effects of degrees of floods and time variation. *Sci. Total Environ.* 698, 134179.
- Zhang, P., Zhou, Z., Feng, Z., Wang, J., Zhang, Y., 2023. Inference and analysis on the evidential reasoning rule with time-lagged dependencies. *Eng. Appl. Artif. Intell.* 126 (C), 106978.
- Zhang, R., Rong, R., Gan, J.Q., Xu, Y., Wang, H., Wang, X., 2024. Reliable and fast automatic artifact rejection of long-term EEG recordings based on isolation forest. *Med. Biol. Eng. Comput.* 62, 521–535.
- Zio, S., Poan, D.E., Adaman, Y., Bé, K., Noe, Y.A., 2025. Multi-task Gaussian process for imputing missing daily rainfall data using nearby stations: case of Burkina Faso. *J. Sens. Technol.* 15 (1), 1–13.