

Integrating Time Series, Remote Sensing and Computational Intelligence for Flood Prediction in a Data-Scarce Tropical Floodplain of the Lower Magdalena River Basin, Colombia

Jaime Alberto Parra-Plazas

jarrap@udistrital.edu.co

Doctoral Programme in Engineering, Faculty of Engineering, Universidad Distrital Francisco José de Caldas, Bogotá, Colombia; Mechanical Engineering, Faculty of Engineering, Escuela Colombiana de Carreras Industriales (ECCI), Bogotá, Colombia. <https://orcid.org/0000-0002-4992-5077>

Juan Sebastian Parra-Perdomo

Department of Electrical Engineering, Universidade Federal do Espírito Santo, Vitória, Espírito Santo, Brazil

Leonardo Plazas-Nossa

Doctoral Programme in Engineering, Faculty of Engineering, Universidad Distrital Francisco José de Caldas, Colombia

Research Article

Keywords: Flood prediction, computational intelligence, principal component analysis, artificial neural networks, support vector regression, spatial interpolation, remote sensing, tropical floodplains

Posted Date: October 16th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10374442/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Integrating Time Series, Remote Sensing and Computational Intelligence for Flood Prediction in a Data-Scarce Tropical Floodplain of the Lower Magdalena River Basin, Colombia

Jaime Alberto Parra-Plazas^{1*}, Juan Sebastián Parra-Perdomo² and Leonardo Plazas-Nossa³

1. Doctoral Programme in Engineering, Faculty of Engineering, Universidad Distrital Francisco José de Caldas, Bogotá, Colombia; Mechanical Engineering, Faculty of Engineering, Escuela Colombiana de Carreras Industriales (ECCI), Bogotá, Colombia.

**jarrap@udistrital.edu.co / jparrap@ecci.edu.co . +573112579388*

2. Department of Electrical Engineering, Universidade Federal do Espírito Santo, Vitória, Espírito Santo, Brazil, *juan.perdomo@edu.ufes.br*

3. Doctoral Programme in Engineering, Faculty of Engineering, Universidad Distrital Francisco José de Caldas, Bogotá, Colombia, *lpplazasn@udistrital.edu.co*

ABSTRACT

Flood prediction in environmentally vulnerable and data-constrained regions remains a major challenge, particularly in tropical floodplains where hydrometeorological variability, sparse monitoring networks, and uneven spatial coverage limit the reliability of conventional forecasting approaches. This study proposes a hybrid and modular framework for flood prediction in the lower Magdalena River basin, Colombia, integrating long-term rainfall and river-level time series from IDEAM stations with spatial contextual information from remote sensing products, digital elevation data, spatial interpolation, principal component analysis (PCA), and computational intelligence models. Point-based observations were transformed into continuous spatial fields using inverse distance weighting (IDW) and ordinary kriging, and PCA was subsequently applied to reduce dimensionality and improve training efficiency. Artificial neural networks (ANN) and support vector regression (SVR) were then used to reconstruct monthly predictive maps, which were evaluated against 2022 reference patterns using visual comparison and mean absolute percentage error (MAPE). The results show that predictive performance depended strongly on the interaction between the interpolation method, the learning model, and the target variable. The ordinary kriging–PCA–ANN configuration provided the most hydrologically coherent performance for water-level prediction, preserving regional continuity and the dominant spatial structure. In contrast, IDW-based configurations were more suitable for preserving local rainfall variability, although they also showed greater spatial error dispersion. A comparative synthesis of representative-month and annual MAPE evidence across the eight evaluated configurations confirmed that no single configuration was universally superior, reinforcing the need for variable-sensitive model selection. Overall, the proposed framework provides a reproducible and transferable basis for spatially explicit flood prediction in flood-prone and data-scarce environments, with clear relevance for environmental monitoring and future flood early-warning systems.

KEYWORDS

Flood prediction; computational intelligence; principal component analysis; artificial neural networks; support vector regression; spatial interpolation; remote sensing; tropical floodplains

1. INTRODUCTION

Flooding is increasingly recognised as one of the most recurrent and destructive natural hazards worldwide, with its severity amplified by climate change and growing hydroclimatic variability. This problem is

particularly critical in tropical regions such as Colombia, where pluvial and fluvial floods are strongly controlled by marked rainfall seasonality and large-scale climatic drivers, especially ENSO, which alter rainfall distribution, runoff generation and river discharge behaviour (Jiménez-Saenz *et al.*, 2026). Despite substantial progress in large-scale hydrometeorological modelling and the global documentation of flood impacts in terms of mortality and economic losses, a major challenge remains in converting these advances into effective local early warning and decision-support systems, particularly in regions characterised by sparse monitoring networks and limited forecasting capacity (Rodríguez-Espinosa *et al.*, 2025; Dasgupta *et al.*, 2025).

To address this challenge, recent studies have increasingly emphasised the integration of heterogeneous data sources, including in situ observations, satellite remote sensing products, and climate reanalysis datasets, to improve the spatial and temporal representation of flood dynamics. Remote sensing platforms such as MODIS and Sentinel-1 support flood detection and inundation mapping, while digital elevation models (DEMs) provide the topographic context required to improve forecast reliability. Likewise, climate reanalysis datasets such as ERA5, particularly when combined with high-resolution precipitation products, contribute to a more accurate characterisation of hydrometeorological drivers over complex terrain. Within this context, data driven and hybrid hydrological frameworks have gained increasing relevance, as they strengthen early warning systems and support risk informed decision-making, particularly in developing regions where the growing frequency and severity of extreme events have exposed the limitations of purely deterministic and conventional statistical approaches (Jayashree *et al.*, 2017; Brunner *et al.*, 2021; Khan *et al.*, 2024).

A key methodological advance in this transition has been the integration of hydrological modelling with artificial intelligence (AI), particularly machine learning (ML), to capture complex non-linear relationships within large and heterogeneous datasets. Hybrid hydrological ML models have generally shown greater robustness and adaptability than conventional approaches, especially in regions with limited or discontinuous hydrometeorological records (Maity *et al.*, 2024; Liu *et al.*, 2025; Mai *et al.*, 2025). Within this broader context, computational intelligence (CI) techniques have emerged as effective tools for accurate and near-real-time flood prediction in complex environments. This study focuses on two such families: artificial neural networks and support vector regression.

Effective flood prediction increasingly relies on integrative frameworks that combine hydrometeorological time series with geospatial information to capture flood dynamics across temporal and spatial scales. In this context, digital elevation models (DEMs) improve catchment delineation and runoff representation, while synthetic aperture radar (SAR) imagery combined with interpolation methods such as ordinary kriging and inverse distance weighting (IDW) enhances the spatial characterisation of flood-prone areas (Majone *et al.*, 2022; Surampudi & Kumar, 2023; Adesina *et al.*, 2025; Xie *et al.*, 2025; Azarm *et al.*, 2025). At the same time, advances in CI have strengthened predictive performance under noisy and incomplete data conditions, particularly using principal component analysis (PCA) to improve the efficiency and generalisation of hybrid models based on artificial neural networks (ANNs) and related approaches (Shojaeian *et al.*, 2024; Mereu *et al.*, 2025). From an operational perspective, the integration of geographic information systems (GIS), DEMs, SAR imagery and machine learning has also enhanced flood early warning systems through improved identification of critical rainfall thresholds and topographically vulnerable areas (Qatrinnada *et al.*, 2024; Amatebelle *et al.*, 2025).

Accordingly, this study proposes a methodology that integrates weather-station time series, remote sensing products, interpolated spatial fields, and terrain elevation data within a computational intelligence framework to develop and validate a robust flood prediction strategy. Designed for regions with fragmented data infrastructure, the approach combines ground-based records, SAR imagery, DEMs and spatial interpolation methods such as IDW and ordinary kriging, followed by outlier detection, missing data imputation and dimensionality reduction through PCA to support the efficient training and evaluation of ANN and SVR based predictive models.

This article is structured as follows: Section 2 presents the theoretical background and key concepts; Section 3 describes the study area, data sources, preprocessing procedures and the proposed computational intelligence framework; Section 4 presents the results and comparative analysis of ANN- and SVR-based models under different interpolation scenarios; Section 5 discusses the results in depth; Section 6 presents the conclusions; and Section 7 outlines future research directions.

2. BACKGROUND

This section summarises the main conceptual elements supporting the proposed framework, with emphasis on hydrometeorological time series, remote sensing inputs, spatial interpolation, and computational intelligence for flood prediction.

A. Time Series

Hydrometeorological time series, particularly rainfall, river level, and discharge, are inherently non-stationary because they are influenced by seasonality, interannual variability, and extreme events. In flood-prone environments, their predictive value depends not only on temporal continuity, but also on appropriate conditioning, including the treatment of missing data, outliers, and noise. Reliable preprocessing is therefore essential for subsequent interpolation, dimensionality reduction, and predictive modelling, especially when time series are later integrated with spatial datasets and computational intelligence methods (Chauhan & Mathews, 2020; Parra-Plazas *et al.*, 2023; Shojaeian *et al.*, 2024; Mereu *et al.*, 2025).

B. Remote Sensing

Remote sensing products provide essential spatial information for flood-prone environments. DEMs describe terrain morphology and flow controls, whereas SAR and optical imagery support the detection of surface water and inundation patterns. However, these datasets are affected by artefacts, speckle, cloud interference, and resolution-related limitations, which makes preprocessing a necessary step before predictive use. Noise reduction, spatial harmonisation, and consistency control are therefore critical for preserving physically meaningful patterns and supporting robust multi-source flood modelling (Byaruhanga *et al.*, 2024; Chatrabhuj *et al.*, 2024; Adesina *et al.*, 2025; Zhang *et al.*, 2025; Amatebelle *et al.*, 2025; Mereu *et al.*, 2025).

C. Spatial interpolation

Spatial interpolation is a key step in the proposed framework because it converts point-based hydrometeorological observations into continuous spatial fields suitable for map-based prediction. Methods such as inverse distance weighting (IDW) and kriging are widely used for this purpose, especially in data-scarce regions where continuous observational coverage is unavailable. Their usefulness, however, depends strongly on data quality, station density, and the assumptions of the interpolation method, meaning that interpolation uncertainty can propagate directly into subsequent predictive modelling. This makes bias correction, harmonisation, and methodological consistency essential for robust hydrological applications (Boumpoulis *et al.*, 2023; Wang *et al.*, 2024; Workneh *et al.*, 2024).

D. Computational Intelligence

Computational intelligence methods provide a flexible framework for modelling the non-linear and heterogeneous behaviour of flood-related variables. ANN and SVR are well suited to high-dimensional prediction problems, although their performance depends strongly on input conditioning, feature representation, and scale consistency. For river streamflow and water-level forecasting, recent comparative studies have shown that advanced machine-learning techniques can improve predictive performance by

capturing non-linear temporal relationships in hydrological time series (Abdoulhalik & Ahmed, 2024). In rainfall forecasting, artificial neural networks have been widely used to model non-linear relationships between climatic predictors and precipitation response, particularly when input variables are carefully selected and optimised for monthly prediction tasks (Abbot & Marohasy, 2014).

In flood forecasting, this is especially important because hydrometeorological time series, Earth observation imagery, and geospatial covariates must be harmonised before they can support reliable predictive learning. Previous studies have shown that machine learning performance improves when forecasting workflows combine appropriate feature extraction, temporal representation, and uncertainty-aware preprocessing under non-stationary environmental conditions (Fu *et al.*, 2024; Ashok & Pekkat, 2024; Rana *et al.*, 2024; Refadah, 2025; Dasgupta *et al.*, 2025).

Within this framework, spatial interpolation and PCA should not be understood as merely preliminary operations, but as integral parts of the predictive chain. Interpolators such as IDW and kriging convert point-based observations into continuous raster inputs, while PCA reduces dimensionality and helps structure the feature space on which ANN and SVR based models operate. However, these steps also introduce uncertainty associated with measurement error, spatial representativeness, variogram assumptions, anisotropy, and covariate availability. As a result, the reliability of computational intelligence-based flood forecasting depends not simply on the choice of learning model, but on careful conditioning, feature reduction, scale harmonisation, and transparent validation throughout the workflow (Khan *et al.*, 2024; Liu *et al.*, 2025; Refadah, 2025; Dasgupta *et al.*, 2025; Shao *et al.*, 2026).

3. DATA AND METHODOLOGY

The following section describes the data sources and the methodological framework adopted in this study. It includes the characterisation of long-term hydrometeorological time series, remote sensing products, topographic datasets, and the interpolated image outputs generated through inverse distance weighting (IDW) and ordinary kriging. In addition, it presents the procedures used for data preprocessing, spatial integration and predictive modelling within the proposed flood forecasting framework.

A. Time Series

The study focuses on the lower Magdalena River basin, particularly Santa Cruz de Mompós, a flood-prone area characterised by low relief, marked rainfall seasonality, and recurrent inundation, making it suitable for multitemporal flood analysis using computational intelligence approaches (Jiménez-Saenz *et al.*, 2026; IFRC, 2025). The temporal component of the analysis was based on precipitation and river water-level records from approximately 68 IDEAM stations covering the period 1980–2022. Because these datasets contained missing values, outliers, and transmission inconsistencies, preprocessing was necessary to support reliable interpolation, data fusion, and flood forecasting (Parra-Plazas *et al.*, 2018; Parra-Plazas *et al.*, 2023; Rodriguez-Espinosa *et al.*, 2025). Figures 1 and 2 show the location of the study area and its surrounding topographic context.

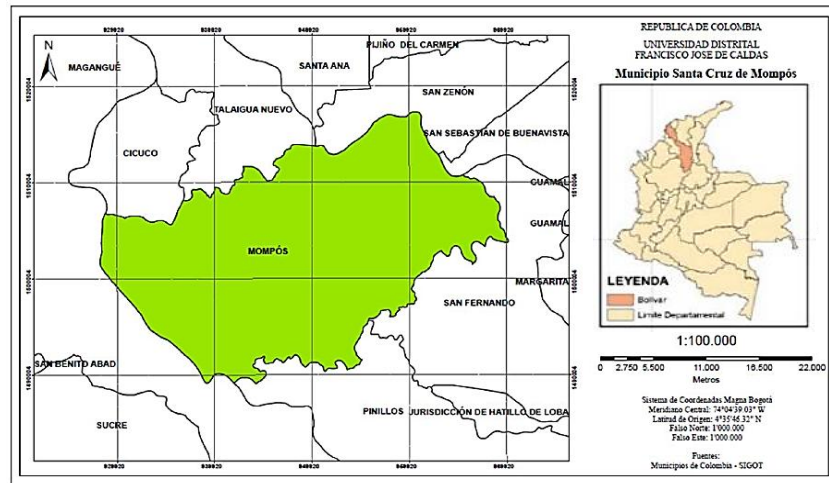


Fig. 1 Location of the municipality of Mompós, Bolívar. Source: Based on data from SIGOT and the DANE Geoportal

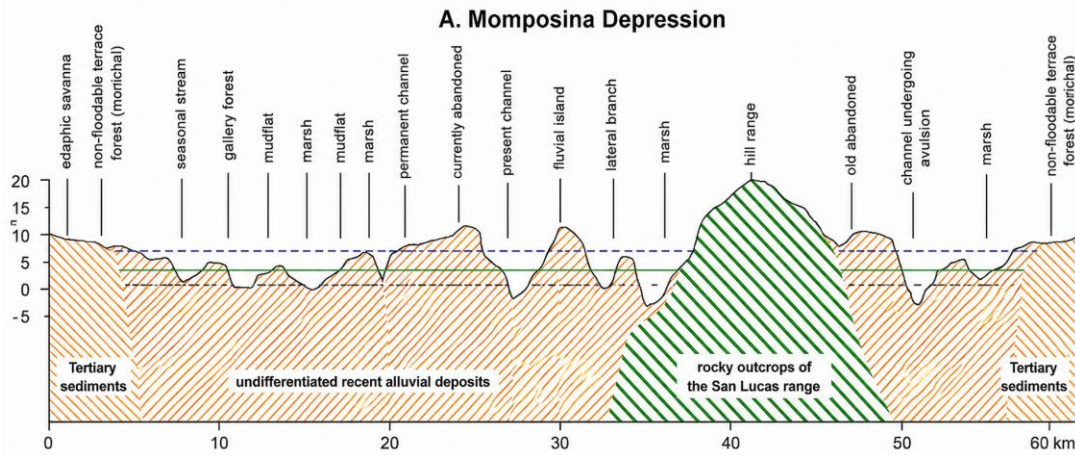


Fig. 2 Typical profiles of the alluvial plain in the Mompós region. Source: García-Lozano (2001)

The monthly rainfall pattern shown in Figure 3 confirms the seasonal behaviour of the Mompós floodplain. Rainfall remains low during January–March, increases progressively from April onwards, reaches its highest values around September–October, and decreases again towards the end of the year. This temporal structure provided the hydroclimatic basis for selecting January, March, September and November as representative months for the detailed visual validation described in the validation and performance assessment subsection. These months were therefore chosen to reflect contrasting seasonal conditions rather than to replace the full twelve-month evaluation.

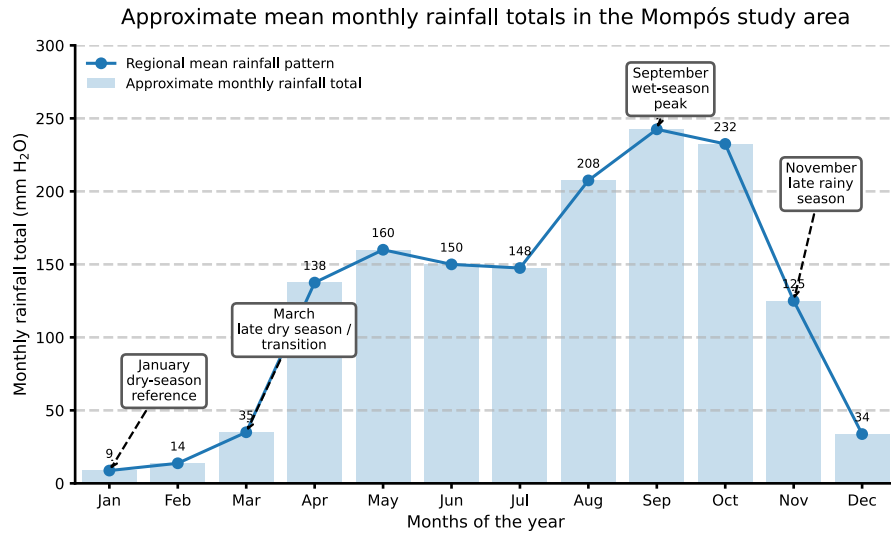


Fig. 3 Approximate mean monthly rainfall pattern in the Mompós study area, derived from regional weather-station records and using Parra-Plazas *et al.* (2023) as a reference basis for the temporal characterisation of Colombian weather-station data. The selected months highlight contrasting hydroclimatic conditions used for detailed map-based validation: January as a dry-season reference, March as a late dry-season or transition month, September as a wet-season peak, and November as a late rainy-season condition.

B. Remote Sensing

Remote sensing products were incorporated as contextual spatial information to support the characterisation of the flood-prone environment. Synthetic aperture radar (SAR) imagery was considered because of its capacity to provide surface-water and inundation-related information under adverse atmospheric conditions, while digital elevation models (DEMs) provided information on terrain structure, flow paths, low-relief areas and potential water-accumulation zones. In this study, these spatial datasets supported the interpretation of the floodplain context and the assessment of physically meaningful spatial patterns, whereas the predictive modelling stage was based on interpolated rainfall and water-level image banks derived from station records. Both SAR and DEM products are affected by uncertainties related to speckle noise, geometric distortions, vertical accuracy, spatial resolution and preprocessing choices, making systematic conditioning and harmonisation essential before their integration into predictive workflows (Adesina *et al.*, 2025; Zhang *et al.*, 2025; Soltani *et al.*, 2026).

C. Methodology

The proposed framework was designed as a sequential and reproducible pipeline integrating data conditioning, spatial interpolation, dimensionality reduction, predictive modelling and map-based validation. First, the hydrometeorological time series were cleaned, imputed and normalised to reduce the influence of missing values, outliers and transmission inconsistencies. In parallel, the spatial datasets were processed to derive terrain-related attributes and to provide floodplain context, following the increasing use of GIS-based and DEM-supported workflows in flood assessment and environmental modelling (Ruiz-Lendínez *et al.*, 2023; Copăcean *et al.*, 2025).

As illustrated in Figure 4, point-based rainfall and water-level observations were transformed into monthly spatial fields using inverse distance weighting (IDW) and ordinary kriging. These methods were selected because they are widely used to convert sparse hydrometeorological observations into continuous surfaces, although their performance depends on station density, spatial variability and the assumptions of each

interpolation method (Wang *et al.*, 2024; Workneh *et al.*, 2024). The resulting interpolated fields were organised as spatiotemporal image banks for each variable and interpolation method.

The modelling dataset covered a 23-year period, from 2000 to 2022. Images from 2000 to 2021 were used for training, while 2022 was reserved for out-of-sample comparison against the predictive outputs generated by the computational intelligence models. After interpolation, the image banks were transformed into a reduced PCA space to decrease dimensionality while preserving the dominant spatial structure of the data. ANN and SVR models were then trained on the retained principal-component scores, and the predicted outputs were reconstructed into the original image domain through inverse PCA. This design allowed the predictive models to operate in a more compact feature space while preserving direct spatial comparability between reconstructed predictions and 2022 reference maps (Xie *et al.*, 2021; Yeşilköy *et al.*, 2024).

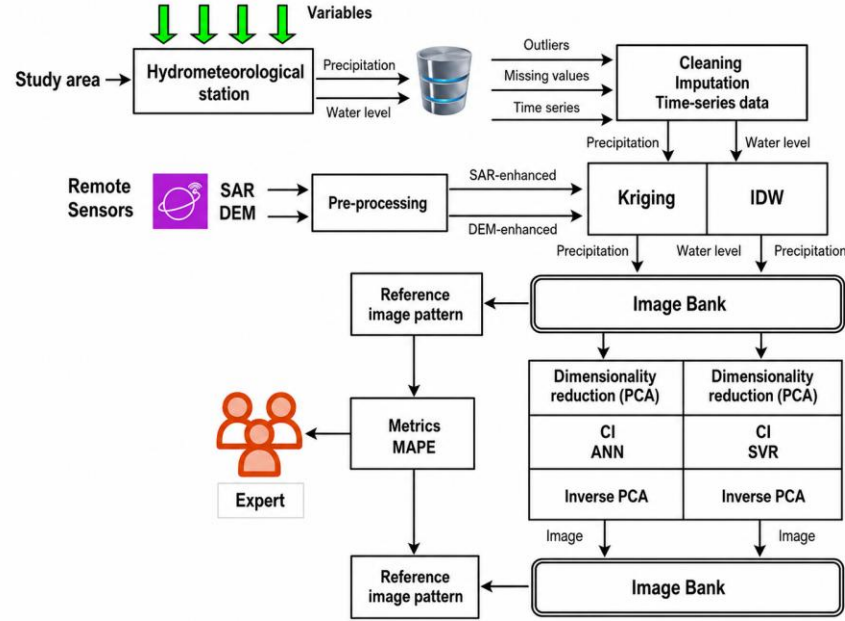


Fig. 4 Methodological workflow used to integrate hydrometeorological time-series data, SAR/DEM-derived spatial information, IDW and ordinary kriging interpolation, PCA-based dimensionality reduction, ANN/SVR prediction, inverse PCA reconstruction and map-based validation.

Figure 4 also clarifies the role of each methodological block within the proposed framework. The hydrometeorological and spatial datasets were first conditioned and transformed into comparable raster inputs; these inputs were then organised into image banks and projected into a reduced PCA space for predictive modelling. The ANN and SVR models operated only on the retained PCA scores, and the predicted outputs were subsequently reconstructed through inverse PCA to recover spatially explicit maps. This structure ensured traceability from the original data sources to the final validation products, while preserving the separation between the training period and the 2022 reference maps used for evaluation.

D. PCA-based dimensionality reduction

Because the interpolated images generated high-dimensional matrices, principal component analysis (PCA) was incorporated as an intermediate dimensionality-reduction stage. Its role was to preserve the most relevant variance structure of the spatial fields while improving computational efficiency and reducing multicollinearity effects before predictive training (Jia *et al.*, 2022; Shojaeian *et al.*, 2024; Yang *et al.*, 2020). The use of PCA was motivated by three main factors: the computational burden associated with high-resolution image inputs, the strong spatial correlation between neighbouring pixels, and the need to reduce

the effect of measurement and interpolation noise in the feature space (Westra *et al.*, 2007; Hasan & Abdulazeez, 2021; Jia *et al.*, 2022; Shojaeian *et al.*, 2024).

The number of retained components, (k), was selected using a dynamic cumulative-variance criterion, defined as the minimum number of principal components required to explain at least 95% of the total variance. This threshold was adopted because it is widely used in hydrological and environmental studies and provides a suitable balance between preserving the dominant information in the data and maintaining computational tractability (Westra *et al.*, 2007; Jia *et al.*, 2022; Shojaeian *et al.*, 2024).

In this study, the criterion also allowed the number of components to adapt to the complexity of each dataset, while reducing redundancy, noise and multicollinearity in the spatial feature space. Therefore, PCA was not used merely as a compression technique, but as a stabilising step that linked the high dimensionality of the interpolated spatial images with the training requirements of the predictive models (Chauhan & Mathews, 2020; Rostampour *et al.*, 2025).

In matrix form, this can be expressed as follows:

- $X \in R^{n \times m}$ is the centred data matrix, where n is the number of observations (images) and m is the number of original variables (pixels).
- $C = \frac{1}{n-1} X^T X$ is the covariance matrix of size $m \times m$.
- $W_k \in R^{m \times k}$ is the column matrix whose columns are the k eigenvectors associated with the k largest eigenvalues of C .

The PCA transformation is therefore defined as Equation 1:

$$Z = XW_k \quad \text{Equation (1)}$$

Where:

- $Z \in R^{n \times k}$ is the projection of the data onto the reduced subspace.

The inverse transformation for reconstruction in the original space is given by Equation 2:

$$\hat{Z} = Z_{pred}W_k^T + 1_n\mu^T \quad \text{Equation (2)}$$

- where Z_{pred} represents the predictions produced by ANN or SVR in the reduced space.
- is the original mean vector used for centring, ensuring that the predictions preserve their original spatial structure and allowing direct comparison with the reference images generated during the interpolation stage;
- and 1_n is the column vector of ones used to replicate μ_{across} all observations

This matrix formulation facilitates computational implementation and ensures that both the projection and its inverse preserve the orthogonal structure defined by W_k .

The cumulative-variance criterion used to select the retained components is expressed in Equation 3.

$$\frac{\sum_{i=1}^k l_i}{\sum_{i=1}^m l_i} \geq 0.95 \quad \text{Equation (3)}$$

E. Predictive modelling

Flood prediction was performed using artificial neural networks (ANN) and support vector regression (SVR), selected for their capacity to model non-linear relationships under high spatial and temporal complexity (Tang *et al.*, 2023; Zhou & Kang, 2023; Dasgupta *et al.*, 2025; Shao *et al.*, 2026). Because the predictive task involved the estimation of continuous pixel-based rainfall and water-level values rather than

class labels, the support vector machine framework was implemented in its regression form, hereafter referred to as SVR.

In the proposed workflow, the interpolated images were first transformed into a reduced PCA space, after which ANN and SVR were trained and the outputs were reconstructed through the inverse PCA transformation for comparison with the reference patterns. This strategy enabled predictive learning to be carried out in a more compact and computationally efficient feature space while retaining the spatial interpretability of the reconstructed maps (Hasan & Abdulazeez, 2021; Jia *et al.*, 2022; Rana *et al.*, 2024).

F. Validation and performance assessment

The final stage of the workflow focused on the validation and visualisation of the reconstructed predictive maps. Model performance was assessed using the mean absolute percentage error (MAPE) and direct visual comparison between the predicted maps and the corresponding 2022 reference maps. This evaluation was designed to assess not only numerical accuracy, but also the spatial coherence of the reconstructed rainfall and water-level patterns.

For model evaluation, the full modelling dataset covered the period 2000–2022. Images from 2000 to 2021 were used for training, while the year 2022 was reserved for out-of-sample evaluation. The evaluation was performed for the twelve months of 2022 in order to characterise the annual behaviour of each modelling configuration. Therefore, the monthly MAPE values represent the complete annual validation evidence, rather than a partial assessment based only on selected months.

In addition to the twelve-month quantitative evaluation, four representative months, namely January, March, September and November, were selected for detailed visual comparison. This selection was based on the seasonal hydroclimatic behaviour of the Mompós floodplain. January represents dry-season conditions, March corresponds to a late dry-season or transitional stage, September represents a wet-season peak, and November reflects late rainy-season conditions, when floodplain wetness may persist. These months were therefore used to illustrate contrasting hydrological conditions, not to replace the full twelve-month performance assessment.

The visual assessment was based on a three-map comparison comprising the 2022 reference map, the reconstructed prediction and the corresponding MAPE distribution. This structure made it possible to examine the spatial distribution of the error, the preservation of dominant gradients, local maxima, transition zones and hydrologically meaningful spatial patterns. In this way, the validation considered both the magnitude of the numerical error and the physical plausibility of the reconstructed maps.

MAPE was calculated as shown in Equation (4):

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad \text{Equation (4)}$$

Where y_i is the reference value at the valid grid cell (i), $\hat{y}_i$ is the predicted value at the same grid cell, and (n) is the total number of valid grid cells within the study-area mask. Cells with zero or near-zero reference values were treated consistently to avoid artificial inflation of percentage errors, particularly in low-rainfall areas or dry months.

Taken together, this validation strategy allowed the predictive outputs to be assessed from both quantitative and spatial perspectives, while preserving traceability between the reconstructed maps, the 2022 reference patterns and the hydrological interpretation of the results.

4. Results and comparative analysis

A 23-year dataset covering the period 2000–2022 was assembled from rainfall and water-level station records and used to generate monthly spatiotemporal surfaces through IDW and ordinary kriging interpolation. Images from 2000 to 2021 were used for model training, while 2022 was reserved as the out-of-sample reference year for evaluating predictive performance. Eight modelling configurations were established by combining two target variables, rainfall and water level, with two interpolation methods, IDW and ordinary kriging, and two predictive models, ANN and SVR, as summarised in Table 1.

Variable	Interpolator Method	Predictive model
Rainfall	Ordinary kriging	Artificial Neural Networks (ANN)
		Support Vector Regression (SVR)
	IDW	Artificial Neural Networks (ANN)
		Support Vector Regression (SVR)
Water level	Ordinary kriging	Artificial Neural Networks (ANN)
		Support Vector Regression (SVR)
	IDW	Artificial Neural Networks (ANN)
		Support Vector Regression (SVR)

Table 1. Modelling configurations evaluated in the comparative analysis.

Each configuration produced 12 monthly forecasts, resulting in 96 prediction cases for the 2022 validation year. These outputs were assessed using MAPE and visual comparison against the corresponding 2022 reference maps. The analysis was organised at two complementary levels: first, a seasonal visual evaluation using January, March, September and November to represent dry-season, transitional, wet-season peak and late rainy-season conditions; and second, a configuration-based comparison to identify which combinations of interpolation method and predictive model were more suitable for rainfall and water-level prediction.

Methodological framework for PCA-based image forecasting

Training and forecasting are conducted in the reduced PCA space, followed by inverse PCA reconstruction in the original image domain.

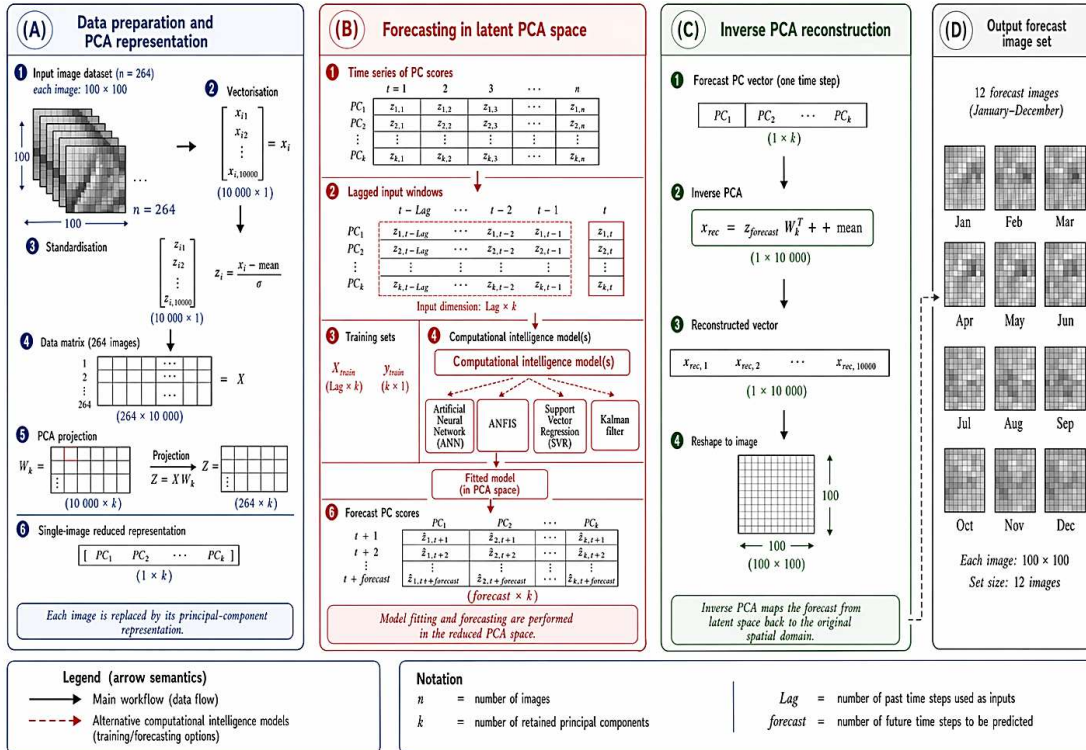
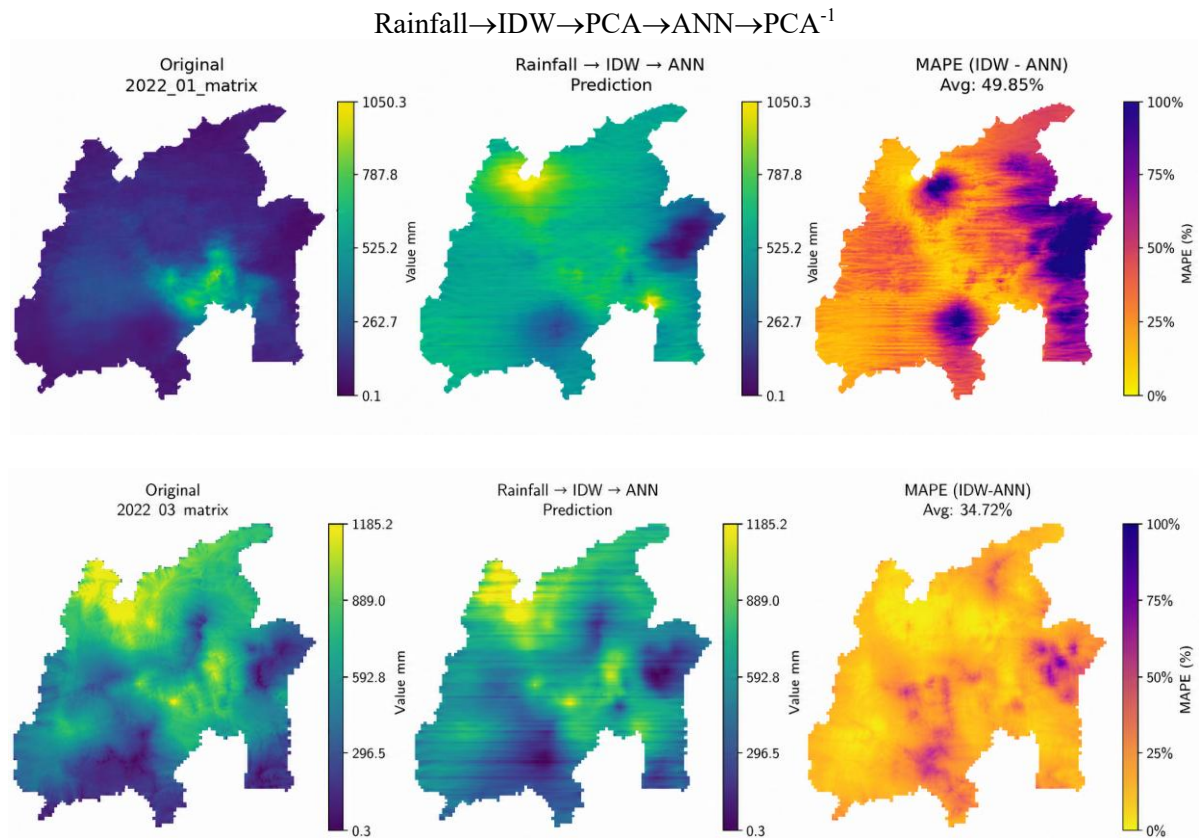


Fig. 5 PCA-based image forecasting workflow used for ANN and SVR prediction.

Figure 5 summarises the PCA-based image forecasting workflow used in this study. The original image sequence was standardised, vectorised and projected into a reduced latent space. Forecasting was then performed on the retained principal-component scores using ANN and SVR models. The predicted PCA scores were subsequently transformed back into the original image domain through inverse PCA, producing monthly forecast maps that could be directly compared with the 2022 reference images. This structure allowed the predictive task to be carried out in a computationally efficient latent space while preserving the spatial interpretability of the reconstructed outputs.

Because all evaluated configurations followed the same PCA-based forecasting workflow, the differences observed in the results can be interpreted in terms of the target variable, the interpolation method and the predictive model used after dimensionality reduction.

To evaluate whether the reconstructed maps were consistent across contrasting hydroclimatic conditions, the visual analysis focused on four representative months selected from the seasonal rainfall pattern shown in Figure 3. January represents dry-season conditions, March captures the transition from the dry season, September represents the wet-season peak, and November reflects late rainy-season conditions. Figures 6 and 7 present the seasonal visual validation for the most representative configurations of each target variable: IDW–PCA–ANN for rainfall and ordinary kriging–PCA–ANN for water level.



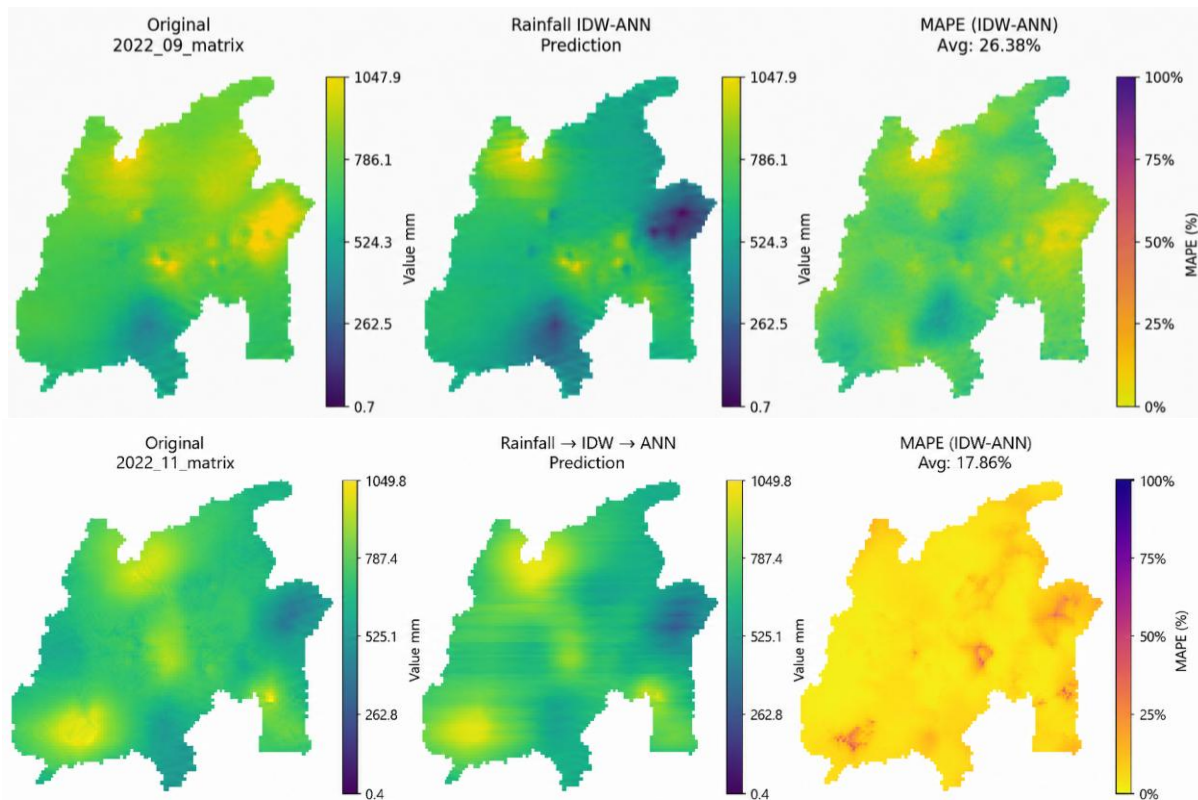


Fig. 6 Seasonal visual validation of the rainfall IDW-PCA-ANN configuration for January, March, September and November 2022. Each row compares the 2022 reference rainfall map, the reconstructed prediction and the corresponding spatial MAPE distribution

The seasonal validation of the rainfall IDW-PCA-ANN configuration shows that this modelling chain was able to preserve relevant spatial contrasts across contrasting hydroclimatic conditions. January and March represent low-rainfall and transitional conditions, where percentage errors may become more sensitive because reference rainfall values are smaller. In contrast, September and November represent wetter conditions, where the main challenge is to reproduce local rainfall maxima and sharper spatial gradients. The reconstructed maps indicate that the IDW-PCA-ANN configuration preserved the heterogeneous structure of rainfall reasonably well, particularly in the wetter months, although this advantage also implies greater sensitivity to station distribution and possible interpolation artefacts inherited from IDW.

Within the IDW based rainfall configurations, ANN and SVR should not be interpreted as equivalent alternatives. The ANN based reconstruction was selected for visual presentation because it provided a more flexible response to local rainfall variability after PCA-based dimensionality reduction. The corresponding IDW-PCA-SVR configuration remained relevant as a comparative alternative, particularly for assessing whether a more conservative regression model could reproduce broad rainfall tendencies. However, its role is better represented in the configuration based MAPE synthesis rather than through additional map panels, to avoid overloading the main manuscript with repetitive visual outputs.

Figure 7 presents the corresponding seasonal validation for water level using the ordinary kriging-PCA-ANN configuration. This configuration was selected because water-level fields are more spatially continuous and hydrologically integrated than rainfall fields, making ordinary kriging a more suitable interpolation basis for preserving regional structure before PCA-based prediction.

Water level→Kriging→PCA→ANN→PCA⁻¹

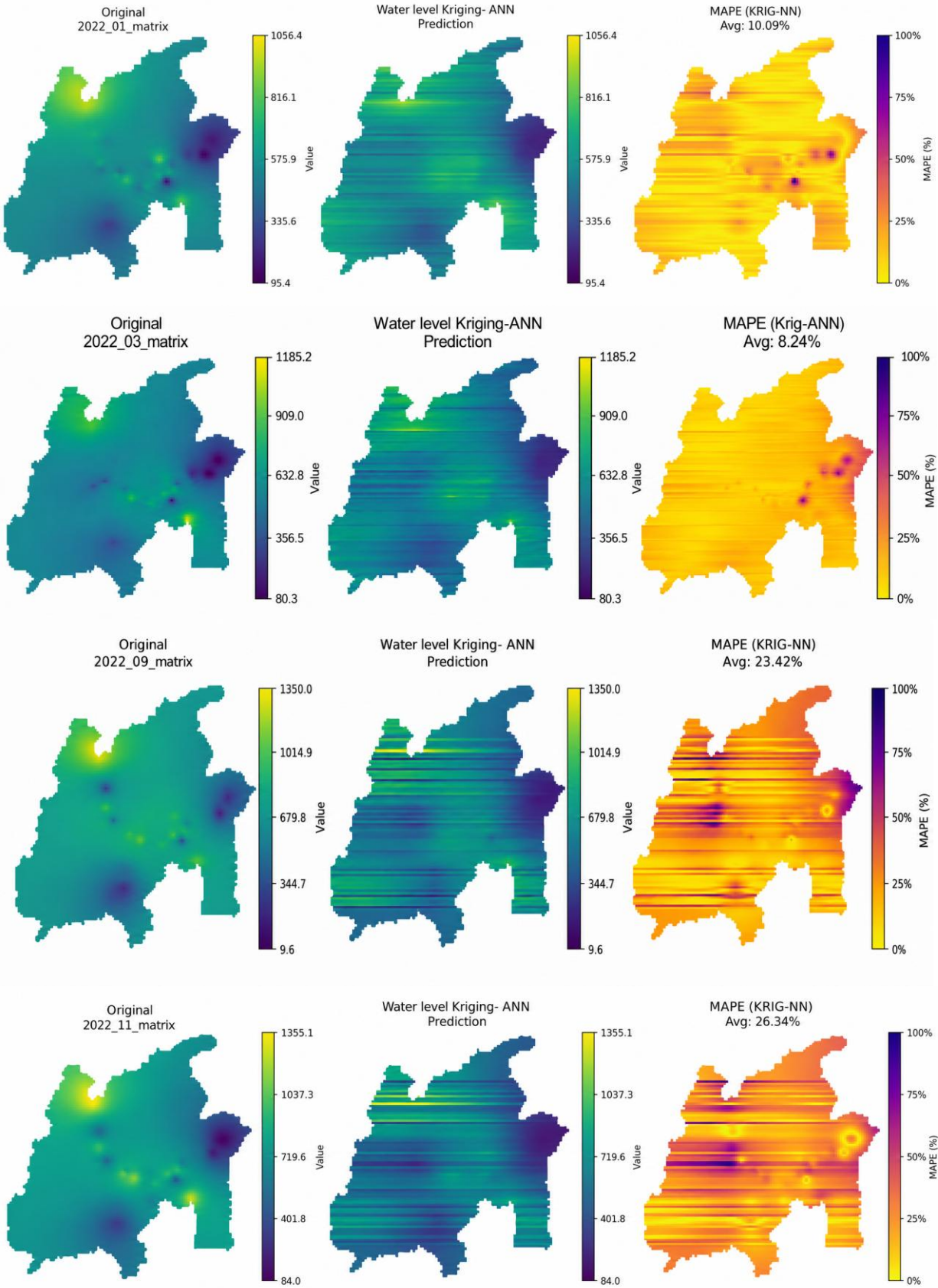


Fig. 7 Seasonal visual validation of the water-level ordinary kriging–PCA–ANN configuration for January, March, September and November 2022. Each row compares the 2022 reference water-level map, the reconstructed prediction and the corresponding spatial MAPE distribution

The seasonal validation of the water-level ordinary kriging–PCA–ANN configuration shows a more spatially continuous reconstruction pattern than that observed for rainfall. This behaviour is consistent with the physical nature of river-level fields, which are governed by regional hydrological connectivity rather than isolated local maxima. Across the representative months, the reconstructed maps preserved the dominant water-level structure under both low-water and wetter-season conditions. Although local discrepancies remained in transition or high-water periods, the ordinary kriging–PCA–ANN configuration provided a coherent spatial representation of the main hydrological gradients, supporting its suitability for water-level prediction.

The comparison between Figures 6 and 7 confirms that rainfall and water level required different modelling configurations. Rainfall prediction benefited from the local sensitivity of IDW combined with the nonlinear reconstruction capacity of ANN, whereas water-level prediction was better represented by ordinary kriging–PCA–ANN because this configuration preserved regional continuity and hydrological coherence. These two figures therefore provide representative visual evidence for the variable-specific behaviour of the framework, while the remaining configurations are summarised quantitatively in the subsequent comparative table.

Based on the representative visual examples in Figures 6 and 7 and the full twelve-month MAPE assessment, the comparative analysis focused on three aspects: the influence of the interpolation method on the reconstructed maps, the different behaviour of ANN and SVR after PCA-based dimensionality reduction, and the degree to which each configuration preserved physically meaningful rainfall and water-level patterns. The representative maps were interpreted using the same three-panel structure: the 2022 reference map, the reconstructed prediction and the spatial MAPE distribution. This allowed the numerical error to be read together with spatial agreement, including the preservation of dominant gradients, local maxima, transition zones and hydrologically meaningful spatial patterns.

The cross-variable comparison also showed that rainfall and water level should not be interpreted through the same spatial criterion. Rainfall prediction required sensitivity to local maxima and heterogeneous spatial patterns, whereas water-level prediction required stronger regional continuity and hydrological coherence. Therefore, the comparative analysis did not seek a single universal configuration, but rather evaluated which interpolation–prediction chain was more suitable for each target variable.

Taken together, the visual results indicate that interpolation quality strongly conditioned predictive behaviour. IDW tended to preserve local rainfall variability more effectively, but also introduced stronger spatial error dispersion. Ordinary kriging provided a more stable basis for water-level reconstruction because it favoured spatial continuity. ANN and SVR also contributed differently after PCA-based dimensionality reduction: ANN provided greater flexibility for reconstructing spatial detail, whereas SVR was more conservative and better suited to broad spatial tendencies. These patterns are further supported by the selected monthly MAPE comparisons shown in Figure 8 and are synthesised across all evaluated configurations in Table 2.

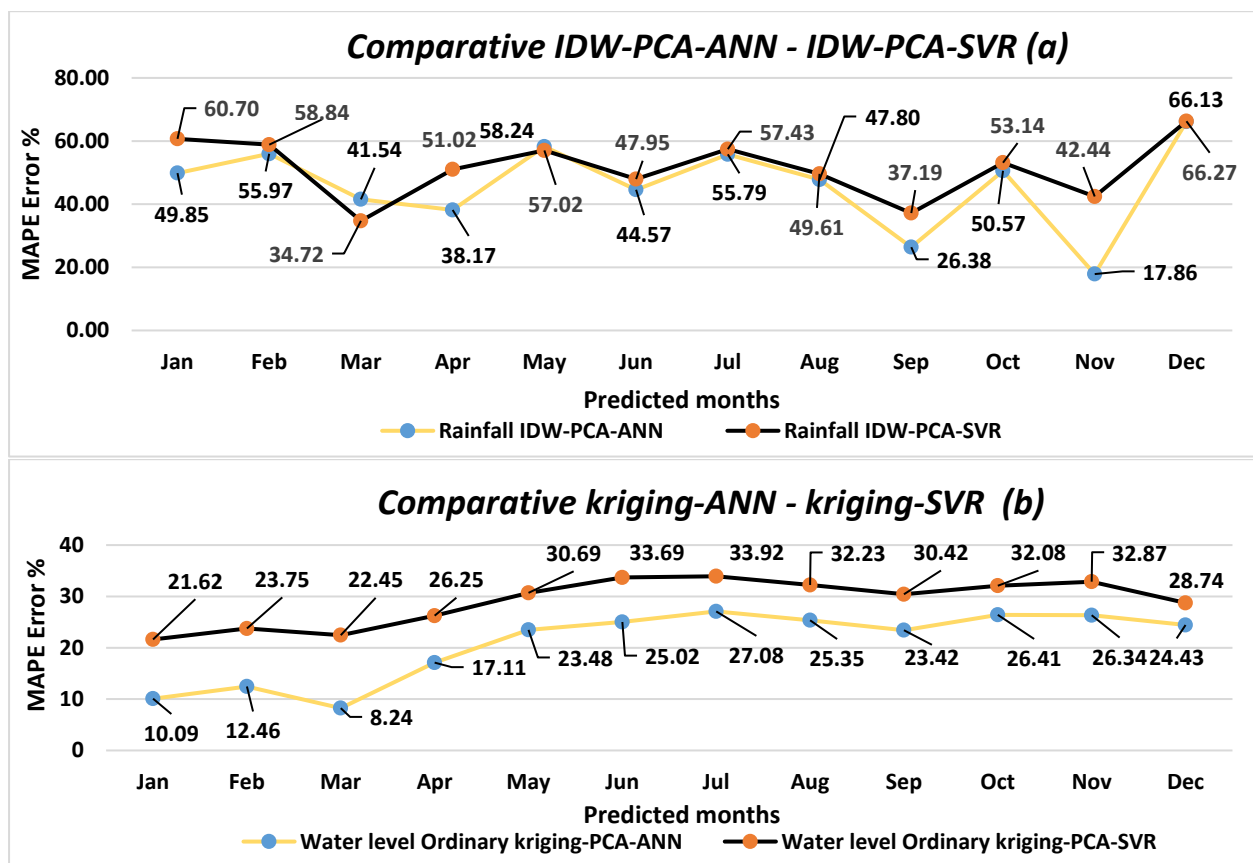


Fig. 8 Selected monthly MAPE comparison for representative variable-specific configuration pairs during the 2022 validation year. (a) Rainfall prediction using IDW inputs, comparing IDW-PCA-ANN and IDW-PCA-SVR. (b) Water-level prediction using ordinary kriging inputs, comparing ordinary kriging-PCA-ANN and ordinary kriging-PCA-SVR. The selected panels complement the spatial validation shown in Figures 6 and 7 and provide monthly quantitative support for the configuration-based synthesis reported in Table 2.

Figure 8 reinforces the variable-dependent behaviour of the evaluated framework. For rainfall under IDW inputs, ANN and SVR showed similar annual behaviour, with mean MAPE values of 52.86% and 53.78%, respectively. This indicates that the main advantage of the IDW-based rainfall configurations was not a large numerical separation between ANN and SVR, but their ability to retain local rainfall variability compared with smoother ordinary kriging-based alternatives. For water level under ordinary kriging inputs, ANN consistently produced lower MAPE values than SVR, with annual means of 20.79% and 29.06%, respectively. This supports the selection of ordinary kriging-PCA-ANN as the representative water-level configuration because it combined lower error under this interpolation setting with stronger spatial continuity and hydrological coherence.

In summary, the monthly MAPE curves support the map-based interpretation: IDW-based rainfall configurations retained local variability with similar ANN and SVR behaviour, whereas ordinary kriging-PCA-ANN was clearly more favourable than ordinary kriging-PCA-SVR for water-level prediction under a spatial-continuity criterion.

Table 2 was constructed using two complementary levels of evidence. The main quantitative comparison is based on the twelve monthly MAPE values obtained for each configuration during the 2022 evaluation year. Therefore, the reported mean and range describe the full out-of-sample annual behaviour of each model configuration. In addition, the four representative months shown in Figures 6 and 7, January, March, September and November are reported separately to support the visual interpretation of contrasting hydrological conditions. These four months were not used as a substitute for the twelve-month performance

assessment, but as illustrative cases for analysing spatial error distribution, local maxima, transition zones and hydrologically meaningful gradients.

Variable	Configuration	MAPE evidence and spatial interpretation
Rainfall	IDW-PCA-ANN	Rep. MAPE (%): Jan 49.85; Mar 41.54; Sep 26.38; Nov 17.86. Annual: mean 52.86; range 34.24–66.27. Preserved local rainfall contrasts and several spatial maxima; most representative rainfall configuration because it retained local variability, although it remained sensitive to station distribution and inherited IDW artefacts.
Rainfall	IDW-PCA-SVR	Rep. MAPE (%): Jan 60.70; Mar 41.54; Sep 37.19; Nov 42.44. Annual: mean 53.78; range 37.19–66.13. Reproduced broad rainfall tendencies with similar annual behaviour to IDW–PCA–ANN, but with a more conservative response and greater tendency to smooth local extremes.
Rainfall	Ordinary kriging-PCA-ANN	Rep. MAPE (%): Jan 39.61; Mar 62.16; Sep 75.54; Nov 49.32. Annual: mean 54.50; range 25.09–75.84. Produced spatially coherent rainfall fields but smoothed local rainfall maxima; useful for broad rainfall structure, but less effective for heterogeneous rainfall peaks.
Rainfall	Ordinary kriging-PCA-SVR	Rep. MAPE (%): Jan 70.63; Mar 38.41; Sep 90.98; Nov 71.91. Annual: mean 69.01; range 38.41–94.68. Maintained broad spatial continuity but showed excessive smoothing and weak sensitivity to local rainfall contrasts; least favourable rainfall configuration among the evaluated alternatives.
Water level	Ordinary kriging-PCA-ANN	Rep. MAPE (%): Jan 10.09; Mar 8.24; Sep 23.42; Nov 26.34. Annual: mean 20.79; range 8.24–27.08. Preserved the dominant water-level structure, regional continuity and main hydrological gradients; the most hydrologically coherent water-level configuration.
Water level	Ordinary kriging-PCA-SVR	Rep. MAPE (%): Jan 21.62; Mar 22.45; Sep 30.42; Nov 32.87. Annual: mean 29.06; range 21.62–33.92. Reproduced broad spatial continuity but with consistently higher errors than ANN; conservative water-level alternative with stronger smoothing.
Water level	IDW-PCA-ANN	Rep. MAPE (%): Jan 10.17; Mar 10.04; Sep 21.05; Nov 22.15. Annual: mean 18.48; range 10.04–25.78. Performed well in simpler months and produced competitive MAPE values but was more sensitive to local artefacts under complex spatial conditions; moderate and variable for water-level prediction.
Water level	IDW-PCA-SVR	Rep. MAPE (%): Jan 14.28; Mar 11.50; Sep 14.53; Nov 17.58. Annual: mean 16.22; range 10.12–22.37. Produced the lowest annual mean MAPE for water level, but spatial interpretation should be cautious because SVR and IDW can smooth peaks and retain directional interpolation artefacts.

Table 2. Comparative synthesis of representative-month MAPE values, annual MAPE evidence and spatial behaviour for the eight evaluated configurations.

The comparison shows that the most favourable configuration depended on both the target variable and the evaluation criterion considered. For rainfall, the four configurations produced relatively high and variable MAPE values, reflecting the heterogeneous and locally variable nature of rainfall fields. IDW–PCA–ANN and IDW–PCA–SVR showed similar annual mean errors, but their relevance lies in the better preservation

of local rainfall variability compared with the smoother ordinary kriging-based alternatives. Ordinary kriging–PCA–SVR produced the highest annual mean error for rainfall, indicating that the combination of a smoothing interpolator and a conservative regression model was less suitable for local rainfall extremes.

For water-level prediction, the numerical and spatial criteria did not lead to exactly the same ranking. IDW–PCA–SVR produced the lowest annual mean MAPE among the water-level configurations, indicating that SVR can be numerically competitive when applied to smoother IDW-derived fields. However, this result should be interpreted together with the spatial behaviour of the reconstructed maps, because IDW-based water-level predictions may retain local or directional artefacts inherited from the interpolation stage. In contrast, ordinary kriging–PCA–ANN provided the most hydrologically coherent reconstruction, preserving regional continuity and the dominant water-level structure across the representative months.

The comparison between ANN and SVR also showed that both models contributed differently after PCA-based dimensionality reduction. ANN reconstructions retained more spatial detail and dominant hydrological gradients, whereas SVR reconstructions were smoother and, in several cases, produced lower numerical errors. For this reason, MAPE values were interpreted together with the spatial plausibility of the maps, since a lower error value did not necessarily correspond to a more hydrologically meaningful prediction.

These results indicate that the proposed framework should not be interpreted as a single fixed model. Its main value lies in identifying variable-sensitive modelling strategies: IDW-based configurations were more suitable for representing local rainfall variability, whereas ordinary kriging-based configurations were more appropriate for hydrologically coherent water-level reconstruction. The following discussion addresses interpolation effects, model behaviour after dimensionality reduction, and the broader implications of the framework for flood prediction in data-scarce floodplains.

5. Discussion

The comparative patterns observed across the evaluated configurations indicate that flood prediction in data-scarce environments depends on the interaction between spatial representation, dimensionality reduction and predictive learning, rather than on the isolated performance of a single algorithm. In this study, interpolation, PCA and ANN/SVR modelling operated as interdependent stages of the same predictive chain. Therefore, the results should not be interpreted as a simple comparison between ANN and SVR or between IDW and ordinary kriging, but as evidence that each methodological block shaped the final hydrometeorological response according to the physical structure of the target variable (Aguasca-Colomo *et al.*, 2019; Jayathilake *et al.*, 2023; Rostampour *et al.*, 2025).

The contrast between IDW and ordinary kriging reflects a fundamental trade-off between local sensitivity and regional continuity. IDW was more suitable for rainfall fields because it retained local spatial contrasts and rainfall maxima, which are important in heterogeneous precipitation patterns. However, this same local sensitivity also increased spatial error dispersion and made the reconstructed maps more dependent on station distribution and inherited interpolation artefacts. Ordinary kriging, in contrast, produced smoother and more spatially coherent fields, which made it more appropriate for water-level prediction, where regional hydrological continuity is more relevant than isolated local maxima. This explains why ordinary kriging–PCA–ANN was not necessarily the lowest-MAPE water-level configuration but still provided the most hydrologically coherent reconstruction (Goovaerts, 2000; Majone *et al.*, 2022; Workneh *et al.*, 2024).

This interpretation is consistent with previous studies showing that deterministic methods such as IDW may be useful under strong local heterogeneity, whereas ordinary kriging is generally more appropriate when spatial dependence and regional continuity are dominant features of the target variable (Goovaerts, 2000; Li & Heap, 2011; Kwon *et al.*, 2020; Majone *et al.*, 2022; Workneh *et al.*, 2024). However, the smoother

behaviour of kriging may also reduce sensitivity to abrupt local extremes, which remains a relevant limitation in flood-related applications (Ikechukwu *et al.*, 2017; Fattorini *et al.*, 2023).

A central implication of these results is that MAPE should not be interpreted independently from spatial plausibility. Percentage-based errors are useful for comparing configurations, but they may favour smoother reconstructions or become sensitive to low reference values, particularly in dry rainfall months. For this reason, the evaluation combined monthly MAPE evidence with map-based interpretation of spatial gradients, local maxima and hydrologically meaningful structures. This combined assessment is especially important in flood-prone floodplains, where a numerically acceptable prediction may still be of limited operational value if it fails to preserve the spatial organisation of water-level or rainfall patterns.

To synthesise these interpretative differences, Table 3 compares the main modelling configurations in terms of methodological behaviour, limitations and hydrological implications. Unlike Table 2, which reports representative and annual MAPE evidence, Table 3 focuses on how each configuration should be interpreted when the prediction task prioritises local rainfall variability, regional water-level continuity or broader flood-related coherence.

Configuration	Methodological behaviour and limitation	Hydrological implication and Supporting references
IDW + PCA + ANN	Main behaviour: Preserves local rainfall contrasts and adapts to heterogeneous spatial variability after PCA-based reduction. Main limitation: Sensitive to station distribution, local error dispersion and artefacts inherited from IDW.	Best suited condition: Rainfall prediction under locally heterogeneous conditions. Hydrological implication: Useful when capturing local rainfall maxima is more important than enforcing spatial smoothness. Suggested support: Li & Heap (2011); Workneh <i>et al.</i> (2024); Jayathilake <i>et al.</i> (2023).
IDW + PCA + SVR	Main behaviour: Reproduces broad tendencies and can produce competitive numerical errors under smoother fields. Main limitation: More conservative; can smooth local extremes and retain IDW-related directional artefacts.	Best suited condition: Intermediate scenarios with moderate spatial variability. Hydrological implication: Useful as a conservative alternative, but spatial interpretation should remain cautious. Suggested support: Jia <i>et al.</i> (2022); Shojaeian <i>et al.</i> (2024); Rostampour <i>et al.</i> (2025).
Ordinary kriging + PCA + ANN	Main behaviour: Preserves regional continuity and dominant spatial structure, especially for water-level fields. Main limitation: May smooth abrupt local responses or local extremes because of the interpolation stage.	Best suited condition: Water-level prediction and spatially coherent hydrological fields. Hydrological implication: Suitable for variables governed by integrated basin-scale processes and regional continuity. Suggested support: Goovaerts (2000); Majone <i>et al.</i> (2022); Workneh <i>et al.</i> (2024).
Ordinary kriging + PCA + SVR	Main behaviour: Represents broad spatial continuity under stable input conditions. Main limitation: Conservative response; less sensitive to local maxima, sharper gradients and transition zones.	Best suited condition: Smoother water-level fields or denser station networks. Hydrological implication: Favours spatial coherence but may underrepresent sharp flood-related responses. Suggested support: Goovaerts (2000); Li & Heap (2011); Kwon <i>et al.</i> (2020).
Variable-sensitive configuration selection	Main behaviour: Adapts the modelling chain to the physical behaviour of each target variable instead of assuming a universal best model. Main limitation: Requires joint interpretation of numerical error, spatial plausibility and hydrological meaning.	Best suited condition: Data-scarce floodplains with heterogeneous rainfall and spatially integrated water levels. Hydrological implication: Supports a flexible framework rather than a fixed algorithmic preference. Suggested support: Khan <i>et al.</i> (2024); Liu <i>et al.</i> (2025); Dasgupta <i>et al.</i> (2025).

Table 3. Interpretative comparison of the main modelling configurations in terms of methodological behaviour, limitations and hydrological implications.

The comparison makes clear that no single configuration was universally superior. Instead, model suitability depended on whether the prediction task prioritised local rainfall peaks, broader water-level continuity or conservative reconstruction of spatial tendencies. This reinforces the need to select interpolation and learning strategies according to the physical structure of the target variable rather than by algorithmic preference alone.

The joint interpretation of rainfall and water level adds an important hydrological dimension to the framework. Their relationship is not immediate, because it is mediated by lag effects, soil moisture

conditions, storage processes and floodplain connectivity. Even so, the results suggest that the predictive chain can preserve physically plausible behaviour across both variables when the interpolation method is selected according to the spatial structure of the target field. This is particularly relevant in wetter months, when rainfall variability and water-level response become more closely associated within the floodplain system. However, this study did not explicitly model rainfall–runoff lag relationships, and future work should evaluate time-lagged and multivariable structures to better represent coupled flood dynamics (Rana *et al.*, 2024; Liu *et al.*, 2025; Zhou *et al.*, 2026).

More broadly, the results show that robust flood-related prediction in data-scarce tropical floodplains depends on methodological balance rather than on the isolated superiority of a single model. The proposed framework combines long-term station records, SAR- and DEM-derived spatial information, IDW and ordinary kriging interpolation, PCA-based feature reduction, ANN/SVR prediction and map-based validation into a transparent workflow. Its practical value lies in combining temporal continuity with spatially explicit environmental context, producing predictive outputs that can support environmental monitoring, risk interpretation and the future development of flood early-warning tools in regions with fragmented monitoring networks (Abdoulhalik & Ahmed, 2024; Li *et al.*, 2023; Zhang *et al.*, 2025).

Several limitations should also be acknowledged. The reliability of the reconstructed maps depends on station density, spatial representativeness and the assumptions of the interpolation method. In addition, PCA reduces dimensionality by preserving dominant variance structures, but may attenuate local extremes that are relevant for flood applications (Shojaeian *et al.*, 2024). Finally, MAPE is sensitive to low reference values, especially in dry rainfall months, which reinforces the need to interpret numerical errors together with spatial patterns. These limitations do not invalidate the framework, but indicate where future improvements should focus, particularly on uncertainty quantification, time-lagged rainfall–water-level coupling and the integration of additional remote-sensing or hydrological covariates.

6. CONCLUSIONS

This study developed and evaluated a hybrid framework for spatially explicit flood-related prediction in a data-scarce tropical floodplain. The framework integrated long-term rainfall and water-level station records, SAR- and DEM-derived spatial information, IDW and ordinary kriging interpolation, PCA-based dimensionality reduction, ANN/SVR prediction, inverse PCA reconstruction and map-based validation. Its main contribution lies in demonstrating that point-based hydrometeorological observations can be transformed into continuous predictive maps through a reproducible and interpretable modelling chain, rather than through an isolated forecasting algorithm.

The results show that predictive performance was strongly conditioned by the interaction between the target variable, the interpolation method and the learning model. For rainfall, IDW-based configurations were more suitable for preserving local spatial variability and rainfall contrasts, although they also produced greater spatial error dispersion and remained sensitive to station distribution. IDW–PCA–ANN and IDW–PCA–SVR showed similar annual behaviour, indicating that the main advantage of the IDW-based rainfall configurations was related more to the preservation of local spatial structure than to a large numerical separation between ANN and SVR.

For water-level prediction, ordinary kriging–based configurations provided stronger regional continuity and more hydrologically coherent reconstructions. Although IDW–PCA–SVR produced the lowest annual mean MAPE among the water-level configurations, its spatial interpretation required caution because IDW-derived inputs and SVR-based smoothing may retain directional artefacts or attenuate local hydrological detail. In contrast, ordinary kriging–PCA–ANN provided the most physically coherent water-level reconstruction by preserving the dominant spatial structure, regional gradients and hydrological continuity across the representative months.

The findings also confirm that MAPE should not be interpreted as the only criterion for model selection. A lower percentage error did not always imply a more hydrologically meaningful prediction, particularly when the reconstructed maps smoothed local extremes or retained artefacts inherited from the interpolation stage. The combined use of annual MAPE evidence, representative seasonal months and spatial error maps allowed the evaluation to consider both numerical performance and physical plausibility.

Overall, the proposed framework provides a transferable methodological basis for flood prediction in data-constrained floodplains. Its value lies in identifying variable-sensitive modelling strategies: IDW-based chains are more appropriate when local rainfall variability must be retained, whereas ordinary kriging-based chains are more suitable when regional water-level continuity and hydrological coherence are prioritised. This supports the use of hybrid temporal-spatial workflows as a practical foundation for environmental monitoring, risk interpretation and the future development of flood early-warning tools in vulnerable tropical regions.

7. FUTURE WORK

Future research should first address the temporal resolution of the proposed framework. Although monthly aggregation provided stable spatial patterns and allowed a consistent twelve-month evaluation, it also limited the representation of short-duration processes such as intense rainfall pulses, rapid water-level rises and rainfall-runoff delays. Future implementations should therefore incorporate higher-frequency station records and locally bias-corrected satellite rainfall products to evaluate whether the framework can capture sub-monthly hydrological variability without losing spatial interpretability.

A second priority is to strengthen the treatment of interpolation uncertainty and validation design. Because the predictive models operate downstream of IDW- and ordinary kriging-derived surfaces, part of the interpolation error may be propagated into the PCA space and then into the reconstructed forecast maps. Future studies should therefore combine blocked spatial cross-validation, rolling-origin temporal validation and uncertainty quantification to separate interpolation error from predictive error more clearly. Interpolation could also be improved through terrain-aware covariates, anisotropy analysis, regression-kriging and topography-dependent weighting strategies, particularly in areas where station density is low or where floodplain morphology controls spatial variability.

A third line of development should focus on time-lagged and multivariable modelling. The present study evaluated monthly reconstructed rainfall and water-level maps, but it did not explicitly model the lagged response between rainfall, soil moisture, storage processes and river-level dynamics. Future work should therefore test sequential and hybrid architectures, such as LSTM, GRU or attention-based models, to represent temporal dependence and rainfall-to-level delays more directly. These models should be evaluated not only by numerical error, but also by their ability to preserve spatial coherence in the reconstructed maps.

Finally, future applications should evaluate the framework in operational and transferability contexts. This includes testing the workflow in other data-scarce tropical floodplains, integrating additional hydrological and environmental covariates such as soil moisture, discharge, land cover and SAR-derived inundation indicators, and assessing how forecast maps can be translated into decision-support products. In this sense, future development should not focus only on improving predictive accuracy, but also on determining how spatially explicit outputs can support flood monitoring, risk communication and early-warning procedures in regions with fragmented observational networks.

8. REFERENCES

Abbot J, Marohasy J (2014) Input selection and optimisation for monthly rainfall forecasting in Queensland, Australia using artificial neural networks. *Atmos Res* 138:166–178. <https://doi.org/10.1016/j.atmosres.2013.11.002>

803 Abdoulhalik A, Ahmed AA (2024) A comparative analysis of advanced machine learning techniques for river streamflow time-
804 series forecasting. *Sustainability* 16(10):4005. <https://doi.org/10.3390/su16104005>

805 Adesina EA, Odumosu JO, Ajayi OG, Musa A, Onuigbo IC, Adesiji AR (2025) Evaluating the Impact of the Spatial Resolution of
806 Digital Elevation Models on Flood Modelling. *Water Resour Manag* 39(11):5359–5390. [https://doi.org/10.1007/s11269-025-](https://doi.org/10.1007/s11269-025-04206-6)
807 04206-6

808 Aguasca-Colomo R, Castellanos-Nieves D, Méndez M (2019) Comparative analysis of rainfall prediction models using machine
809 learning in islands with complex orography: Tenerife Island. *Appl Sci* 9(22):4931. <https://doi.org/10.3390/APP9224931>

810 Amatebelle CE, Owolabi ST, Ogundeji AA, Okolie CC (2025) A systematic analysis of remote sensing and geographic information
811 system applications for flood disaster risk management. *J Spat Sci* 70(3):441–467. <https://doi.org/10.1080/14498596.2025.2476973>

812 Ashok SP, Pekkat S (2024) Performance assessment of rainfall forecasting models for urban Guwahati City using machine learning
813 techniques and singular spectrum analysis. *J Water Clim Change* 15(4):1565–1587. <https://doi.org/10.2166/wcc.2024.465>

814 Azarm Z, Mehrabi H, Nadi S (2025) Enhanced land subsidence interpolation through a hybrid deep convolutional neural network
815 and InSAR time series. *Geosci Model Dev* 18(19):6903–6919. <https://doi.org/10.5194/gmd-18-6903-2025>

816 Boumpoulis V, Michalopoulou M, Depountis N (2023) Comparison between different spatial interpolation methods for the
817 development of sediment distribution maps in coastal areas. *Earth Sci Inform* 16(3):2069–2087. [https://doi.org/10.1007/s12145-](https://doi.org/10.1007/s12145-023-01017-4)
818 023-01017-4

819 Brunner MI, Slater L, Tallaksen LM, Clark M (2021) Challenges in modeling and predicting floods and droughts: A review. *WIREs*
820 *Water* 8(3):e1520. <https://doi.org/10.1002/wat2.1520>

821 Byaruhanga N, Kibirige D, Gokool S, Mkhonta G (2024) Evolution of flood prediction and forecasting models for flood early
822 warning systems: a scoping review. *Water* 16(13):1763. <https://doi.org/10.3390/w16131763>

823 Chauhan D, Mathews R (2020) Review on dimensionality reduction techniques. In: Pandian A, Palanisamy R, Ntalianis K (eds)
824 *Proceeding of the International Conference on Computer Networks, Big Data and IoT (ICCBi 2019). Lecture Notes on Data*
825 *Engineering and Communications Technologies*, vol 49. Springer, Cham, pp 356–362. [https://doi.org/10.1007/978-3-030-43192-](https://doi.org/10.1007/978-3-030-43192-1_41)
826 1_41

827 Chatrabhuj, Meshram K, Mishra U, Omar PJ (2024) Integration of remote sensing data and GIS technologies in river management
828 system. *Discov Geosci* 2:67. <https://doi.org/10.1007/s44288-024-00080-8>

829 Copăcean L, Man ET, Cojocariu LL, Popescu CA, Vilceanu CB, Beilicci R, Crețan A, Herbei MV, Cuzic O, Herban S (2025) GIS-
830 Based Flood Assessment Using Hydraulic Modeling and Open Source Data: An Example of Application. *Appl Sci* 15(5):2520.
831 <https://doi.org/10.3390/app15052520>

832 Dasgupta A, Arnal L, Emerton R, Harrigan S, Matthews G, Muhammad A, O'Regan K, Pérez-Ciria T, Valdez E, van Osnabrugge
833 B, Werner M, Buontempo C, Cloke H, Pappenberger F, Pechlivanidis IG, Prudhomme C, Ramos MH, Salamon P (2025)
834 Connecting hydrological modelling and forecasting from global to local scales: Perspectives from an international joint virtual
835 workshop. *J Flood Risk Manag* 18(1):e12880. <https://doi.org/10.1111/jfr3.12880>

836 Fattorini L, Franceschi S, Marcheselli M, Pisani C, Pratelli L (2023) Design-based spatial interpolation with data driven selection
837 of the smoothing parameter. *Environ Ecol Stat* 30(1):103–129. <https://doi.org/10.1007/s10651-023-00555-w>

838 Fu JC, Su MP, Liu WC, Huang WC, Liu HM (2024) Water Level Forecasting Combining Machine Learning and Ensemble Kalman
839 Filtering in the Danshui River System, Taiwan. *Water* 16(23):3530. <https://doi.org/10.3390/w16233530>

840 García-Lozano (2001) Región de Mompox: síntesis de estudios de evaluación ambiental regional para el sector transporte.
841 Undergraduate thesis, Universidad Distrital Francisco José de Caldas, Bogotá, Colombia

842 Goovaerts P (2000) Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *J Hydrol* 228(1–
843 2):113–129. [https://doi.org/10.1016/S0022-1694\(00\)00144-X](https://doi.org/10.1016/S0022-1694(00)00144-X)

844 Hasan BMS, Abdulazeez AM (2021) A Review of Principal Component Analysis Algorithm for Dimensionality Reduction. *J Soft*
845 *Comput Data Min* 2(1):20–30. <https://doi.org/10.30880/jscdm.2021.02.01.003>

846 International Federation of Red Cross and Red Crescent Societies (IFRC) (2025) Colombia: floods—DREF final report
847 (MDRCO024). ReliefWeb, 25 March 2025. <https://reliefweb.int/report/colombia/colombia-floods-dref-final-report-mdrc024>.
848 Accessed 11 Jun 2026.

849 Ikechukwu MN, Ebinne E, Idorenyin U, Raphael NI (2017) Accuracy assessment and comparative analysis of IDW, spline and
850 kriging in spatial interpolation of landform (topography): An experimental study. *J Geogr Inf Syst* 9(3):354–371.
851 <https://doi.org/10.4236/jgis.2017.93022>

852 Jayashree S, Sarika S, Solai AL, Prathibha S (2017) A novel approach for early flood warning using android and IoT. In: 2017 2nd
853 International Conference on Computing and Communications Technologies (ICCCCT), Chennai, India, pp 339–343.
854 <https://doi.org/10.1109/ICCCCT2.2017.7972302>

855 Jayathilake T, Sarukkalige R, Hoshino Y, Rathnayake U (2023) Wetland water level prediction using artificial neural networks—
856 A case study in the Colombo flood detention area, Sri Lanka. *Climate* 11(1):1. <https://doi.org/10.3390/cli11010001>

857 Jia W, Xiang C, Peng X, Wang L (2022) Feature dimensionality reduction: A review. *Complex Intell Syst* 8(3):2663–2693.
858 <https://doi.org/10.1007/s40747-021-00637-x>

859 Jiménez-Saenz C, Pulgarin-Morales L, Krueger T (2026) The water-energy nexus under ENSO variability in four Colombian
860 hydropower basins. *Int J River Basin Manag* 24(3):445–456. <https://doi.org/10.1080/15715124.2024.2428383>

861 Khan M, Khan AU, Ullah B, Khan S (2024) Developing a machine learning-based flood risk prediction model for the Indus Basin
862 in Pakistan. *Water Pract Technol* 19(6):2213–2225. <https://doi.org/10.2166/wpt.2024.151>

863 Kwon M, Kwon H, Han D (2020) A hybrid approach combining conceptual hydrological models, support vector machines and
864 remote sensing data for rainfall-runoff modeling. *Remote Sens* 12(11):1801. <https://doi.org/10.3390/rs12111801>

865 Li J, Heap AD (2011) A review of comparative studies of spatial interpolation methods in environmental sciences: Performance
866 and impact factors. *Ecol Inform* 6(3–4):228–241. <https://doi.org/10.1016/j.ecoinf.2010.12.003>

867 Li W, Wu J, Liu Q, Zhang Y, Cui B, Jia Y, Gui G (2023) An Effective Multimodel Fusion Method for SAR and Optical Remote
868 Sensing Images. *IEEE J Sel Top Appl Earth Obs Remote Sens* 16:5881–5892. <https://doi.org/10.1109/JSTARS.2023.3288143>

869 Liu B, Li Y, Ma M, Mao B (2025) A comprehensive review of machine learning approaches for flood depth estimation. *Int J*
870 *Disaster Risk Sci* 16(3):433–445. <https://doi.org/10.1007/s13753-025-00639-0>

871 Mai G, Xie Y, Jia X, Lao N, Rao J, Zhu Q, Liu Z, Chiang YY, Jiao J (2025) Towards the next generation of Geospatial Artificial
872 Intelligence. *Int J Appl Earth Obs Geoinf* 136:104368. <https://doi.org/10.1016/j.jag.2025.104368>

873 Maity R, Srivastava A, Sarkar S, Khan MI (2024) Revolutionizing the future of hydrological science: impact of machine learning
874 and deep learning amidst emerging explainable AI and transfer learning. *Appl Comput Geosci* 24:100206.
875 <https://doi.org/10.1016/j.acags.2024.100206>

876 Majone B, Avesani D, Zulian P, Fiori A, Bellin A (2022) Analysis of high streamflow extremes in climate change studies: how do
877 we calibrate hydrological models? *Hydrol Earth Syst Sci* 26(14):3863–3883. <https://doi.org/10.5194/hess-26-3863-2022>

878 Mereu I, Natale M, Piconi M, Troiani A, Suriani V, Bloisi DD, Burghignoli P, Costarelli D, Veneri A, Comite D (2025)
879 Interpolation Theory and Artificial Intelligence: A Roadmap for Satellite Data Augmentation. *IEEE J Sel Top Appl Earth Obs*
880 *Remote Sens* 18:17422–17448. <https://doi.org/10.1109/JSTARS.2025.3577534>

881 Parra-Plazas J, Gaona-Garcia P, Plazas-Nossa L (2023) Time series outlier removal and imputing methods based on Colombian
882 weather stations data. *Environ Sci Pollut Res* 30:72319–72335. <https://doi.org/10.1007/s11356-023-27176-x>

883 Parra-Plazas JA, Gaona-García PA, Marin Montenegro CE (2018) Proposal of a computational intelligence prediction model based
884 on Internet of Things technologies. *Proceedings - 2018 IEEE International Conference on Smart Internet of Things, SmartIoT 2018*,
885 186–191. <https://doi.org/10.1109/SmartIoT.2018.000-5>

886 Qatrinnada WFP, Hidayah E, Halik G, Wiyono RUA (2024) A literature review: rainfall thresholds as flash flood monitoring for
887 an early warning system. *Water Pract Technol* 19(11):4486–4498. <https://doi.org/10.2166/wpt.2024.271>

888 Rana VK, Pham QB, Granata F, Di Nunno F, Dang TD (2024) Fusion of diverse data sources for flood extent mapping and risk
889 assessment in Sindh: A comparative study of inundation mapping approaches. *Adv Space Res* 74(3):1140–1163.
890 <https://doi.org/10.1016/j.asr.2024.05.001>

891 Refadah SS (2025) Development in flood forecasting: A comprehensive review of complex and machine learning models. *Phys*
892 *Chem Earth* 139:103975. <https://doi.org/10.1016/j.pce.2025.103975>

893 Rodriguez-Espinosa CA, Valencia CF, Ramirez CG, Valencia CD (2025) Dynamic effect of climate change on flood damage cost
894 in the Andean region of Colombia using an ARDL-ECM model and climate change projections. *Sustain Cities Soc* 125:106309.
895 <https://doi.org/10.1016/j.scs.2025.106309>

896 Rostampour J, Mollazadeh M, Pourreza-Bilondi M, Sahranavard H (2025) Implementation of a machine learning approach to
897 predict streamflows in watersheds: A case study of the Leaf River catchment. *Water Supply* 25(1):20–33.
898 <https://doi.org/10.2166/ws.2025.001>

899 Ruiz-Lendínez JJ, Ariza-López FJ, Reinoso-Gordo JF, Ureña-Cámara MA, Quesada-Real FJ (2023) Deep learning methods applied
900 to digital elevation models: state of the art. *Geocarto Int* 38(1):2252389. <https://doi.org/10.1080/10106049.2023.2252389>

- Shao H, Chau KY, Zaman A, Moslehpour M, Pan X (2026) Machine learning frameworks to analyze climate change impact on hydropower productivity. *Appl Water Sci* 16(3):54. <https://doi.org/10.1007/s13201-025-02677-x>
- Shojaeian A, Shafizadeh-Moghadam H, Sharafati A, Shahabi H (2024) Extreme flash flood susceptibility mapping using a novel PCA-based model stacking approach. *Adv Space Res* 74(11):5371–5382. <https://doi.org/10.1016/j.asr.2024.08.004>
- Soltani SS, Belleflamme A, Goergen K, Kollet S (2026) Improving real-time flood forecasting: probabilistic validation of assimilated remotely-sensed soil moisture data. *Earth Syst Environ* 10:1961–1986. <https://doi.org/10.1007/s41748-025-00726-8>
- Surampudi S, Kumar V (2023) Flood Depth Estimation in Agricultural Lands from L and C-Band Synthetic Aperture Radar Images and Digital Elevation Model. *IEEE Access* 11:3241–3256. <https://doi.org/10.1109/ACCESS.2023.3234742>
- Tang Y, Wu Q, Soomro S, Li X, Sun Y, Hu C (2023) Comparison of different ensemble precipitation forecast system evaluation, integration and hydrological applications. *Acta Geophys* 71(1):405–421. <https://doi.org/10.1007/s11600-022-00877-6>
- Wang Y, Liu X, Liu R, Zhang Z (2024) Research Progress on Spatiotemporal Interpolation Methods for Meteorological Elements. *Water* 16(6):818. <https://doi.org/10.3390/w16060818>
- Westra S, Brown C, Lall U, Sharma A (2007) Modeling multivariable hydrological series: Principal component analysis or independent component analysis? *Water Resour Res* 43(6):W06429. <https://doi.org/10.1029/2006WR005617>
- Workneh HT, Chen X, Ma Y, Bayable E, Dash A (2024) Comparison of IDW, Kriging and orographic based linear interpolations of rainfall in six rainfall regimes of Ethiopia. *J Hydrol Reg Stud* 52:101696. <https://doi.org/10.1016/j.ejrh.2024.101696>
- Xie J, Ding L, Wang X, Qin W, Xu H, Zhang M (2025) Comparison of spatial interpolation methods for rainfall erosivity in a typical basin in the Hengduan Mountain region, Southwest China. *Ecol Indic* 174:113451. <https://doi.org/10.1016/j.ecolind.2025.113451>
- Xie S, Wu W, Mooser S, Wang QJ, Nathan R, Huang Y (2021) Artificial neural network based hybrid modeling approach for flood inundation modeling. *J Hydrol* 592:125605. <https://doi.org/10.1016/j.jhydrol.2020.125605>
- Yang W, Zhao Y, Wang D, Wu H, Lin A, He L (2020) Using principal components analysis and idw interpolation to determine spatial and temporal changes of Surfacewater quality of Xin’Anjiang river in huangshan, china. *Int J Environ Res Public Health* 17(8):2942. <https://doi.org/10.3390/ijerph17082942>
- Yeşilköy S, Baydaroglu Ö, Singh N, Sermet Y, Demir I (2024) A contemporary systematic review of cyberinfrastructure systems and applications for flood and drought data analytics and communication. *Environ Res Commun* 6(10):102003. <https://doi.org/10.1088/2515-7620/ad85c4>
- Zhang M, Chen Z, Wang J, Kar B, Pierce M, Tiampo K, Eguchi R, Glasscoe M (2025) Optical Remote Sensing for Global Flood Disaster Mapping: A Critical Review Towards Operational Readiness. *Remote Sens* 17(11):1886. <https://doi.org/10.3390/rs17111886>
- Zhou L, Kang L (2023) A Comparative Analysis of Multiple Machine Learning Methods for Flood Routing in the Yangtze River. *Water* 15(8):1556. <https://doi.org/10.3390/w15081556>
- Zhou Y, Liang X, Shan N, Yang C, Yuan H, Zheng C (2026) Temporal scaling in hydrological variables: methods, patterns, and comparative insights. *Sustain Horiz* 17:100170. <https://doi.org/10.1016/j.horiz.2025.100170>

9. ACKNOWLEDGEMENTS

The development of this research was carried out within the academic and research environment of the Centro de Investigaciones y Desarrollo Científico (CIDC) at the Universidad Distrital Francisco José de Caldas. The CIDC provided institutional and infrastructural support that facilitated the development of the study but did not provide direct financial funding for the preparation of this manuscript.

10. STATEMENTS AND DECLARATIONS

Funding

The authors did not receive funds, grants or other direct financial support from any organisation for the preparation of this manuscript. The development of the research was facilitated by the academic and infrastructural environment of the Centro de Investigaciones y Desarrollo Científico (CIDC) at the Universidad Distrital Francisco José de Caldas, but no direct financial funding was provided by the CIDC for this study.

Author Contributions

All authors contributed to the conception and design of the study. Material preparation, data collection, preprocessing, modelling and analysis were performed by Jaime Alberto Parra-Plazas, Juan Sebastián Parra-Perdomo and Leonardo Plazas-Nossa. The first draft of the manuscript was prepared by Jaime Alberto Parra-Plazas, Leonardo Plazas-Nossa and Juan Sebastián Parra-Perdomo. All authors commented on previous versions of the manuscript, revised the final version critically and approved the manuscript for submission.

Ethics Approval

This study did not involve human participants, human data, human biological material or animals. Therefore, ethics approval was not required.

Consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors have no relevant financial or non-financial interests to disclose.

Data Availability Statement

The data supporting the findings of this study are described in the manuscript. The study used hydrometeorological station records from IDEAM and spatial information derived from institutional and publicly available sources, as described in the Data and Methodology section. Because some source datasets are subject to institutional or third-party data-use conditions, the raw data are not redistributed directly by the authors. Processed datasets generated during the study may be made available by the corresponding author upon reasonable request, subject to institutional, third-party data-use and reproducibility constraints.