Hydrology Research

 Hydrology Research Vol No doi:

Spatiotemporal Prediction of Monthly Precipitation: A Systematic Review of Hybrid Models

Manuel, R, Pérez [a], Marco, J, Suárez [b], Oscar, J, García [c]

[a]Doctoral Program in Engineering, Pedagogical and Technological University of Colombia, Sogamoso, 152210, Colombia.
[b]School of Systems and Computing Engineering, Pedagogical and Technological University of Colombia, Sogamoso, 152210, Colombia.
[c]School of Geological Engineering, Pedagogical and Technological University of Colombia, Sogamoso, 152210, Colombia.
*Corresponding author. E-mail: manuelricardo.perez@uptc.edu.co

MP, 0009-0003-2963-1631; MS, 0000-0003-1656-4452; OG, 0000-0002-7396-8915

ABSTRACT

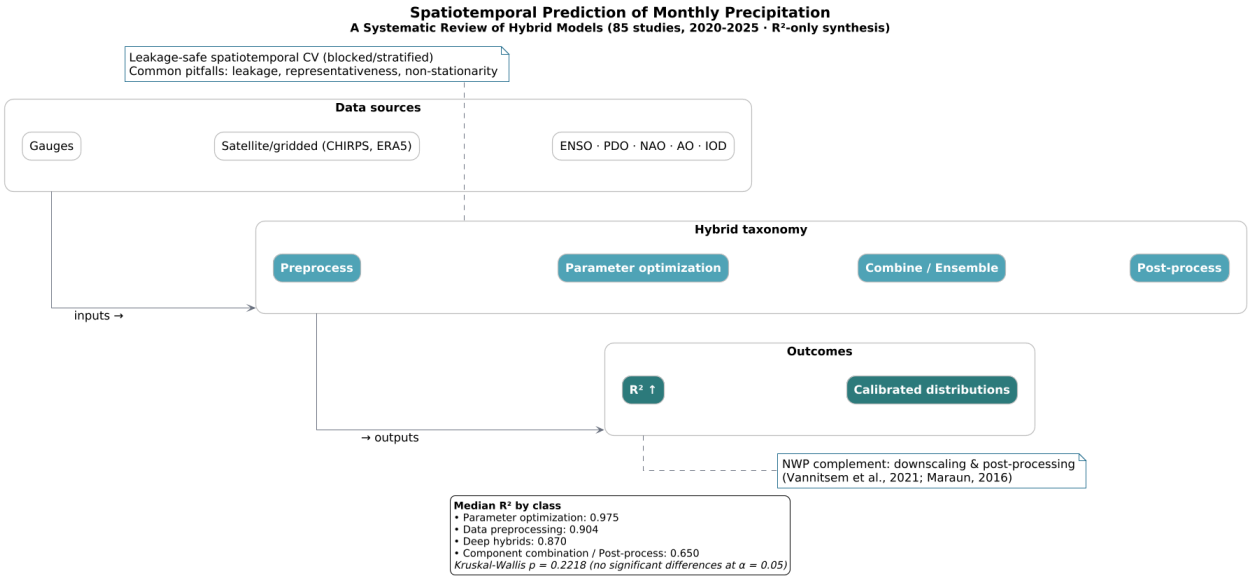
Hybrid and ensemble machine learning models are increasingly used to predict monthly precipitation. This review synthesizes 85 studies (published between 2020 and 2025) using a unified R^2-based framework. For each study, the hybrid model's R^2 is contrasted with the best non-hybrid baseline, and results are summarized by hybrid class. Median R^2 is highest for parameter optimization and data preprocessing hybrids, followed by deep hybrids; component combination and post-processing configurations show more variable performance. A Kruskal–Wallis test on R^2 across classes indicates no statistically significant differences at $\alpha = 0.05$ ($p = 0.2218$). The review clarifies complementarity with physics-based NWP (downscaling and post-processing), highlights recurrent pitfalls (leakage, representativeness, non-stationarity), and provides practical guidance for R^2 sound evaluation and reporting.

Key Words: Hybrid models, machine learning, monthly precipitation prediction, spatiotemporal prediction.

HIGHLIGHTS

- A systematic review of hybrid models for monthly precipitation prediction with a spatiotemporal focus is presented.
- Classification of hybridization approaches based on preprocessing, parameter optimization, component combination, and postprocessing stages.
- The tables included in the study offer a detailed synthesis of the reviewed hybrid models, their configurations, evaluation metrics, and geographic areas of application.
- A comprehensive performance improvement analysis was conducted for the top two models from each study, with a percentage comparison of their respective accuracy. A log-transformed RMSE and MAE graph is presented, offering a clear visualization of the models' performance in predicting monthly precipitation.

GRAPHICAL ABSTRACT



1. INTRODUCTION

Precipitation, derived from the condensation of atmospheric water vapor, is the primary source of freshwater on Earth. It includes forms like drizzle, rain, sleet, snow, and hail. With less than 1% of Earth's water being fresh and accessible, mainly replenished through precipitation, accurate prediction is essential for effective water resource management (and Space Administration, n.d.). In recent years, climate change has intensified the challenges associated with precipitation prediction by altering global patterns, making dry regions drier in subtropical zones and wet regions wetter in mid to high latitude areas (Ali et al., 2020; Foufoula-Georgiou et al., 2020; M. Wang et al., 2024; Yaseen et al., 2019). These shifts significantly impact agriculture, the economy, water supply, and energy production, highlighting the need for effective predictive modeling to mitigate potential impacts (Cronin et al., 2018; Kang et al., 2009;Liu et al., 2022).

In response to these challenges, machine learning (ML) has emerged as a powerful tool, enabling computers to detect patterns and excel in prediction and classification tasks. Among its subfields, Deep Learning (DL) an advanced form of Artificial Neural Networks (ANN) has demonstrated particular effectiveness in modeling complex, nonlinear systems. These models simulate brain-like processes to handle data across various levels of abstraction (Rahman *et al.*, 2022; Anwar *et al.*, 2018;LeCun et al., 2015;Mukherjee et al., 2019; X. Zhang et al., 2021; Lindsay et al., 2022;W. Zhang et al., 2020). DL techniques have been successfully applied to precipitation forecasting across daily, weekly, monthly, and annual timescales, supporting applications in flood and landslide prediction, as well as in water supply and quality management (Giannaros et al., 2022; X. Zhang & Wu, 2023).

Traditional statistical models, which explore the relationship between precipitation and variables such as humidity, temperature, wind speed, and atmospheric pressure, often fall short due to the inherent complexity of hydrological systems and the influence of geomorphological and climatic variability (Barrera-Animas et al., 2022; Wu and Chau, 2013; Li et al., 2023). This has led to the increased application of ML models, such as backpropagation neural networks (Singh & Borah, 2013), Long short-term memory (LSTM) networks (Endalie et al., 2021; Kim & Bae, 2017), random forests (Chao et al., 2018), and support vector machines (SVM) some enhanced with generic algorithms for parameter optimization (Chen, 2009; Tamang & Shukla, 2019; L. Zhang et al., 2008). These models are commonly evaluated using metrics like RMSE, MAE and NSE to ensure prediction reliability (Rainio et al., 2024;Steurer et al., 2021).

Among the most widely applied techniques, Artificial Neural Networks (ANNs) have become essential tools for precipitation forecasting across timescales. For instance, (Y. Tao et al., 2016) used deep neural networks to estimate

precipitation from remote sensing data, demonstrating their effectiveness in handling complex datasets. Likewise, Fredyan & Kusuma (2022) integrated convolutional neural networks, LSTM, and attention mechanisms to predict monthly precipitation using CHIRPS data at high spatial (0.05° * 0.05°) and temporal resolution, achieving superior RMSE and MAE results compared to Gated Recurrent Unit (GRU) and LSTM-AT models. Further research on ANN applications includes monsoon rainfall prediction in India using backpropagation, which significantly reduced prediction errors compared to traditional methods (Chattopadhyay, 2007). In Kerala, an Adaptive Basis Function Neural Network (AB-FNN) a backpropagation algorithm variation outperformed Fourier analysis, achieving a stabilized mean squared error of 0.085 during training (Philip et al., 2001).

In addition, soft computing techniques applied during the Indian monsoon, such as feedforward neural networks (FFNNs) and multilayer perceptrons (MLPs), reduced prediction errors from 18.3% to 10.2% relative to persistence models, indicating substantial gains in accuracy (Chattopadhyay & Chattopadhyay, 2007).

The STConvS2S model proposed by (Castro et al., 2021), which uses 3D convolutional neural networks within a sequence-to-sequence framework, has shown a 23% improvement in predicting future sequences and operates five times faster than RNN- based models. By leveraging a sequence-to-sequence network based on 3D convolutional neural networks, excels in climate prediction, including daily precipitation. It demonstrates a 23% improvement in predicting future sequences, operating five times faster than RNN-based models. Similarly, Nasseri et al., (2008) demonstrated that a hybrid model combining ANN with backpropagation and generic algorithms significantly reduced mean squared error and improved the coefficient of determination compared to standard ANN models.

Other innovative approaches include focused time-delay neural network models (Htike & Khalifa, 2010), which dynamically adapt to temporal data by optimizing delay times through trial and error. These models demonstrate high accuracy for annual precipitation, although performance declines at shorter temporal resolutions. Modular ANNs incorporating preprocessing techniques such as Moving Averages (MA), Principal Component Analysis (PCA), and Singular Spectrum Analysis (SSA) have outperformed conventional models, yielding superior performance in terms of Coefficient of Efficiency (CE), RMSE, and Prediction Intervals (PI) (C. L. Wu et al., 2010; C. L. Wu & Chau, 2013) .

Modular systems that combine support vector regression (SVR) with ANN have also improved accuracy for stations like Zhenwan and Wuxi (Chao et al., 2018). Meanwhile Salimi et al., (2019) applied ANN for downscaling and evaluating satellite-based precipitation estimates, leveraging Particle Swarm Optimization (PSO), imperialist competitive algorithm (ICA), and Genetic Algorithm (GA) to enhance performance, demonstrating improved prediction quality.

Recent advancements have focused on hybrid models that integrate time series decomposition techniques with machine learning algorithms. For instance, Ardabili et al.,)2020) showed that combining Complementary Ensemble Empirical Mode Decomposition (CEEMD) with wavelets significantly improves forecasting at annual, monthly, and daily levels. Ensemble methods such as bagging, boosting, and stacking have further enhanced prediction accuracy by aggregating model outputs through a meta-learner (Niazkar et al., 2024). The CEEMD-FCMSE (Fuzzy Comprehensive Multi-Criteria Evaluation)-Stacking model, for instance, significantly reduces error metrics, demonstrating the superior capabilities of such advanced models (X. Zhang et al., 2022).

A detailed assessment of hybrid and ensemble models is therefore essential to better understand their advantages and limitations. This systematic review consolidates recent developments in hybrid precipitation forecasting by analyzing the integration of ANNs, random forests, and SVMs with statistical methods. It highlights the substantial error reductions achieved by incorporating CEEMD, SSA, wavelet transforms, and various optimization algorithms compared to standalone models (Dotse et al., 2024). Moreover, ensemble strategies such as model averaging, stacking, bagging, boosting, and dagging have further improved model robustness and reliability (Zounemat-Kermani et al., 2021). Unlike prior reviews, this study offers a more comprehensive systematic analysis by explicitly comparing specific hybrid approaches based on key performance metrics and identifying critical knowledge gaps. These contributions lay the groundwork for future research aimed at addressing the growing challenges posed by climate change in hydrological prediction.

1.1 Hydrologic data & modeling context

Monthly precipitation prediction operates at the interface of hydrologic data limitations and climate dynamics. Station networks are sparse and biased toward populated valleys; gauges in complex terrain undersample windward–leeward contrasts. Satellite/gridded products (e.g., CHIRPS, ERA5) alleviate gaps but introduce retrieval and orographic biases, scale mismatches, and temporal discontinuities (Funk et al., 2015; Hersbach et al., 2020). These factors yield heteroskedastic, spatially autocorrelated targets. Hybrid designs that address spatial structure and multi-scale variability are especially relevant at monthly horizons; reported R^2 should reflect leakage safe validation and cross-climate robustness (Pagliero et al., 2019).

1.2 Objectives and contributions

- Taxonomy: provide a clear functional taxonomy of hybrid monthly precipitation predictors (preprocessing, optimization, component combination/ensemble, post-processing).
- Quantitative synthesis: derive $\nabla R^2 = R^2_{hyb} - R^2$ per study and compare hybrid classes using non-parametric tests.
- Practice-oriented guidance: identify validation pitfalls and provide leakage-safe, spatiotemporal CV recommendations for monthly horizons.
- Positioning with physics: clarify how ML hybrids complement physical NWP (downscaling, bias correction/post-processing, surrogates) and how R^2-based verification should be embedded in monthly workflows.

2. MATERIALS AND METHODS

This section describes the stages of the systematic review using the PRISMA methodology (Page et al., 2021). A literature search on hybrid and ensemble ML (machine learning) for precipitation prediction was conducted in major databases, including Scopus and Science Direct. The systematic search strategy was developed using the ScienceDirect, Scopus, and IEEE Xplore databases. The search was conducted between January 1, 2020, and March 31, 2025. Only articles published in English or Spanish in peer-reviewed journals were included. Both original research articles and review papers were considered. The Boolean search combinations used were:

- (“machine learning” AND precipitation AND monthly AND hybrid)
- (“machine learning” AND precipitation AND monthly AND hybrid AND spatiotemporal)
- (“hybrid model” AND prediction AND precipitation AND wavelet)
- (“hybrid model” AND prediction AND precipitation AND LSTM)

The results were filtered to include only articles with full-text availability, and open access articles were prioritized whenever possible. Technical reports, book chapters, preprints, and grey literature were excluded. The selection process and eligibility criteria are presented in Table 1 and Figure 1. At the outset, 452 articles were identified, of which 130 were duplicates and subsequently. After reviewing titles and abstracts, 160 articles were excluded for not meeting the criteria. Finally, 162 articles were assessed in full text, and 85 were included in the final analysis, as detailed in Figure 1. ScienceDirect offers critical environmental and hydrological journals such as Journal of Hydrology, Environmental Modelling & Software, and Expert Systems with Applications, which publish foundational studies on hybrid modeling approaches (Ardabili et al., 2020; M. Wang et al., 2024). Scopus was chosen for its comprehensive interdisciplinary indexing, including prominent journals like Water Resources Management, Hydrological Sciences Journal, and Remote Sensing, renowned for publishing innovative precipitation modeling research (Ali et al., 2020; D. Kumar et al., 2021). IEEE Xplore offers specialized coverage in computational intelligence and machine learning, indexing influential IEEE journals such as IEEE Transactions on Neural Networks and Learning Systems and IEEE Transactions on Geoscience and Remote Sensing, which are essential for advanced studies on hybrid and ensemble precipitation prediction models (Tang et al., 2022).

Together, these databases ensured a rigorous and comprehensive capture of relevant, high-impact literature, significantly enhancing the depth and credibility of this systematic review. Figure 1 shows the identification stage. In this stage, a preliminary analysis was conducted to establish the information sources and search strategies, and the eligibility and inclusion criteria were defined. The first stage included searching for information sources. Subsequently, duplicate articles were removed, and the snowball technique was applied to identify additional studies, and the duplicate articles were filtered out completely. The screening stage was divided into two reviews. The first review examined the title, abstract, and keywords to determine whether the article met the inclusion and exclusion criteria. The selected articles were then reviewed again, considering the introduction and conclusions. Subsequently, in the eligibility stage Figure 1, a full review of the articles selected during the screening stage was conducted. In the eligibility stage, trends and opportunities were carefully observed. Finally, the set of articles and research studies to be included in the review was obtained.

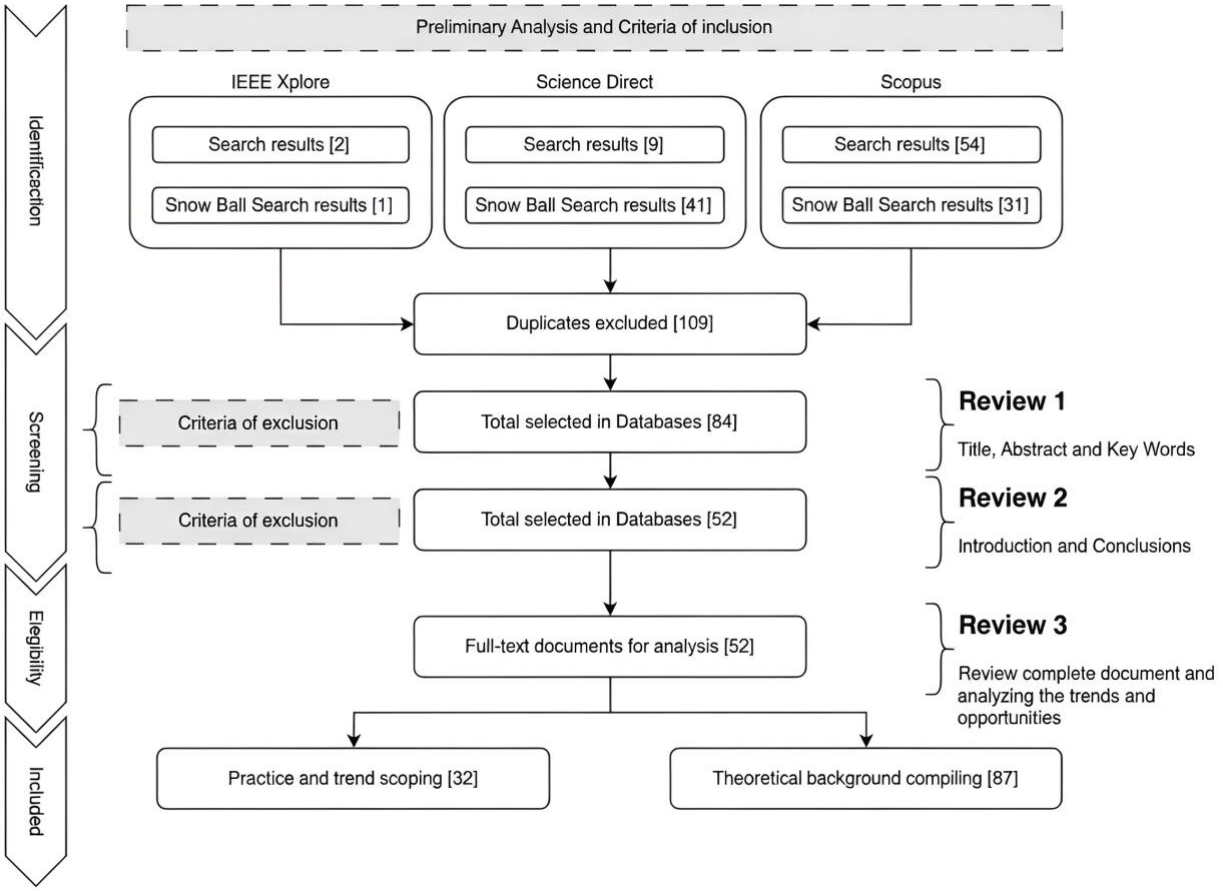


Figure 1 | PRISMA flow diagram for study selection. The diagram shows the number of records identified, excluded, and ultimately included in the review.

2.1 Literature Analysis before Evaluation

This section defines the scope of the systematic review and establishes the inclusion and exclusion criteria. It also develops a detailed search strategy, including keywords and time ranges. This preliminary analysis is essential to ensure a solid, well-defined foundation that guarantees the process is systematic, comprehensive, and replicable. A series of central themes is proposed to address the research limitations, informing the search for the most closely related research studies for doctoral work.

The study identifies five key keywords that represent central themes of the research: Machine Learning, Precipitation, Monthly, Hybrid, and Spatiotemporal. Each of these terms encapsulates essential aspects of the study's focus. Machine Learning is highlighted due to its foundational role in the analytical approach, while Precipitation represents a core meteorological variable under investigation. The keyword Monthly reflects the temporal resolution used in the data analysis. The term Hybrid denotes the integration of multiple methods or models, and Spatiotemporal emphasizes the study's focus on variations across both space and time. All of these keywords are not only frequently used but also serve as the conceptual pillars of the research.

The limitations of the study were organized into three exclusion criteria that help define the scope and focus of the analysis. The first criterion (CE1) excludes extreme events, as these are beyond the intended scope of the research. The second criterion (CE2) refers to the exclusion of other geographical regions and phenomena not considered within the proposal. Finally, the third criterion (CE3) excludes topics that are explicitly outside the scope of the study's objectives. These classification criteria facilitate the filtering of information and the delineation of relevant content.

A wide range of topics was excluded from the study's scope as part of the established exclusion criteria. Under CE2, excluded themes include typhoons, cyclones, monsoons, drought, heavy rainfall, extreme rainfall, slope failures, standardized precipitation indices, data intelligence models, and other geographical phenomena such as debris, air pollution, and aquifer levels. CE3 groups exclusions that are beyond the study's objective, such as landscape, runoff, streamflow, groundwater, flood, wildfire, water footprint, and rainfall. This structured classification ensured a focused and coherent research framework by eliminating topics not directly aligned with the study's purpose.

2.2 Literature Analysis after Evaluation

The systematic review of bibliographic sources followed the guidelines proposed by Higuera Martínez et al., (2021), selecting two bibliographic databases and a subsequent document filtering process through identification, screening, eligibility, and inclusion. The snowball technique was applied using the software VOS Viewer, which adds search results from databases to obtain data views that improved article searches with terms not initially considered but highly valuable for the ongoing research. Then, duplicates were excluded in both databases for the original searches and the snowball method. The next step involved screening, where articles were filtered by reviewing titles, abstracts, introductions, and conclusions and where the inclusion criteria.

$$CI1\ machine\ learning\ AND\ precipitation\ AND\ monthly\ AND\ Hybrid \quad (1)$$

$$CI2\ machine\ learning\ AND\ precipitation\ AND\ monthly\ AND\ Hybrid\ AND\ spatiotemporal \quad (2)$$

The VOSviewer analysis included a variety of terms categorized according to their conceptual and computational approaches, along with the time of their appearance. Terms such as data, model, and precipitation were associated with general, machine learning (ML), and climatic approaches, all appearing early in the timeline (mostly in 2022). Metric-based terms like MAE and RMSE were classified under the metric approach and emerged in mid to late 2022 and early 2023, respectively. More advanced machine learning terms, such as LSTM, prediction, and wavelet, corresponded to higher computational levels, particularly levels 2 and 3, reflecting increasing complexity in the methodology. The temporal distribution of these terms, from early 2022 to early 2023, demonstrates a progressive refinement in the use of analytical tools within the study.

By filtering only by the study focus on ML techniques and precipitation, a subset of key terms was identified. Terms such as hybrid model and model were classified under ML approaches and grouped into category 1, both appearing in early 2022. More advanced ML techniques, including LSTM, prediction, and *wavelet,* were placed in groups 2 and 3, also within the same year, indicating a progression in methodological complexity. The term precipitation was categorized under the climatic approach and assigned to group 2, reinforcing its relevance as a climatic variable within the scope of this filtered analysis. This focused selection highlights the thematic convergence of machine learning methodologies with precipitation-centered studies in the given timeframe. New search conditions could be added for the databases based on the terms found at the computational level and machine learning algorithms, and this study's central climatic variable, precipitation. To improve searches, conditions focused specifically on the words extracted in Table 1, focusing the searches on the hybrid model by omitting the word model as it is compound, and performing

separate searches for the words Wavelet and LSTM as they are not complementary, with data window of 2020 to 2025 and research and review type articles that are open access.

Methodological quality criteria were applied to assess the risk of bias in the included studies, based on the robustness of the reported validations and the transparency in model descriptions. Studies with insufficient information were marked for qualitative rather than quantitative analysis.

Table 1 | Search Query Results for Hybrid Precipitation Prediction Models Using Wavelet and LSTM, Input for VOSviewer Analysis.

Title, abstract, keywords	Source	Result
hybrid model AND prediction AND precipitation AND wavelet	Science Direct	6
hybrid model AND prediction AND precipitation AND lstm	Science Direct	35
hybrid model AND prediction AND precipitation AND wavelet	Scopus	18
hybrid model AND prediction AND precipitation AND lstm	Scopus	17
("All Metadata":hybrid mode) AND ("All Metadata":prediction) AND ("All Metadata":precipitation) AND ("All Metadata":wavelet)	IEEE	0
("All Metadata":hybrid mode) AND ("All Metadata":prediction) AND ("All Metadata":precipitation) AND ("All Metadata":lstm)	IEEE	1

2.3 Quantitative synthesis protocol

This synthesis uses R^2 as the sole comparison metric. For each study at a monthly resolution, the hybrid R^2 (R^2_hyb) and the best non-hybrid baseline R^2 (R^2_{base}) were extracted from leakage safe validation or independent test splits when available. The study level contrast is $\nabla R^2 = R^2_{hyb} - R^2_{base}$ interpreted as an absolute gain in explained variance (percentage points). R^2 values are summarized by hybrid class (median, IQR). Kruskal–Wallis tests compare classes ($\alpha = 0.05$), with planned Dunn post hoc tests performed if the result is significant. Given heterogeneous designs, no inverse variance pooling is attempted; robust non-parametric summaries are emphasized. All figures use R^2 in [0, 1] with the axis label "R^2 (coefficient of determination)."

3. HYBRID PREDICTIVE MODELS

Hybrid predictive models combine two or more machine learning or soft computing methods to achieve better performance by leveraging the advantages and purposes of each method (Hajirahimi & Khashei, 2019). Often, one method is used for prediction and another to optimize that prediction, functioning like a company with employees from diverse specialties working towards a common goal. This approach produces more robust models, improving accuracy and surpassing the performance of individual models by capitalizing on the strengths of each component. It encompasses all computational phases, from data normalization to decision-making (Ardabili et al., 2020; Woźniak et al., 2014). A categorization proposed by for hybrid models applied to time series, potentially generalizable to other contexts, identifies four main components:

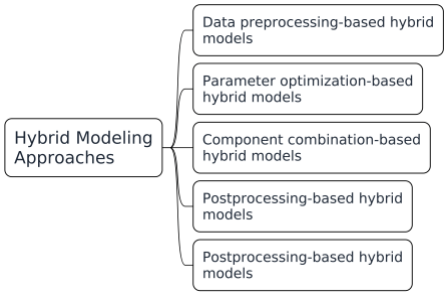


Figure 2 | Classification of hybrid modeling approaches based on functional integration.

Figure 2 presents a classification of hybrid modeling approaches into four main categories, each corresponding to a different stage of the modeling process. First, data preprocessing-based hybrid models aim to improve input quality through techniques such as normalization, decomposition, or noise reduction. Second, parameter optimization-based hybrid models employ metaheuristic algorithms (such as PSO, GA, or HBA, the Honey Badger Algorithm) to fine-tune the hyperparameters of base models, enhancing their performance. The third category, component combination-based hybrid models, focuses on integrating different parts of multiple models—for example, combining the predictive structure of a statistical model with the learning capabilities of a neural network. Finally, postprocessing-based hybrid models operate on the model outputs to correct errors or refine predictions, using techniques like calibration or ensemble post-adjustments.

3.1 Data Preprocessing-based Hybrid Models

Data preprocessing refers to a set of techniques applied to raw data prior to modeling, with the goal of improving data quality, consistency, and analytical value. These techniques may involve normalization, outlier removal, missing value imputation, feature scaling, or more complex transformations such as Principal Component Analysis (PCA) or clustering. By aligning data distributions, reducing dimensionality, and eliminating noise, preprocessing lays the foundation for more robust and accurate predictive modeling. In contrast, data preprocessing-based hybrid models represent a specific modeling architecture in which the preprocessing stage is explicitly integrated as a core component of a hybrid framework. Rather than treating preprocessing as an independent or auxiliary step, these hybrid models are designed to combine preprocessing techniques with one or more learning algorithms in a structured and interdependent manner. The goal is not only to clean or transform data, but also to reshape the input space in ways that enhance the predictive engine's learning capacity.

For example, clustering meteorological stations using K-means into climatically homogeneous regions does not modify raw precipitation values but restructures the input feature space. When such preprocessing is tightly coupled with machine learning models, such as Support Vector Regression (SVR) or Group Method of Data Handling (GMDH), the resulting hybrid model performs better in handling spatial heterogeneity and temporal variability (Parviz et al., 2023). Empirical evidence indicates that hybrid model configurations incorporating preprocessing, such as residual decomposition coupled with machine learning algorithms like Support Vector Regression (SVR), Gene Expression Programming (GEP), and the Group Method of Data Handling (GMDH), can substantially enhance forecasting accuracy. Specifically, studies report up to a 67% reduction in Mean Square Error (MSE) and a 5% improvement in Nash-Sutcliffe Efficiency when such hybrid models are further optimized using genetic algorithms. Moreover, the application of Savitzky-Golay filtering as a noise-reduction preprocessing step has been shown to increase prediction accuracy in daily precipitation estimation via Adaptative Neuro Fuzzy Interference System – Artificial Bee Colony (ANFIS-ABC) models, with the correlation coefficient (R) rising from 0.78 to 0.85 (Pham et al., 2024).

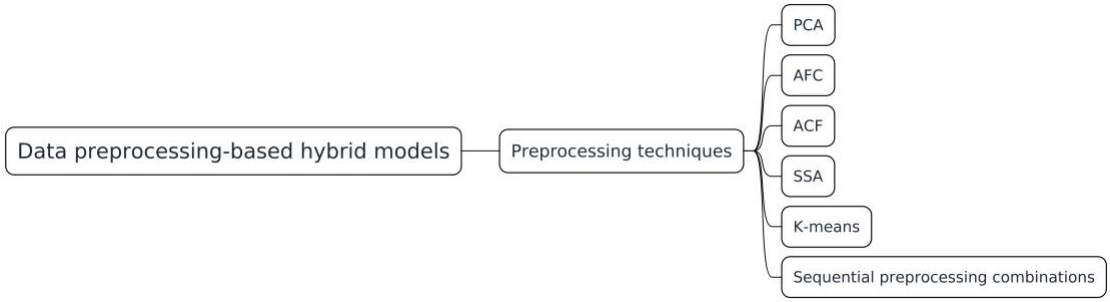


Figure 3 | Classification of hybrid models based on data preprocessing. The main methods used in the preprocessing stage are illustrated, along with their impact on the predictive performance of hybrid models.

The following figure is a classification of hybrid models based on data preprocessing techniques, emphasizing methods designed to enhance data quality before predictive modeling. This classification includes Principal Component (PCA), which effectively reduces data dimensionality; Canonical Factor Analysis (CFA) and the Autocorrelation Function

(ACF), both of which help identify statistical relationships between temporal variables; and Singular Spectrum Analysis (SSA), a powerful tool for decomposing complex time series. The K-means algorithm is also widely used as a method for unsupervised data clustering. Moreover, Figure 3 highlights the use of sequential preprocessing strategies, where multiple techniques are applied in series to extract different latent structures or patterns from the data.

Furthermore, the approach of sequential preprocessing combinations is highlighted, allowing multiple techniques to be applied in series to extract various hidden patterns or structures within the data. For instance, (Tang et al., 2022) employed the K-means algorithm to enhance data structure through clustering, while (Ridwan et al., 2021) utilized the ACF to analyze temporal dependencies within the precipitation time series.

3.2 Parameter Optimization-based Hybrid Models.

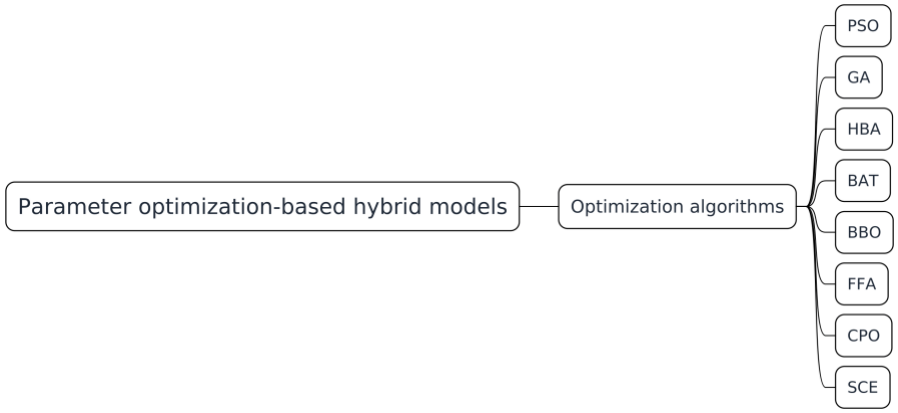


Figure 4 | Optimization algorithms commonly used in parameter optimization-based hybrid models.

These hybrid models incorporate a systematic optimization layer that autonomously tunes hyperparameters and internal configurations of forecasting algorithms. Unlike traditional grid or manual search, modern hybrid systems utilize metaheuristic algorithms, such as (GA), (PSO), and Ant Colony Optimization (ACO), to effectively explore high-dimensional parameter spaces (Pham et al., 2024).

In recent applications, GA has been employed to optimize the weight combination of (Gene Expression Programming) GEP, SVR, and GMDH outputs, leading to drastic improvements in model accuracy. At the Rasht station, this strategy reduced MSE by 67% and increased forecast skill significantly compared to SVR-based hybrids (Parviz et al., 2023). Furthermore, novel techniques such as the Brownian Motion-based Pelican Optimization Algorithm (BMPOA) have demonstrated superior performance in fine-tuning hybrid Deep Belief Networks (DBNs) and (LSTM) models (Anuradha et al., 2024).

The diagram (Figure 4) illustrates the optimization algorithms employed to enhance the performance of predictive models. These algorithms, predominantly metaheuristic and bio-inspired in nature, are designed to fine-tune the hyperparameters of base models, enabling the identification of optimal configurations that maximize system accuracy. Bio-inspired algorithms draw inspiration from natural processes and behaviors, allowing them to efficiently explore complex solution spaces. Among the widely used methods are Particle Swarm Optimization (PSO), which simulates the social behavior of bird flocking to guide search agents; Genetic Algorithm (GA), based on the principles of natural selection and genetics; Bat Algorithm (BAT), inspired by the echolocation behavior of bats; and Honey Badger Algorithm (HBA), which mimics the foraging and digging behavior of honey badgers. Additionally, the diagram includes less conventional techniques such as Biogeography-Based Optimization (BBO), which models the migration patterns of species across habitats; Fruit Fly Optimization Algorithm (FFOA), based on the food-searching behavior of fruit flies; Firefly Algorithm (FFA), inspired by the flashing patterns and attraction mechanisms of fireflies; Chimp Optimization Algorithm (CPO), which emulates the intelligent hunting strategies of chimpanzees; and Shuffled Complex Evolution (SCE), a population-based algorithm that combines complex shuffling and evolutionary principles to enhance convergence.

These algorithms have demonstrated their effectiveness in recent studies. For instance, in the article Zerouali et al., (2023)", techniques such as PSO and GA to optimize artificial intelligence models for precipitation prediction, resulting in notable improvements in accuracy. Similarly, the study "(R. Kumar et al., 2021)" implemented several metaheuristic algorithms, including HBA, to optimize both deep neural networks and Extreme Learning Machines (ELM), achieving superior performance in long-term precipitation forecasting.

3.3 Component Combination-based Hybrid Models

Component combination-based hybrid models integrate multiple forecasting algorithms into a unified framework to exploit their complementary strengths. Each component is designed to capture distinct statistical or dynamic patterns within the data, thereby improving the model's overall robustness and accuracy (Figure 5).

These models frequently employ ensemble learning techniques such as Bagging, Boosting, and Stacking—to combine the outputs of diverse base learners. In Bagging (e.g., Random Forests), models are trained in parallel on bootstrapped subsets, and predictions are aggregated (usually via voting or averaging), which helps reduce variance. Boosting, by contrast, trains models sequentially, where each new learner corrects the residuals of its predecessors, thereby reducing bias (Zhou, 2012). Stacking goes a step further by combining the predictions of multiple base models using a meta-learner, which is trained to optimize final predictions based on the outputs of the base models, often improving performance on complex, nonlinear tasks (Sagi & Rokach, 2018).

This hybrid architecture has shown significant improvements in precipitation forecasting, especially under noisy, multiscale, or nonstationary conditions. For instance, integrating decision trees, SVR, and ANN into a single ensemble has yielded lower RMSE values and better bias calibration compared to individual models (Parviz et al., 2023). Furthermore, hybrid frameworks integrating Ant Colony Optimization (ACO) with Gradient Boosted Decision Trees (GBDT) have achieved prediction accuracies exceeding 93% for monthly precipitation forecasts, surpassing the performance of either component alone (Sharma et al., 2024).

Further expanding on ensemble predictive models, these combine multiple training algorithms to improve predictive accuracy, robustness, and overall performance. By aggregating the strengths of diverse models, these methods reduce errors and enhance generalization. Ensemble techniques are widely regarded as a key approach within supervised learning frameworks, offering significant advantages in handling complex datasets and achieving superior results (Ardabili et al., 2020). There are five main techniques for developing ensemble models: model averaging, stacking, bagging, boosting, and dagging (Zounemat-Kermani et al., 2021).

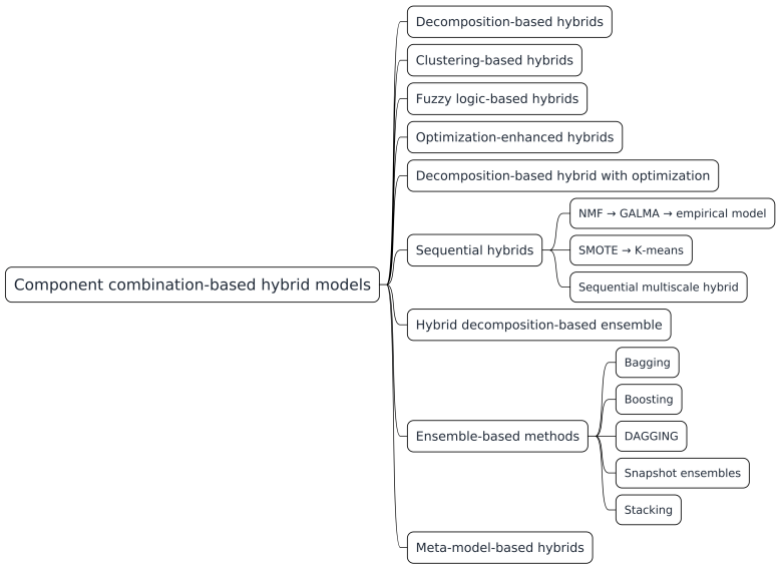


Figure 5 | Taxonomy of component combination-based hybrid modeling structure.

3.4 Model Averaging

This method involves training multiple models independently and averaging their predictions to reduce variance and improve generalization (Ganaie et al., 2022).

For models $f_1, f_2, \cdots, f_n$, the averaged prediction $\hat{y}$ is:

$$\hat{y} = \frac{1}{n}\sum_{i=1}^{n} f_i(x) \tag{3}$$

The Model Averaging (MA) approach has become increasingly popular as a straightforward and effective Ensemble Machine Learning (EML) method for hydrological process modeling. This approach encompasses a variety of techniques, including Simple Arithmetic Mean (SAM), Bayesian Model Averaging (BMA), Equal Weights Averaging (EWA), Bates-Granger Model Averaging (BGA), Averaging based on Akaike's Information Criterion (AICA), Bayes' Information Criterion (BICA), Mallows Model Averaging (MMA), and Granger-Ramanathan Averaging (GRA) (Arsenault et al., 2015). BMA is an exceptionally efficient and widely recognized algorithm because it addresses model uncertainty. In BMA, predictions are generated through a weighted averaging process, where the weights are derived from the posterior probabilities of the competing models. Below is a concise explanation of the BMA algorithm (Duan et al., 2007), $M = \{M_1, \cdots, M_k\}$ which is a set k of models; the posterior probability (ρ) for the model M_k is defined as the dataset D:

$$\rho(M_k|D) = \frac{\rho(D|M_k)\rho(M_k)}{\sum_{l=1}^{K}\rho(D|M_l)\rho(M_l)} \tag{4}$$

Where:

$$\rho(D|M_k) = \int \rho(D|\theta_k, M_k)\rho(\theta_k|M_k)d\theta_k \tag{5}$$

Where D denotes the observed dataset $\rho(D|\mathrm{M_k})$ likelihood of model prediction, $\mathrm{M_k}$, $\rho(\theta_\mathrm{k}|\mathrm{M_k})$ the prior density of the vector parameter, and $\rho(D|\theta_\mathrm{k}, \mathrm{M_k})$ the likelihood.

3.5 Bagging (Bootstrap Aggregating)

Bagging involves creating multiple subsets of the training data through bootstrapping (random sampling with replacement), training a model on each subgroup, and aggregating their predictions. This technique reduces variance and helps prevent overfitting (Ganaie et al., 2022); similar to averaging, the aggregated prediction is:

$$\hat{y} = \frac{1}{n}\sum_{i=1}^{n} f_i(x) \tag{6}$$

3.6 Stacking

Stacking combines multiple models by training a meta-model (or second-level model) to make final predictions based on the outputs of base models. This approach leverages the strengths of diverse models to improve overall performance [15]. If $f_1(x), f_2(x), \cdots, f_n(x)$ our base model predictions, the meta-model g predicts is:

$$\hat{y} = g(f_1(x), f_2(x), \cdots, f_n(x)) \tag{7}$$

3.7 Boosting

Boosting trains models sequentially, with each new model focusing on correcting the errors of its predecessors. By adjusting the weights of misclassified instances, boosting aims to reduce bias and build a strong predictive model from weak learners (Ghojogh & Crowley, 2019). The combined model is:

$$\hat{y} = \sum_{i=1}^{n} \alpha_i f_i(x) \tag{8}$$

Where $\alpha_i$represents the weight assigned to the model f_i based on its accuracy.

3.8 Dagging (Disjoint Aggregating)

Dagging involves partitioning the training data into disjoint subsets, training a model on each subgroup, and aggregating the predictions from these models. This method aims to reduce variance and improve model robustness (Arsov et al., 2019):

$$\hat{y} = \frac{1}{n}\sum_{i=1}^{n} f_i(x) \quad (9)$$

The applications of ensemble methods are diverse; for example, (Y. Wang et al., 2022) present a probabilistic precipitation forecasting scheme using an ensemble model (EM) based on CEEMDAN and AM-MCMC. First, precipitation series are decomposed using signal decomposition techniques (CEEMDAN). Then, empirical models (TSAM, GSM, and LSTM) produce quantitative forecasts. Finally, an ensemble model combines these forecasts with weights determined by AM-MCMC. The CEEMDAN-EM-AM scheme provides more accurate predictions and reduces uncertainty through 90% confidence intervals. Compared to individual models, the ensemble model shows greater accuracy.

In their study, (Chhetri et al., 2020) evaluated six individual and ensemble models using the stacking technique. The ensemble model outperformed the individual models, achieving an RMSE of 17.5 mm. The model integrated a decision tree, KNN, and LSTM as the base learners, with XGBoost as the second-level learner.

Detailed taxonomy of component combination-based hybrid models, which integrate multiple techniques or sub-models to enhance predictive performance. This category is subdivided according to the nature of the components being combined. For instance, decomposition-based hybrids apply transformations such as Empirical Mode Decomposition (EMD) or Discrete Wavelet Transform (DWT) to break down time series into simpler components, which are then modeled using machine learning algorithms. Clustering-based and fuzzy logic-based models aim to uncover hidden structures or manage uncertainty in data. Moreover, optimization-enhanced hybrids employ algorithms like Particle Swarm Optimization (PSO) or Genetic Algorithms (GA) to fine-tune parameters across various modeling stages.

A particularly notable subcategory is that of sequential hybrids, where techniques are applied in consecutive stages. (Song et al., 2021) exemplify this approach, using EMD followed by machine learning to handle the nonlinear nature of precipitation data. Meanwhile, ensemble-based hybrids—such as bagging, boosting, or snapshot ensembles—combine multiple predictors to reduce variance and improve model stability. Finally, meta-model-based hybrids, like stacking, integrate the outputs of several base models through a higher-level meta-model. This is demonstrated by (Gu et al., 2022), where stacking significantly improved monthly precipitation prediction accuracy.

3.9 Postprocessing-based Hybrid Models

Researchers typically apply postprocessing as a set of statistical or probabilistic techniques after the primary model generates its predictions. These methods aim to correct systematic biases, calibrate output distributions, or enhance the interpretability of forecasts. Common independent techniques such as bias correction, Quantile Mapping (QM), and the Probability Integral Transformation (PIT) are widely used in climatology and hydrology to refine model outputs, especially in ensemble systems or physical models that often produce distributional mismatches.

In contrast, hybrid models based on postprocessing integrate these techniques directly into a two-stage predictive architecture. Rather than applying corrections after model execution, these hybrids embed postprocessing modules such as quantile regression forests, copula-based recalibrators, or neural calibration layers biases during training rather than as an isolated step.

For example, when researchers integrate Quantile Regression Forests (QRF) into a hybrid modeling pipeline, they achieve substantial improvements in daily precipitation forecasts by addressing underdispersion and distributional skewness issues that standard machine learning models often neglect (Taillardat et al., 2016). Likewise, Ensemble Copula Coupling (ECC), which reshapes forecast ensembles while maintaining dependencies between variables, significantly improves uncertainty quantification in high-impact weather events (Schefzik et al., 2013).

Another hybrid approach combines neural networks with embedded probabilistic calibration layers, allowing postprocessing to be learned jointly with the main forecasting model. These neural hybrid systems outperform traditional statistical postprocessing methods in terms of both sharpness and reliability of probabilistic forecasts (Rasp & Lerch, 2018). Furthermore, comprehensive reviews have emphasized that integrating postprocessing directly within ensemble or ML architectures yields forecasts that are more adaptive to regime shifts, temporal non-stationarity, and nonlinear dependencies (Van Schaeybroeck & Vannitsem, 2018).

Classification of postprocessing-based hybrid models, which incorporate additional techniques following the initial prediction stage to enhance accuracy, correct systematic errors, or integrate multiple outputs, as indicated in Figure 6. This category is divided into three primary approaches:

3.9.1 **Statistical correction**, which includes methods such as bias correction and Bayesian Joint Probability (BJP). These techniques systematically adjust prediction errors to improve the reliability and accuracy of model outputs.

3.9.2 **Probabilistic model fusion**, represented by techniques like Bayesian Model Averaging (BMA) and Probabilistic Model Averaging, seeks to combine outputs from multiple models based on their probability distributions, producing more robust and uncertainty-aware predictions. For instance, the study *"*(Chhetri et al., 2020)*"* implemented probabilistic model fusion using BMA to enhance the accuracy and stability of precipitation predictions.

3.9.3 **Averaging-based postprocessing**, which encompasses methods such as prediction averaging, feature averaging, model averaging, and prediction selection. These approaches aggregate results from multiple models or configurations to improve generalization, reduce overfitting, and enhance model robustness. A clear example is presented in (Ehteram et al., 2024) who employed prediction averaging as a postprocessing strategy to refine model outputs.

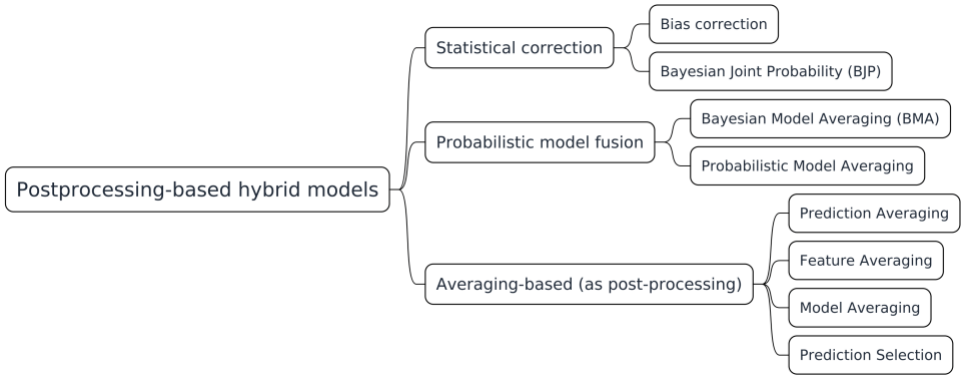


Figure 6 | Structure and categorization of postprocessing-based hybrid models.

4. METRICS FOR PERFORMANCE EVALUATION MODELS

Performance evaluation is based on comparing quality metrics among the models used in the analyzed studies Figure 7. These metrics assess the accuracy and reliability of model predictions by comparing the observed or actual values with the predicted results. This evaluation is conducted at various stages of predictive model development, including preparation, validation, and testing processes. Overall, these quality metrics provide essential insights into the model's effectiveness.

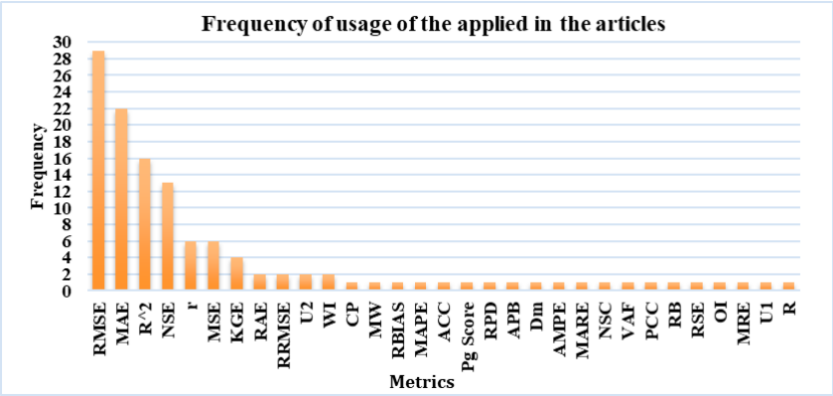


Figure 7 | Frequency of use distribution of performance metrics used in hybrid model evaluation.

Figure 7 shows the analysis of performance metrics, which reveals a clear preference for traditional error-based indicators in the assessment of monthly precipitation forecasting models. Root Mean Square Error (RMSE) is the most frequently utilized metric, appearing in 17 studies, followed by Mean Absolute Error (MAE) with 13 occurrences and Nash–Sutcliffe Efficiency (NSE) in 10 studies. These metrics dominate due to their intuitive interpretation and their sensitivity to deviations between observed and predicted values.

Deterministic correlation-based measures such as the coefficient of determination (R^2) and Pearson's correlation coefficient (R) are used less frequently, with 8 and 5 appearances, respectively. In contrast, advanced or normalized metrics, such as Kling-Gupta Efficiency (KGE), Nash–Sutcliffe Coefficient (NSC), and relative indicators like MAPE, RAE, or MARE, are used sporadically, suggesting a lower level of adoption in the literature. Metrics such as Bias Ratio (RBIAS), Volume Agreement Fraction (VAF), and Overall Index (OI) are rarely reported, each appearing in only one study. This pattern highlights the community's reliance on error magnitude metrics over normalized or distribution-sensitive measures, potentially limiting comparative analyses across heterogeneous climatic regions. Future studies would benefit from incorporating a broader set of performance indicators to enhance the robustness and comparability of hybrid model evaluations across diverse geographical and climatic contexts.

4.1 Root Mean Square Error (RMSE)

The metric measures the average magnitude of the squared errors between predicted and observed values. It heavily penalizes large errors.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2} \tag{10}$$

4.2 Mean Absolute Error (MAE)

MAE calculates the average of the absolute differences between predicted and observed values. It provides a more balanced view of model errors.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|x_i - y_i| \tag{11}$$

4.3 Coefficient of Determination (R^2)

R^2 measures the proportion of variance in the observed data that is explained by the model.

$$\mathbf{R^2 = 1 - \frac{\sum_{i=1}^{N}(x_i - y_i)^2}{\sum_{i=1}^{N}(x_i - \bar{y})^2}} \quad (12)$$

4.4 Nash–Sutcliffe Efficiency (NSE)

NSE evaluates the predictive skill of a model by comparing it to the performance of the mean of the observed data. Values close to 1 indicate high accuracy.

$$NSE = 1 - \frac{\sum_{i=1}^{N}(x_i - \bar{y}_i)^2}{\sum_{i=1}^{N}(x_i - \bar{y})^2} \quad (13)$$

4.5 Pearson Correlation Coefficient (r)

This metric r measures the strength and direction of the linear relationship between predicted and observed values.

$$r = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2 \sum_{i=1}^{N}(y_i - \bar{y})^2}} \quad (14)$$

5. REVIEW OF THE HYBRID MONTHLY PRECIPITATION PREDICTION MODELS

Hybrid models have gained increasing attention in recent years due to their ability to capture the complex, nonlinear nature of monthly precipitation patterns by combining statistical, machine learning, and signal decomposition methods. To illustrate this trend, Figure 8 shows the annual distribution of publications on hybrid monthly precipitation forecasting from 1996 to 2025, across three major databases: ScienceDirect, Scopus, and IEEE Xplore.

The results highlight a clear increase in research activity since 2017, with a significant surge from 2020 onward, peaking in 2023 with 14 publications. IEEE has become the most prominent source in recent years, reflecting a growing emphasis on AI-driven approaches. This trend confirms the rising relevance and scientific interest in hybrid forecasting models for climate and water resource applications.

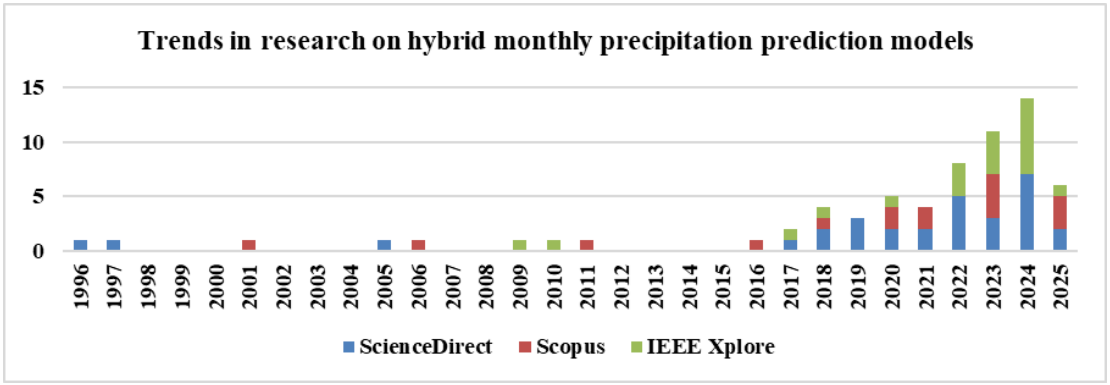


Figure 8 | Trends in research on hybrid monthly precipitation prediction models. The result for the year 2025 is partial; that is why its value is lower.

Table 2 | Main metrics used to evaluate the accuracy of hybrid models for monthly precipitation prediction.

Ref	Study Zone	Resolution Temporal	Spatial Resolution	Variables	Hybrid Type	Based Algorithm

(Gu et al., 2022)	Yangtze River Delta, on the southeast coast of China	Monthly	1°x 1°	Number station, Name Station, Latitude, Longitude, Altitude, Mean Monthly Precipitation, Maximun Monthly Precipitation, Coefficient of Variation of Monthly Precipitation, Southern Oscillation Index, Western Pacific Subtropical High Intensity, See Level Pressure, Maximum Monthly Air Temperature, Minimum Monthly Air Temperature, Mean Monthly Pressure, Mean Vapor Pressure, Relative Humidity, Sunshine Duration.	Meta model based hybrids	KNN, XGB, SVR, ANN
(Pakdaman et al., 2022)	Region of southwest Asia	Monthly	1°x 1°	Mean Monthly Precipitation, Latitude, Longitude.	Meta model based hybrids	CanCM4i, COLA-RSMAS-CCSM4, GEM-NEMO, NASA-GEOSS2S
(L. Tao et al., 2021)	Yangtze River basin (YRB)	Monthly	Regional (129 stations)	Mean Monthly Precipitation, Atlantic Multidecadal Oscilation, Sea Surface Temperature, Indian Ocean Dipole, Ocean Niño Index, Pacific Decadal Oscillation, Transition Niño Index, Western Hemisphere Warm Pool, Atlantic Meridional Mode, Arctic Oscillation, North Atlantic Oscillation, Eastern Atlantic/ Western Russia, North Pacific Pattern, Pacific North American Index, Quasi-Bienn North Atlantic Oscillation, Southern Os Eastern Atlantic/Western Russia, Western Pac North Pacific Pattern, Biavariate EN: Pacific North American Index, Multivariate ENSO Index, Global Average Temperature Anomalies, Solar Flux.	Sequential multiscale hybrid	LSTM, MLR
(Bojang et al., 2020)	Deji Reservoir Basins in Taiwan	Monthly	Regional (22 stations)	Mean Monthly Precipitation,	Sequential multiscale hybrid	LS-SVR, RF
	Shihmen Reservoir Basins in Taiwan	Monthly			Sequential multiscale hybrid	
					Sequential Hybrid	
(Luo et al., 2022)	Kunming Station, China	Monthly	Point (1 station)	Precipitation Products, Atmospheric Pressure, Sea Surface Temperature.	Hybrid decompositio n-based ensemble Sequential Hybrid	ARIMA, SVR, LSTM
(Shen & Ban, 2023)	Lanzhou City, China	Monthly	Point (1 station)	Monthly Precipitation.		SVM, LSTM, XGB, ARIMA

					Descomposition based hybrid	
(Z. Zhou et al., 2021)	Eastern China	Monthly	Regional (25 stations)	Mean Wind Speed, Maximum Monthly Precipitation, Minimum Monthly Temprerature, Mean Monthly Pressure, El Niño Southern Oscillation, North Atlantic Oscillation.	Stacking	RF, ARIMA,ANN SVR, RNN
(Zandi et al., 2022)	Northwest Iran	Monthly	0,25°x0,25°	Cloud Cover Data, Relative Humidity, Wind Speed and Direction	Meta model based hybrids	SVM, RF, MLP
(Kumar *et al.*, 2021)	India	Monthly	Regional (36 stations)	Precipitation Products, Atmospheric Pressure, Sea Surface Temperature.	Optimization enhanced hybrid	DNN, ELM, WDNN, BLR
(Song et al., 2021)	China	Monthly	Point (6 stations)	Precipitation Products, Atmospheric Pressure, North Atlatic Oscillation.	Sequential hybrid	CEEMD, WT
(X. Zhang & Wu, 2023)	Zhengzhou City	Monthly	0,25°x0,25°	Precipitation Products, Atmospheric Pressure, Sea Surface Temperature, North Atlatic Oscillation.	Sequential multiscale hybrid with optimization	ELM, EMD-HHT, FFOA, CEEMD
(Tang et al., 2022)	Danjiangkou River Basin	Monthly	0,25°x0,25°	Precipitation Products, Sea Surface Temperature, Atmospheric Pressure, El Niño Southern Oscillation, North Atlatic Oscillation, Standardized Precipitation Index.	Sequential Hybrid	RF, XGB, RNN, LSTM
(Yeditha et al., 2023)	Krishna River Basin	Monthly	0,25°x0,25°	Atmospheric Pressure, Maximum Monthly Precipitation, Mean Monthly Precipitation.	Sequential multiscale hybrid	ELM, FFBN-NN, MLR
(Papacharalampous et al., 2023)	EE. UU	Monthly	0,25°x0,25°	Altitude.	Meta model based hybrids	RF, XGB, LR, BRNN, MARS, GBM
(Parviz et al., 2021)	Tabríz, Rasht, Iran	Monthly	Point (2 stations)	Maximum Monthly Precipitation, Indian Ocean Dipole, El Niño Southern Oscillation, North Atlantic Oscillation, Sea Surface Temperature.	Stacking	SVR, GEP, GMDH, SARIMA
(Esmaeili et al., 2021)	Ardabil City	Monthly	Point (1 station)	Altitude, Maximum Monthly Precipitation, Mean Monthly Precipitation.	Sequential hybrid	ELM, WT
(Y. Li et al., 2021)	China	Monthly	0,5°x0,5°	Maximum Monthly Precipitation, Mean Monthly Precipitation, Relative Humidity, Wind Speed and Direction, El Niño Southern Oscillation, North Atlantic Oscillation.	Probabilistic Model Averaging	LGB, XGB
		Monthly				

(X. Zhang et al., 2022)	Zhongwei City, China		Point (2 stations)	Altitude, Maximum Monthly Precipitation, Mean Monthly Precipitation, Pacific Decadal Oscillation, Southern Oscillation Index.	Probabilistic Model Averaging	KNN, RF, SVR, ANN, LSTM
(Zerouali et al., 2023)	Sebaou River Basin	Monthly	Point (3 stations)	Mean Monthly Precipitation, Maximum Monthly Precipitation, Wind Speed and Direction, Relative Humidity, Sea Level Pressure.	Sequential hybrid	MLP, ELM
(Priestly et al., 2023)	Idukki, Kerala, India	Monthly	Point (Geographic coordinates)	Precipítation Products, Atmospheric Pressure, Sea Surface Temperature, Standardized Precipitation Index.	Optimization enhanced hybrid	SVM, ANN, KNN, MLR, GP
(Ridwan et al., 2021)	Tasik Kenyir, Terengganu, Malaysia	Monthly	Point (10 stations)	Mean Monthly Precipitation, Maximum Monthly Precipitation, Wind Speed and Direction, Relative Humidity, Sea Level Pressure, Relative Humidity.	Probabilistic Model Averaging	RNN, BDTR, DFR, BLR
(Chhetri et al., 2020)	Simtokha, Thimphu, Bhutan	Monthly	Point (1 station)	Relative Humidity, Wind Speed and Direction, Sea Level Pressure, Mean Monthly Precipitation, Meridional Wind, Altitude, El Niño Southern Oscillation.	Probabilistic Model Averaging	MLP,CNN, LSTM, LR, GRU, BLSTM
(Coşkun & Citakoglu, 2023)	Sakarya, Turkiye	Monthly	Point (1 station)	Altitude, Maximum Monthly Precipitation, North Atlantic Oscillation, Sea Level Pressure, Sea Surface Temperature, Southern Oscillation Index, Wind Speed and Direction.	Descomposition based hybrid	LSTM, ELM, EMD
(Ehteram et al., 2024)	Kashan Plain, Irán	Monthly	Point (5 Stations)	Longitude, Mean Monthly Precipitation, Maximum Monthly Precipitation, Relative Humidity, Atmospheric Pressure, Latitude, Satellite Derived Rainfall Estimates, Sea Surface Temperature	Prediction Averaging (FR), Feature Averaging (DRSN-TCN-RF) Prediction Averaging (FR), Feature Averaging (DRSN-TCN-RF) Descomposition based hybrid	RF, TCN, DRSN, OPA
(Ahmadi et al., 2024)	Sindh River basin, India	Monthly	Pointl (8 stations)	Precipitation Products, Standard Deviation, Coefficient of Variation, Approximate Coefficient (Level 3)	Descomposition based hybrids	RF, Kstar, GPR
(Parviz, 2020)	Rasht, Gorgan, Irán	Monthly	Point (2 stations)	Altitude, Atmospheric Pressure, Latitude, Longitude, Mean Monthly Precipitation, Solar Flux, Seasonal Cycle, Streamflow, Altitude, Atmospheric Pressure.	Sequential hybrid	SARIMA, ANN, SVM

(Guo et al., 2024)	Jinan, China	Monthly	Point (1 station)	Mean Monthly Precipitation, Maximum Monthly Precipitation, Relative Humidity, Maximum Monthly Temperature, Minimum Monthly Temperature.	Sequential hybrid	ANN, RNN, LSTM, CNN
(Hou et al., 2024)	Lanzhou, Gansu, China	Monthly	Point (1 station)	Atmospheric Pressure, Relative Humidity, Wind Spedd and Direction, Mean Wind Speed.	Descompositi on based hybrid with optimization	LSTM, CPO-LSTM
(X. Wu et al., 2021)	Jilin, China	Monthly	Point (3 stations)	Altitude, Longitude, Latitude.	Descompositi on based hybrid	ARIMA. LSTM, GM, DGM, GM
(H. Zhang et al., 2020)	Hubei, China	Monthly	Point (27 stations)	Mean Monthly Precipitation, Altitude, Relative Humidity, Latitude, Terrain Slope, Normalizaed Diference Vegetation Index, Daytima Surface Temperature.	Sequential hybrid	BP-ANN, MLR, RF, SVR, GPR
(Latif et al., 2024)	Sulaymaniyah, Iraq	Monthly	Point (1 station)	Mean Monthly Precipitation, Time Lags.	Sequential hybrid	ANN, SARIMA
(Katipoğlu & Keblouti, 2024)	El Kerma, Algeria	Monthly	Point (1 station)	Maximum Monthly Precipitation.	Descompositi on based hybrid	SVM, ANN, GPR

Table 3 | Comparative Performance of Hybrid and Ensemble Models for Monthly Precipitation Prediction: Evaluation Metrics and Percent Performance Improvement.

Ref	Based Algorithm	Bets Model	Evaluation Metrics	% Performance
(Gu et al., 2022)	KNN, XGB, SVR, ANN	ANN, Stacking	**ANN**: MAE=42,47; R^2= 0,532; RMSE= 60,92; **Stacking;** MAE=41,65; R^2= 0,526; RMSE= 61,51	0,96
(Pakdaman et al., 2022)	CanCM4i, COLA-RSMAS-CCSM4, GEM-NEMO, NASA-GEOSS2S	RF, ANN	**RF**: KGE= 0,68 ; NSE= 0,6; r= 0,77; RMSE= 24,83; **ANN**: KGE= 0,64 NSE= 0,52; r= 0,74; RMSE= 26,08	4,21
(L. Tao et al., 2021)	LSTM, MLR	MLSTM-AM, LSTM	**MLSTM-AM:** NSE=0,46; RAE= 0,41; **LSTM;**NSE= 0,44; RAE= 0,44	6,81
(Bojang et al., 2020)	LS-SVR, RF	SSA-LSSVR , SSA-RF	**SSA-LSSVR:** NSE= 0,86; RMSE= 75,29; **SSA-RF:** NSE= 0,63; RMSE= 121,76	38,16
		SSA-RF, SSA-LSSVR	**SSA-RF:** NSE= 0,82; RMSE= 98,75; **SSA-LSSVR:** NSE= 0,67; RMSE= 132,81	25,64
(Luo et al., 2022)	ARIMA, SVR, LSTM	EEMD-BMA, EEMD-LSTM	**EEMD.BMA** MSE= 298,6359; CP=93,75; RMSE=17,28; MAE=12,7; R^2=0,9573 MW=60,315 **EEMD-LSTM:** MSE=301,94; RMSE= 17,37; MAE= 12,77; R^2= 0,956	0,5
				57,54

(Shen & Ban, 2023)	SVM, LSTM, XGB, ARIMA	CEEMDAN-SVM-LSTM, CEEDMAN-LSTM	**CEEMDAN-SVM-LSTM:** RMSE=7,68; MSE=58,98; MAE=5,58; R^2=0,95; **CEEDMAN-LSTM:** RMSE=18,09; MSE= 327,25; MAE=12,43; R^2=0,71	
(Z. Zhou et al., 2021)	RF, ARIMA,ANN SVR, RNN	RF, ANN	**RF:** MAE=25,7; RMSE=40,8; R^2=0,82; **ANN:** MAE=26,9; RMSE=41,7; R^2=0,82	2,16
(Zandi et al., 2022)	SVM, RF, MLP	Stck-Las-Rsc, Stck-Las	**Stck-Las-Rsc:** RMSE=25,4; MAE=15,1; RBIAS= -1,4; KGE=0,88; **Stck-Las:** RMSE=25,7; MAE=15,2; RBIAS= -1,6; KGE=0,89	1,16
(R. Kumar et al., 2021)	DNN, ELM, WDNN, BLR	WDNN, WBBO-ELM	**WDNN:** RMSE=185,9; NSE=0,96; MAE=130,86; R^2= 0,98; **WBBO-ELM:** RMSE=194,4; NSE=0,957; MAE=132,47; R^2= 0,979	4,37
	CEEMD, WT	TVF-EMD-ENN, WT-ENN	**TVF-EMD-ENN:** RMSE=23,775; MAE=18,474; NSE= 0,97; **WT-ENN:** RMSE=35,5; MAE=28; NSE= 0,95	33,03
(X. Zhang & Wu, 2023)	ELM, EMD-HHT, FFOA, CEEMD	CEEMD-ELM-FFOA, CEEMD-ELM	**CEEMD-ELM-FFOA;** MAE=0,55; RMSE=0,81; MAPE=1,39; **CEEDM-ELM:** MAE=0,63; RMSE=0,92; MAPE=3,23	11,96
(Tang et al., 2022)	RF, XGB, RNN, LSTM	SMOTE-km-XGB, SMOTE-km-RF	**SMOTE-km-XGB:** ACC=05; Pg score=80; **SMOTE-km-RF:** ACC=05; Pg score=70	12,5
(Yeditha et al., 2023)	ELM, FFBN-NN, MLR	WT-ELM, WT-FFBP-NN	**WT-ELM:** RMSE=0,03; MAE=0,052; r=0,925; NSE=0,855; **WT-FFBP-NN:** RMSE=0,033; MAE=0,055; r=0,892; NSE=0,796	9,1
(Papacharalampous et al., 2023)	RF, XGB, LR, BRNN, MARS, GBM	LR, polyMARS	**LR:** MSE= 39,77; **ployMARS:** MSE= 39,18	1,48
(Parviz et al., 2021)	SVR, GEP, GMDH, SARIMA	GA, SARIMA-SVR	**GA:** NSE=0,97; RMSE=11,48; RRMSE=0,1; MSE=131,84; RPD=6,57; APB=0,078; Dm=0,99; U2=0,088;AMPE=4,52; **SARIMA-SVR:** NSE=0,92; RMSE=20,02; RRMSE=0,187; MSE=400,9; RPD=3,77; APB=0,142; Dm=0,97; U2=0,15; AMPE= 4,58	42,66
(Esmaeili et al., 2021)	ELM, WT	WORELM 14, WORELM 6	**WORELM 14:** MARE=4,345 ; r=0,965; NSC=0,965; VAF=93,16; **WORELM 6:** MARE=6,5; r=0,92; NSC=0,90; VAF=92,5	33,15
(Y. Li et al., 2021)	LGB, XGB	SEAS4 & SEAS5	**SEAS 5:** MAE=0,35; PCC=0,9; **SEAS 4:** MAE=0,65; PCC=0,8;	11,11
(X. Zhang et al., 2022)	KNN, RF, SVR, ANN, LSTM	CEEMD-FCMSE-Stacking, CEEDM-LSTM	**CEEMD-FCMSE-Stacking:** MAE=4,26; RMSE=7,71; R^2=0,923; **CEEDM-LSTM:** MAE= 7,62; RMSE= 10,16; R^2=0,88	24,11
(Zerouali et al., 2023)	MLP, ELM	Bat-ELM, MLP-PSO	**Bat-ELM:** RMSE=11,89; NSE=0,985; r=0,993; R^2=0,986; **MLP-PSO:** RMSE=30,50; NSE=0,879; r=0,951; R^2=0,905;	61,31
(Priestly et al., 2023)	SVM, ANN, KNN, MLR, GP	LSO-ABR Model 3, LSO-ABR Model 1	**LSO-ABR Model 3:** RMSE=52,39; MAE=39,89; **LSO-ABR Model 1:** RMSE=53,29; MAE=40,33	1,69
(Ridwan et al., 2021)	RNN, BDTR, DFR, BLR	BDTR, DFR	**BDTR;** MAE=0,0123; RMSE=0,0335; R^2=0,999 **DFR:** MAE=0,0946; RMSE=0,1565; R^2=0,7623	78,84
(Chhetri et al., 2020)	MLP, CNN, LSTM, LR, GRU, BLSTM	BLSTM-GRU, LSTM	**BLSTM-GRU:** RMSE=0,087; MSE=0,0075; R^2=0,87; PCC= 0,938; **LSTM:** MSE=0,0075; R^2=0,87; PCC=0,90	4,05
(Coşkun & Citakoglu, 2023)	LSTM, ELM, EMD	LSTM & EMD-ELM	**LSTM:** MAE=0,11; NSE=0,97; R^2=0,97; RMSE=0,17; OI=0,97;	32

(Ehteram et al., 2024)	RF, TCN, DRSN, OPA	DTR, TCN-RF	**EMD-ELM:** MAE=0,22; NSE=0,95; R^2=0,96; RMSE=0,25 OI=0,95 **DTR:** RMSE=0,15; MAE=0,195; NSE=0,96; R^2=0,992; **TCN-RF:** RMSE=0,19; MAE=0,195; NSE=0,94; R^2=0,985	21,05
(Ahmadi et al., 2024)	RF, Kstar, GPR	Wavelet-RF, Wavelet-Kstar	**Wavelet-RF:** RMSE=34,47; MAE=20,53; KGE=0,917; WI=0,884; **Wavelet-Kstar:** RMSE=36,59; MAE=20,339; KGE=0,878; WI=0,871	5,79
(Parviz, 2020)	SARIMA, ANN, SVM	SARIMA-SVM, ANN	**SARIMA-SVM:** RMSE=14,18; MAE=9,55; RRMSE=0,39; d= 0,84 ; U1=0,16; **ANN:** RMSE=23,88; MAE= 18,93; RRMSE=0,66; d= 0,43; U1=0,26	40,62
(Guo et al., 2024)	ANN, RNN, LSTM, CNN	CNN-LSTM, CNN	**CNN-LSTM:** RMSE=0,6292; MAE=0,5048; r=0,998; **CNN:** RMSE=0,8015; MAE=0,6680; r=0,996	21,5
(Hou et al., 2024)	LSTM, CPO-LSTM	VMD-CPO-LSTM, VDM-LSTM	**VMD-CPO-LSTM:** RMSE=15,52; MAE=9,98; R^2=0,9039; NSE=0,9039; **VMD-LSTM:** RMSE=22,02; MAE=13,74; R^2=0,806; NSE=0,806;	29,52
(X. Wu et al., 2021)	ARIMA. LSTM, GM, DGM, GM	W-AL (Wavelet-ARIMA-LSTM), LSTM	**W-AL:** RMSE=31,308; MAE=22,853; R^2=0,712 **LSTM:** RMSE=36,024; MAE=25,865; R^2=0,618	13,09
(H. Zhang et al., 2020)	BP-ANN, MLR, RF, SVR, GPR	BP-ANN, Rattional Quadratic GPR	**BP-ANN:** RMSE=43,32; MAE=31,9; R^2=0,64; MSE=1876,5; **Rattional Quadratic GPR:** RMSE=51,22; MAE=33,13; R^2=0,59; MSE=2623,8	15,42
(Latif et al., 2024)	ANN, SARIMA	SARIMA-ANN, ANN Model 4	**SARIMA-ANN:** R^2=0,98; RMSE= 11,5; **ANN Model 4:** R^2=0,43; RMSE= 51,002	77,45
(Katipoğlu & Keblouti, 2024)	SVM, ANN, GPR	W-LSVM, LSVM	W-LSVM: R^2=0,78; RMSE=8,1; MAE=5,9; LSVM: R^2=0,78; RMSE=8,18; MAE=5,9	0,98

(RMSE) and (MAE) stand out among the six most used evaluation metrics. However, it is essential to note that (NSE), (R2), and (MSE) are also widely used alongside RMSE and MAE. Other evaluation metrics include CC, RBIAS, ACC, and OR CRPS. As shown in Figure 7, there is a wide variety of metrics, and choosing the most appropriate one for the model being evaluated is essential to ensure rigor, efficiency, and accuracy. Most review articles compare simple models with hybrid or ensemble models Table 2, demonstrating the superior performance of hybrid and ensemble models over simple models. However, few studies compare hybrid models with other hybrid or ensemble models. In the models where such comparisons were made, it was observed that some models perform better for a specific geographic area. The forms of hybridization (Hybrid predictive models) include Hybridization by Preprocessing (HP), Hybridization of Neural Networks with Optimization Algorithms (HRNAO), Hybridization of Machine Learning Models with Signal Decomposition (HMMLDS), and Hybridizations of Statistical Time Series Models with Neural Networks (HMESTRN). Hybrid and ensemble models present various variations and construction methods, adapting to different contexts and specific needs by leveraging the strengths of different techniques, methods, algorithms, and models. As shown in Table 4 across 34 unique studies, data preprocessing–based and component combination–based hybrids dominate (10 and 9 studies, respectively), followed by other hybrids (7), while parameter optimization (4) and deep hybrids (3) are less common, and post-processing hybrids are rare (1).

Table 4 | R^2 summary by class.

Model Type	Studies
Parameter Optimization	4
Data Preprocessing	10
Deep Hybrid	3
Other Hybrid	7
Component Combination	9
Post-processing	1
Parameter Optimization	4
Data Preprocessing	10
Deep Hybrid	3

6. DISCUSSION

Hybrid and ensemble models consistently achieve higher explained variance (R^2) than their non-hybrid counterparts (Table 2 - Table 3**Error! Reference source not found.**; Figure 3). This pattern supports the premise that combining signal-processing or data-conditioning stages with advanced learners improves skill at the monthly scale. The subsections below compare hybrid families on a common R^2 basis, emphasize protocol-sensitive factors that drive gains, and identify techniques with the most robust performance. Differences observed across regions are plausibly linked to data density, terrain complexity, and prevailing circulation regimes, which together condition the attainable out-of-sample R^2. While most primary studies compare a single hybrid against simple baselines, systematic head-to-head tests among hybrids remain comparatively scarce; expanding these comparisons is essential to understand which design choices generalize.

6.1 Systematic Comparison of Specific Hybrid Techniques

The taxonomy-based synthesis indicates clear differences in central R^2 across hybrid families (Table 2; Figure 3). Parameter-optimization hybrids display the highest median ($R^2 \approx 0.975$), followed by data-preprocessing hybrids ($R^2 \approx 0.904$) and deep hybrids ($R^2 \approx 0.870$). Other hybrids center around 0.820, whereas component-combination/stacking and post-processing-only configurations show lower central tendencies (both ≈ 0.650; the latter from a single study). A Kruskal–Wallis test using study-level R^2 does not identify statistically significant differences among families at $\alpha = 0.05$ ($p = 0.2218$), reflecting heterogeneity in datasets and validation protocols and the limited sample in some categories. Even so, the concentration of higher medians in parameter-optimization and data-preprocessing hybrids suggests that feature/parameter search and multi-scale signal conditioning are recurrent ingredients of high-R^2 designs.

6.2 Comparative analysis of performance in terms of R^2

Across studies, R^2 improves when hybrids (i) reduce non-stationary variance before learning (e.g., EMD/CEEMDAN/VMD, wavelets, SSA) and/or (ii) search the hyper-parameter/feature space with metaheuristics (e.g., GA, PSO, BBO, BAT, LSO). Reported examples include increases from $R^2 \approx 0.60$ to ≈ 0.95 when CEEMDAN-based preprocessing is paired with LSTM/SVM in highly seasonal settings (Shen & Ban, 2023), and ~0.12–0.20 absolute gains in R^2 for WDNN versus conventional baselines under national-scale evaluation (Kumar et al., 2021). Studies combining standardization or SSA with SVM/ensemble learners also report sizable ΔR^2 in monsoonal and subtropical climates (Gu et al., 2022; Pakdaman et al., 2022). Where authors emphasize error metrics, the accompanying narratives almost invariably note higher R^2 for the hybrid relative to its baseline; the synthesis here retains only the R^2 figures to enable cross-study comparability.

Two methodological issues dominate the spread in R^2: (1) validation design, where random k-folds can inflate R^2 by mixing spatial neighbors or seasons, versus blocked spatiotemporal CV that yields more realistic estimates; and (2) data regime, where complex orography and sparse gauges depress attainable R^2 unless hybrids explicitly encode topography or perform region-wise modeling. Consequently, high-median families in Figure 3 should be interpreted as promising defaults, not guarantees—protocols and covariates remain decisive.

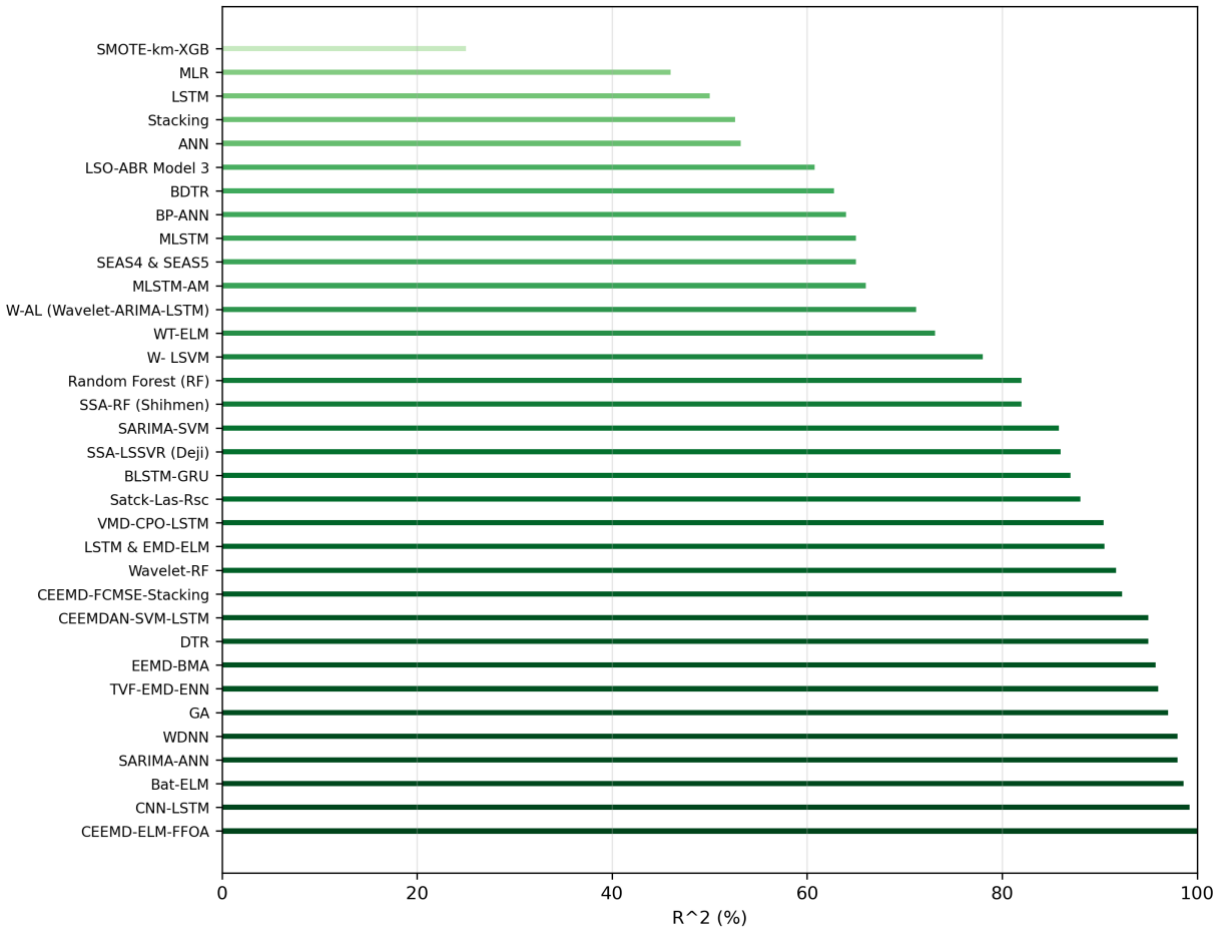


Figure 9 | Comparative performance of hybrid models according to R^2:7

The model-wise ranking in Figure 9 (R^2 on a 0–1 scale) exhibits a clear gradient from plain learners to hybrids with explicit signal conditioning and/or parameter optimization. Three tiers emerge:

- High R^2 ($\geq\approx 0.90$): hybrids that couple decomposition (CEEMD/EMD/VMD/TVF-EMD) with DL/ELM, often augmented by metaheuristic tuning or informed combination (e.g., CEEMD-ELM-FFOA, CNN-LSTM, Bat-ELM, WDNN, EEMD-BMA, CEEMDAN-SVM-LSTM, VMD-CPO-LSTM).
- Mid R^2 (≈0.70–0.85): pipelines using wavelets/SSA + RF/SVM/ELM and SARIMA–SVM; gains are consistent yet smaller and sensitive to the downstream regressor.
- Low R^2 ($<\approx 0.65$): plain stacking, ANN/LSTM without preprocessing, MLR, SMOTE-km-XGB; aggregation alone does not offset non-stationarity or regime shifts.

These patterns are consistent with Table 3, where the parameter-optimization and data-preprocessing families have the highest medians (≈0.975 and ≈0.904, respectively), whereas the component-combination and post-processing-only configurations center near ≈0.650. Near-unit R^2 values, while achievable in strongly seasonal settings, warrant caution because they may reflect optimistic validation. Consequently, comparative reporting should emphasize $\nabla R^2 = R^2_{hyb} - R^2_{base}$ under-blocked spatiotemporal cross-validation, accompanied by uncertainty bands bootstrap, to ensure robust and out-of-sample interpretability.

6.3 Limitations and sources of uncertainty in ML precipitation prediction (R^2 focus)

Key limitations include: (i) historical dependence and non-stationarity—relationships may drift under climate variability; (ii) initial-state sensitivity at monthly leads; (iii) sparse or biased observations and representativeness errors; (iv) shortcut learning/overfitting with limited samples; (v) data leakage from random CV that mixes space or

seasons, typically inflating R^2; and (vi) limited uncertainty quantification. Mitigations include blocked spatial/temporal CV (leave-region/month-out), hindcast evaluation, explicit leakage checks, feature ablation, and routine reporting of R^2 on independent splits with uncertainty bands as bootstrap CIs.

6.4 Complementarity of physical NWP and ML

Machine learning complements—rather than replaces—physical numerical weather prediction (NWP) along three main axes: (i) data-assimilation error correction/emulation, (ii) downscaling and bias-correction/post-processing, and (iii) surrogate modeling of expensive components (Chantry et al., 2021; Farchi et al., 2021; Geer, 2021; Maraun, 2016; Vannitsem et al., 2021). At monthly horizons, most hybrid pipelines reviewed here operate in (ii) and, to a lesser extent, (iii), whereas (i) is primarily relevant to short-range NWP but increasingly informed by ML.

Evidence from this review is consistent with that positioning Figure 10. Categories that most closely mirror NWP post-processing workflows—namely Parameter Optimization and Data Preprocessing—concentrate the highest median R^2 (≈0.975 and ≈0.904, respectively), indicating that multi-scale signal conditioning (EMD/CEEMD/VMD/TVF-EMD, SSA, standardization) combined with principled hyper-parameter/feature search yields the largest gains in explained variance when refining predictors derived from large-scale circulation. Deep hybrids follow (median R^2 ≈0.870), particularly when coupled with decomposition, while component-combination and post-processing-only configurations center around ≈0.650 (the latter represented by a single study), suggesting that aggregation without explicit conditioning/tuning seldom suffices to raise out-of-sample R^2. The global distribution is right-skewed with a median R^2 of 0.859, yet near-unit values warrant caution as they may reflect optimistic validation; blocked spatiotemporal CV and reporting of $\nabla R^2 = R^2_{hyb} - R^2_{base}$ on independent splits remain essential.

Operationally, the results support a pragmatic division of labor: use NWP for dynamics and conservation, then apply ML for spatial refinement and statistical calibration, including EMOS/QRF/ECC or neural calibration, and quantify gains with ΔR^2 under leakage-safe protocols. Recent advances in ML-augmented global forecasting and generative ensemble emulation (Bi et al., 2023; L. Li et al., 2024) further underline the promise of hybrid NWP–ML systems, provided that verification adheres to robust, distribution-aware standards.

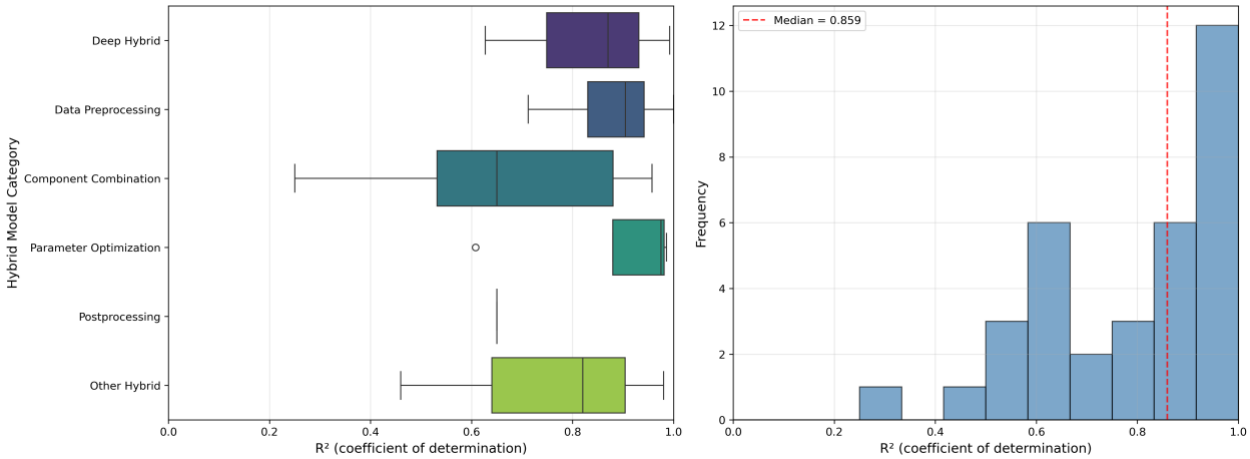


Figure 10 | R^2 performance of hybrid model categories for monthly precipitation. Left: boxplots of R^2 (coefficient of determination, 0–1) by hybrid category (Deep Hybrid, Data Preprocessing, Component Combination, Parameter Optimization, Post-processing, Other Hybrid). Boxes show median and IQR; whiskers denote 1.5×IQR; points are outliers. Right: histogram of study-level R^2 with the global median = 0.859 (red dashed line). The concentration of higher medians in Parameter Optimization and Data Preprocessing is consistent with ML's complementary role to NWP in downscaling/bias-correction and tuning, whereas Component Combination and Post-processing-only configurations show lower central tendencies.

6.5 Promising areas and under-explored hybrid techniques

Several directions remain under-tested in R^2-centric evaluations: post-processing hybrids (P(M)A variations), bio-inspired optimizers coupled to deep learners (HBA, BAT), and sequential preprocessing (SMOTE-K-means before learning) for regime balancing. Incorporating topography-aware clustering or elevation-stratified modeling is

particularly promising in complex terrain. For example, in a mountainous Andean case study (Niño Medina et al., 2024), an LSTM achieved $R^2 \approx 0.92$ for 16-month monthly forecasts using CHIRPS inputs; explicit altitude-based clustering (k-means/DBSCAN) could plausibly further stabilize R^2 by enforcing spatial homogeneity. Future studies should report $\nabla R^2 = R^2_{hyb} - R^2_{base}$ under a blocked CV and include uncertainty bands, enabling rigorous head-to-head comparisons among hybrid families.

Hybrid and ensemble models have consistently demonstrated better performance compared to individual models, as shown in. This finding reinforces the idea that integrating signal decomposition techniques with advanced machine learning algorithms significantly enhances predictive capability. The following subsections provide a more detailed comparative analysis, exploring specific techniques and highlighting key metrics to identify particularly robust methodologies. These findings are consistent with previous studies that have demonstrated the advantages of hybrid and ensemble approaches. This could be attributed to differences in the datasets used and the specific climatic conditions of each region studied. Most studies have focused on developing and evaluating simple machine learning models compared with hybrid models; however, there has been sufficient development and evaluation among hybrid or ensemble models themselves. To continue improving prediction accuracy, it is advisable to investigate new combinations of signal decomposition techniques and machine learning algorithms. It is important to develop and evaluate hybrid models with small variations, whether in the signal decomposition stage, training, result combination, or other aspects, to determine which combinations may prove to be more beneficial. Most of the reviewed studies were conducted in mountainous regions or areas combining mountains and plains, featuring diverse terrains such as mountain plateaus, river valleys, and coastal plains. Elevations ranged widely from -431 meters below sea level to a maximum of 8,848 meters, reflecting a broad geographical scope. Eastern and northwestern China, as well as the mountainous regions of Iran, were the most studied locations. Regarding data sources, the most frequently used variables included mean monthly precipitation, obtained from both local meteorological stations and satellite products. Some studies also incorporated topographic variables like altitude and atmospheric pressure, alongside global climate phenomena such as ENSO.

7. ABBREVIATIONS

This section provides a comprehensive list of abbreviations used throughout the article. **Error! Reference source not found.** summarizes the key terms, including abbreviations related to predictive models, data preprocessing techniques, data postprocessing techniques, and optimization parameters. This table aims to facilitate the reader´s understanding by clarifying the terminology used in the methodological and analytical discussions.

Table 5 | Abbreviations of Models, Techniques, and Optimization Parameters.

Abbreviation	Description	Abbreviation	Description
ACF	Autocorrelation Function	GP	Genetic Programming
ACF and CCF	Autocorrelation Function / Cross-Correlation Function	GPR	Gaussian Process Regression
ADF	Augmented Dickey-Fuller Test	GRU	Gated Recurrent Unit
ANFIS	Adaptive Neuro-Fuzzy Inference System	HBA	Honey Badger Algorithm
ANN	Artificial Neural Network	HYPERPARAMETER TUNING	The process of fine-tuning model parameters.
ARIMA	AutoRegressive Integrated Moving Average	IMFs	Intrinsic Mode Functions
ARIMA	AutoRegressive Integrated Moving Average	K MEANS	K-means Clustering
ARIMA-STL	Autoregressive integrated moving average descompistion with STL	KNN	K-Nearest Neighbors
BAT	Bat Algorithm	LGB	Light Gradient Boosting Machine
BBO	Biogeography-Based Optimization	LOMOCV	Leave-One-Month-Out Cross-Validation
BBO-ELM	Biogeography-based Extreme Learning Machine	LR	Linear Regression

BDTR	Boosted Decision Tree Regression
Bias Correction	A method to adjust systematic biases in prediction models
BJP	Bayesian Joint Probability
BLR	Bayesian Linear Regression
BLSTM	Bidirectional Long Short-Term Memory
BLSTM-GRU	Bidirectional LSTM combined with GRU
BPNN	Backpropagation Neural Network
BRFFNN	Bayesian regularized Feedforward Neural Network
BRNN	Bidirectional Recurrent Neural Network
CanCM4i	Canadian Climate Model version 4
CEEMD	Complete Ensemble Empirical Mode Decomposition
CEEMDAN	Complete Ensemble Empirical Mode Decomposition with Adaptive Noise
CNN	Convolutional Neural Network
COLA-RSMAS-CCSM4	Community Climate System Model by COLA-RSMAS
CPO	The Crested Porcupine Optimization (CPO)
CPO-LSTM	The crested porcupine optimization algorithm for LSTM
DFR	Decision Forest Regression
DNN	Deep Neural Network
DRSN	Deep Residual Shrinkage Network
DSABASRNN	Deep Self Attention-Based Augmented SRNN
DWT	Discrete Wavelet Transform
EEMD	Ensemble Empirical Mode Decomposition
EEMD	Ensemble Empirical Mode Decomposition
EEMD-ARIMA	EEMD combined with ARIMA
EEMD-LSTM	EEMD combined with LSTM
EEMD-SVR	EEMD combined with Support Vector Regression
ELM	Extreme Learning Machine
EMD	Empirical Mode Decomposition
EMD	Empirical Mode Decomposition
EMD-ELM	EMD combined with ELM
EMD-HHT	EMD using Hilbert–Huang Transform
FBP	Facebook Prophet ModeL
FCMSE	Fuzzy Comprehensive Mean Square Error
FFA	The Firefly Algorithm (FFA)
FFBP-NN	Feedforward Backpropagation Neural Network
LSO	Leave-Some-Out
LS-SVR	Least Squares Support Vector Regression
LSTM	Long Short-Term Memory
LTP	Lead times processing
MARS	Multivariate Adaptive Regression Splines
MIN-MAX NORMALIZATION	Scales data to a range between 0 and 1
MLP	Multilayer Perceptron
MLR	Multiple Linear Regression
MLSTM	Multiscale LSTM
MLSTM-AM	Multiscale LSTM with Attention Mechanism
MODWT	Maximal Overlap Discrete Wavelet Transform
NASA-GEOSS2S	NASA Global Earth Observing System Subseasonal to Seasonal
NMF	Non-negative Matrix Factorization
OPA	Optimization algorithm (unspecified)
PACF	Partial Autocorrelation Function
PCA	Principal component analysis (PCA) method
PI	Predictor Importance
PSO	Particle Swarm Optimization:
PSO-ELM	Particle Swarm Optimization optimized ELM
R/S Analysis	Rescaled Range Analysis
RF	Random Forest
RNN	Recurrent Neural Network
RT	Regression Trees
SARIMA	Seasonal ARIMA
SARIMA	Seasonal ARIMA: An extension of ARIMA that incorporates seasonal components into time series modeling.
SCE	The Shuffled Complex Evolution (SCE)
SHANNON ENTROPY	Is a measure of uncertainty or randomness in a dataset.
SMOTE	Synthetic Minority Oversampling Technique
SMOTE + KMEANS	Sequential Preprocessing combinations
SSA	Singular Spectrum Analysis
SSA-LSSVR	Singular Spectrum Analysis combined with LSSVR
SSA-RF	Singular Spectrum Analysis combined with Random Forest
SVM	Support Vector Machine
SVN/ANN	Support Vector Network / Artificial Neural Network
SVR	Support Vector Regression

FFNN	Feedforward Neural Network	TCN	Temporal Convolutional Network
FFOA	Fruit Fly Optimization Algorithm	TPS	Thin Plate Spline
FFOA	The Fruit Fly Optimization Algorithm (FFOA)	TVF	Time-Varying Filtering
GA	Genetic Algorithm	TVF-EMD	Time-Varying Filtering with EMD
GBM	Gradient Boosting Machine	VMD	Variational Mode Decomposition
GCM SEAS4 y SEAS5	Outputs from global climate models (GCMs) of the SEAS4/SEAS5 systems from ECMWF	WDNN	Weighted Deep Neural Network
GEM-NEMO	Global Environmental Multiscale model – NEMO	WELM	Weighted Extreme Learning Machine
GEP	Gene Expression Programming	WORELM	Weighted Online Random ELM
GMDH	Group Method of Data Handling	WT	Wavelet Transform
GM-OPA	Generalized Model with Optimal Predictor Algorithm	XGB	Extreme Gradient Boosting

8. CONCLUSIONS

This review shows that hybrid and ensemble pipelines consistently explain more variance than non-hybrid baselines for monthly precipitation prediction. In a harmonized R^2-only synthesis, study-level medians concentrate in Parameter-Optimization (≈0.975) and Data-Preprocessing hybrids (≈0.904), followed by Deep Hybrids (≈0.870); Component-Combination and Post-processing-only solutions center near ≈0.650. The global median $R^2 = 0.859$ indicates broadly strong performance across contexts. Differences among families are not statistically significant under Kruskal–Wallis ($p = 0.2218$), reflecting heterogeneous datasets and protocols; medians should therefore be read as tendencies, not rankings. Two design elements recur in high-R^2 systems: (i) multi-scale signal conditioning (EMD/CEEMD/VMD/TVF-EMD, SSA, standardization) that reduces non-stationary variance before learning, and (ii) principled search over features and hyper-parameters (GA/PSO/BBO/BAT/LSO). Pipelines relying on aggregation alone, or deep learning without explicit preprocessing, typically occupy the mid–low R^2 tiers. Results align with the complementary role of ML relative to NWP: physics-based models supply dynamical consistency, while ML is most effective for downscaling and statistical calibration of large-scale drivers or NWP-derived predictors.

A key barrier to comparability across studies is the patchwork of evaluation metrics. Most articles report RMSE/MAE in millimetres and rarely document how errors are spatially aggregated (areal mean vs. mean of points; area-weighting; gridding). Unit-dependent and aggregation-dependent choices obscure the relationship between error magnitude and local climatology and impede meta-analysis. To remove this ambiguity, this review adopts R^2 as the primary metric, as it is dimensionless, scale-invariant, and directly interpretable as the explained variance. Going forward, credible reporting should pair R^2 with $\Delta R^2 = R^2\_hyb - R^2\_base$, blocked spatiotemporal cross-validation, leakage checks, and uncertainty bands (e.g., bootstrap CIs), ensuring that reported skill reflects model quality rather than methodological or hydrologic scale effects.

Standardize evaluation by adopting shared spatial-aggregation protocols (areal mean vs. pointwise mean, area-weighting, gridding) and by reporting unit-free errors alongside R^2. Studies should always provide $\Delta R^2 = R^2\_hyb - R^2\_base$ with uncertainty bands under leakage-safe spatiotemporal validation. Build open, multi-climate benchmarks with identical partitions to compare hybrid families, including computational cost and data requirements, so that generalizable—not merely well-tuned—designs can be identified.

Investigate the synergy between multi-scale preprocessing and metaheuristic optimization via controlled ablations; advance topography-aware modeling (elevation/cluster stratification, hierarchical or graph-based formulations) to stabilize out-of-sample R^2 and improve transfer across basins. Expand distributional calibration within hybrid pipelines (EMOS, QRF, ECC, neural calibration) and quantify gains with ΔR^2 and distribution-aware diagnostics. Finally, monitor robustness and drift using rolling hindcasts and drift tests, and enhance transparency and reproducibility by releasing code, splits, and extraction sheets, while documenting preprocessing, aggregation, and hyperparameter search. Collectively, these actions will clarify when and why particular hybrid designs deliver durable R^2 gains across climates, data regimes, and operational settings.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The extracted datasets and R^2 tables generated during the current study, along with analysis scripts, are available from the corresponding author on reasonable request.

REFERENCES

Ahmadi, F., Mirabbasi, R., Kumar, R., & Gajbhiye, S. (2024). Prediction of precipitation using wavelet-based hybrid models considering the periodicity. *Neural Computing and Applications*, *36*(26), 16345–16364. https://doi.org/10.1007/s00521-024-10006-7

Ali, M., Deo, R. C., Xiang, Y., Li, Y., & Yaseen, Z. M. (2020). Forecasting long-term precipitation for water resource management: a new multi-step data-intelligent modelling approach. *Hydrological Sciences Journal*, *65*(16), 2693–2708. https://doi.org/10.1080/02626667.2020.1808219

and Space Administration, N. A. (n.d.). *Understanding Earth: Whats Up With Precipitation? Precipitation Education*. https://gpm.nasa.gov/education/articles/understanding-earth-whats-precipitation

Anuradha, T., Aruna Sri Formal, P. S. G., & RamaDevi, J. (2024). Hybrid model for rainfall prediction with statistical and technical indicator feature set. *Expert Systems with Applications*, *249*. https://doi.org/10.1016/j.eswa.2024.123260

Ardabili, S., Mosavi, A., & Várkonyi-Kóczy, A. R. (2020). *Advances in Machine Learning Modeling Reviewing Hybrid and Ensemble Methods* (A. R. Várkonyi-Kóczy, Ed.; pp. 215–227). Springer International Publishing. https://doi.org/10.1007/978-3-030-36841-8_21

Arsenault, R., Gatien, P., Renaud, B., Brissette, F., & Martel, J.-L. (2015). A comparative analysis of 9 multi-model averaging approaches in hydrological continuous streamflow simulation. *Journal of Hydrology*, *529*, 754–767. https://doi.org/10.1016/j.jhydrol.2015.09.001

Arsov, N., Pavlovski, M., & Kocarev, L. (2019). *Stacking and stability*. arXiv. https://doi.org/10.48550/arXiv.1901.09134

Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., & Tian, Q. (2023). Accurate medium-range global weather forecasting with 3D neural networks. *Nature*, *619*(7970), 533–538. https://doi.org/10.1038/s41586-023-06185-3

Bojang, P. O., Yang, T. C., Pham, Q. B., & Yu, P. S. (2020). Linking Singular Spectrum Analysis and Machine Learning for Monthly Rainfall Forecasting. *Applied Sciences 2020, Vol. 10, Page 3224*, *10*(9), 3224. https://doi.org/10.3390/APP10093224

Castro, R., Souto, Y. M., Ogasawara, E., Porto, F., & Bezerra, E. (2021). STConvS2S: Spatiotemporal Convolutional Sequence to Sequence Network for weather forecasting. *Neurocomputing*, *426*, 285–298. https://doi.org/10.1016/j.neucom.2020.09.060

Chantry, M., Christensen, H., Dueben, P., & Palmer, T. (2021). Opportunities and challenges for machine learning in weather and climate modelling: Hard, medium and soft AI. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *379*(2194). https://doi.org/10.1098/rsta.2020.0083

Chao, Z., Pu, F., Yin, Y., Han, B., & Chen, X. (2018). Research on Real-Time Local Rainfall Prediction Based on MEMS Sensors. *Journal of Sensors*, *2018*(1), 6184713. https://doi.org/10.1155/2018/6184713

Chattopadhyay, S. (2007). Feed forward Artificial Neural Network model to predict the average summer-monsoon rainfall in India. *Acta Geophysica*, *55*(3), 369–382. https://doi.org/10.2478/s11600-007-0020-8

Chattopadhyay, S., & Chattopadhyay, M. (2007). *A Soft Computing Technique in rainfall forecasting*. files/921/Chattopadhyay y Chattopadhyay - 2007 - A Soft Computing Technique in rainfall forecasting.pdf

Chen, X. (2009). Prediction of Urban Water Demand Based on GA-SVM. *2009 ETP International Conference on Future Computer and Communication (FCC 2009)*, 285–288. https://doi.org/10.1109/FCC.2009.82

Chhetri, M., Kumar, S., Roy, P. P., & Kim, B. G. (2020). Deep BLSTM-GRU model for monthly rainfall prediction: A case study of Simtokha, Bhutan. *Remote Sensing*, *12*(19), 1–13. https://doi.org/10.3390/rs12193174

Coşkun, Ö., & Citakoglu, H. (2023). Prediction of the standardized precipitation index based on the long short-term memory and empirical mode decomposition-extreme learning machine models: The Case of Sakarya, Türkiye. *Physics and Chemistry of the Earth*, *131*. https://doi.org/10.1016/j.pce.2023.103418

Cronin, J., Anandarajah, G., & Dessens, O. (2018). Climate change impacts on the energy system: a review of trends and gaps. *Climatic Change*, *151*(2), 79–93. https://doi.org/10.1007/s10584-018-2265-4

Dotse, S.-Q., Larbi, I., Limantol, A. M., & De Silva, L. C. (2024). A review of the application of hybrid machine learning models to improve rainfall prediction. *Modeling Earth Systems and Environment*, *10*(1), 19–44. https://doi.org/10.1007/s40808-023-01835-x

Duan, Q., Ajami, N. K., Gao, X., & Sorooshian, S. (2007). Multi-model ensemble hydrologic prediction using Bayesian model averaging. *Advances in Water Resources*, *30*(5), 1371–1386. https://doi.org/10.1016/j.advwatres.2006.11.014

Ehteram, M., Afshari Nia, M., Panahi, F., & Shabanian, H. (2024). Gaussian mutation–orca predation algorithm–deep residual shrinkage network (DRSN)–temporal convolutional network (TCN)–random forest model: an advanced machine learning model for predicting monthly rainfall and filtering irrelevant data. *Environmental Sciences Europe*, *36*(1). https://doi.org/10.1186/s12302-024-00841-9

Endalie, D., Haile, G., & Taye, W. (2021). Deep learning model for daily rainfall prediction: case study of Jimma, Ethiopia. *Water Supply*, *22*(3), 3448–3461. https://doi.org/10.2166/ws.2021.391

Esmaeili, F., Shabanlou, S., & Saadat, M. (2021). A wavelet-outlier robust extreme learning machine for rainfall forecasting in Ardabil City, Iran. *Earth Science Informatics*, *14*(4), 2087–2100. https://doi.org/10.1007/s12145-021-00681-8

Farchi, A., Laloyaux, P., Bonavita, M., & Bocquet, M. (2021). Using machine learning to correct model error in data assimilation and forecast applications. *Quarterly Journal of the Royal Meteorological Society*, *147*(739), 3067–3084. https://doi.org/10.1002/qj.4116

Foufoula-Georgiou, E., Guilloteau, C., Nguyen, P., Aghakouchak, A., Hsu, K.-L., Busalacchi, A., Turk, F. J., Peters-Lidard, C., Oki, T., Duan, Q., Krajewski, W., Uijlenhoet, R., Barros, A., Kirstetter, P., Logan, W., Hogue, T., Gupta, H., & Levizzani, V. (2020). Advancing Precipitation Estimation, Prediction, and Impact Studies. *Bulletin of the American Meteorological Society*, *101*(9), E1584–E1592. https://doi.org/10.1175/BAMS-D-20-0014.1

Fredyan, R., & Kusuma, G. (2022). Spatiotemporal convolutional LSTM with attention mechanism for monthly rainfall prediction. *Communications in Mathematical Biology and Neuroscience*. https://doi.org/https://doi.org/10.28919/cmbn/7761

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., & Michaelsen, J. (2015). The climate hazards infrared precipitation with stations - A new environmental record for monitoring extremes. *Scientific Data*, *2*. https://doi.org/10.1038/sdata.2015.66

Ganaie, M. A., Hu, M., Malik, A. K., Tanveer, M., & Suganthan, P. N. (2022). Ensemble deep learning: A review. *Engineering Applications of Artificial Intelligence*, *115*, 105151. https://doi.org/10.1016/j.engappai.2022.105151

Geer, A. J. (2021). Learning earth system models from observations: Machine learning or data assimilation? In *Philosophical Transactions of the Royal Society A: Mathematical, Physical*

and Engineering Sciences (Vol. 379, Issue 2194). Royal Society Publishing. https://doi.org/10.1098/rsta.2020.0089

Ghojogh, B., & Crowley, M. (2019). *The Theory Behind Overfitting, Cross Validation, Regularization, Bagging, and Boosting: Tutorial*. http://arxiv.org/abs/1905.12787

Giannaros, C., Dafis, S., Stefanidis, S., Giannaros, T. M., Koletsis, I., & Oikonomou, C. (2022). Hydrometeorological analysis of a flash flood event in an ungauged Mediterranean watershed under an operational forecasting and monitoring context. *Meteorological Applications*, *29*(4), e2079. https://doi.org/10.1002/met.2079

Gu, J., Liu, S., Zhou, Z., Chalov, S. R., & Zhuang, Q. (2022). A Stacking Ensemble Learning Model for Monthly Rainfall Prediction in the Taihu Basin, China. *Water*, *14*(3), 492. https://doi.org/10.3390/w14030492

Guo, Q., He, Z., & Wang, Z. (2024). Monthly climate prediction using deep convolutional neural network and long short-term memory. *Scientific Reports*, *14*(1). https://doi.org/10.1038/s41598-024-68906-6

Hajirahimi, Z., & Khashei, M. (2019). Hybrid structures in time series modeling and forecasting: A review. *Engineering Applications of Artificial Intelligence*, *86*, 83–106. https://doi.org/10.1016/j.engappai.2019.08.018

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., … Thépaut, J. N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Higuera Martínez, O. I., Fernández-Samacá, L., & Serrano Cárdenas, L. F. (2021). Trends and opportunities by fostering creativity in science and engineering: a systematic review. *European Journal of Engineering Education*, *46*(6), 1117–1140. https://doi.org/10.1080/03043797.2021.1974350

Hou, Y., Deng, X., & Xia, Y. (2024). Precipitation prediction based on variational mode decomposition combined with the crested porcupine optimization algorithm for long short-term memory model. *AIP Advances*, *14*(6). https://doi.org/10.1063/5.0204644

Htike, K. K., & Khalifa, O. O. (2010). Rainfall forecasting models using focused time-delay neural networks. *2010 International Conference on Computer and Communication Engineering (ICCCE)*, 1–6. https://doi.org/10.1109/ICCCE.2010.5556806

Kang, Y., Khan, S., & Ma, X. (2009). Climate change impacts on crop yield, crop water productivity and food security – A review. *Progress in Natural Science*, *19*(12), 1665–1674. https://doi.org/10.1016/j.pnsc.2009.08.001

Katipoğlu, O. M., & Keblouti, M. (2024). Comparative analysis of data-driven models and signal processing techniques in the monthly maximum daily precipitation prediction of El Kerma

station Northeast of Algeria. *Soft Computing*, *28*(17–18), 10751–10765. https://doi.org/10.1007/s00500-024-09860-3

Kim, H.-U., & Bae, T.-S. (2017). Preliminary Study of Deep Learning-based Precipitation Prediction. 한국측량학회지, *35*(5), 423–429. https://www.dbpia.co.kr

Kumar, D., Roshni, T., Singh, A., Himayoun, D., & Samui, P. (2021). A Simplified Approach for Rainfall-runoff Modeling Using Advanced Soft-computing Methods. *JORDAN JOURNAL OF CIVIL ENGINEERING*, *15*(3), 378–392.

Kumar, R., Singh, M. P., Roy, B., & Shahid, A. H. (2021). A Comparative Assessment of Metaheuristic Optimized Extreme Learning Machine and Deep Neural Network in Multi-Step-Ahead Long-term Rainfall Prediction for All-Indian Regions. *Water Resources Management*, *35*(6), 1927–1960. https://doi.org/10.1007/s11269-021-02822-6

Latif, S. D., Mohammed, D. O., & Jaafar, A. (2024). Developing an innovative machine learning model for rainfall prediction in a semi-arid region. *Journal of Hydroinformatics*, *26*(4), 904–914. https://doi.org/10.2166/hydro.2024.014

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Li, L., Carver, R., Lopez-Gomez, I., Sha, F., & Anderson, J. (2024). Generative emulation of weather forecast ensembles with diffusion models. In *Sci. Adv* (Vol. 10). https://www.science.org

Li, Y., Xu, B., Wang, D., Wang, Q. J., Zheng, X., Xu, J., Zhou, F., Huang, H., & Xu, Y. (2021). Deterministic and probabilistic evaluation of raw and post-processing monthly precipitation forecasts: a case study of China. *Journal of Hydroinformatics*, *23*(4), 914–934. https://doi.org/10.2166/hydro.2021.176

Lindsay, G. W., Hasson, U., Norman, K. A., & Baldassano, C. (2022). Deep Learning for Cognitive Neuroscience. In D. Poeppel, G. R. Mangun, & M. S. Gazzaniga (Eds.), *The Cognitive Neurosciences, 6th Edition*. MIT Press. https://direct.mit.edu/books/edited-volume/5456/chapter-abstract/3967020/Deep-Learning-for-Cognitive-Neuroscience?redirectedFrom=PDF

Liu, X., Liu, W., Tang, Q., Liu, B., Wada, Y., & Yang, H. (2022). Global Agricultural Water Scarcity Assessment Incorporating Blue and Green Water Availability Under Future Climate Change. *Earth's Future*, *10*(4), e2021EF002567. https://doi.org/10.1029/2021EF002567

Luo, S., Zhang, M., Nie, Y., Jia, X., Cao, R., Zhu, M., & Li, X. (2022). Forecasting of monthly precipitation based on ensemble empirical mode decomposition and Bayesian model averaging. *Frontiers in Earth Science*, *10*. https://doi.org/10.3389/feart.2022.926067

Maraun, D. (2016). Bias Correcting Climate Change Simulations - a Critical Review. In *Current Climate Change Reports* (Vol. 2, Issue 4, pp. 211–220). Springer. https://doi.org/10.1007/s40641-016-0050-x

Mukherjee, P., Das, A., Bhunia, A. K., & Roy, P. P. (2019). *Cogni-Net: Cognitive Feature Learning through Deep Visual Perception*. arXiv. https://doi.org/10.48550/arXiv.1811.00201

Nasseri, M., Asghari, K., & Abedini, M. J. (2008). Optimized scenario for rainfall forecasting using genetic algorithm coupled with artificial neural network. *Expert Systems with Applications*, *35*(3), 1415–1421. https://doi.org/10.1016/j.eswa.2007.08.033

Niazkar, M., Menapace, A., Brentan, B., Piraei, R., Jimenez, D., Dhawan, P., & Righetti, M. (2024). Applications of XGBoost in water resources engineering: A systematic literature review (Dec 2018–May 2023). *Environmental Modelling & Software*, *174*, 105971. https://doi.org/10.1016/J.ENVSOFT.2024.105971

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Systematic Reviews*, *10*(1). https://doi.org/10.1186/s13643-021-01626-4

Pagliero, L., Bouraoui, F., Diels, J., Willems, P., & McIntyre, N. (2019). Investigating regionalization techniques for large-scale hydrological modelling. *Journal of Hydrology*, *570*, 220–235. https://doi.org/10.1016/j.jhydrol.2018.12.071

Pakdaman, M., Babaeian, I., & Bouwer, L. M. (2022). Improved Monthly and Seasonal Multi-Model Ensemble Precipitation Forecasts in Southwest Asia Using Machine Learning Algorithms. *Water*, *14*(17), 2632. https://doi.org/10.3390/w14172632

Papacharalampous, G., Tyralis, H., Doulamis, N., & Doulamis, A. (2023). Ensemble Learning for Blending Gridded Satellite and Gauge-Measured Precipitation Data. *Remote Sensing*, *15*(20), 4912. https://doi.org/10.3390/rs15204912

Parviz, L. (2020). Comparative Evaluation of Hybrid SARIMA and Machine Learning Techniques Based on Time Varying and Decomposition of Precipitation Time Series. *J. Agr. Sci. Tech*, *22*(2), 563–578.

Parviz, L., Rasouli, K., & Torabi, A. (2021). *Improving Hybrid Models For Precipitation Forecasting By Combining Nonlinear Machine Learning Methods*. https://doi.org/10.21203/RS.3.RS-779973/V1

Parviz, L., Rasouli, K., & Torabi Haghighi, A. (2023). Improving Hybrid Models for Precipitation Forecasting by Combining Nonlinear Machine Learning Methods. *Water Resources Management*, *37*(10), 3833–3855. https://doi.org/10.1007/s11269-023-03528-7

Pham, B. T., Bui, K. T. T., Prakash, I., & Ly, H. B. (2024). Hybrid artificial intelligence models based on adaptive neuro fuzzy inference system and metaheuristic optimization algorithms for prediction of daily rainfall. *Physics and Chemistry of the Earth*, *134*. https://doi.org/10.1016/j.pce.2024.103563

Philip, N. S., Joseph, K. B., Goos, G., Hartmanis, J., & Van Leeuwen, J. (2001). On the Predictability of Rainfall in Kerala - An Application of ABF Neural Network. In V. N. Alexandrov, J. J. Dongarra, B. A. Juliano, R. S. Renner, & C. J. K. Tan (Eds.), *Computational Science - ICCS 2001* (Vol. 2074, pp. 400–408). Springer Berlin Heidelberg. http://link.springer.com/10.1007/3-540-45718-6_44

Priestly, S. E., Raimond, K., Cohen, Y., Brema, J., & Hemanth, D. J. (2023). Evaluation of a novel hybrid lion swarm optimization – AdaBoostRegressor model for forecasting monthly precipitation. *Sustainable Computing: Informatics and Systems*, *39*, 100884. https://doi.org/https://doi.org/10.1016/j.suscom.2023.100884

Rainio, O., Teuho, J., & Klén, R. (2024). Evaluation metrics and statistical tests for machine learning. *Scientific Reports*, *14*(1), 6086. https://doi.org/10.1038/s41598-024-56706-x

Rasp, S., & Lerch, S. (2018). *Neural Networks for Postprocessing Ensemble Weather Forecasts*. https://doi.org/10.1175/MWR-D-18

Ridwan, W. M., Sapitang, M., Aziz, A., Kushiar, K. F., Ahmed, A. N., & El-Shafie, A. (2021). Rainfall forecasting model using machine learning methods: Case study Terengganu, Malaysia. *Ain Shams Engineering Journal*, *12*(2), 1651–1663. https://doi.org/10.1016/j.asej.2020.09.011

Sagi, O., & Rokach, L. (2018). Ensemble learning: A survey. In *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* (Vol. 8, Issue 4). Wiley-Blackwell. https://doi.org/10.1002/widm.1249

Salimi, A. H., Masoompour Samakosh, J., Sharifi, E., Hassanvand, M. R., Noori, A., & Von Rautenkranz, H. (2019). Optimized Artificial Neural Networks-Based Methods for Statistical Downscaling of Gridded Precipitation Data. *Water*, *11*(8), 1653. https://doi.org/10.3390/w11081653

Schefzik, R., Thorarinsdottir, T. L., & Gneiting, T. (2013). Uncertainty quantification in complex simulation models using ensemble copula coupling. *Statistical Science*, *28*(4), 616–640. https://doi.org/10.1214/13-STS443

Sharma, P. K., Shireesha, D. L., Prashantha Kumar, K., Patil, P. A., Shrivastav, A. K., & Kaliappan, S. (2024). Advancing Monthly Rainfall Predictions: Multi-Step Forecasting with Hybrid ACO and GBDT Model. *International Conference on Intelligent Algorithms for Computational Intelligence Systems, IACIS 2024*. https://doi.org/10.1109/IACIS61494.2024.10721823

Shen, Z., & Ban, W. (2023). Machine learning model combined with CEEMDAN algorithm for monthly precipitation prediction. *Earth Science Informatics*. https://doi.org/10.1007/s12145-023-01011-w

Singh, P., & Borah, B. (2013). Indian summer monsoon rainfall prediction using artificial neural network. *Stochastic Environmental Research and Risk Assessment*, *27*(7), 1585–1599. https://doi.org/10.1007/s00477-013-0695-0

Song, C., Chen, X., Wu, P., & Jin, H. (2021). Combining time varying filtering based empirical mode decomposition and machine learning to predict precipitation from nonlinear series. *Journal of Hydrology*, *603*, 126914. https://doi.org/https://doi.org/10.1016/j.jhydrol.2021.126914

Steurer, M., Hill, R. J., & Pfeifer, N. (2021). Metrics for evaluating the performance of machine learning based automated valuation models. *Journal of Property Research*, *38*(2), 99–129. https://doi.org/10.1080/09599916.2020.1858937

Taillardat, M., Mestre, O., Zamo, M., & Naveau, P. (2016). Calibrated ensemble forecasts using quantile regression forests and ensemble model output statistics. *Monthly Weather Review*, *144*(6), 2375–2393. https://doi.org/10.1175/MWR-D-15-0260.1

Tamang, A., & Shukla, S. (2019). Water Demand Prediction Using Support Vector Machine Regression. *2019 International Conference on Data Science and Communication (IconDSC)*, 1–5. https://doi.org/10.1109/IconDSC.2019.8816969

Tang, T., Jiao, D., Chen, T., & Gui, G. (2022). Medium- and Long-Term Precipitation Forecasting Method Based on Data Augmentation and Machine Learning Algorithms. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *15*, 1000–1011. https://doi.org/10.1109/JSTARS.2022.3140442

Tao, L., He, X., Li, J., & Yang, D. (2021). A multiscale long short-term memory model with attention mechanism for improving monthly precipitation prediction. *Journal of Hydrology*, *602*. https://doi.org/10.1016/j.jhydrol.2021.126815

Tao, Y., Gao, X., Ihler, A., Hsu, K., & Sorooshian, S. (2016). Deep neural networks for precipitation estimation from remotely sensed information. *2016 IEEE Congress on Evolutionary Computation (CEC)*, 1349–1355. https://doi.org/10.1109/CEC.2016.7743945

Van Schaeybroeck, B., & Vannitsem, S. (2018). Chapter 10 - Postprocessing of Long-Range Forecasts. In S. Vannitsem, D. S. Wilks, & J. W. Messner (Eds.), *Statistical Postprocessing of Ensemble Forecasts* (pp. 267–290). Elsevier. https://doi.org/https://doi.org/10.1016/B978-0-12-812372-0.00010-8

Vannitsem, S., Bremnes, J. B., Demaeyer, J., Evans, G. R., Flowerdew, J., Hemri, S., Lerch, S., Roberts, N., Theis, S., Atencia, A., Bouallègue, Z. Ben, Bhend, J., Dabernig, M., de Cruz, L., Hieta, L., Mestre, O., Moret, L., Plenković, I. O., Schmeits, M., … Ylhaisi, J. (2021). Statistical postprocessing for weather forecasts review, challenges, and avenues in a big

data world. In *Bulletin of the American Meteorological Society* (Vol. 102, Issue 3, pp. E681–E699). American Meteorological Society. https://doi.org/10.1175/BAMS-D-19-0308.1

Wang, M., Yan, B., Zhang, Y., Zhang, L., Wang, P., Huang, J., Shan, W., Liu, H., Wang, C., & Wen, Y. (2024). Optimizing Precipitation Forecasting and Agricultural Water Resource Allocation Using the Gaussian-Stacked-LSTM Model. *Atmosphere*, *15*(11), 1308. https://doi.org/10.3390/atmos15111308

Wang, Y., Yuan, Z., Liu, H., Xing, Z., Ji, Y., Li, H., Fu, Q., & Mo, C. (2022). A new scheme for probabilistic forecasting with an ensemble model based on CEEMDAN and AM-MCMC and its application in precipitation forecasting. *Expert Systems with Applications*, *187*, 115872. https://doi.org/10.1016/j.eswa.2021.115872

Woźniak, M., Graña, M., & Corchado, E. (2014). A survey of multiple classifier systems as hybrid systems. *Information Fusion*, *16*, 3–17. https://doi.org/10.1016/j.inffus.2013.04.006

Wu, C. L., & Chau, K. W. (2013). Prediction of rainfall time series using modular soft computing methods. *Engineering Applications of Artificial Intelligence*, *26*(3), 997–1007. https://doi.org/10.1016/j.engappai.2012.05.023

Wu, C. L., Chau, K. W., & Fan, C. (2010). Prediction of rainfall time series using modular artificial neural networks coupled with data-preprocessing techniques. *Journal of Hydrology*, *389*(1–2), 146–167. https://doi.org/10.1016/j.jhydrol.2010.05.040

Wu, X., Zhou, J., Yu, H., Liu, D., Xie, K., Chen, Y., Hu, J., Sun, H., & Xing, F. (2021). The development of a hybrid wavelet-arima-lstm model for precipitation amounts and drought analysis. *Atmosphere*, *12*(1). https://doi.org/10.3390/ATMOS12010074

Yaseen, Z. M., Sulaiman, S. O., Deo, R. C., & Chau, K. W. (2019). An enhanced extreme learning machine model for river flow forecasting: State-of-the-art, practical applications in water resource engineering area and future research direction. In *Journal of Hydrology* (Vol. 569, pp. 387–408). Elsevier B.V. https://doi.org/10.1016/j.jhydrol.2018.11.069

Yeditha, P. K., Anusha, G. S., Nandikanti, S. S. S., & Rathinasamy, M. (2023). Development of Monthly Scale Precipitation-Forecasting Model for Indian Subcontinent using Wavelet-Based Deep Learning Approach. *Water (Switzerland)*, *15*(18). https://doi.org/10.3390/w15183244

Zandi, O., Zahraie, B., Nasseri, M., & Behrangi, A. (2022). Stacking machine learning models versus a locally weighted linear model to generate high-resolution monthly precipitation over a topographically complex area. *Atmospheric Research*, *272*, 106159. https://doi.org/https://doi.org/10.1016/j.atmosres.2022.106159

Zerouali, B., Santos, C. A. G., Farias, C. A. S. de, Muniz, R. S., Difi, S., Abda, Z., Chettih, M., Heddam, S., Anwar, S. A., & Elbeltagi, A. (2023). Artificial intelligent systems optimized by metaheuristic algorithms and teleconnection indices for rainfall modeling: The case of a

humid region in the mediterranean basin. *Heliyon*, *9*(4), e15355. https://doi.org/https://doi.org/10.1016/j.heliyon.2023.e15355

Zhang, H., Loáiciga, H. A., Ren, F., Du, Q., & Ha, D. (2020). Semi-empirical prediction method for monthly precipitation prediction based on environmental factors and comparison with stochastic and machine learning models. *Hydrological Sciences Journal*, 1928–1942. https://doi.org/10.1080/02626667.2020.1784901

Zhang, L., Wei, Y., & Wang, Z. (2008). Prediction on Ecological Water Demand Based on Support Vector Machine. *2008 International Conference on Computer Science and Software Engineering*, 1032–1035. https://doi.org/10.1109/CSSE.2008.442

Zhang, W., Zhao, S., Hu, X., Dong, Q., Huang, H., Zhang, S., Zhao, Y., Dai, H., Ge, F., Guo, L., & Liu, T. (2020). Hierarchical Organization of Functional Brain Networks Revealed by Hybrid Spatiotemporal Deep Learning. *Brain Connectivity*, *10*(2), 72–82. https://doi.org/10.1089/brain.2019.0701

Zhang, X., Wang, K., & Zheng, Z. (2022). A novel integrated learning model for rainfall prediction CEEMD-FCMSE -Stacking. *EARTH SCIENCE INFORMATICS*, *15*(3), 1995–2005. https://doi.org/10.1007/s12145-022-00819-2

Zhang, X., & Wu, X. (2023). Combined Forecasting Model of Precipitation Based on the CEEMD-ELM-FFOA Coupling Model. *Water*, *15*(8), 1485. https://doi.org/10.3390/w15081485

Zhang, X., Yao, L., Wang, X., Monaghan, J., Mcalpine, D., & Zhang, Y. (2021). A survey on deep learning-based non-invasive brain signals: recent advances and new frontiers. In *Journal of Neural Engineering* (Vol. 18, Issue 3). IOP Publishing Ltd. https://doi.org/10.1088/1741-2552/abc902

Zhou, Z., Ren, J., He, X., & Liu, S. (2021). A comparative study of extensive machine learning models for predicting long-term monthly rainfall with an ensemble of climatic and meteorological predictors. *Hydrological Processes*, *35*(11). https://doi.org/10.1002/hyp.14424

Zhou, Z.-H. (2012). *Ensemble Methods: Foundations and Algorithms* (1st ed.). Chapman and Hall/CRC. https://doi.org/10.1201/b12207

Zounemat-Kermani, M., Batelaan, O., Fadaee, M., & Hinkelmann, R. (2021). Ensemble machine learning paradigms in hydrology: A review. In *Journal of Hydrology* (Vol. 598). Elsevier B.V. https://doi.org/10.1016/j.jhydrol.2021.126266