sustainability

MDPI

Article

A Digital Twin Framework Integrating Multiscale Signal Decomposition and Hybrid Deep Learning for Intelligent Water Storage Control on Remote Islands

Xubin Wu, Tao Chen, Yurun Chen, Guoqiang Wu, Qiliang Yang, Xuewei Zhang and Liqiang Xie *

College of Defense Engineering, Army Engineering University of People's Liberation Army, Nanjing 210007, China; xubinwu1002@163.com (X.W.); ct2268662256@163.com (T.C.); 15251757992@163.com (Y.C.); wuguoqiang0909@163.com (G.W.); yql@893.com.cn (Q.Y.); zxwlucky@163.com (X.Z.)

* Correspondence: xielq@aeu.edu.cn

Abstract

Precipitation on remote offshore islands is characterized by non-stationarity and strong stochasticity. However, rainwater harvesting systems require accurate precipitation forecasts to ensure effective water storage and discharge and a stable freshwater supply. Here, we propose an intelligent framework integrating precipitation signal decomposition, a hybrid neural network for predicting daily precipitation over the next 30 days, and digital twin technology to optimize water storage and discharge strategies for rainwater collection pools. For signal preprocessing, variational mode decomposition decomposes the original daily precipitation series into multiple intrinsic mode functions, a chaotic sparrow search optimization algorithm performs global parameter optimization, and singular spectrum analysis is applied twice to denoise high-frequency mode components. The hybrid prediction model combines a convolutional neural network to extract local temporal features, bidirectional long short-term memory to capture bidirectional long-term dependencies, and a multilayer perceptron for nonlinear mapping and regression output. Finally, this model is integrated into a digital-twin-enabled decision framework to automatically implement water storage control through water balance calculations and threshold comparisons. The proposed method achieves a test mean absolute error of 2.49 mm, a root mean square error of 5.72 mm, and an R^2 of 0.83; digital twin control significantly reduces overflow losses and, during typical typhoon months, maintains the water level in the rainwater collection pool below the warning threshold. In summary, we present a simulation-based pathway for optimizing small-scale intelligent rainwater harvesting systems, which offers a practical pathway to enhance the sustainability of freshwater resources on remote offshore islands by improving the efficiency of rainwater utilization and reducing dependence on external water supplies.

Keywords: precipitation forecasting; variational mode decomposition; singular spectrum analysis; CNN–BiLSTM–MLP; digital twin; rainwater harvesting; remote offshore island; deep learning; water resource sustainability

Check for updates

Academic Editors: Sergio Zubelzu and Dariusz Młyński

Received: 16 July 2026
Revised: 16 October 2026
Accepted: 3 November 2026
Published: 5 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Remote offshore islands are highly vulnerable to long-term water shortages owing to their geographical isolation, weak surface runoff, and limited natural freshwater reserves [1]. Sansha City in the South China Sea is a typical remote offshore island region characterized

by a lack of stable rivers, lakes, and underground freshwater aquifers. Freshwater resources are primarily maintained through two artificial replenishment methods: rainwater harvesting and seawater desalination. From an engineering perspective, seawater desalination ensures a relatively stable water supply but entails high energy consumption and equipment maintenance costs. Additionally, brine discharge can disturb the local marine ecology [2]. Therefore, for offshore stations, island reefs, and small-scale residential water-use scenarios, seawater desalination is often used only for emergency replenishment or maintaining backup supplies during drought periods [3]. In contrast, low-energy, sustainable, and simple rainwater harvesting systems have become integral to ensuring a reliable daily water supply [4].

However, the performance of rainwater harvesting systems depends on effective storage allocation in rainwater collection pools and appropriate operational strategies. On remote offshore islands, such systems must retain as much usable water as possible during dry periods while maintaining sufficient reserve capacity before typhoons and heavy rainfall events to reduce overflow losses and engineering risks. Thus, system managers must forecast future precipitation conditions and monthly rainfall trends to determine appropriate storage, release, maintenance, and warning strategies [5]. If managers rely solely on annual average precipitation or simple empirical rules for scheduling, they cannot effectively address common precipitation anomalies, such as droughts and offshore typhoons; this limitation can lead to insufficient storage or overflow losses, reducing freshwater resource utilization efficiency and threatening water security [6].

Precipitation forecasting is a key interdisciplinary field linking hydrological and meteorological research. Short-term or near-term precipitation forecasting on hourly to daily scales is important for weather process identification, flood early warning, and agricultural scheduling [7]. Conversely, statistical climate prediction on monthly, seasonal, and annual scales typically addresses large-scale atmospheric circulation and anomalous climate responses [8]. However, cross-scale precipitation forecasting remains limited, and the resulting predictions are seldom directly incorporated into water system regulation. Thus, a coupling framework that integrates forecasting and decision-making for engineering applications is currently lacking.

Precipitation sequences on remote offshore islands exhibit typical non-stationarity and strong randomness, with clear transitions between dry and rainy seasons, significant intraseasonal fluctuations, abundant high-frequency disturbances, and extreme pulse events due to tropical weather systems, land–sea–air interactions, and typhoons [9]. Precipitation signals comprise a mixture of trend components, low-frequency seasonal oscillations, high-frequency weather noise, and sparsely distributed extreme values [10]. Traditional deep learning models have difficulty simultaneously addressing multiscale feature extraction, noise suppression, and nonlinear mapping. Thus, for regions with limited samples and single monitoring stations, directly modeling the original daily precipitation sequence may introduce noise contamination and interference from extreme values, reducing the generalizability of prediction models.

To address these challenges, we develop a hybrid precipitation forecasting model tailored to optimize rainwater harvesting on remote offshore islands. The proposed framework integrates three aspects. First, we perform signal decomposition using variational mode decomposition (VMD) to split the original precipitation sequence into multiscale intrinsic mode function (IMF) components, followed by singular spectrum analysis (SSA) for secondary noise reduction in high-frequency modes. Compared with single deep learning models, VMD and SSA denoising improve the interpretability of complex non-stationary precipitation signals. Second, we combine a multilayer perceptron (MLP) with a convolutional neural network and bidirectional long short-term memory network (CNN–BiLSTM)

to construct a hybrid deep learning architecture for predicting low- and high-frequency components, respectively. The convolutional neural network–bidirectional long short-term memory–multilayer perceptron (CNN–BiLSTM–MLP) hybrid model balances model capacity with signal complexity to reduce the risk of overfitting under limited-sample conditions. Third, the forecasting results are embedded within a digital twin decision framework to drive intelligent water storage control and improve the efficiency of rainwater harvesting systems.

The remainder of this paper is organized as follows. Section 2 reviews current research on multisource collaborative water supply systems, precipitation forecasting methods, and digital twins. Section 3 details the data, preprocessing workflow, proposed model architecture, and integration methods for digital twin control. Section 4 introduces the experimental setup and presents and discusses the results. Finally, Section 5 summarizes the study findings and discusses potential limitations and directions for future research.

2. Literature Review

Globally, freshwater scarcity has become a critical issue constraining socioeconomic development and ecological security. Owing to population growth, urbanization, climate change, and frequent extreme hydrological events, traditional single-source water supply models cannot meet the demand for stable, secure, and economical water supplies [11,12]. Efforts to optimize water resource systems focus on aspects such as resource development, demand forecasting, operational scheduling, and intelligent control. Recently, statistical analysis, machine learning, and deep learning methods have been introduced to enhance the forecasting and decision-making capabilities of water systems. However, remote offshore islands face severe freshwater security challenges owing to their geographical isolation, limited natural water storage capacity, inadequate infrastructure, and difficulties in obtaining external water supplies [13]; thus, conventional terrestrial water supply frameworks are unsuitable.

For remote offshore islands, freshwater supply typically relies on two approaches: rainwater harvesting and seawater desalination. Seawater desalination is characterized by high supply stability and minimal influence from short-term climate variability, making it a key means of ensuring water supply for many islands [14]. Following advances in reverse osmosis technology, seawater desalination systems have been widely applied in island regions. Although system reliability has been improved by optimizing energy consumption, controlling membrane fouling, and reducing water production costs [15–17], seawater desalination remains an inherently energy- and maintenance-intensive method of ensuring water supply. Rigid water production strategies implemented before and after the rainy season can easily lead to energy waste or inefficient allocation of water storage capacity. Additionally, excessive reliance on desalination increases dependence on electrical infrastructure and specialized maintenance capabilities, weakening the overall economy and resilience of the water supply system.

Compared with seawater desalination, rainwater harvesting enables on-site resource acquisition and low water extraction costs, making it an important component of island water supply systems [18]. Combining rainwater harvesting systems with urban water system technologies can enhance rainwater utilization rates and partly alleviate water supply pressures during dry seasons [19,20]. Water balance models have confirmed the practical value of rainwater harvesting systems in alleviating urban water resource shortages [21]. However, rainwater collection pool scheduling models must simultaneously consider physical boundary conditions, such as overflow risk thresholds, minimum guaranteed water levels, and evaporation and seepage losses, which place greater demands on the lead time and reliability of prediction models. Traditional methods are generally based on long-term

average rainfall conditions or statistically "typical" years and therefore cannot reflect the strong seasonal oscillations and interannual variability in precipitation on offshore islands. Consequently, their operational strategies often lack sufficient adaptability during extreme drought periods or unexpected heavy rainfall events.

Some researchers have implemented multisource collaboration and integrated design in water systems. Marwein et al. achieved complementary water production from multiple sources under alternating sunny and rainy conditions by integrating rainwater collection troughs and photovoltaic-driven wind cooling devices into traditional solar distillers [22]. Papathanasiou et al. constructed a wind-powered desalination–pumped storage system for Mykonos Island, Greece, and verified the technical and economic feasibility of achieving almost 100% water security on seasonal tourist islands by comparing different excess electricity allocation strategies [23]. These studies emphasize that multisource collaboration is essential for improving water supply security and controlling water production costs. However, most studies rely on offline optimization or preset operational rules, which limit their responsiveness to abrupt weather changes and short- to medium-term rainfall uncertainties [24], thereby failing to support real-time decision-making during actual operation.

To address the shortcomings of traditional forecasting and scheduling methods, machine learning and deep learning techniques, such as support vector machines, random forests, XGBoost (eXtreme Gradient Boosting), artificial neural networks, and other deep learning models, have been applied to water resource forecasting and control [25]. Zhao et al. used random forest regression models to significantly reduce the false alarm rate in storm-level precipitation forecasting [26]. Sun et al. utilized various CNN configurations to statistically downscale daily precipitation and temperature in China, thereby improving the accuracy of monthly precipitation forecasts [27]. In terms of radar-based near-term forecasting, Kim et al. developed a radar rainfall forecasting method with lead times ranging from 10 min to 4 h based on conditional generative adversarial networks (CGANs) [28]. Ko et al. proposed a pretraining scheme and a novel loss function for the U-Net model, achieving an improvement of up to 95.7% in the critical success index for 5-h heavy rainfall forecasting and reducing weak rainfall estimation errors by 10.7% [29]. Additionally, Rojas-Campos et al. used complete variable fields from numerical weather prediction models as inputs to train deep learning models such as U-Net, deconvolutional networks, and CGANs, generating high-resolution short-term precipitation prediction maps while effectively correcting overforecasting biases in the original inputs [30].

A more in-depth understanding of the non-stationarity of hydrological time series has revealed feature aliasing in methods that rely solely on raw sequence inputs to neural networks. To address this issue, researchers have begun to couple signal decomposition techniques with deep learning. Farman et al. constructed VMD–LSTM and VMD–GRU (variational mode decomposition–gated recurrent unit) hybrid frameworks, demonstrating that deep learning models enhanced with VMD exhibit greater consistency in temporal learning and robustness to noise [31]. Wang et al. proposed an SSA–BiTCN–Self-Attention (singular spectrum analysis–bidirectional temporal convolutional network–self-attention) model, which reduced the root mean square error (RMSE) and mean absolute error (MAE) by approximately 58% and 64%, respectively, compared with a standalone TCN model [32]. Johny et al. introduced an adaptive EEMD–ANN (ensemble empirical mode decomposition–artificial neural network) hybrid framework for forecasting Indian summer monsoon precipitation, which significantly outperformed EEMD–ANN control experiments in capturing extreme hydrological events [33]. Mir et al. used EMD (empirical mode decomposition) to decompose annual precipitation sequences from 61 meteorological stations in Iran and identified decadal and multidecadal cycles relevant to regional drought early warning [34]. However, high-frequency modes are often directly input into complex networks without

secondary denoising, allowing residual noise components to adversely affect prediction accuracy. In addition, most existing coupled models adopt a unified network architecture to process all components and fail to design heterogeneous networks based on the intrinsic differences between low- and high-frequency signals.

Intelligent control and real-time scheduling are also important aspects of recent water resource systems research. Ulusoy et al. achieved water supply recovery and pumping cost control under pipeline burst conditions using adaptive model predictive control (MPC) [35]. Liu et al. used distributed regional MPC to satisfy pressure constraints in large-scale pipeline networks and achieve coordinated convergence among subsystems [36]. Velarde et al. implemented coordinated scheduling of multiple systems driven by runoff forecasting using hierarchical distributed MPC [37]. However, these approaches often require accurate acquisition of system states and future input information, limiting their adaptability to complex and uncertain environments.

Digital twins offer new approaches to integrating perception, prediction, and decision-making into water resource systems [38]. Digital twins integrate physical entities, sensor networks, data models, and virtual spaces, enabling continuous mapping, real-time monitoring, and dynamic inference of the actual system state. Digital twins have been applied to urban water management [39], water supply pipeline networks [40], and watershed flood management [41], demonstrating significant advantages in anomaly diagnosis, operational optimization, and risk warning. Unlike static modeling approaches, digital twins can continuously adjust model parameters using real-time observational data [42]. This capability is particularly important for remote offshore islands that are strongly affected by meteorological disturbances. However, existing digital twin research primarily focuses on large-scale urban infrastructure, with limited research on multisource water supply systems for remote islands.

Although recent advances in digital twin frameworks for water systems have combined sensors and visualization platforms to achieve monitoring and warning capabilities, most applications remain at the level of operational visualization. In current digital twin systems, precipitation prediction modules are often driven directly by raw observation sequences and fail to fully account for the inherent nonlinear and non-stationary dynamics of precipitation on offshore islands. Moreover, insufficient preprocessing, such as signal decomposition and denoising, reduces prediction accuracy, especially in remote island water supply scenarios. This limitation hampers the ability to translate short- and medium-term precipitation forecasts into actionable operational decisions, significantly undermining the value of digital twins. Therefore, simply constructing visualization- or monitoring-oriented twin platforms will not fundamentally alleviate the excessive reliance of offshore island freshwater systems on energy-intensive supplementary water sources.

3. Methods

Figure 1 illustrates the complete precipitation forecasting process using the proposed framework. First, missing values in the collected precipitation data were interpolated, and outliers were removed. Subsequently, VMD was used to decompose the data, and the decomposition parameters were optimized using a chaotic sparrow search optimization algorithm (CSSOA). Then, SSA was used for secondary denoising of high-frequency IMF components. Subsequently, CNN–BiLSTM was employed to predict high-frequency components, whereas MLP was used to predict low-frequency components. Finally, the prediction results for the high- and low-frequency components were linearly superimposed to reconstruct the complete precipitation predictions, which were then fed into the digital twin framework.

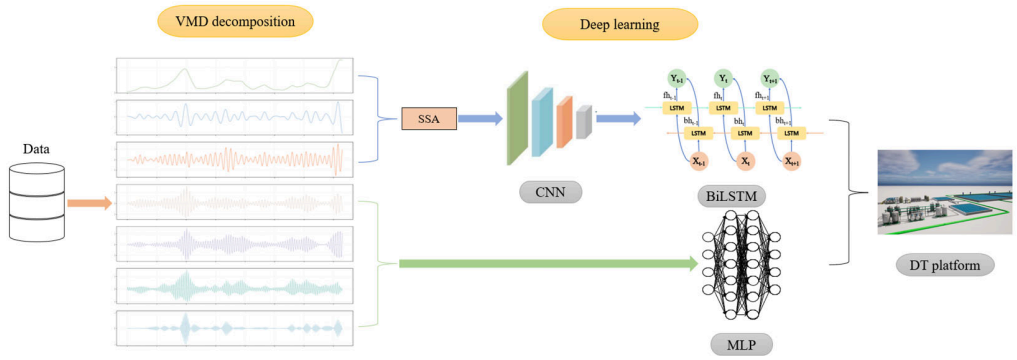


Figure 1. Schematic of the proposed precipitation forecasting workflow: signal decomposition via variational mode decomposition (VMD) and singular spectrum analysis (SSA); a hybrid deep learning convolutional neural network–bidirectional long short-term memory–multilayer perceptron (CNN–BiLSTM–MLP) architecture for 30-day precipitation prediction; and integration within a digital twin (DT) framework. (Source: Original figure by the authors).

3.1. Study Area, Data Collection, and Data Preprocessing

We used daily observational precipitation records from 2000 to 2019 collected at a meteorological station in Sansha City, specifically the NANSHA DAO station (10.38° N, 114.37° E). The original data were sourced from the public database of the United States National Oceanic and Atmospheric Administration. To ensure the reliability of the forecasting results, systematic corrections were made to the original observational data prior to modeling to address data incompleteness. Missing values caused by incomplete records or transmission failures were interpolated using synchronous observational data from four nearby meteorological stations surrounding the target site, following the method described in [43]. These reference stations are, from west to east, HUYEN TRAN (8.02° N, 110.62° E), TRUONG SA (8.65° N, 111.92° E), YONGSHUJIAO (9.53° N, 112.88° E), and SONG TU TAY SOUTH WEST CAY (11.42° N, 114.33° E). This approach minimizes the impact of missing data on subsequent modeling and forecasting accuracy. The cleaned dataset included the time and precipitation amount, as summarized in Figure 2.

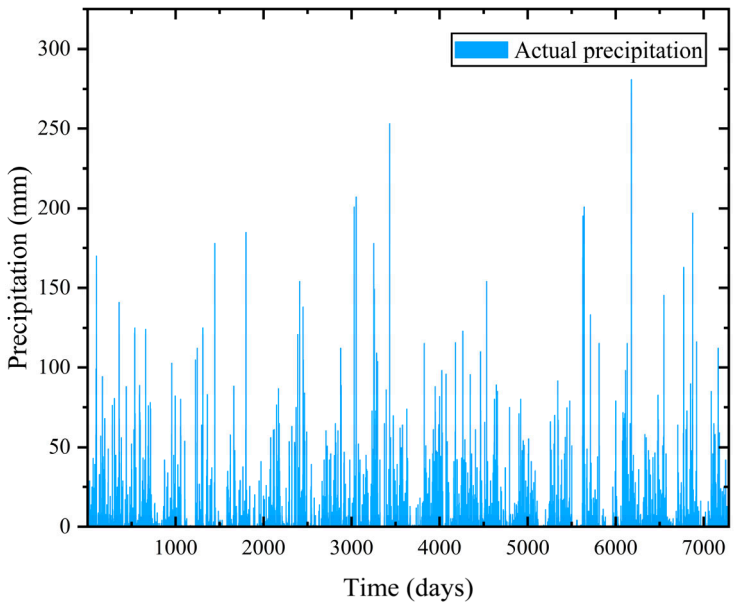


Figure 2. Observed daily precipitation observations at the target site from 2000 to 2019. (Source: Original figure by the authors).

It should be emphasized that interpolation using data from nearby stations was strictly confined to the historical data preprocessing stage and is not a prerequisite for the online application of the forecasting system. For actual deployment on remote islands, the training data should be derived from locally measured precipitation records collected by in situ observation stations or real-time monitoring networks, without relying on estimated or interpolated values. In our historical dataset, the annual rate of missing data remained below 5%, rendering the effect of interpolation on the overall modeling results negligible.

3.2. Normalization and Data Division

To improve the stability of neural network training and accelerate convergence, min–max normalization was applied to both the input features and forecasting targets, linearly scaling the original values to the range $[0, 1]$. For any variable z, the normalized form z' was defined as

$$z' = \frac{z - z_{\min}}{z_{\max} - z_{\min}} \quad (1)$$

where $z_{\min}$ and $z_{\max}$ represent the minimum and maximum values of the variable within the reference dataset, respectively.

Following the chronological order of the observations, the dataset was divided without random shuffling to preserve the temporal structure of the original sequences. Data from 2000 to 2013 were used as the training set, data from 2014 to 2015 as the validation set for early stopping, and data from 2016 to 2019 were used as the test set, accounting for approximately 70%, 10%, and 20% of the total data, respectively. For time-series tasks, this strategy more effectively reflects the temporal generalization capability of the model than random splitting.

It should be emphasized that all data-dependent preprocessing parameters and any optimization procedures that require fitting, including normalization statistics, hyperparameter optimization, and the configuration of signal decomposition, were determined using only the training set. The validation set was used only for early stopping and was not involved in fitting or optimizing these parameters or configurations. The resulting configurations were then fixed and applied unchanged to the validation and test sets. This measure ensured that information from future observations did not leak into the model development stage. Throughout the entire process, a walk-forward rolling procedure was implemented: for each prediction time t, only the preceding 90-day observation window (from t − 89 to t) was extracted. The windowed data were directly filtered through the pre-configured VMD decomposer, whose center frequencies and mode shapes were frozen from the training set to obtain the corresponding IMF components. This guarantees that the spectral bands of the decomposed components remain identical across all rolling test windows.

3.3. Sequence Decomposition and Denoising

3.3.1. VMD

The original daily precipitation sequence is influenced by seasonal fluctuations between the dry and rainy seasons, as well as extreme precipitation events caused by typhoons. Directly feeding this sequence into a neural network would likely result in noise dominating the model learning process. Therefore, we used VMD [44] to perform multiscale decomposition of the original sequence by dividing the data into multiple IMFs. These IMF components differ in terms of central frequency and bandwidth, enabling a finer representation of the signal structure and characteristics. Compared with empirical mode

decomposition, VMD exhibits greater robustness to anomalies and noise. The complete VMD optimization process is expressed as

$$\min_{u_k,w_k}\left\{\sum_{k=1}^{K}\left\|\partial_t[(\delta(t)+\frac{j}{\pi t})*u_k(t)]e^{-jw_kt}\right\|_2^2\right\} \tag{2}$$

$$s.t.\sum_{k=1}^{K}u_k(t)=f(t) \tag{3}$$

where $f(t)$ represents the original input signal; $u_k=\{u_1,u_2,...,u_k\}$ indicates the K IMFs decomposed from $f(t)$; w_k represents the central frequency of the Kth IMF; ∂_t is the derivative with respect to t; and * denotes the convolution operation.

To obtain the optimal solution, the Lagrange multiplier, λ, and a quadratic penalty term, α, were introduced to convert the above constrained optimization problem into an unconstrained optimization problem:

$$\begin{aligned}L(\{u_k\},\{w_k\},\lambda)&=\alpha\sum_{k=1}^{K}\left\|\partial_t[(\delta(t)+\frac{j}{\pi t})*u_k(t)]e^{-jw_kt}\right\|_2^2\\&+\left\|f(t)-\sum_{k=1}^{K}u_k(t)\right\|+\left\langle\lambda(t),f(t)-\sum_{k=1}^{K}u_k(t)\right\rangle\end{aligned} \tag{4}$$

Here, $\delta(t)$ represents the impulse function. The components and their corresponding central frequencies can be optimized using the alternating direction method of multipliers, with the update formulas given as follows:

$$\widehat{u}_k^{(n+1)}(\omega)=\frac{\widehat{f}(\omega)-\sum_{i\neq k}\widehat{u}_k^{(n)}(\omega)+\frac{\widehat{\lambda}^{(n)}(\omega)}{2}}{1+2\alpha(\omega-\omega_k^{(n)})^2} \tag{5}$$

$$\omega_k^{(n+1)}=\frac{\int_0^\infty\omega\left|\widehat{u}_k^{(n+1)}(\omega)\right|^2d\omega}{\int_0^\infty\left|\widehat{u}_k^{(n+1)}(\omega)\right|^2d\omega} \tag{6}$$

where $\hat{f}(\omega)$, $\hat{u}_k(\omega)$, and $\hat{\lambda}_k(\omega)$ denote the Fourier transforms of $f(t)$, $u(t)$, and $\lambda(t)$, respectively; ω represents frequency; and n is the iteration count.

Through this decomposition, VMD converts the original non-stationary signal into components with specific frequency characteristics, significantly reducing the complexity of the original signal. This decomposition process enhances the identifiability of the input sequence and provides more structured data for subsequent deep learning modeling. This enables neural networks to extract effective features and address forecasting challenges resulting from the inherent non-stationarity of daily precipitation time series.

The decomposition effectiveness of VMD is significantly influenced by the number of modes, K, and the penalty factor, α. We adopted CSSOA [45] for global optimization of these parameters; this algorithm utilizes chaotic mapping for population initialization, discoverer–joiner collaborative updates, and stagnation disturbance mechanisms, thereby balancing global search capability with convergence speed. Specifically, K and α were treated as optimization variables with search ranges $K\in[4,10]$ and $\alpha\in[500,5000]$. The fitness function was defined as the combination of the average envelope entropy of the IMFs, the relative reconstruction error, and a light K-complexity penalty and was minimized. The population size was set to 6, the maximum number of iterations to 5, and the random seed to 42. In each iteration, candidate solutions were sorted by fitness; the top 20% acted as discoverers, while the remaining individuals were updated through a discoverer–joiner collaborative strategy. In addition, 20% of individuals were randomly selected as scouts and reinitialized by Logistic chaotic mapping to implement stagnation disturbance and avoid local optima. The internal VMD maximum iteration number was 80 and the convergence

tolerance was 10^{-6}. After 5 iterations, the optimal parameter configuration was determined to be $K = 7$ and $\alpha = 2000$. The resulting IMF components are shown in Figure 3. Each IMF component contains distinct frequency information, and the submodes generated after decomposition exhibit prominent frequency-domain characteristics. This demonstrates the effectiveness of VMD, which significantly reduces the complexity of internal feature extraction for forecasting daily precipitation over the subsequent 30 days, thereby laying the groundwork for improved forecasting accuracy.

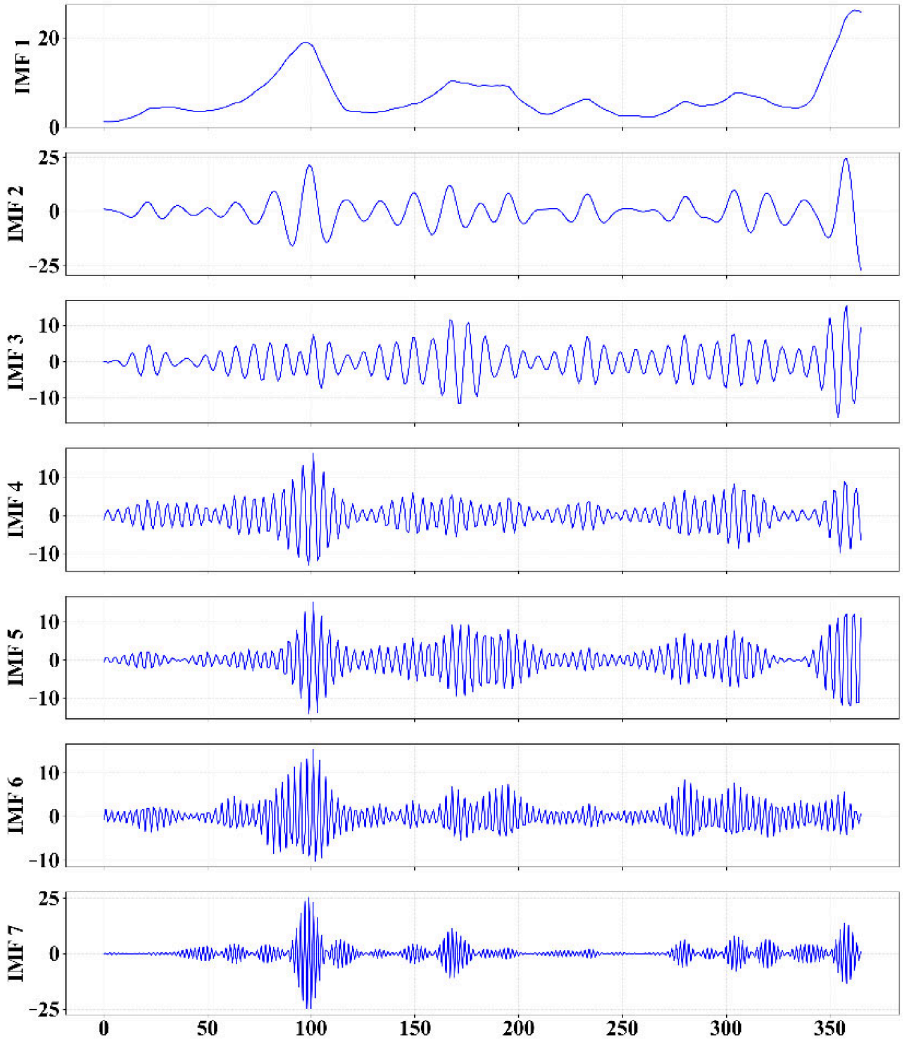


Figure 3. VMD results (using data from 2000 as an example). (Source: Original figure by the authors).

3.3.2. Singular Spectrum Analysis

Although VMD initially decomposes the original precipitation sequence into different frequency bands, the high-frequency IMF components still contain complex fluctuations caused by the stochasticity of typhoon processes and observational errors. If directly fed into deep learning models, these components introduce noise into the training process, weakening the model's ability to learn dominant seasonal signals. To suppress redundant noise while preserving meaningful high-frequency information, SSA was introduced to perform secondary decomposition and reconstruction of selected high-frequency IMF components. The SSA procedure included four steps: embedding, singular value decomposition, grouping, and diagonal averaging.

(1) For embedding, we constructed the trajectory matrix T for the IMF1–IMF3 sequences, where N is the sequence length, L is the window length, and $K = N - L + 1$:

$$T = \begin{bmatrix} x_1 & x_2 & \cdots & x_K \\ x_2 & x_3 & \cdots & x_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & \cdots & x_N \end{bmatrix} \quad (7)$$

(2) We performed singular value decomposition of the trajectory matrix T:

$$T = \sum_{i=1}^{r} \sigma_i u_i v_i^T \quad (8)$$

where r is the rank of matrix T; σ_i represents the ith singular value, sorted in descending order ($\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$); u_i is the ith left singular vector of length L; and v_i is the ith right singular vector of length K.

(3) For principal component selection, we used the cumulative variance contribution ratio to distinguish signal from noise. To determine the 85% threshold, the variance contribution ratio of the ii-th singular value was first calculated as $r_i = \sigma_i^2 / \sum_{j=1}^{d} \sigma_j^2$, where d is the total number of singular values. The cumulative variance contribution ratio was then computed as $C_m = \sum_{i=1}^{m} \sigma_i^2 / \sum_{j=1}^{d} \sigma_j^2$ and plotted against the number of retained singular values m. The cumulative contribution curve was examined; it became relatively flat when C_m reached approximately 85%, indicating that the remaining components mainly represented noise. Accordingly, a threshold of 85% was applied by selecting the first m singular values for which the cumulative sum of squares reached or exceeded 85% of the total sum of squared singular values:

$$\frac{\sum_{i=1}^{m} \sigma_i^2}{\sum_{i=1}^{r} \sigma_i^2} \geq 0.85 \quad (9)$$

Here, the numerator represents the total energy of the first m principal components, and the denominator represents the total energy of all components.

(4) For sequence reconstruction, we retained the singular vectors corresponding to the first m singular values and then reconstructed a new trajectory matrix via truncated singular value decomposition:

$$T_{denoised} = \sum_{i=1}^{m} \sigma_i u_i v_i^T \quad (10)$$

3.4. Prediction Model Architecture

Considering the intrinsic differences among the components in different frequency bands decomposed by VMD, we adopted a heterogeneous prediction strategy: low-frequency components were fitted using an MLP to avoid overfitting associated with complex networks under limited-sample conditions, whereas high-frequency components were predicted using the CNN–BiLSTM model after SSA denoising, with the CNN extracting short-term local patterns and the BiLSTM capturing bidirectional temporal dependencies.

3.4.1. MLP Layer

An MLP is a feedforward neural network that achieves input-to-output mapping through multilayer nonlinear transformations comprising input, hidden, and output layers. As the low-frequency components exhibit substantial smoothness and long-term trend characteristics over time, an MLP is advantageous for two reasons. First, the simple structure and relatively few parameters of an MLP help prevent overfitting under limited-sample conditions, allowing stable capture of deterministic trends in low-frequency components.

Second, an MLP uses fully connected layers to comprehensively capture global input information, avoiding the gradient decay problem associated with recurrent structures in long sequences and improving the efficiency of fitting slowly varying signals.

The MLP training process followed a standard error backpropagation mechanism. In this study, we used the rectified linear unit activation function for the hidden layers and linear activation for the output layer to accommodate the regression task of precipitation prediction. During forward propagation, the input data sequentially passed through linear transformations and nonlinear activations in the hidden layers, generating predictions at the output layer. Subsequently, mean squared error was used as the loss function to quantify the deviation between the predictions and the true values. During backpropagation, the output-layer error was propagated backward through the network, and the gradients of the loss function with respect to the weights and biases were calculated using the chain rule. The Adam optimizer was used to update the network parameters to minimize the loss function. The local gradient of the output layer, δ_j, was calculated as follows:

$$\delta_j = e_j \cdot \sigma'(v_j) \tag{11}$$

where e_j is the error signal of the jth output neuron; v_j is the weighted sum of the inputs to the jth neuron in this layer; and $\sigma'(\cdot)$ represents the derivative of the activation function.

3.4.2. CNN Layer

A CNN is a typical feedforward neural network comprising convolutional, pooling, and fully connected layers. The convolutional layer performs local feature extraction from the input data using convolution kernels to identify latent features. The pooling layer downsamples the convolutional output and performs secondary feature extraction to accelerate network convergence while retaining key information. The fully connected layer integrates the outputs of the convolutional and pooling layers and produces the final predictions. The formulas for the convolutional and fully connected layers are as follows:

$$x_j^l = \sigma\left(\sum_{i \in M_j} x_i^{l-1} \times w_{ij}^l + b_j^l\right) \tag{12}$$

$$x^{(l+1)(j)} = \sum_{i=1}^{n} w_{ij}^l x_j^l + b_j^l \tag{13}$$

where x_j^l is the input to the jth convolution kernel in layer l; $x^{(l+1)(j)}$ is the output value of the jth neuron in layer $l+1$; σ is the activation function; Mj is the set of input feature maps; w_{ij}^l is the weight matrix of the jth convolution kernel in layer l; and b_j^l is the bias term.

By sliding convolution kernels of different sizes across the time dimension, the CNN effectively captures local correlations between neighboring time steps and short-term fluctuation patterns, providing more abstract and information-dense representations for subsequent time-series modeling.

3.4.3. BiLSTM Layer

The BiLSTM simultaneously models past and future information in sequences using forward and backward long short-term memory layers (Figure 4). After the local feature sequences extracted by the CNN were input into the BiLSTM, memory retention and noise filtering were controlled through the coordinated operation of the forget, input, and output gates, enabling the model to capture periodic patterns in precipitation sequences. In offshore precipitation processes, current precipitation is influenced by water vapor accumulation and weather system evolution in previous days, as well as precursor signals of subsequent weather patterns. Through the coordinated operation of the forget, input, and

output gates, the BiLSTM retains relevant long-term information and filters irrelevant noise, comprehensively capturing periodic patterns, sustained dry–wet trends, and precursors of extreme events in precipitation sequences. Thus, the model remains sensitive to peak precipitation values during events such as rapidly approaching typhoons.

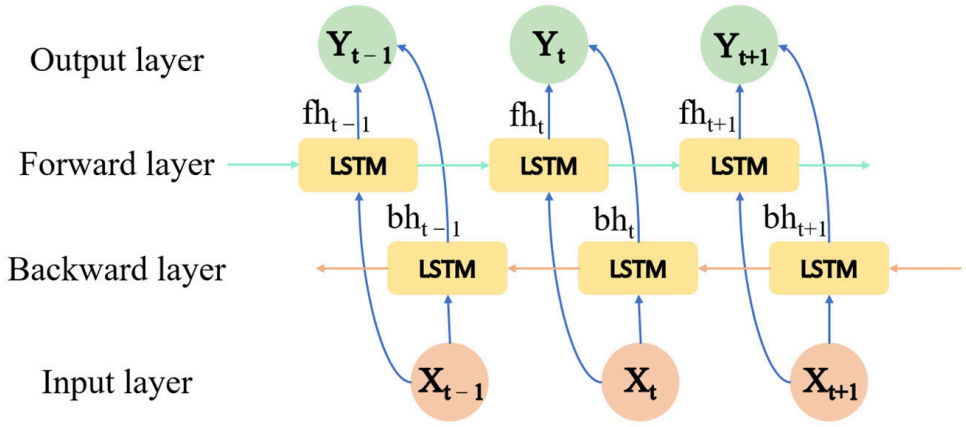


Figure 4. BiLSTM architecture: x_{t-1}, x_t, and x_{t+1} represent the inputs; y_{t-1}, y_t, and y_{t+1} represent the outputs; and fh_{t-1}, fh_t, and fh_{t+1} and bh_{t-1}, bh_t, and bh_{t+1} represent the hidden states in the forward and backward long short-term memory layers, respectively. (Source: Original figure by the authors).

3.5. Performance Evaluation Metrics

An effective model requires good predictive capability and stability. Therefore, multiple evaluation metrics—MAE, RMSE, and coefficient of determination (R^2)—were used to assess model performance, as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - f_i| \tag{14}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{n=1}^{n}(y_i - f_i)^2} \tag{15}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(f_i - y_i)^2}{\sum_{i=1}^{n}\left(f_i - \bar{f}_i\right)^2} \tag{16}$$

where y_i is the predicted value, and f_i and $\bar{f}$ are the observed value and the mean of the observed values, respectively. Lower MAE and RMSE values indicate smaller prediction errors, whereas higher R^2 values indicate a greater proportion of variance in the observed data explained by the model and better goodness of fit. RMSE has the same units as the original data, facilitating intuitive interpretation of the magnitude of prediction errors; values closer to zero indicate more accurate predictions. As an absolute error metric, MAE is less sensitive to outliers than RMSE and reflects the average magnitude of prediction errors, with values closer to zero indicating greater predictive accuracy. R^2 ranges from $\bar{f}_i$ to 1, with values closer to 1 indicating better goodness of fit.

3.6. Digital Twin-Enabled Decision Framework for Intelligent Water Storage Control

To support intelligent water storage operations and enhance freshwater security on remote offshore islands, the digital twin framework is formulated in this study as a physically consistent decision system that integrates a discrete-time rolling-horizon optimization scheme with a continuous-time conceptual dynamic model. The state variable of the system is defined as the reservoir storage volume S (m^3), with precipitation treated as an exogenous input variable. At each discrete decision instant t_s, the digital twin assimilates the

measured storage state, recent hydrometeorological observations, and precipitation forecast information over the forecast horizon and evaluates system responses under different control strategies within the virtual space.

The core of the decision-making process is governed by a discrete-time water-balance formulation. Let S_t denote the storage volume at the end of day t and P_t denote the precipitation depth on day t; the daily storage recursion is then given as follows:

$$S_t = S_{t-1} + \eta \cdot A \cdot P_t \times 10^{-3} - C_t - E_t - O_t - R_t \quad (17)$$

Here, S_{t-1} denotes the storage volume at the end of day $t-1$ (m^3); P_t (mm) is the precipitation depth on day t; A (m^2) represents the catchment area; and the term $\eta \cdot A \cdot P_t \times 10^{-3}$ converts the precipitation depth into a volume (m^3), with η being the runoff coefficient. C_t (m^3) is the water consumption on that day; E_t (m^3) is the evaporation loss; O_t (m^3) is the overflow loss, and R_t (m^3) is the controlled release volume (m^3), which serves as the primary decision variable.

The overflow term is computed from the volume exceeding the maximum effective storage capacity:

$$O_t = \max(0, S_{t-1} + \eta \cdot A \cdot P_t \times 10^{-3} - C_t - R_t - E_t - S_{\max})) \quad (18)$$

Here, S_{max} denotes the maximum effective storage capacity of the rainwater collection pool. The system is additionally subject to the physical constraint $0 \leq S_t \leq S_{max}$.

4. Results and Discussion

4.1. Model Parameter Configuration

A robust forecasting model depends not only on its architectural design but also on careful hyperparameter configuration. Selecting appropriate hyperparameters helps the model capture complex patterns in the data, producing accurate predictions while reducing the computational burden.

For the proposed CNN–BiLSTM–MLP hybrid prediction model, the required hyperparameters include (1) the number of channels in the convolutional layer; (2) the number of BiLSTM and MLP layers; (3) the learning rate; and (4) the activation function. An excessive number of convolution kernels, overly deep network architectures, or excessively large hidden-unit dimensions can lead to a substantial increase in the number of model parameters, increasing not only computational overhead but also the risk of overfitting. Conversely, an overly shallow network architecture may fail to effectively capture the inherent multi-scale oscillations and abrupt features in rainfall sequences, reducing prediction reliability and accuracy. In addition, the total number of training epochs is crucial. We adopted an early stopping strategy that monitors validation loss and dynamically terminates training when model performance no longer improves, thereby reducing computational time while ensuring sufficient convergence. As defined in Section 3.2, the training set was used for gradient-based parameter updates, while the validation set was reserved exclusively for early stopping monitoring. In compliance with the rolling-origin protocol, the validation subset was preprocessed independently using the frozen parameters derived from the training subset. Training was halted if the validation loss showed no improvement for 15 consecutive epochs.

To reduce the influence of random weight initialization and assess statistical robustness, all experiments were independently repeated 10 times using different random seeds. Based on these multiple runs, the optimal model parameters (Table 1), the software and hardware environment used for training, and the optimizer-related hyperparameters (Table 2) were determined, along with the MAE, RMSE and R^2 on the test set and the standard deviations of these metrics (Table 3).

Table 1. Optimal configuration of the hybrid model parameters.

Module	Parameter	Value/Configuration
CNN	Number of layers	2
	Kernel size	3
	Channels (layer 1 → layer 2)	32 → 64
	Activation function	ReLU
BiLSTM	Number of layers	2
	Hidden units per direction (layer 1 → layer 2)	64 → 32
FC layer after BiLSTM	Number of layers	1
	Number of neurons	32
	Activation function	ReLU
MLP branch	Number of layers	2
	Neurons in hidden layers (layer 1 → layer 2)	64 → 32
	Number of neurons in the output layer	30
	Activation function	ReLU
Final fusion MLP	Number of layers	1
	Number of hidden neurons	16
	Activation function	ReLU
Output layer	Output neurons	30

Table 2. Optimal configuration of model training hyperparameters and runtime environment.

Parameter	Value
GPU specifications	NVIDIA GeForce RTX 4090
Python version	3.8.5
CUDA version	11.8
Deep learning framework	PyTorch 2.1.1
Epochs	80
Batch size	64
Learning rate	0.0003
Weight decay	0.0001
Patience	15
Optimizer	Adam

Table 3. Comparison of the mean values and standard deviations of MAE, RMSE, and R^2 for the model.

Metric	Mean	Std. Dev.
MAE (mm)	2.49	0.23
RMSE (mm)	5.72	0.45
R^2	0.83	0.03

4.2. Model Performance Evaluation and Extreme Weather Validation

To further validate the effectiveness of the proposed model, we compared it with four commonly used baseline models, SARIMA, LSTM, CNN–LSTM, and the climatological mean. All baseline models were trained and tested using exactly the same experimental protocol as the proposed model and were evaluated using identical metrics. Figure 5 presents a comparison of precipitation forecasting performance across all models for the 2016–2019 period. The figure shows that the proposed model achieves the highest correlation coefficient and the lowest normalized RMSE, indicating that its predictions are in the closest

agreement with the observed values; its normalized standard deviation is also closest to that of the observations, demonstrating superior preservation of variability compared with the other methods. In contrast, although LSTM and CNN–LSTM show some predictive capability, their errors are markedly higher than those of the proposed model; SARIMA and the climatological mean perform considerably worse, capturing little of the variability in the precipitation series. The proposed model outperforms all baseline models across all metrics, with particularly small prediction biases during months with extreme precipitation, which may be attributable to the ability of its hybrid architecture to capture multiscale features. By contrast, SARIMA struggles to capture nonlinear abrupt changes; LSTM and CNN–LSTM, although capable of learning temporal dependencies, lack differentiated treatment of low-frequency trends and high-frequency oscillations; and the climatological mean is overly smooth and fails to reflect interannual variability.

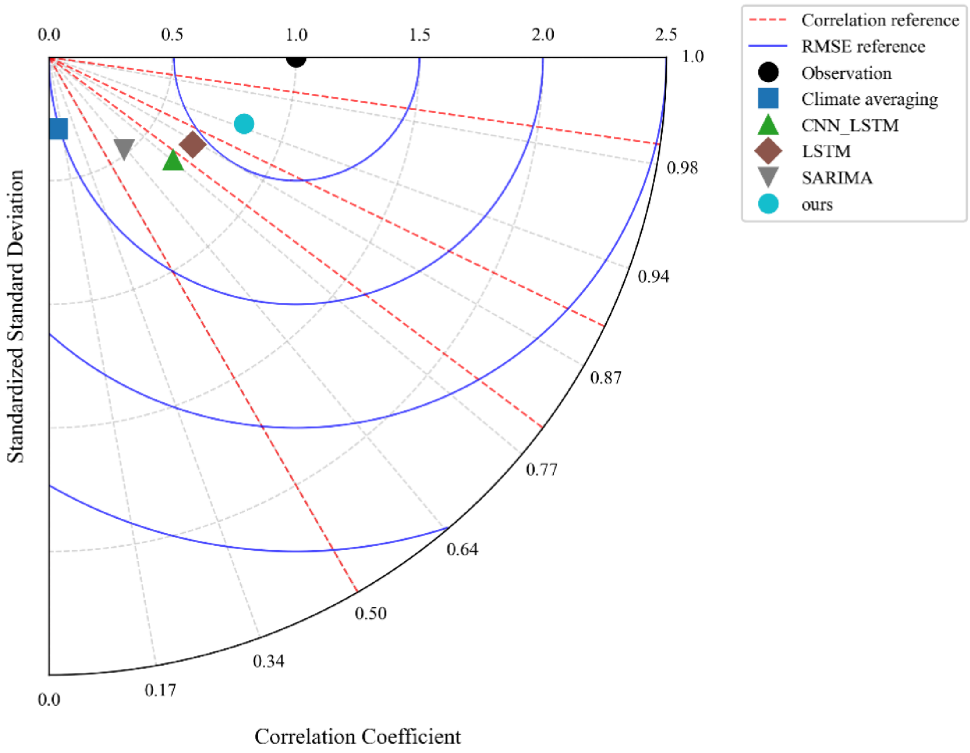


Figure 5. Taylor diagram comparing the statistical performance of different models for precipitation prediction during 2016–2019. (Source: Original figure by the authors).

To further validate the predictive capability of the proposed model for extreme weather events, we compared the precipitation forecasts from each model with the observed precipitation during both dry and wet seasons, using forecasts for April and December 2017 as illustrative examples, as shown in Figures 6 and 7. These figures reveal substantial differences among the models in their ability to capture extreme precipitation events. The SARIMA model predicts peak values during the heavy rainfall event in December that reach only 40–50% of the observed precipitation, with peak timing delayed by 1–2 days. The LSTM model performs relatively consistently for the April forecasts yet still exhibits a pronounced smoothing effect during extreme events in the rainy season. In December, when heavy rainfall exceeding 100 mm occurs, the peaks predicted by the LSTM are underestimated by more than 50%, indicating that the unidirectional LSTM has limited sensitivity to abrupt extreme signals that are not captured by long-term dependencies. The CNN–LSTM model shows improved fitting performance during continuous rainfall processes; however, it still underestimates extreme daily precipitation peaks by approximately 65%, and the fluctuation amplitude during

high-frequency oscillations is notably smaller than that of the LSTM, suggesting that its nonlinear mapping of extreme values remains inadequate. The climatological mean performs worst among all comparison models; although it captures the basic dry–wet seasonal pattern, it fails to reproduce daily-scale extreme fluctuations. In contrast, the proposed model maintains high fitting accuracy in both dry and wet seasons. These comparative results demonstrate that the proposed model effectively extracts precipitation oscillation features at different scales through VMD–SSA signal decomposition. Combined with the CNN–BiLSTM–MLP hybrid prediction architecture, this multiscale decomposition enables the model to preserve its overall trend-fitting capability while enhancing its sensitivity and peak prediction accuracy for extreme precipitation events, thereby supporting its potential practical application under the climatic conditions of remote offshore islands.

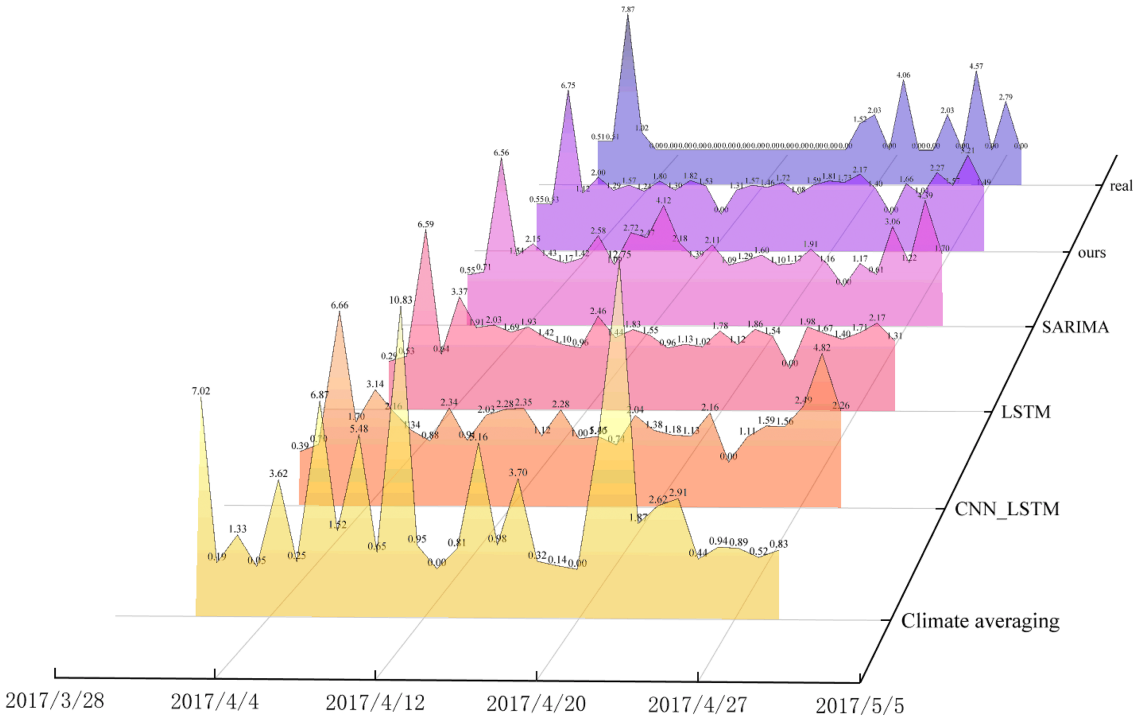


Figure 6. Comparison of daily precipitation predictions from different models with observations in April 2017. (Source: Original figure by the authors).

To further validate the reliability of the proposed model under heavy rainfall conditions, an additional analysis of the forecast results was conducted, with the outcomes summarized in Table 4. The evaluation employed daily precipitation thresholds of 25 mm/d (heavy rainfall) and 50 mm/d (typhoon-induced extreme rainfall). At the heavy-rainfall threshold, the model achieved a relatively high probability of detection (POD = 0.80) and a low false alarm rate (FAR = 0.16). At the typhoon-extreme threshold, however, the POD dropped to 0.59 and the FAR rose to 0.30, indicating that while the model demonstrates strong capability in identifying routine heavy rainfall events, its forecasting performance for extreme typhoon-related precipitation remains limited, likely attributable to the insufficient proportion of typhoon samples in the training dataset. From a seasonal perspective, the model accuracy remained stable, with a mean absolute error (MAE) of 2.93 mm during the rainy season (June to November) and 2.05 mm during the dry season (December to May). Collectively, these metrics suggest that the model holds promising practical application potential for intelligent water storage regulation under extreme hydrological events.

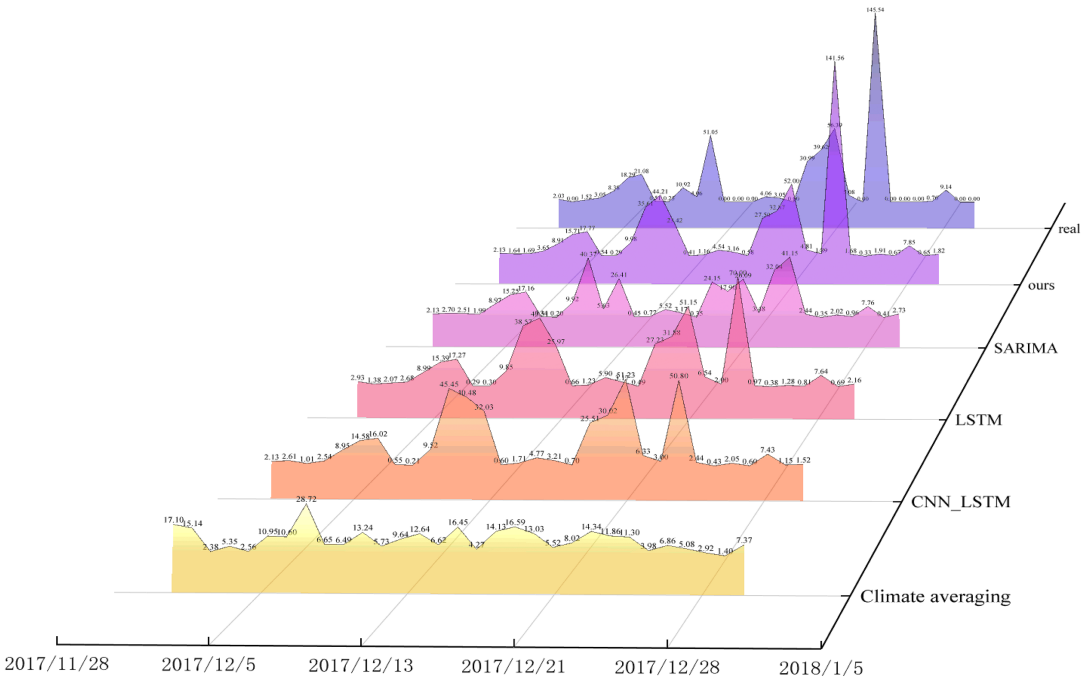


Figure 7. Comparison of daily precipitation predictions from different models with observations in December 2017. (Source: Original figure by the authors).

Table 4. Performance evaluation of the model on the prediction results.

Evaluation Category	Subcategory	POD	FAR	MAE (mm)
Threshold evaluation	Rainstorm	0.80	0.16	—
	Typhoon rainstorm	0.59	0.30	—
Seasonal evaluation	Wet season	—	—	2.93
	Dry season	—	—	2.05

4.3. Evaluation of the Digital Twin Control Effect

Based on the precipitation prediction results, we further conducted simulations to evaluate the regulation performance of the proposed digital twin decision framework. The simulations were based on the designed physical rainwater collection pool system and compared the operational performance of the traditional empirical strategy with that of the strategy generated by the digital twin framework. The rainwater collection pool had a maximum effective storage capacity S_{max} of 2000 m^3, a catchment area A of 8000 m^2, a runoff coefficient of 0.85, an initial storage volume of 800 m^3, a minimum allowable storage volume S_{min} of 250 m^3 to prevent pump cavitation and dry running, domestic water consumption of 80 m^3 per day, and an evaporation loss of 1 L/m^2 under no-rainfall conditions, which was considered negligible under rainy conditions. The maximum daily release volume R_{max} was 400 m^3.

The conventional control strategy uses heuristic control, whereby water is released when the storage volume exceeds 80% of the total capacity. In contrast, the digital twin system determines the optimal daily release volume based on the complete forecast precipitation sequence. The daily planned release R_t is taken as the decision variable. The primary objective is to minimize the combined losses due to water shortage and overflow. To achieve this, we define a water shortage variable U_t, which represents the volume of unmet demand on day t.

$$U_t = \max(0, C_t - (S_{t-1} + \eta \cdot A \cdot P_t \times 10^{-3} - R_t - E_t - S_{\min})) \quad (19)$$

The objective function for the 30-day planning horizon is then formulated as follows:

$$J = \sum_{t=1}^{H} (\lambda_1 U_t + \lambda_2 O_t) \tag{20}$$

Here, $H = 30$ days is the receding prediction horizon. The penalty weights λ_1 and λ_2 are assigned to water shortage and overflow, respectively. To prioritize supply reliability, we set $\lambda_1 > \lambda_2$. In this study, we chose λ_1 =10 and λ_2 =3 after sensitivity analysis. The release decision is constrained by the outlet capacity and the minimum storage reserve:

$$0 \le R_t \le R_{max}, S_t \ge S_{min} \tag{21}$$

The digital twin assimilates the observed storage state and the updated precipitation forecast for the next 30 days and then solves the finite-horizon optimization problem using dynamic programming to obtain the optimal release sequence:

$$\left\{R^*_{k|k}, R^*_{k+1|k}, \ldots, R^*_{k+H-1|k}\right\}$$

Only the first control action $R^*_{k|k}$ is applied to the physical system. On the following day, the state is updated using new observations and revised forecasts, and the optimization is repeated over a shifted horizon.

Daily precipitation data from October 2016 were analyzed to verify the effectiveness of the intelligent control strategy. During this month, five concentrated heavy rainfall events occurred under the influence of Typhoon Sarika, with daily precipitation peaks reaching 113 mm. The preceding period was characterized by a continuous mild drought lasting 25 days. Figures 8 and 9 present the complete time series of storage volume, release volume, inflow volume, and overflow volume of the rainwater collection pool under the two control strategies, alongside the observed and predicted precipitation. Figure 8 shows the results of an experience-based strategy without precipitation forecasting, in which threshold-based decisions are made solely on the basis of the real-time storage state and daily precipitation. Figure 9 shows the results obtained using the digital twin framework driven by the prediction model proposed in this study, which preemptively reserves storage capacity or releases water based on the 30-day precipitation forecast.

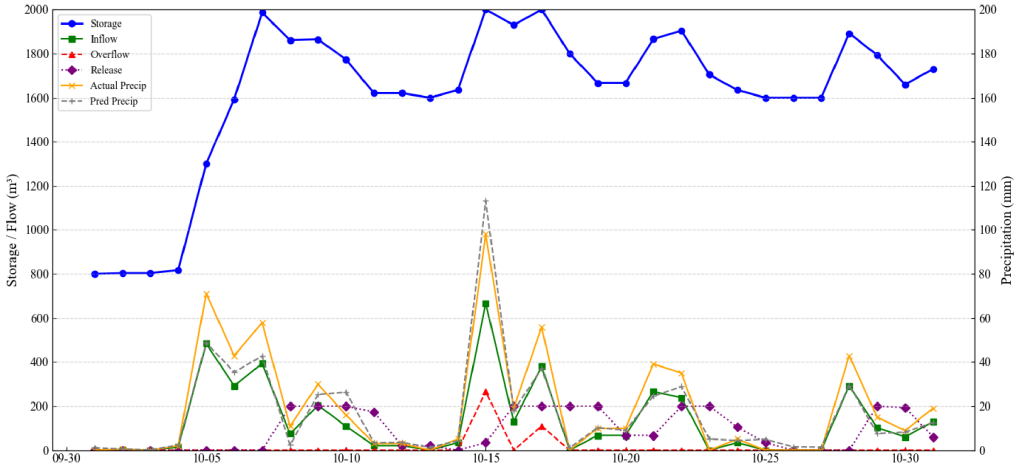


Figure 8. Time series of storage, inflow, overflow, and release volumes, along with observed and predicted precipitation, for the rainwater collection pool under conventional control. (Source: Original figure by the authors).

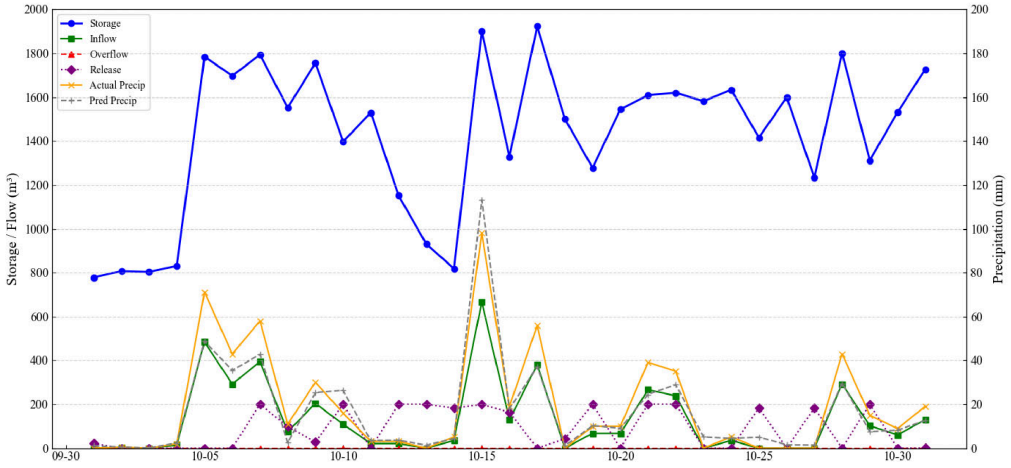


Figure 9. Time series of storage, inflow, overflow, and release volumes, along with observed and predicted precipitation, for the rainwater collection pool under digital twin control. (Source: Original figure by the authors).

The simulation results demonstrate that without digital twin control, the rainwater collection pool experienced two overflow events, and during heavy precipitation events, the storage volume repeatedly approached the overflow warning threshold. Owing to the absence of precipitation forecasts, preventive measures could not be implemented before precipitation peaks, consequently leading to the loss of freshwater resources. In contrast, under digital twin control, the storage volume of the rainwater collection pool never exceeded 90% of the total storage capacity; overflow losses were effectively avoided, and storage capacity was proactively created before heavy precipitation events, thereby maximizing freshwater conservation.

Quantitatively, under the conventional reservoir operation strategy, two overflow events occurred, causing a total overflow loss of 376 m^3. During the rainy season, the average stored volume reached 1617 m^3, remaining at a high level and repeatedly approaching the warning threshold. In contrast, under the precipitation-forecast-informed digital twin control, the average rainy-season stored volume was 1380 m^3, and no overflow occurred. This corresponds to a reduction of 237 m^3 (approximately 14.7%) in average stored volume and a decrease in overflow loss from 376 m^3 to zero. These results indicate that the improvement in water management efficiency arises from the integration of precipitation forecasting with digital twin-based scheduling and optimization, which enables proactive storage regulation and more effective rainwater harvesting while reducing overflow losses.

Thus, embedding the proposed hybrid prediction model into the digital twin decision framework can effectively enhance the adaptive control capability of rainwater collection systems under alternating conditions of extreme rainfall and drought, providing a feasible technical approach for intelligent freshwater management on offshore islands.

5. Conclusions

In this study, we propose a novel framework for 30-day precipitation forecasting and water storage control decision-making that integrates VMD–CSSOA–SSA signal preprocessing, CNN–BiLSTM–MLP hybrid modeling, and a digital twin framework to address the need for a stable freshwater supply on remote offshore islands. The major conclusions of this study are as follows.

1. At the signal preprocessing level, the coupled VMD–CSSOA–SSA strategy effectively addresses the multiscale complexity of precipitation sequences on offshore islands. VMD decomposes the original precipitation sequence into modal components at different

frequencies, reducing the effects of non-stationarity and mode mixing on subsequent modeling. SSA provides secondary denoising of high-frequency modes, suppressing random noise while preserving extreme precipitation signals and thereby providing clearly structured inputs for deep learning models.

2. At the forecasting and modeling levels, the CNN–BiLSTM–MLP hybrid architecture extracts multiscale temporal features from precipitation sequences. The CNN module captures short-term fluctuation patterns, whereas the BiLSTM module processes the feature sequences extracted by the CNN and learns long-term dependencies in both forward and backward directions. This design not only enhances the interpretability of the model but also improves its suitability for forecasting daily precipitation sequences over a 30-day horizon.
3. At the application level, the storage strategies of rainwater collection pools are dynamically optimized by integrating the precipitation forecasting model into the digital twin framework. Compared with experience-based rules without forecasting, digital twin control with forecasting effectively improves the utilization of rainwater collection pools by preemptively reserving storage capacity or releasing water, thereby reducing overflow losses.

Within this application framework, the 30-day forecast serves as a medium-term basis for scheduling valve closure to retain water during dry periods and valve opening to create storage capacity during wet periods, rather than as the sole trigger for emergency actions. For extreme events, the medium-term projection provided by the 30-day forecast should be integrated with short-term warnings issued by meteorological and oceanographic agencies. Operators can then execute valve operations based on a combined assessment of both sources of information, while fixed storage thresholds serve as binding constraints to ensure that rainwater collection pool operations remain within manually controllable safety limits. These operational procedures are conceptual and remain to be tested in real-world applications.

This study contributes to the sustainable development of freshwater systems on remote offshore islands. By reducing overflow losses and enhancing rainwater utilization efficiency, the reliance on energy-intensive seawater desalination technologies is diminished, consequently lowering operational costs and alleviating the dependence of island communities on external freshwater supplies. The proactive storage regulation enabled by the digital twin framework also helps ensure a stable water supply during drought periods, thereby supporting local livelihoods.

The framework proposed in this study is data-driven and relies solely on single-station observations, without incorporating surrounding atmospheric forcing fields, satellite precipitation products, or ocean dynamic variables. This configuration stems partly from the scarcity of long-term, continuous, multi-source observational data in remote island environments, and partly from an intention to assess whether precipitation data alone can effectively support water storage management. However, this data-driven attribute also gives rise to notable limitations in physical interpretability: the framework cannot represent the internal dynamical processes of weather systems directly, thereby constraining its capability to depict the evolution of regional weather patterns. Beyond the constraints at the data input level, the current study remains focused on framework development and simulation-based validation of the digital twin decision system at the validation level, with no physical sensors deployed and no closed-loop field control experiments conducted. The simulated scenarios were designed around the dry–wet alternation and rainfall intensities observed at the study site, covering drought periods, moderate rainfall, and heavy precipitation events, which provides a reasonable basis for evaluating the framework's performance under the predominant local conditions, yet

does not encompass all possible seasonal, extreme, or water-demand-related combinations. Consequently, the results of this paper are most directly applicable to sites with similar climatic and storage characteristics; their extension to other operating conditions would require recalibration of the control parameters and thresholds. It should be noted that the current analysis is based on a single station and selected simulation scenarios; although it offers a reference for assessing the feasibility of the proposed framework, the uncertainty of 30-day precipitation forecasts and the influence of site-specific conditions still warrant further consideration in practical deployment.

To address these limitations, future work will integrate multisource meteorological data, including wind speed, atmospheric pressure, humidity, sea surface temperature, and satellite cloud imagery, to construct a multimodal prediction model with stronger physical constraints. Furthermore, after the on-site retrofitting of the rainwater harvesting system is completed, we plan to establish a complete digital twin platform with bidirectional physical–virtual interaction and conduct long-term field validation, thereby further evaluating the practical engineering applicability of this framework. The scenario set will also be expanded to cover a wider range of hydrometeorological conditions, seasonal transitions, and water-demand levels, so that the transferability of the framework to other remote islands can be assessed more systematically. Overall, the proposed framework should be regarded as a conceptual and simulation-based decision-support tool, and its practical effectiveness requires further validation through field deployment and long-term operational testing.

Author Contributions: Conceptualization, X.W. and L.X.; methodology, X.W.; software, X.W. and T.C.; validation, T.C., G.W. and Y.C.; formal analysis, X.Z.; investigation, Q.Y.; resources, L.X.; data curation, X.W.; writing—original draft preparation, X.W.; writing—review and editing, L.X. and Q.Y.; visualization, X.W.; supervision, L.X.; project administration, Q.Y.; funding acquisition, Q.Y. and L.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key R&D Program of China (Key Special Project for Marine Environmental Security and Sustainable Development of Coral Reefs 2023-3.2), grant number 2023YFC3107100.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to a confidentiality agreement with the company that restricts the public release of part of the code.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Papathanasiou, A.F.; Ntemiri, K.; Baltas, E. Sustainable water resources management and energy production in Small Mediterranean islands. *Environ. Process.* **2025**, *12*, 19. [CrossRef]
2. Roberts, D.A.; Johnston, E.L.; Knott, N.A. Impacts of desalination plant discharges on the marine environment: A critical review of published studies. *Water Res.* **2010**, *44*, 5117–5128. [CrossRef]
3. Elimelech, M.; Phillip, W.A. The future of seawater desalination: Energy, technology, and the environment. *Science* **2011**, *333*, 712–717. [CrossRef] [PubMed]
4. Dallman, S.; Chaudhry, A.; Muleta, M.; Lee, J. The Value of Rain: Benefit-Cost Analysis of Rainwater Harvesting Systems. *Water Resour. Manag.* **2016**, *30*, 4415–4428. [CrossRef]
5. Campisano, A.; Modica, C. Optimal sizing of storage tanks for domestic rainwater harvesting in Sicily. *Resour. Conserv. Recycl.* **2012**, *63*, 9–16. [CrossRef]
6. Basinger, M.; Montalto, F.; Lall, U. A rainwater harvesting system reliability model based on nonparametric stochastic rainfall generator. *J. Hydrol.* **2010**, *392*, 105–118. [CrossRef]

7. de A. Coelho, G.; Ferreira, C.M.; Kinter, J.L.; Cassalho, F. Real-time short-range flood forecasting based on a watershed scale 2-D hydrodynamic model and high-resolution precipitation forecast ensemble. *J. Hydrol.* **2025**, *650*, 132564. [CrossRef]
8. Faikrua, A.; Torsri, K.; Dike, V.N.; Peangta, P.; Sawangwattanaphaibun, R. 20–60-day intraseasonal variation of summer rainfall in Thailand and its associated large-scale atmospheric moisture circulation. *Theor. Appl. Climatol.* **2023**, *154*, 1165–1178. [CrossRef]
9. Sharifazari, S.; McCallum, J.; Meredith, K.; Johnson, F.; Palmer, J.G.; Turney, C.S.M.; Andersen, M.S. The impact of declining rainfall and ocean forcing on morphology and dynamics of an island fresh groundwater lens, South-West Western Australia. *J. Hydrol.* **2025**, *663*, 134294. [CrossRef]
10. Liu, H.; Zheng, X.; Yuan, J.; Frederiksen, C.S. Potential Predictability of Seasonal Global Precipitation Associated with ENSO and MJO. *Atmosphere* **2023**, *14*, 695. [CrossRef]
11. He, C.; Liu, Z.; Wu, J.; Pan, X.; Fang, Z.; Li, J.; Bryan, B.A. Future global urban water scarcity and potential solutions. *Nat. Commun.* **2021**, *12*, 4667. [CrossRef] [PubMed]
12. Roshan, A.; Kumar, M. Water End-use estimation can support the urban water crisis management: A critical review (Review). *J. Environ. Manag.* **2020**, *268*, 110663. [CrossRef]
13. Ferguson, C.; Guerra, R.; Cooke, S.; Proaño, D. Trial of fog and rainwater harvesting to alleviate water scarcity in the agricultural highlands of the Galapagos Islands. *J. Water Sanit. Hyg. Dev.* **2025**, *15*, 1–7. [CrossRef]
14. Padrón, I.; Avila, D.; Marichal, G.N.; Rodríguez, J.A. Assessment of hybrid renewable energy systems to supplied energy to autonomous desalination systems in two islands of the canary archipelago. *Renew. Sustain. Energy Rev.* **2019**, *101*, 221–230. [CrossRef]
15. Sohrabi, A.; Meratizaman, M.; Liu, S. Comparative analysis of integrating standalone renewable energy sources with brackish water reverse osmosis plants: Technical and economic perspectives. *Desalination* **2024**, *571*, 117106. [CrossRef]
16. Avila Prats, D.; San Luis Gutiérrez, F.; Hernández López, Á.; Marichal Plasencia, G.N. Optimal Arrangements of Renewable Energy Systems for Promoting the Decarbonization of Desalination Plants. *J. Mar. Sci. Eng.* **2024**, *12*, 1193. [CrossRef]
17. Soleimanzade, M.A.; Kumar, A.; Sadrzadeh, M. Novel data-driven energy management of a hybrid photovoltaic-reverse osmosis desalination system using deep reinforcement learning. *Appl. Energy* **2022**, *317*, 119184. [CrossRef]
18. Feloni, E.; Nastos, P.T. Evaluating rainwater harvesting systems for water scarcity mitigation in Small Greek islands under climate change. *Sustainability* **2024**, *16*, 2592. [CrossRef]
19. Zang, J.; Kumar, M.; Werner, D. Real-world sustainability analysis of an innovative decentralized water system with rainwater harvesting and wastewater reclamation. *J. Environ. Manag.* **2021**, *280*, 111639. [CrossRef]
20. de Sá Silva, A.C.R.; Bimbato, A.M.; Balestieri, J.A.; Vilanova, M.R.N. Exploring environmental, economic and social aspects of rainwater harvesting systems: A review. *Sustain. Cities Soc.* **2022**, *76*, 103475. [CrossRef]
21. Kakoulas, D.A.; Golfinopoulos, S.K.; Koumparou, D.; Alexakis, D.E. The effectiveness of rainwater harvesting infrastructure in a Mediterranean island. *Water* **2022**, *14*, 716. [CrossRef]
22. Marwein, D.; Yadav, O.P.; Dkhar, D.H.P.; Yadav, P.; Debnath, B.K.; Das, R.S.; Das, B. Performance of a rainwater harvester and double slope desalinator with booster mirror and cooling fan. *J. Sol. Energy Eng.* **2026**, *148*, 031002. [CrossRef]
23. Papathanasiou, A.F.; Platanitis, I.; Baltas, E. Investigation of water supply through renewable energy during the peak tourist season on Mykonos island. *Water* **2026**, *18*, 259. [CrossRef]
24. Nagai, M.; Bormudoi, A. A spatiotemporal analysis of heterogeneity and non-stationarity of extreme precipitation in the Ayeyarwady River Basin, myanmar, and their linkages to global climate variability. *Water* **2026**, *18*, 227. [CrossRef]
25. Amini, A.; Dolatshahi, M.; Kerachian, R. Adaptive precipitation nowcasting using deep learning and ensemble modeling. *J. Hydrol.* **2022**, *612*, 128197. [CrossRef]
26. Zhao, H.; Huang, X.; Huang, Y.; Jin, J. A revised forecast method of ECMWF precipitation based on CNN feature extraction. *IOP Conf. Ser. Earth Environ. Sci.* **2021**, *658*, 012025. [CrossRef]
27. Sun, L.; Lan, Y. Statistical downscaling of daily temperature and precipitation over China using deep learning neural models: Localization and comparison with other methods. *Int. J. Climatol.* **2021**, *41*, 1128–1147. [CrossRef]
28. Kim, Y.; Hong, S. Very short-term rainfall prediction using ground radar observations and conditional generative adversarial networks. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 4104308. [CrossRef]
29. Ko, J.; Lee, K.; Hwang, H.; Oh, S.G.; Son, S.W.; Shin, K. Effective training strategies for deep-learning-based precipitation nowcasting and estimation. *Comput. Geosci.* **2022**, *161*, 105072. [CrossRef]
30. Rojas-Campos, A.; Langguth, M.; Wittenbrink, M.; Pipa, G. Deep learning models for generation of precipitation maps based on numerical weather prediction. *Geosci. Model Dev.* **2023**, *16*, 1467–1480. [CrossRef]
31. Farman, H.; Hussain, M.A.; Shaikh, S.; Hassan, S.; Hassan, S.R. Dual framework for rainfall prediction: A multi-seed machine and deep learning evaluation across Pakistan's climatic regimes. *Sci. Rep.* **2026**, *16*, 20031. [CrossRef]
32. Wang, W.-C.; Ye, F.-R.; Wang, Y.-Y.; Gu, M. A singular spectrum analysis-enhanced BiTCN-selfattention model for runoff prediction. *Earth Sci. Inform.* **2024**, *18*, 31. [CrossRef]

33. Johny, K.; Pai, M.L.; Adarsh, S. Adaptive EEMD-ANN hybrid model for Indian summer monsoon rainfall forecasting. *Theor. Appl. Climatol.* **2020**, *141*, 1–17. [CrossRef]
34. Mir, F.; Mahmoudi, P.; Khosravi, M.; Amir Jahanshani, S.M. Detecting intra- and extra-decade nature of droughts in Iran using empirical mode decomposition (EMD). *Iran. J. Sci. Technol. Trans. Civ. Eng.* **2026**, *50*, 893–904. [CrossRef]
35. Ulusoy, A.J.; Nerantzis, D.; Stoianov, I. Adaptive MPC for burst incident management in water distribution networks. *IEEE Trans. Control Netw. Syst.* **2023**, *10*, 2060–2071. [CrossRef]
36. Liu, D.; Li, S.; Zheng, Y.; Wu, J. Distributed zone MPC of pressure management for water distribution network systems. *IET Control Theory Appl.* **2019**, *13*, 1704–1717. [CrossRef]
37. Velarde, P.; Tian, X.; Sadowska, A.D.; Maestre, J.M. Scenario-based hierarchical and distributed MPC for water resources management with dynamical uncertainty. *Water Resour. Manag.* **2019**, *33*, 677–696. [CrossRef]
38. Divya, A.; Ramesh, S.; Chakrasali, S.V.; Kumar, M.K.; Taj, M.; Ganiya, R.K.; Nageswara Rao, J.; Balasubramanian, K.; Qamar, S. Federated digital-twin–empowered blockchain architecture for autonomous and resilient smart water management in sustainable cities. *Sustain. Comput. Inform. Syst.* **2026**, *50*, 101303. [CrossRef]
39. Xu, Y.; Hui, M.; Qu, H. Design of a 3D platform for the evaluation of water quality in urban rivers based on a digital twin model. *Water* **2024**, *16*, 3668. [CrossRef]
40. Alenezi, A.; Alabaiadly, Y. Digital twins and edge AI for small/off-grid water systems: A review. *IEEE Access* **2026**, *14*, 39960–39978. [CrossRef]
41. Nguyen, T.H.; Bhattacharya, S.; Wong, J.S.; Didry, Y.; Phan, L.D.; Tamisier, T.; Maguire, B.; Paolucci, J.B.; Matgen, P. Towards digital twin in flood forecasting with data assimilation satellite earth observations—A proof-of-concept. *Remote Sens.* **2026**, *18*, 685. [CrossRef]
42. Green, A.C.; Lewis, E.; Tong, X.; Wardle, R. A framework for incorporating rainfall data into a flooding digital twin. *J. Hydrol.* **2025**, *656*, 132893. [CrossRef]
43. Başkesen, K.M.; Güçlü, Y.S.; Yakut Şevik, S.E.; Şahin, A.D. Comparative analysis of interpolation methods for missing daily precipitation data by suggesting an alternative inverse distance weighted model. *Theor. Appl. Climatol.* **2025**, *156*, 659. [CrossRef]
44. Dragomiretskiy, K.; Zosso, D. Variational Mode Decomposition. *IEEE Trans. Signal Process.* **2014**, *62*, 531–544. [CrossRef]
45. Lyu, X.; Mu, X.D.; Zhang, J.; Wang, Z. Chaos sparrow search optimization algorithm. *J. Beijing Univ. Aeronaut. Astronaut.* **2021**, *47*, 1712–1720. (In Chinese) [CrossRef]

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