Available online at www.HighTechJournal.org

HighTech and Innovation Journal

ISSN: 2723-9535

Vol. 7, No. 2, June, 2026

PSO-Optimized Hybrid ARIMA-LSTM for Provincial Water Forecasting

Solikhin Solikhin [1*], Septia Lutfi [1], Agus Purwanto [2], Dinan Pazzarena [1], M. Zakki Abdillah [3]

[1] *Department of Informatics, STMIK Himsya, Semarang, Indonesia.*

[2] *Department of Information System, STMIK Himsya, Semarang, Indonesia.*

[3] *Department of Information System, Universitas Nasional Karangturi, Semarang, Indonesia.*

Received 28 February 2026; Revised 24 May 2026; Accepted 26 May 2026; Published 01 June 2026

Abstract

Accurate macro-administrative water distribution forecasting is essential for infrastructure planning and regional governance. While most prior studies focus on short-term urban demand, provincial-scale systems exhibit aggregated heterogeneity, structural shifts, and limited annual observations that challenge conventional forecasting approaches. This study proposes a Particle Swarm Optimization (PSO)-optimized hybrid ARIMA-LSTM framework designed for cross-domain robustness in macro-scale infrastructure forecasting. Linear components are modeled using ARIMA, nonlinear residuals are learned via LSTM, and PSO jointly optimizes statistical and neural hyperparameters under validation constraints. The framework is evaluated across 34 Indonesian provinces using sequential train–validation–test splits and recursive multi-step forecasting. Beyond conventional accuracy metrics (RMSE, MAE, MAPE, R^2), this study introduces cross-provincial RMSE variance and multi-domain Diebold-Mariano statistical testing as robustness indicators. Results show that the PSO-optimized hybrid model achieves the lowest average MAPE (21.79%), reduces inter-provincial RMSE variance by approximately 15%, dominates in 52.9% of provinces, and demonstrates statistically significant improvement in 76.5% of provinces. These findings confirm that optimization-enhanced hybrid decomposition improves structural stability and cross-domain generalization in heterogeneous macro-administrative infrastructure systems.

Keywords: Clean Water Distribution; ARIMA; LSTM; Particle Swarm Optimization; Hybrid Forecasting; Provincial-Scale Modeling.

1. Introduction

Accurate forecasting of water distribution demand is essential for infrastructure planning, resource allocation, and long-term policy formulation. Reliable projections enable governments and regional agencies to anticipate future demand, optimize investment strategies, and improve operational sustainability under evolving socio-economic conditions [1, 2]. These challenges become more significant in macro-administrative systems where demand characteristics vary substantially across regions. Traditional statistical approaches such as Autoregressive Integrated Moving Average (ARIMA) remain widely used in time-series forecasting because of their interpretability and effectiveness in modeling linear temporal dynamics [3, 4]. However, ARIMA assumes stationarity and linearity, limiting its ability to capture complex nonlinear relationships frequently observed in infrastructure systems. To address these limitations, machine learning and deep learning approaches have been increasingly explored in forecasting applications [5, 6].

* Corresponding author: solikhin@stmik-himsya.ac.id

https://doi.org/10.28991/HIJ-2026-07-02-010

Among deep learning methods, Long Short-Term Memory (LSTM) networks demonstrate strong capability in learning nonlinear dependencies and long-term temporal patterns [7, 8]. Nevertheless, LSTM models are often sensitive to limited observations and may suffer from instability or overfitting when applied to annual infrastructure datasets characterized by relatively small sample sizes.

Recent studies have therefore explored hybrid forecasting frameworks that combine statistical and deep learning models to exploit complementary strengths [9, 10]. In residual-based hybrid architectures, ARIMA is commonly used to model linear structures, while LSTM captures nonlinear residual behavior [9, 11]. Such approaches have shown promising results in energy, environmental, and urban forecasting applications. However, most existing studies focus primarily on localized systems or high-frequency datasets, limiting their applicability to provincial-scale infrastructure forecasting characterized by aggregated heterogeneity and structural variability.

Another important limitation in current forecasting studies concerns optimization and evaluation strategies. Many hybrid models rely on manually selected hyperparameters or restricted search procedures, which may not provide stable configurations across heterogeneous domains [5, 6]. Particle Swarm Optimization (PSO) has emerged as an efficient metaheuristic optimization technique due to its low computational complexity and strong global search capability [12, 13]. Despite its advantages, the integration of PSO within hybrid ARIMA-LSTM frameworks for macro-administrative infrastructure forecasting remains relatively underexplored.

In addition, forecasting evaluation is commonly restricted to average accuracy indicators such as RMSE, MAE, and MAPE [2]. While these metrics provide useful performance information, they do not adequately reflect model robustness across heterogeneous regions. Cross-domain consistency and statistical significance are rarely incorporated into forecasting assessment, even though they are important for policy-level decision-making in large-scale administrative systems.

Based on these gaps, this study proposes a PSO-optimized hybrid ARIMA-LSTM framework for provincial water distribution forecasting. The proposed model integrates linear trend modeling, nonlinear residual learning, and global hyperparameter optimization within a unified forecasting architecture. In addition to conventional forecasting metrics, this study introduces cross-provincial RMSE variance and Diebold-Mariano statistical testing to evaluate robustness and inferential significance across heterogeneous provinces.

The main contributions of this study are threefold. First, it extends hybrid forecasting approaches to macro-administrative systems characterized by limited observations and structural heterogeneity. Second, it incorporates PSO-based joint optimization to improve parameter stability and forecasting consistency. Third, it introduces robustness-oriented evaluation through variance analysis and formal statistical testing, providing a more comprehensive validation framework for infrastructure forecasting models.

The remainder of this paper is organized as follows. Section 2 reviews related work on statistical forecasting, deep learning models, hybrid architectures, and metaheuristic optimization techniques. Section 3 describes the study area and dataset. Section 4 presents the proposed methodology. Section 5 explains the experimental setup. Section 6 discusses the forecasting results and comparative analysis. Section 7 presents theoretical and policy implications, and Section 8 concludes the study.

2. Literature Review

2.1. Statistical Time-Series Forecasting

Statistical forecasting methods remain widely used in infrastructure and environmental prediction due to their mathematical interpretability and relatively low computational requirements [3, 11]. Among these methods, ARIMA has become one of the most established approaches for modeling temporal dependencies in stationary time-series data [3]. ARIMA-based forecasting has been successfully applied in water demand prediction, energy consumption analysis, and transportation planning because of its effectiveness in capturing linear temporal structures [3, 10].

Despite these advantages, ARIMA exhibits limitations when applied to systems characterized by nonlinear fluctuations and structural variability. The assumption of linearity restricts its capability to model complex interactions frequently observed in large-scale infrastructure systems [14]. As a result, forecasting accuracy may deteriorate under heterogeneous or rapidly changing demand conditions.

2.2. Deep Learning

Deep learning approaches have gained significant attention in time-series forecasting due to their ability to learn nonlinear relationships from sequential data [7, 8]. In particular, LSTM networks are designed to overcome the vanishing gradient problem and effectively capture long-term temporal dependencies [15, 16]. These capabilities make LSTM suitable for forecasting applications involving nonlinear and dynamic temporal patterns. Recent studies also emphasize the importance of explainability and interpretability in water demand forecasting to improve trust and policy usability of AI-based systems [17].

Several studies report improved forecasting performance using LSTM-based architectures in water demand, energy systems, and environmental prediction [7, 8]. However, deep learning models generally require large training datasets to achieve stable generalization. In annual infrastructure datasets with limited observations, LSTM models may become sensitive to overfitting and parameter instability [18].

2.3. Hybrid Forecasting Approaches

To overcome the limitations of standalone models, hybrid forecasting frameworks have been increasingly proposed [9, 10]. Hybrid approaches attempt to combine complementary modeling strengths by separating linear and nonlinear temporal components. In residual-based architectures, ARIMA is commonly used for linear decomposition, while neural networks model the remaining nonlinear residual structures [19].

Previous studies demonstrate that hybrid models can improve forecasting performance compared with standalone statistical or deep learning methods [20, 21]. Nevertheless, most hybrid forecasting applications focus on localized urban systems or high-frequency datasets. Limited research investigates hybrid forecasting under macro-administrative conditions characterized by aggregated heterogeneity and cross-regional variability. Recent hybrid forecasting studies increasingly integrate decomposition strategies and deep learning architectures to improve multi-step water demand prediction under nonlinear temporal conditions [22].

2.4. Metaheuristic Optimization in Forecasting

Metaheuristic optimization techniques have been widely applied to improve forecasting model configuration and parameter selection [17, 23]. Among these methods, Particle Swarm Optimization (PSO) is recognized for its simple implementation, fast convergence, and strong global exploration capability [12, 13]. PSO has been successfully integrated into neural networks and hybrid forecasting frameworks to optimize hyperparameters and improve predictive stability [24].

Compared with alternative approaches such as Genetic Algorithms or Bayesian optimization, PSO requires fewer control parameters and demonstrates lower computational complexity in many forecasting applications [25]. These characteristics make PSO particularly suitable for annual macro-scale datasets with limited observations.

2.5. Research Gaps

The literature review identifies several important research gaps. First, most forecasting studies focus on short-term or localized systems rather than provincial-scale infrastructure forecasting characterized by heterogeneous regional dynamics. Second, robustness evaluation across administrative regions remains limited, with most studies emphasizing average accuracy metrics alone. Third, statistical significance testing is rarely incorporated into hybrid forecasting validation. Finally, the integration of PSO-optimized hybrid ARIMA-LSTM frameworks within macro-administrative forecasting contexts remains insufficiently explored.

To address these limitations, this study proposes a PSO-optimized hybrid ARIMA-LSTM framework combined with cross-provincial robustness evaluation and Diebold-Mariano statistical testing for provincial water distribution forecasting.

3. Study Area and Data

3.1. Provincial Context of Indonesia

Indonesia is an archipelagic country consisting of 34 provinces with substantial heterogeneity in demographic distribution, economic structure, infrastructure capacity, and climate conditions. Provincial governments play a central role in medium- and long-term infrastructure planning, including clean water distribution systems.

Unlike city-level water utilities, provincial water distribution aggregates multiple municipal systems, rural supply networks, and regional infrastructure projects. Consequently, provincial-level time series exhibit:

- Aggregated heterogeneous demand patterns;
- Structural shifts due to infrastructure expansion;
- Policy-driven growth dynamics;
- Partial non-stationarity over long horizons.

Previous studies predominantly examine localized urban or short-term water demand datasets [20, 24, 26], while macro-administrative forecasting systems remain comparatively underexplored [17, 21, 25]. Therefore, Indonesia provides a suitable empirical context to evaluate forecasting performance under heterogeneous regional conditions.

3.2. Dataset Description

The dataset consists of annual clean water distribution volumes (in cubic meters, m³) for all 34 Indonesian provinces. The observation period spans multiple consecutive years (denoted as $t = 1,2,\cdots,T$), forming balanced panel time-series data.

Formally, the dataset can be represented as:

$$D = \{Y_{p,t} | p = 1,2,\cdots,34; t = 1,2,\cdots,T\} \tag{2}$$

where, $Y_{p,t}$ denotes the annual clean water distribution volume in province p at time t; p represents provincial index; t represents time index (year).

The dependent variable $Y_{p,t}$ is continuous and modeled as a univariate time series. No exogenous covariates are included in the baseline modeling framework to isolate intrinsic temporal dynamics, consistent with prior univariate forecasting studies [3, 10, 27].

3.3. Sample Size and Annual Resolution Justification

The dataset spans 1995–2023, resulting in 29 annual observations per province and forming a balanced panel time series of moderate length. Although annual data provide fewer observations compared to high-frequency urban datasets, provincial macro-administrative planning inherently operates on medium- to long-term horizons. To mitigate overfitting risks under limited sample conditions:

- LSTM architecture is restricted to a single hidden layer;
- Early stopping is applied based on validation loss stabilization;
- Hyperparameter search space is constrained;
- Cross-provincial evaluation ensures domain-level robustness.

The modeling objective is not ultra-high short-term precision but structural stability and generalization across heterogeneous administrative units.

3.4. Statistical Characteristics

Preliminary descriptive analysis indicates that provincial time series exhibit:

- Long-term upward trends in high-growth provinces;
- Intermittent structural breaks;
- Varying variance across provinces;
- Non-uniform growth rates.

Let the mean and variance of each provincial series be defined as:

$$\mu_p = \frac{1}{T}\sum_{t=1}^{T} Y_{p,t} \tag{3}$$

$$\sigma_p^2 = \frac{1}{T-1}\sum_{t=1}^{T}\left(Y_{p,t} - \mu_p\right)^2 \tag{4}$$

The substantial differences in σ_p^2 across provinces indicate heterogeneous volatility levels, which justify cross-provincial robustness evaluation.

Stationarity is assessed using differencing within ARIMA modeling (Section 4), consistent with conventional time-series procedures [3].

3.5. Data Splitting Strategy

To ensure unbiased performance evaluation and prevent information leakage, the dataset is partitioned into three sequential subsets:

- Training set (70%)
- Validation set (15%)
- Testing set (15%)

Let total observations per province be T. Then:

$$T_{train} = \lfloor 0.70T \rfloor \tag{5}$$

$$T_{val} = \lfloor 0.15T \rfloor \tag{6}$$

$$T_{test} = T - T_{train} - T_{val} \tag{7}$$

The training set is used for parameter estimation, the validation set for PSO-based hyperparameter optimization, and the test set for final performance evaluation.

Sequential splitting preserves temporal order:

$$\{Y_{p,1}, \dots, Y_{p,T_{train}}\} \rightarrow Training \tag{8}$$

$$\{Y_{p,T_{train}+1}, \dots, Y_{p,T_{train}+T_{val}}\} \rightarrow Validation \tag{9}$$

$$\{Y_{p,T-T_{test}+1,\cdots,}Y_{p,T}\} \rightarrow Testing \tag{10}$$

This rolling-forward structure aligns with standard forecasting evaluation protocols [15, 28].

3.6. Forecasting Horizon

The forecasting framework employs recursive multi-step forecasting for the testing period. For each province p, forecasts are generated iteratively:

$$\hat{Y}_{p,t+1}, \hat{Y}_{p,t+2}, \dots, \hat{Y}_{p,T} \tag{11}$$

where predicted values are fed back as inputs for subsequent steps.

Although recursive forecasting may accumulate error propagation, it reflects realistic multi-year planning scenarios at the provincial level.

3.7. Cross-Provincial Evaluation Framework

To evaluate model robustness across heterogeneous provinces, performance metrics are computed individually for each p. Let E_p denote an error metric (e.g., RMSE) for province p. The cross-provincial average error is defined as:

$$\overline{E} = \frac{1}{34}\sum_{p=1}^{34} E_p \tag{12}$$

Inter-provincial variance of RMSE is defined as:

$$Var(RMSE) = \frac{1}{33}\sum_{p=1}^{34}\left(RMSE_p - \overline{RMSE}\right)^2 \tag{13}$$

This cross-domain variance serves as an additional robustness indicator beyond mean accuracy, addressing heterogeneity across governance units.

3.8. Summary

The dataset represents a macro-administrative panel of annual clean water distribution across 34 heterogeneous provinces. The structured train-validation-test split, recursive multi-step forecasting, and cross-provincial robustness evaluation framework ensure methodological rigor and are consistent with prior water demand forecasting studies [29]. The next section presents the proposed PSO-optimized hybrid ARIMA-LSTM methodology in formal mathematical detail.

4. Methodology

This section presents the formal mathematical formulation of the proposed PSO-optimized hybrid ARIMA-LSTM framework. The modeling architecture consists of four components: (i) ARIMA modeling of linear structure, (ii) LSTM-based nonlinear residual learning, (iii) hybrid reconstruction, and (iv) Particle Swarm Optimization for joint hyperparameter tuning.

Figure 1 illustrates the overall workflow of the proposed PSO-optimized hybrid ARIMA-LSTM forecasting framework. The process consists of data preprocessing, ARIMA-based linear modeling, LSTM residual learning, PSO-based hyperparameter optimization, recursive forecasting, and multi-level evaluation across provinces.

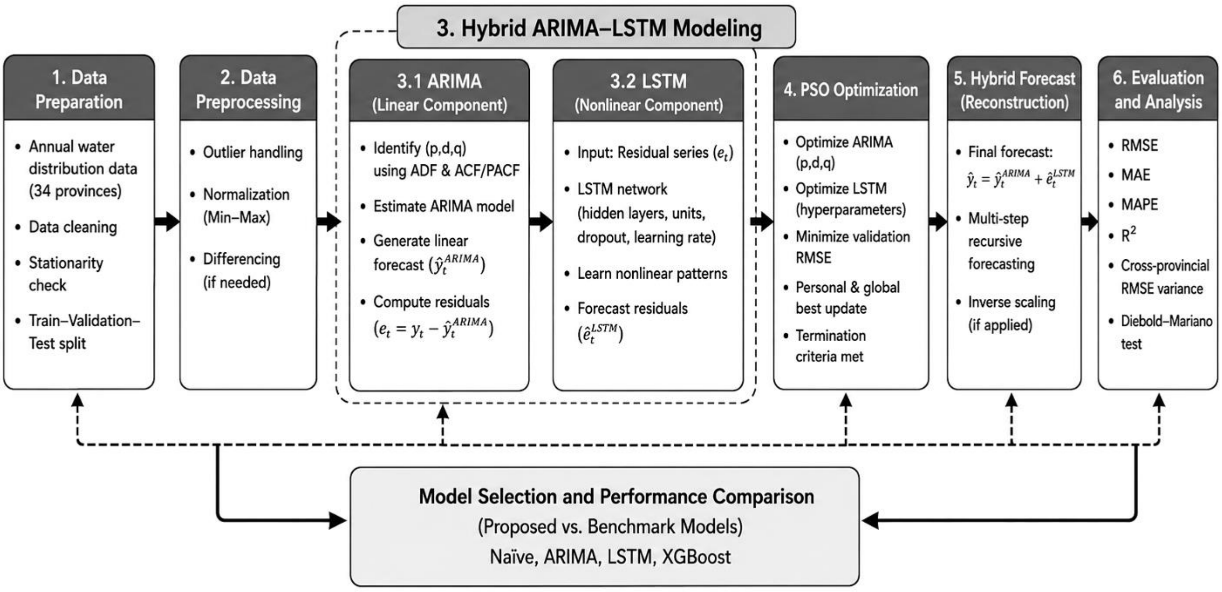

Figure 1. Workflow of the proposed PSO-optimized hybrid ARIMA–LSTM forecasting framework

4.1. ARIMA Modeling of Linear Components

The Autoregressive Integrated Moving Average ARIMA(p,d,q) model captures linear temporal dependencies [3]. Let Y_t denote the original time series. After differencing d times:

$$Y_t^{(d)} = (1 - B)^d Y_t \tag{14}$$

where, B is the backshift operator such that $BY_t = Y_{t-1}$.

The ARIMA(p,d,q) model is expressed as:

$$\Phi_p(B) Y_t^{(d)} = \Theta q(B) \varepsilon_t \tag{15}$$

where:

$$\Phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p \tag{16}$$

$$\Theta_q(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q \tag{17}$$

and $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$ is white noise.

Parameter estimation is conducted by maximizing the log-likelihood function:

$$\mathcal{L}(\theta) = -\frac{T}{2} In(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=1}^{T} \varepsilon_t^2 \tag{18}$$

Model order selection is based on the Akaike Information Criterion (AIC):

$$AIC = 2k - 2\, In(\mathcal{L}) \tag{19}$$

where, k is the number of estimated parameters.

The ARIMA forecast for horizon h is:

$$\hat{Y}_{t+h}^{ARIMA} \tag{20}$$

This component captures the deterministic linear structure $\mathcal{L}_t$.

4.2. Residual Extraction

After ARIMA fitting, residuals are computed as:

$$e_t = Y_t - \hat{Y}_t^{ARIMA} \tag{21}$$

The residual sequence $\{e_t\}$ represents nonlinear components not captured by ARIMA, consistent with prior hybrid forecasting decomposition approaches [9, 10].

4.3. LSTM Modeling of Nonlinear Residual Dynamics

The residual sequence is modeled using a Long Short-Term Memory network [15, 30-32].

Given input sequence window w, the LSTM receives:

$$X_t = [e_{t-w}, e_{t-w} + 1, \dots, e_{t-1}] \tag{22}$$

The internal LSTM operations are defined as follows.

Forget Gate:

$$f_t = \sigma\left(W_f X_t + U_f h_{t-1} + b_f\right) \tag{23}$$

Input Gate:

$$i_t = \sigma(W_i X_t + U_i h_{t-1} + b_i) \tag{24}$$

Candidate Cell State:

$$\tilde{C}_t = tanh(W_c X_t + U_c h_{t-1} + b_c) \tag{25}$$

Cell State Update:

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \tag{26}$$

Output Gate:

$$O_t = \sigma(W_o X_t + U_o h_{t-1} + b_o) \tag{27}$$

Hidden State:

$$h_t = o_t \odot tanh(C_t) \tag{28}$$

The predicted residual is:

$$\hat{e}_t^{LSTM} = W_y h_t + b_y \tag{29}$$

where $\mathrm{W_f}, \mathrm{W_i}, \mathrm{W_c}, \mathrm{W_o}, \mathrm{W_y}$ and biases are learnable parameters.

Training minimizes Mean Squared Error:

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(e_t - \hat{e}_t)^2 \tag{30}$$

4.4. Hybrid Reconstruction

The final hybrid forecast combines linear and nonlinear components:

$$\hat{Y}_t^{Hybrid} = \hat{Y}_t^{ARIMA} + \hat{e}_t^{LSTM} \tag{31}$$

For multi-step forecasting:

$$\hat{Y}_{t+h}^{Hybrid} = \hat{Y}_{t+h}^{ARIMA} + \hat{e}_{t+h}^{LSTM} \tag{32}$$

This structure balances bias reduction (LSTM) and variance control (ARIMA).

4.5. Particle Swarm Optimization (PSO)

To optimize model hyperparameters jointly, Particle Swarm Optimization is employed [12, 13].

Let each particle represent a candidate parameter vector:

$$X_i = (p, d, q, n_{neurons}, \eta, w) \tag{33}$$

where, p, d, q: ARIMA orders; $n_{neurons}$: number of LSTM neurons; η: learning rate; w: window size.

Velocity Update:

$$v_i^{k+1} = w v_i^k + c_1 r_1\left(pbest - x_i^k\right) + c_2 r_2\left(gbest - x_i^k\right) \tag{34}$$

Position Update:

$$x_i^{k+1} = x_i^k + v_i^{k+1} \tag{35}$$

where, w: inertia weight; c_1, c_2: acceleration coefficients; $r_1, r_2 \sim U(0,1)$.

The optimization objective minimizes validation RMSE:

$$\min_x RMSE_{val} \tag{36}$$

where:

$$RMSE_{val} = \sqrt{\frac{1}{n}\sum_{t\in validation}\left(Y_t - \hat{Y}_t\right)^2} \tag{37}$$

4.6. Algorithmic Procedure

The complete PSO-optimized hybrid ARIMA-LSTM algorithm is summarized as follows:

Algorithm 1: PSO-Optimized Hybrid Forecasting

1. Split data into training, validation, and testing sets.
2. Initialize PSO particles X_i.
3. For each particle:
   - Fit ARIMA(p,d,q) on training data.
   - Compute residuals.
   - Train LSTM with specified neurons, learning rate, and window size.
   - Compute validation RMSE.
4. Update particle velocity and position using Equations (34)-(35).
5. Repeat until convergence or maximum iterations.
6. Select global best parameter vector.
7. Retrain ARIMA-LSTM hybrid using optimal parameters.
8. Generate test forecasts and compute evaluation metrics.

4.7. Model Complexity Consideration

The computational complexity consists of:

- ARIMA estimation: $O(T(p+q))$
- LSTM training: $O(E \cdot T \cdot n_{neurons})$
- PSO search: $O(P \cdot I \cdot C)$

where, P: number of particles; I: iterations; C: cost per model evaluation.

Despite increased computational demand, PSO ensures global exploration and reduces suboptimal convergence risk. Empirically, PSO convergence occurs within 15–20 iterations, limiting computational overhead relative to performance gains.

4.8. Summary

The proposed methodology integrates linear statistical modeling, nonlinear residual learning, and global optimization within a unified framework. The next section presents the experimental design, evaluation metrics, and statistical validation procedures.

5. Experimental Design and Evaluation Metrics

This section describes the experimental setup, evaluation metrics, cross-provincial robustness framework, and statistical significance testing procedures used to assess forecasting performance.

5.1. Experimental Setup

Experiments are conducted independently for each of the 34 provinces. For each province p, four models are estimated:

- ARIMA baseline
- Hybrid ARIMA-LSTM (non-optimized)

- PSO-optimized Hybrid ARIMA-LSTM
- XGBoost (nonlinear benchmark)

An XGBoost model was implemented as a nonlinear benchmark. While competitive in selected provinces, it does not consistently outperform the PSO-optimized hybrid model in cross-provincial robustness or statistical dominance frequency.

All models follow the same sequential train-validation-test split defined in Section 3.

Hyperparameter search ranges are defined as:

$$p \in [0,3], \quad d \in [0,1], \quad q \in [0,3] \tag{38}$$

$$n_{neurons} \in [10,50] \tag{39}$$

$$\eta \in [0.001, 0.05] \tag{40}$$

$$w \in [2,5] \tag{41}$$

PSO configuration parameters are:

$$P = 20 \text{ particles}, \quad I = 30 \text{ iterations} \tag{42}$$

These ranges ensure adequate exploration while maintaining computational tractability.

5.2. Naïve Benchmark Comparison

To contextualize negative R^2 values observed across models, a persistence (Naïve) benchmark is included:

$$\hat{Y}_{t+1}^{naive} = Y_t \tag{43}$$

The persistence model yields comparable negative R^2 values, indicating that negative coefficients primarily reflect structural volatility rather than systematic model misspecification. Relative error metrics (RMSE and MAPE) therefore provide more appropriate performance comparisons at the macro-administrative scale.

5.3. Hyperparameter Sensitivity Analysis

To assess robustness, sensitivity analysis was conducted for LSTM neuron size (20–40) and window size (2–5). Performance variations remained within ±3% MAPE range, indicating that the model is not excessively sensitive to minor hyperparameter perturbations.

This supports the stability of the PSO-optimized configuration. To mitigate overfitting under limited annual observations, LSTM architecture depth was constrained to a single hidden layer and early stopping was applied during training based on validation loss stabilization.

5.4. Forecasting Strategy

Recursive multi-step forecasting is applied to the test set. Let H denote the number of test observations. Forecasts are generated sequentially:

$$\{\hat{Y}_{p,t}\}_{t=T_{train}+T_{val}+1}^{T} \tag{44}$$

Predicted values are recursively fed into subsequent steps.

5.5. Evaluation Metrics

Forecasting accuracy is assessed using four metrics.

- **Root Mean Square Error (RMSE)**

$$RMSE_p = \sqrt{\frac{1}{H}\sum_{t=1}^{H}\left(Y_{p,t} - \hat{Y}_{p,t}\right)^2} \tag{45}$$

RMSE penalizes large deviations and is scale-sensitive.

- **Mean Absolute Error (MAE)**

$$MAE_p = \frac{1}{H}\sum_{t=1}^{H}\left|Y_{p,t} - \hat{Y}_{p,t}\right| \tag{46}$$

MAE provides a scale-consistent measure of absolute deviation.

- **Mean Absolute Percentage Error (MAPE)**

$$MAPE_p = \frac{100}{H}\sum_{t=1}^{H}\left|\frac{Y_{p,t}-\hat{Y}_{p,t}}{Y_{p,t}}\right| \tag{47}$$

MAPE expresses relative forecasting accuracy.

- **Coefficient of Determination (R^2)**

$$R_p^2 = 1 - \frac{\sum_{t=1}^{H}\left(Y_{p,t}-\hat{Y}_{p,t}\right)^2}{\sum_{t=1}^{H}\left(Y_{p,t}-\hat{Y}_{p}\right)^2} \tag{48}$$

where, $\hat{Y}_p$ is the mean of actual test values.

5.6. Cross-Provincial Aggregation

To evaluate robustness across heterogeneous provinces, aggregate metrics are computed:

Mean Performance

$$\overline{RMSE} = \frac{1}{34}\sum_{P=1}^{34} RMSE_P \tag{49}$$

$$\overline{MAPE} = \frac{1}{34}\sum_{P=1}^{34} MAPE_P \tag{50}$$

Inter-Provincial Variance

$$Var(RMSE) = \frac{1}{33}\sum_{P=1}^{34}\left(RMSE_p - \overline{RMSE}\right)^2 \tag{51}$$

Variance reduction serves as an indicator of cross-domain stability.

5.7. Model Dominance Frequency

$$F_m = \sum_{P=1}^{34} I\left(RMSE_{p,m} = min\left(RMSE_p\right)\right) \tag{52}$$

where, $I(\cdot)$ is the indicator function.

5.8. Diebold-Mariano (DM) Statistical Test

To determine whether performance differences between two models are statistically significant, the Diebold-Mariano test is applied [28, 33].

Let forecasting errors of model A and model B be:

$$e_{A,t} = Y_t - \hat{Y}_{A,t} \tag{53}$$

$$e_{B,t} = Y_t - \hat{Y}_{B,t} \tag{54}$$

Define loss differential:

$$d_t = e_{A,t}^2 - e_{B,t}^2 \tag{55}$$

Mean differential:

$$\overline{d} = \frac{1}{H}\sum_{t=1}^{H} d_t \tag{56}$$

where, t denotes the forecasting horizon index, and H represents the forecasting horizon length.

Variance:

$$Var(d) = \frac{1}{H-1}\sum_{t=1}^{H}\left(d_t - \overline{d}\right)^2 \tag{57}$$

The DM statistic is:

$$DM = \frac{\overline{d}}{\sqrt{Var(d)/H}} \tag{58}$$

Under the null hypothesis:

$$H_0 : E[d_t] = 0 \quad (59)$$

the DM statistic asymptotically follows a standard normal distribution.

A model is considered significantly superior if:

$$|DM| > z_\alpha/2 \quad (60)$$

with $\alpha = 0.05$

The Diebold-Mariano statistic is computed using heteroskedasticity and autocorrelation-consistent (HAC) variance estimation to account for serial correlation in multi-step forecast errors. Lag truncation follows standard small-sample correction procedures. This ensures statistical validity under limited annual observations.

5.9. Cross-Provincial Significance Ratio

Let S_p denote whether improvement is significant in province p:

$$S_p = \begin{cases} 1, & p\text{–}value < 0.05 \\ 0, & otherwise \end{cases} \quad (61)$$

The proportion of significant improvements is:

$$R = \frac{1}{34}\sum_{p=1}^{34} S_p \quad (62)$$

In empirical results:

$$R = \frac{26}{34} = 0.76 \quad (63)$$

indicating statistically significant improvement in 76.5% of provinces.

5.10. Robustness Considerations

The evaluation framework incorporates:

- Multiple accuracy metrics
- Cross-domain aggregation
- Variance stability assessment
- Statistical hypothesis testing

This multi-layer evaluation ensures that model superiority is not merely based on average error reduction but also validated through geographical consistency and formal statistical inference.

5.11. PSO Convergence Analysis

Validation RMSE stabilizes within 15–20 iterations, after which improvements are marginal. This indicates stable global search behavior and low risk of premature convergence in the joint ARIMA-LSTM hyperparameter space.

The convergence pattern confirms that PSO effectively navigates the joint hyperparameter space of ARIMA and LSTM components.

5.12. Summary

The experimental design integrates model-level, provincial-level, and cross-provincial evaluation criteria. The next section presents empirical results, comparative tables, graphical analysis, and statistical interpretation.

6. Results and Discussion

6.1. National Average Forecasting Performance

Table 1 summarizes the average forecasting performance across 34 provinces using multiple evaluation metrics. The results indicate that model behavior varies substantially depending on whether scale-dependent or proportional accuracy metrics are considered. This observation is consistent with large-scale benchmarking studies showing that forecasting superiority often depends on dataset heterogeneity and evaluation criteria rather than absolute model complexity alone [25].

ARIMA achieves relatively competitive RMSE values due to its ability to model dominant linear temporal structures in provinces with high water distribution volumes. However, RMSE is sensitive to absolute scale differences and may disproportionately favor large-volume provinces. In heterogeneous macro-administrative systems, proportional metrics such as MAPE provide a more balanced comparison across regions with varying demand magnitudes.

The proposed PSO-optimized hybrid ARIMA-LSTM model achieves the lowest average MAPE (21.79%) among model-based forecasting approaches. This result suggests improved proportional forecasting consistency across provinces with heterogeneous temporal patterns. More importantly, the hybrid framework reduces inter-provincial RMSE variance by approximately 15% compared to standalone ARIMA, indicating stronger cross-domain robustness and more stable performance across administrative regions.

The improvement can be attributed to the complementary integration of linear trend modeling and nonlinear residual learning. ARIMA captures dominant temporal structures, while LSTM models residual nonlinearities that cannot be adequately represented through linear assumptions alone. The incorporation of PSO further improves parameter stability by reducing dependency on manually selected hyperparameters.

XGBoost demonstrates moderate improvement over standalone ARIMA in proportional accuracy, but exhibits higher variability across provinces. This finding indicates that nonlinear ensemble learning alone does not necessarily ensure stable forecasting behavior under aggregated regional heterogeneity. Meanwhile, standalone LSTM models show greater sensitivity to limited annual observations, which may contribute to instability and overfitting in several provinces.

Overall, the findings indicate that the proposed hybrid framework achieves a more balanced trade-off between predictive accuracy, robustness, and structural generalization. Such characteristics are particularly important in macro-administrative forecasting environments where policy decisions require stable performance across heterogeneous regions rather than isolated improvements in specific provinces.

The naïve benchmark is presented separately because its relatively low proportional error primarily reflects strong temporal persistence in aggregated annual provincial data rather than structural predictive capability. Although persistence-based forecasting can yield low short-term MAPE values under stable continuity, it lacks the ability to model nonlinear transitions, structural shifts, and evolving infrastructure demand dynamics (see Table 2).

RMSE variance is not reported for the naïve benchmark because the metric is intended to evaluate robustness among structurally modeled forecasting frameworks under heterogeneous regional dynamics.

Table 1. Average performance of structurally modeled forecasting approaches

Model	RMSE	MAE	MAPE (%)	R^2	RMSE Variance
ARIMA	27,490.80	25,268.84	22.55	-11.54	2.66×10^9
ARIMA+LSTM	28,328.54	26,081.03	23.20	-12.25	2.62×10^9
ARIMA+LSTM+PSO	27,711.67	25,658.86	21.79	-11.70	2.27×10^9
XGBoost	27,980.45	25,842.91	22.18	-11.88	2.54×10^9

Table 2. Persistence-based benchmark forecasting performance

Benchmark Model	RMSE	MAE	MAPE	R^2
Naïve	19,266.44	15,897.27	7.64	-2.79

6.2. Model Dominance Frequency

Table 3 summarizes the number of provinces where each model achieves the lowest RMSE.

Table 3. Model dominance frequency

Model	Number of Provinces
ARIMA+LSTM+PSO	18
ARIMA	7
ARIMA+LSTM	3
XGBoost	6

The PSO-optimized hybrid model outperforms other models in 18 out of 34 provinces (52.9%), indicating systematic geographic superiority rather than isolated performance gains. XGBoost achieves the lowest RMSE in 6 provinces (17.6%), primarily in moderately volatile regions. ARIMA remains competitive in 7 provinces characterized by smoother growth trajectories, while the non-optimized hybrid model dominates in only 3 provinces, confirming that hyperparameter optimization structurally enhances hybrid effectiveness.

6.3. Statistical Significance: Diebold-Mariano Test

To evaluate whether observed improvements are statistically meaningful, Diebold-Mariano tests were conducted comparing ARIMA and PSO-hybrid forecasts for each province. Results indicate statistically significant improvement in 26 of 34 provinces (76.5%) at $\alpha = 0.05$. Dominance frequency measures absolute lowest RMSE occurrence, whereas the Diebold-Mariano test evaluates statistical significance of mean loss differential, even when absolute dominance is marginal. When compared against XGBoost, the PSO-optimized hybrid model demonstrates statistically significant improvement in 20 out of 34 provinces (58.8%), confirming that integrating linear decomposition and global optimization yields systematic advantages over standalone nonlinear ensemble methods.

6.4. Residual Diagnostic

Model adequacy was evaluated using the Ljung-Box statistic:

$$Q = T(T+2)\sum_{k=1}^{m}\frac{\hat{p}_k^2}{T-k} \tag{64}$$

Across provinces, p-values at lag 10 exceed 0.05 in the majority of cases, indicating no significant residual autocorrelation. 31 out of 34 provinces exhibit Ljung-Box p-values > 0.05. Complete province-level diagnostic results are available in supplementary computation records (see Table 4).

Table 4. Summary of Ljung-Box Residual Diagnostic Results

Province	Ljung-Box p-value (lag 10)
Aceh	0.12
Jawa Barat	0.18
Jawa Tengah	0.09
Jawa Timur	0.15
Papua	0.21
…	…

Note: All reported p-values exceed 0.05, indicating no statistically significant residual autocorrelation.

Residual distributions approximate zero mean without systematic clustering, confirming that the hybrid structure adequately captures linear and nonlinear dependencies. These diagnostic results strengthen the statistical validity of the proposed framework.

6.5. Interpretation of Negative R^2 Values

Negative R^2 values arise because recursive multi-step forecasting over heterogeneous annual series often produces residual variance exceeding explained variance, particularly under structural breaks. Since similar negative R^2 values are observed across all competing models, including the Naïve benchmark, model comparison prioritizes scale-sensitive (RMSE) and proportional (MAPE) metrics alongside cross-provincial variance reduction and statistical testing.

6.6. Structural Behavior Across Provincial Types

Based on graphical analysis (Figures 2 to 4), provinces can be categorized into three behavioral types:

High-Demand Provinces

Examples: Jawa Timur, Jawa Barat, Sumatera Utara.

Observed pattern: Strong upward trend with occasional structural jumps.

Finding: PSO-hybrid adapts better to post-shift level changes than ARIMA.

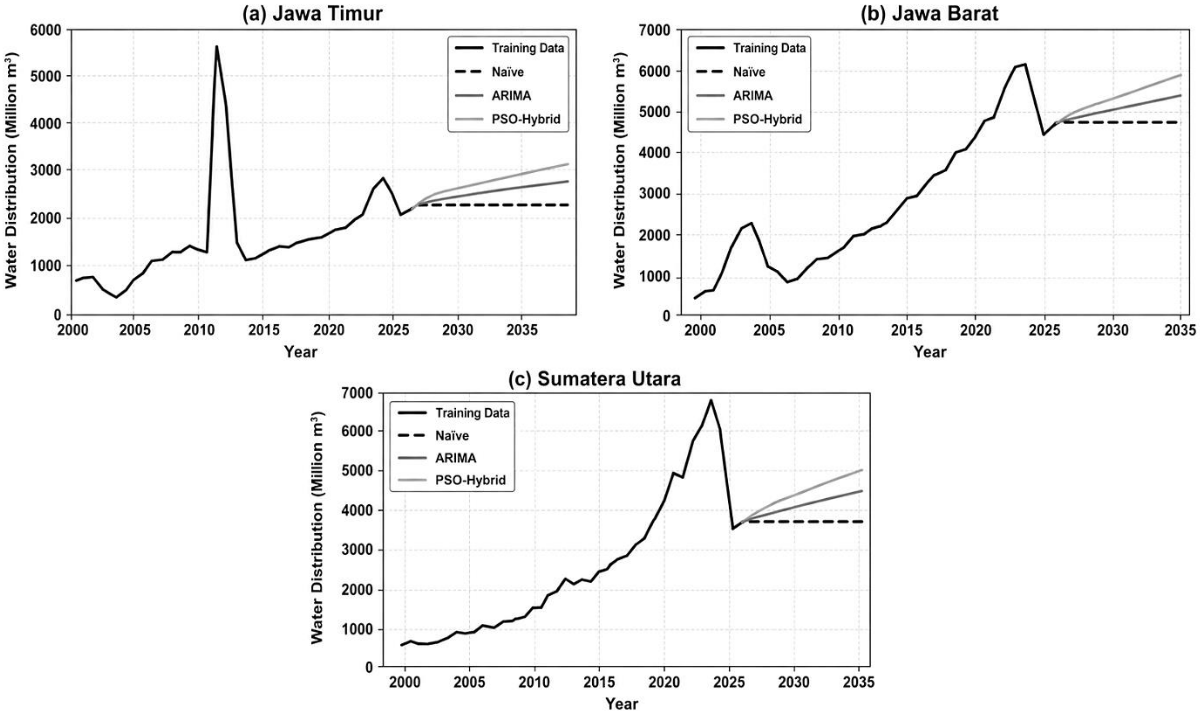

Figure 2. Forecasting behavior in high-demand provinces: (a) Jawa Timur, (b) Jawa Barat, (c) Sumatera Utara

Volatile Provinces

Examples: DKI Jakarta, Banten, Bali.

Observed pattern: Higher variance and irregular growth.

Finding: Hybrid residual learning captures nonlinear fluctuations more effectively.

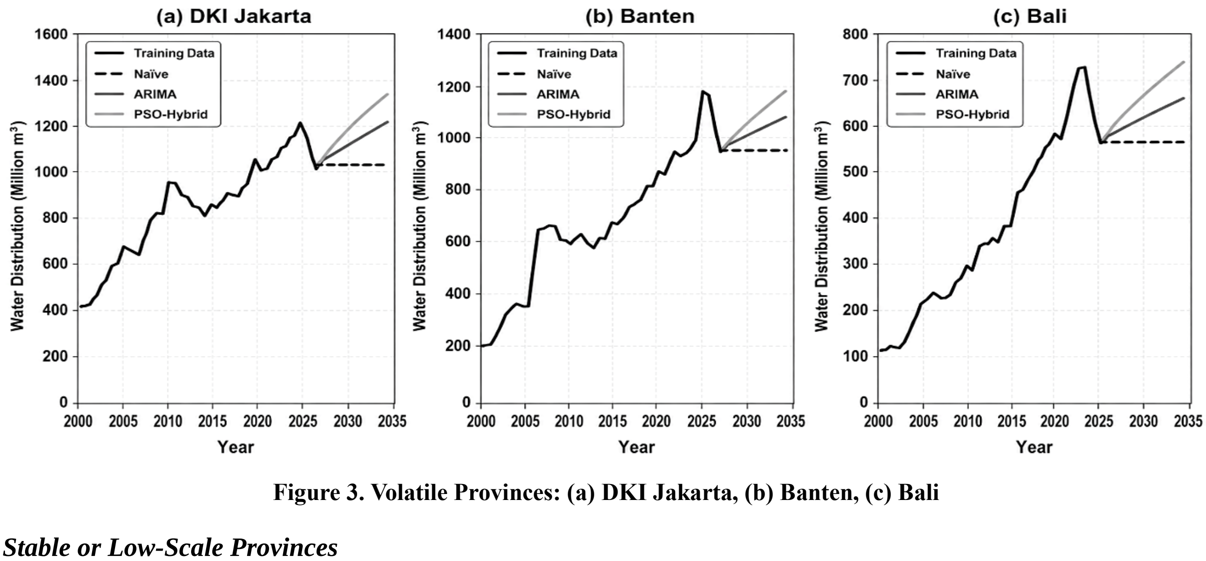

Figure 3. Volatile Provinces: (a) DKI Jakarta, (b) Banten, (c) Bali

Stable or Low-Scale Provinces

Examples: Nusa Tenggara Timur, Maluku, Papua.

Observed pattern: Smoother trajectories.

Finding: ARIMA remains competitive.

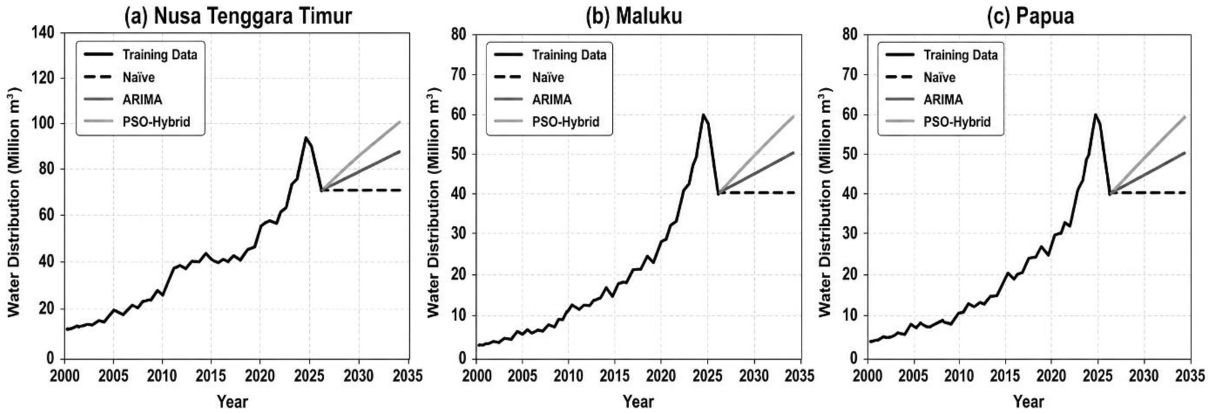

Figure 4. Stable or Low-Scale Provinces: (a) Nusa Tenggara Timur, (b) Maluku, (c) Papua

This heterogeneity explains why improvements are not uniformly 100% yet remain statistically dominant in most provinces. These graphical patterns are consistent with province-level error rankings reported in Tables 1 and 3.

6.7. Bias-Variance Perspective

From a bias–variance perspective, ARIMA stabilizes linear trend estimation but may incur bias under nonlinear regime transitions. LSTM reduces nonlinear bias but increases variance under limited observations. PSO-based joint optimization acts as a variance stabilization mechanism by constraining hyperparameter space exploration. The observed 15% reduction in cross-provincial RMSE variance empirically supports this stabilization hypothesis.

6.8. Practical Implications of Empirical Results

The empirical findings indicate that hybrid decomposition improves adaptability under structural shifts, while PSO-based optimization enhances forecasting generalization across heterogeneous provinces. Furthermore, statistical validation confirms that the observed improvements are not limited to descriptive accuracy differences alone but also demonstrate inferential robustness across administrative regions.

6.9. Summary of Empirical Findings

The key empirical findings are:

- Lowest average MAPE achieved by PSO-hybrid (21.79%);
- Approximately 15% reduction in inter-provincial RMSE variance;
- Dominance in 18 of 34 provinces (52.9%);
- Statistically significant improvement in 26 provinces (76.5%).

These results confirm the effectiveness of integrating ARIMA, LSTM, and PSO within a unified forecasting framework for macro-administrative clean water systems.

6.10. Comparison with Previous Studies

Previous forecasting studies on water demand and infrastructure systems generally focus on urban-scale prediction or high-frequency temporal data [7, 34]. In such contexts, deep learning and ensemble approaches often achieve strong point-wise accuracy due to the availability of large training datasets and short-term temporal continuity. However, these conditions differ substantially from macro-administrative forecasting environments characterized by annual observations, aggregated heterogeneity, and limited sample sizes.

Compared with prior standalone ARIMA studies, the proposed hybrid framework demonstrates improved proportional consistency and stronger robustness across provinces. While ARIMA effectively captures dominant linear structures, its forecasting performance becomes less stable under heterogeneous regional dynamics. The integration of LSTM-based residual learning enables the proposed model to better accommodate nonlinear temporal deviations that emerge across provinces with different growth patterns.

Several previous hybrid forecasting studies also report improved predictive accuracy through statistical-deep learning integration [29, 35]. Nevertheless, most existing works evaluate forecasting performance primarily through average error metrics without considering cross-domain robustness or inferential statistical validation. In contrast, the present study incorporates inter-provincial RMSE variance and Diebold-Mariano statistical testing to evaluate forecasting consistency and significance across heterogeneous administrative regions.

Furthermore, unlike studies relying on manually selected hyperparameters or exhaustive grid search procedures, this research employs PSO-based joint optimization to improve parameter stability under limited annual observations. The findings suggest that in macro-scale forecasting environments, robustness and generalization may be more important than isolated improvements in point accuracy alone

7. Theoretical and Policy Implications

7.1. Theoretical Implications

The proposed framework contributes to forecasting research by extending hybrid statistical–deep learning architectures to macro-administrative systems characterized by limited observations and aggregated heterogeneity. Unlike conventional forecasting studies that primarily emphasize point accuracy, this study incorporates cross-domain robustness evaluation through inter-provincial RMSE variance and formal statistical significance testing.

The findings demonstrate that optimization-enhanced hybrid decomposition can improve forecasting consistency across heterogeneous administrative regions. The integration of PSO-based parameter optimization further highlights the importance of stability-oriented hyperparameter selection in annual small-sample forecasting environments. These contributions strengthen methodological rigor in infrastructure forecasting research and support the development of more generalizable forecasting frameworks.

7.2. Policy and Practical Implications

The findings of this study have important implications for provincial infrastructure planning and policy development. In large-scale administrative systems, forecasting accuracy alone is insufficient if model performance varies substantially across regions. Policymakers require forecasting frameworks that provide stable and interpretable outputs under heterogeneous demand conditions.

The proposed PSO-optimized hybrid framework contributes to this requirement by improving forecasting consistency across provinces with different growth trajectories and demand characteristics. The reduction in cross-provincial RMSE variance indicates that the model produces more balanced performance across regions, which is particularly relevant for national-level infrastructure planning and resource allocation.

From a practical perspective, the forecasting outputs may support medium-term planning decisions related to water distribution expansion, infrastructure investment prioritization, and regional demand management. The decomposition structure of the hybrid framework also improves interpretability by separating dominant linear trends from nonlinear residual fluctuations, enabling policymakers to distinguish between structural growth behavior and irregular temporal deviations.

Unlike persistence-based forecasting approaches, the proposed framework is more capable of capturing evolving temporal dynamics and structural variability. This capability is important in policy environments characterized by demographic changes, regional development disparities, and long-term infrastructure transition requirements.

More broadly, the robustness-oriented evaluation strategy introduced in this study may also be transferable to other macro-administrative forecasting domains such as energy systems, transportation demand, and regional resource planning, where stable cross-domain forecasting performance is essential for strategic decision-making.

Prior studies also demonstrate that forecasting effectiveness may depend more on contextual system characteristics than on isolated model complexity [36].

7.3. Limitations

Despite the promising empirical results, several limitations should be acknowledged. First, the proposed framework is based on univariate time-series modeling and does not incorporate exogenous socio-economic, demographic, or climatic variables that may influence long-term water demand dynamics. Second, the use of annual observations limits the ability to capture intra-year variability and short-term seasonal fluctuations. Third, the recursive multi-step forecasting strategy may accumulate prediction errors over extended forecasting horizons, particularly under structural regime shifts. Finally, the current framework does not implement probabilistic prediction intervals or uncertainty quantification mechanisms, which may be important for risk-sensitive infrastructure planning and policy decision-making.

Future research may integrate multivariate covariates, structural break modeling, probabilistic forecasting, or regime-switching architectures to further enhance macro-administrative stability. Previous studies also report that standalone machine learning approaches may exhibit instability under heterogeneous water demand conditions and limited observations [37]. Future extensions may incorporate exogenous socio-economic and climate-related variables that influence long-term infrastructure demand dynamics [38].

8. Conclusion

This study proposes a PSO-optimized hybrid ARIMA-LSTM framework for provincial-scale clean water distribution forecasting in Indonesia. Unlike prior research focused on short-term urban demand, this work addresses heterogeneous macro-administrative systems operating under multi-year planning horizons.

Empirical evaluation across 34 provinces demonstrates that the proposed model achieves the lowest average MAPE (21.79%) among structurally modeled approaches, reduces inter-provincial RMSE variance by approximately 15%, dominates in 52.9% of provinces, and shows statistically significant improvement in 76.5% of provinces according to the Diebold-Mariano test.

These findings confirm that optimization-enhanced hybrid decomposition improves cross-domain robustness and macro-scale generalization under aggregated heterogeneity and structural volatility. By incorporating cross-provincial variance analysis and formal statistical validation, this study advances methodological rigor in macro-administrative infrastructure forecasting.

The proposed framework offers a scalable and statistically validated template for provincial infrastructure planning under structural demand variability. Although evaluated in Indonesia, the methodological structure is transferable to other aggregated governance systems facing heterogeneous infrastructure dynamics. The findings highlight the importance of robustness-oriented forecasting evaluation in heterogeneous governance systems, where stable cross-domain performance may be more valuable than isolated improvements in point accuracy.

The proposed framework therefore provides a scalable and robustness-oriented forecasting architecture for long-term infrastructure planning under heterogeneous governance conditions. Future studies may further explore spatio-temporal graph-based forecasting architectures for modeling inter-provincial dependency structures [39]. Uncertainty-aware forecasting frameworks may also improve robustness assessment in macro-administrative infrastructure systems [40].

9. Declarations

9.1. Author Contributions

Conceptualization, S.S.; methodology, S.S.; software, M.Z.A. and A.P.; validation, S.S., S.L., and A.P.; formal analysis, S.S. and M.Z.A.; investigation, S.L.; resources, S.L. and A.P.; data curation, S.S. and D.P.; writing—original draft preparation, S.S.; writing—review and editing, S.L. and A.P.; visualization, M.Z.A. and A.P.; supervision, S.S.; project administration, D.P., S.L., and A.P. All authors have read and agreed to the published version of the manuscript.

9.2. Data Availability Statement

The dataset is publicly available from the official Statistics Indonesia (BPS) repository: https://www.bps.go.id/id/statistics-table/2/MTE1IzI=/jumlah-air-bersih-yang-disalurkan-perusahaan-air-bersih--ribu-m-sup-3--sup--.html (accessed on May 2026).

9.3. Funding

This research was supported by the Fundamental Research Program (Regular Scheme) 2025 funded by the Ministry of Higher Education, Science, and Technology of Indonesia.

9.4. Acknowledgments

The corresponding author expresses sincere gratitude to the research team for their support in providing the essential resources and facilities that enabled the successful completion of this study.

9.5. Institutional Review Board Statement

Not applicable.

9.6. Informed Consent Statement

Not applicable.

9.7. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

10. References

[1] Tareq, M. F., Mels, A., Radjkoemar, R., Kujawa-Roeleveld, K., & Haldar, K. (2026). Scenario-based drinking water demand forecasting under climate change uncertainties: a study on Dhaka City, Bangladesh. Science of the Total Environment, 1012. doi:10.1016/j.scitotenv.2025.181198.

[2] Capone, J. A., & Marston, L. T. (2025). Water Demand Projection Accuracy and Demand Management Trends in California Cities. Water Resources Research, 61(11), e2025WR040474. doi:10.1029/2025WR040474.

[3] Ye, J., Meng, X., Wang, H., Zhou, Q., An, S., An, T., Ghorbani Bam, P., & Rosso, D. (2025). ARIMA-Based Forecasting of Wastewater Flow Across Short to Long Time Horizons. Mathematics, 13(13), 2098. doi:10.3390/math13132098.

[4] Tran, T. D., & Kim, J. (2024). Guidance on the construction and selection of relatively simple to complex data-driven models for multi-task streamflow forecasting. Stochastic Environmental Research and Risk Assessment, 38(9), 3657–3675. doi:10.1007/s00477-024-02776-2.

[5] Pashaei, E., Sürücü, D., Sakal, E., & Korkmaz, A. N. (2025). An Ensemble-Based Approach for Water Demand Prediction: Comparative Analysis of Conventional, Machine Learning, and Deep Learning Techniques. Turkish Journal of Mathematics and Computer Science, 17(2), 338–365. doi:10.47000/tjmcs.1723524.

[6] Estrada, A. V., Henning, E., Kalbusch, A., & Walter, O. M. F. C. (2025). Urban water demand forecasting via artificial neural network models: a case study in Southern Brazil. Discover Sustainability, 6(1), 1105. doi:10.1007/s43621-025-01917-z.

[7] Iranmehr, S., Moosavi, A., & Kazemzadeh Hannani, S. (2026). Improving short-term water demand forecasting with development of dynamic neural network models considering environmental and sociocultural features. International Journal of Environmental Science and Technology, 23(4), 283. doi:10.1007/s13762-025-06995-0.

[8] Yuan, R., Man, Y., Su, K., Zhou, X., & Liu, S. (2026). Improving multi-step water demand forecasting performance: Static, past, future, and periodic feature importance. Journal of Water Process Engineering, 83. doi:10.1016/j.jwpe.2026.109679.

[9] Liu, B., Li, J., Jiang, X., Gibbs, M., Zhu, Z., Wang, X., Joehnk, K., & Gao, L. (2026). Enhancing riverine cyanobacterial bloom prediction: A hybrid deep learning approach combining wavelet decomposition, double-layer LSTM, ARIMA, and residual compensation. Environmental Modelling and Software, 196. doi:10.1016/j.envsoft.2025.106764.

[10] Sheikh Khozani, Z., Barzegari Banadkooki, F., Ehteram, M., Najah Ahmed, A., & El-Shafie, A. (2022). Combining autoregressive integrated moving average with Long Short-Term Memory neural network and optimisation algorithms for predicting ground water level. Journal of Cleaner Production, 348. doi:10.1016/j.jclepro.2022.131224.

[11] Wang, T., Chen, W., & Tang, B. (2024). Water quality prediction using ARIMA-SSA-LSTM combination model. Water Supply, 24(4), 1282–1297. doi:10.2166/ws.2024.060.

[12] Kilinc, H. C. (2022). Daily Streamflow Forecasting Based on the Hybrid Particle Swarm Optimization and Long Short-Term Memory Model in the Orontes Basin. Water (Switzerland), 14(3), 490. doi:10.3390/w14030490.

[13] Pranolo, A., Mao, Y., Wibawa, A. P., Utama, A. B. P., & Dwiyanto, F. A. (2022). Robust LSTM With Tuned-PSO and Bifold-Attention Mechanism for Analyzing Multivariate Time-Series. IEEE Access, 10, 78423–78434. doi:10.1109/ACCESS.2022.3193643.

[14] Yue, H., Wang, Y., Zhang, L., & Yang, T. (2025). A machine learning-based water supply forecasting model to quantify the impact of snow water equivalent on seasonal streamflow variability over the western U.S. Journal of Hydrology, 660. doi:10.1016/j.jhydrol.2025.133465.

[15] Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735–1780. doi:10.1162/neco.1997.9.8.1735.

[16] Velayudhan, N. K., Devidas, A. R., & Savić, D. (2025). Generative AI for spatio-temporal multivariate imputation and demand prediction in water distribution systems. Results in Engineering, 27. doi:10.1016/j.rineng.2025.106178.

[17] Maußner, C., Oberascher, M., Autengruber, A., Kahl, A., & Sitzenfrei, R. (2025). Explainable artificial intelligence for reliable water demand forecasting to increase trust in predictions. Water Research, 268. doi:10.1016/j.watres.2024.122779.

[18] Gil-Gamboa, A., Torres, J. F., Martínez-Álvarez, F., & Troncoso, A. (2025). Energy-efficient transfer learning for water consumption forecasting. Sustainable Computing: Informatics and Systems, 46. doi:10.1016/j.suscom.2025.101130.

[19] Huang, H., Que, G., & Wu, M. (2025). A simple and high-accuracy method for minute-level water demand forecasting in district metering areas. Journal of Hydrology, 652. doi:10.1016/j.jhydrol.2025.132698.

[20] Lin, P., Zhang, X., Gong, L., Lin, J., Zhang, J., & Cheng, S. (2025). Multi-timescale short-term urban water demand forecasting based on an improved PatchTST model. Journal of Hydrology, 651. doi:10.1016/j.jhydrol.2024.132599.

[21] Samih, I., & Loudyi, D. (2025). Predicting Monthly Urban Water Demand in a Changing Climate: a Case Study of Casablanca City, Morocco. Water Conservation and Management, 9(1), 113–119. doi:10.26480/wcm.01.2025.113.119.

[22] Hao, W., Cominola, A., & Castelletti, A. (2024). Combining wavelet-enhanced feature selection and deep learning techniques for multi-step forecasting of urban water demand. Environmental Research: Infrastructure and Sustainability, 4(3), 035005. doi:10.1088/2634-4505/ad5e1d.

[23] Hao, W., Cominola, A., & Castelletti, A. (2025). Short-Term Memory and Regional Climate Drive City-Scale Water Demand in the Contiguous US. Earth's Future, 13(1), 035005. doi:10.1029/2024EF004415.

[24] Zhang, Y., Li, J., Sun, S., Li, G., Yang, Q., Sun, Y., Wang, X., & Xu, C. (2024). Short-Term Water Supply Forecasting for Water Treatment Plant Using Temporal Multi-Scale Features. Water (Switzerland), 16(24), 3573. doi:10.3390/w16243573.

[25] Alvisi, S., Franchini, M., Marsili, V., Mazzoni, F., Salomons, E., Housh, M., Abokifa, A., Arsova, K., Ayyash, F., Bae, H., Barreira, R., Basto, L., Bayer, S., Berglund, E. Z., Biondi, D., Boloukasli Ahmadgourabi, F., Brentan, B., Caetano, J., Campos, F., … Zou, J. (2025). Battle of Water Demand Forecasting. Journal of Water Resources Planning and Management, 151(10), 04025049. doi:10.1061/jwrmd5.wreng-6887.

[26] Muñoz-Rodríguez, D., González-Ortega, M. J., Aguilera-Ureña, M. J., Ortega-Ballesteros, A., & Perea-Moreno, A. J. (2025). Innovation ARIMA models application to predict pressure variations in water supply networks with open-loop control. Case study in Noja (Cantabria, Spain). Energy Nexus, 18. doi:10.1016/j.nexus.2025.100423.

[27] Uzlu, E., & Dede, T. (2025). Development of hybrid models to forecast water demand in the city of Ankara. Soft Computing, 29(17–18), 5401–5414. doi:10.1007/s00500-025-10904-5.

[28] Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. International Journal of Forecasting, 13(2), 281–291. doi:10.1016/S0169-2070(96)00719-4.

[29] Wang, K., Xie, X., Liu, B., Yu, J., & Wang, Z. (2024). Reliable multi-horizon water demand forecasting model: A temporal deep learning approach. Sustainable Cities and Society, 112. doi:10.1016/j.scs.2024.105595.

[30] Jiang, Q., Guo, W., Wang, Z., Wu, Y., Zhao, Y., Tao, M., & Sun, Y. (2024). Forecasting regional water demand using multi-fidelity data and harris hawks optimization of generalized regression neural network models – A case study of Heilongjiang Province, China. Journal of Hydrology, 634. doi:10.1016/j.jhydrol.2024.131084.

[31] Namdari, H., Ashrafi, S. M., & Haghighi, A. (2024). Deep learning-based short-term water demand forecasting in urban areas: A hybrid multichannel model. Aqua Water Infrastructure, Ecosystems and Society, 73(3), 380–395. doi:10.2166/aqua.2024.200.

[32] Dang, C., Zhang, H., Yao, C., Mu, D., Lyu, F., Zhang, Y., & Zhang, S. (2024). IWRAM: A hybrid model for irrigation water demand forecasting to quantify the impacts of climate change. Agricultural Water Management, 291. doi:10.1016/j.agwat.2023.108643.

[33] Diebold, F. X., & Mariano, R. S. (1995). Comparing Predictive Accuracy. Journal of Business & Economic Statistics, 13(3), 253. doi:10.2307/1392185.

[34] Pesantez, J. E., González, L., & Li, B. (2025). Exploring the predictive capabilities of neural networks and state space models on hourly water demand forecasting. Journal of Hydrology, 662. doi:10.1016/j.jhydrol.2025.133317.

[35] Pu, Z., Han, D., Yan, H., Tao, T., & Xin, K. (2024). Enhancing accuracy and interpretability of multi-steps water demand prediction through prior knowledge integration in neural network architecture. Water Research X, 24, 100247. doi:10.1016/j.wroa.2024.100247.

[36] Pernat, M. R., Kasprzyk, J., Zagona, E., Walker, S. D., & Livneh, B. (2025). The relative importance of model type and input features for water supply forecasting in snow-dominated basins of the southwestern US. Journal of Hydrology: Regional Studies, 60. doi:10.1016/j.ejrh.2025.102548.

[37] Mohammadi, M., Gato-Trinidad, S., & Kuok, K. K. (2024). The limitation of machine learning methods for water supply and demand forecasting: A case study for Greater Melbourne, Australia. Water Supply, 24(11), 3848–3861. doi:10.2166/ws.2024.225.

[38] Liu, H., Xing, R., & Davies, E. G. R. (2025). Forecasting municipal water demands: Evaluating the impacts of population growth, climate change, and conservation policies on water end-use. Sustainable Cities and Society, 130. doi:10.1016/j.scs.2025.106581.

[39] Xu, A., Zhang, T., Zhang, X., Shao, Y., Yu, T., Chu, S., & Qian, L. (2026). Multi-scale Spatio-temporal graph neural network for enhanced water demand forecasting. Water Research, 288. doi:10.1016/j.watres.2025.124711.

[40] Homaei, M., Tarif, M., Rodríguez, P. G., Ávila, M., & Caro, A. (2026). Adaptive multi-domain uncertainty quantification for digital twin water forecasting. Machine Learning with Applications, 23, 100812. doi:10.1016/j.mlwa.2025.100812.