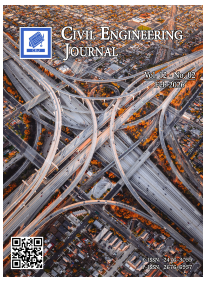
Available online at www.CivileJournal.org

Civil Engineering Journal

(E-ISSN: 2476-3055; ISSN: 2676-6957)

Vol. 12, No. 02, February, 2026



Ensemble and Hybrid Machine Learning Models for Seasonal Water Consumption Forecasting Under Climate Variability

Aruna Rajballie [1*], Vrijesh Tripathi [2], Shikhar Tyagi [2], Amarnath Chinchamee [1]

[1] *Project Management and Civil Infrastructure Systems, The University of Trinidad and Tobago, Port of Spain, Trinidad and Tobago.*

[2] *Department of Mathematics and Statistics, The University of the West Indies, St. Augustine, Trinidad and Tobago.*

Received 03 August 2025; Revised 19 January 2026; Accepted 23 January 2026; Published 01 February 2026

Abstract

The objective of this paper is to improve the forecasting of monthly water consumption under climate variability by combining ensemble and hybrid modelling with a season-aware design. Monthly consumption and meteorological data from 2003 to 2024 were utilized in this study. Four models were evaluated: (i) a stacking ensemble with STL-trend plus residual learning; (ii) a hybrid machine-learning–physics model with differentially-evolved weights; and (iii–iv) season-specific stacked models for wet and dry periods. Robustness was assessed with time-aware validation and residual diagnostics (Shapiro–Wilk, Breusch–Pagan, Durbin–Watson, Ljung–Box). The findings indicate that across models, ensembles captured nonlinear climate–demand variations while maintaining linear structure. The ensemble and hybrid model achieved strong accuracy with low errors while the season-specific models attained high fit (wet $R^2 \approx 0.998$; dry $R^2 \approx 0.991$) with stable residual behavior. Sensitivity to temperature and humidity aligns with expected physical behavior. Precipitation shows a diminishing-returns effect on water use, where moderate rainfall leads to higher consumption, while heavy rainfall tends to reduce demand. The framework innovatively combines decomposition-assisted stacking, physics-informed hybridization, and seasonal ensemble modelling. Overall, the approach provides highly accurate, interpretable, and climate-aware water demand forecasts for tropical regions, offering a practical basis for utility-scale implementation.

Keywords: Hybrid Model; Ensemble Model; Water Consumption; Decomposition-Assisted Stacking; Climate Variability; Tropical Region.

1. Introduction

Urban water-demand forecasting is increasingly challenged by climate variability, especially in tropical and small-island environments. Variables such as temperature, precipitation, humidity, and wind patterns contribute to pronounced seasonal fluctuations in water demand and elevate the risk of supply. Evidence from Trinidad specifically highlights how climate-driven hydrological responses can alter reservoir behavior in tropical contexts. The Navet Reservoir volumes declined during recent meteorological droughts, illustrating the sensitivity of island water systems to climate variability [1]. Research therefore emphasizes climate as a primary driver of urban water consumption, underscoring the need to integrate climate factors into demand forecasting [2].

Accurate monthly forecasts are essential for guiding resource planning, infrastructure development, and climate-adaptation strategies for utilities and government institutions [3-5]. Traditional statistical and time-series methods have set useful benchmarks for forecasting water consumption. However, their assumptions, such as linear responses, stationary error structures, and simplified seasonality, limit their ability to model non-linear, complex climate–demand relationships and patterns observed between wet and dry periods [6, 7]. Over the last decade, machine-learning (ML)

* Corresponding author: aruna.rajballie@utt.edu.tt

https://doi.org/10.28991/CEJ-2026-012-02-019

methods such as Random Forests, gradient boosting, deep neural networks and hybrid models have improved upon conventional methods by capturing complex interactions between climate inputs and water [8]. Kulaczkowski and Lee [9] found that Random Forest was an effective machine learning algorithm for forecasting short-term water demand with high precision in Italian districts. Li et al. [10] documented the use of XGBoost to predict urban water supply quantities, showing how gradient-boosted models can manage practical forecasting goals in water systems.

Liu et al. [11] showed that decomposition-based methods, such as wavelet, variational mode decomposition (VMD), and STL (Seasonal-Trend decomposition using Loess), consistently improved predictive accuracy and robustness under changing climate conditions. Ji [12] identified the hybrid Long Short-Term Memory (LSTM) and Seasonal Autoregressive Integrated Moving Average (SARIMA) models as key contributors to forecasting accuracy because of their ability to capture both linear and non-linear patterns in the data. Rajballie et al. [13] compared the applications of SARIMA models, exponential state space (ETS) models, artificial neural network (ANN) models, and hybrid combinations of them in developing models for all sectors of water consumption in Trinidad. The findings demonstrated the suitability of hybrid models for forecasting water demand for the island. Ensemble methods have also been shown to consistently outperform single learners for urban water demand by reducing variance and increasing robustness across different time horizons. Bakhshipour et al. [14] combined three machine learning models, Neural Hierarchical Interpolation for Time Series forecasting (N-HiTS), XGBoost, and a multi-head convolutional neural network to create an ensemble data-driven model for forecasting water demand within urban areas. The results of this study demonstrated the superior performance of the ensemble model, enhancing the forecasting accuracy and reliability of short-term water demand forecasting in urban environments. Collectively, these studies demonstrate that integrating decomposition, hybridization, and ensemble learning can substantially improve the accuracy and resilience of water demand forecasts under variable climatic conditions.

Despite these advancements, several gaps remain in the current literature. Seasonality is frequently represented as a binary or cyclical factor, which can oversimplify crucial behavioral shifts in water use across wet and dry seasons [15, 16]. This can lead to challenges in capturing hidden patterns in water demand modelling across seasons. While STL/wavelet decomposition is common in hydrometeorological time series, comprehensive methods that incorporate ensemble modelling and provide transparent diagnostics are limited in urban water-demand forecasting [6, 17]. Interest in physics-informed ML is growing, but computationally efficient hybrid models combining lightweight physics and machine learning that are directly compared with modern ensemble methods remain under-researched for monthly demand forecasting [18, 19]. Many studies emphasize metrics such as R^2 and RMSE but provide less on residual diagnostics, which are crucial for comprehensive model evaluation and effective uncertainty quantification. Compared to temperate settings, there remains a scarcity of accurate, interpretable, and implementable monthly demand forecasting approaches for wet and dry seasons and the climate variability characteristic of tropical climates. This study addresses these gaps by proposing and evaluating models that integrate advanced decomposition techniques, such as discrete wavelet transforms, with machine learning and deep learning algorithms to enhance the accuracy and robustness of urban water demand forecasts.

In this study, four forecasting models are developed for monthly water consumption in Trinidad using historical data from 2003 to 2024. Model 1 is a stacked ensemble that integrates Random Forest, XGBoost, and Bayesian Ridge as base learners, with Ridge Regression serving as the meta-learner. The model includes STL decomposition and residual learning to capture intricate seasonal patterns and account for unpredictable fluctuations in urban water demand [20]. Model 2 is a hybrid, non-stacking alternative that combines the same base learners: Random Forest, Bayesian Ridge and a Physics-Based model. The predictions are optimally weighted to generate the ensemble output through a weight-optimization process. This hybrid framework aligns with literature showing that combining data-driven models with physics-based components and optimizing their contributions yields robust climate predictions [21, 22]. To explicitly model seasonal dynamics, Models 3 and 4 are specific to each season and are trained solely on data from the wet and dry seasons, respectively. These season-specific models provide a direct way to test whether explicitly accounting for wet versus dry-season dynamics enhances forecast accuracy compared to a single, globally trained model. This approach is supported by decomposition-based forecasting literature that shows benefits from modeling decomposed components separately [22, 23]. The proposed study contributes to water resources management in Trinidad and Tobago by (i) benchmarking ensemble and hybrid methods for climate-driven monthly demand forecasting, (ii) exploring the value of decomposition- and season-specific approaches in tropical settings, and (iii) providing insights to utilities and policymakers on model choice, deployment and resilience amid climate change.

2. Methodology

2.1. Study Area and Data

The twin-island nation of Trinidad and Tobago is situated in the southernmost reaches of the Caribbean Sea, just off the north-eastern coast of Venezuela. It spans latitudes 10.0°N to 11.5°N and longitudes 60.5°W to 62.0°W. The country's total land area is approximately 5,130 km², with the larger and more populous island of Trinidad accounting for 94% of this territory [24].

The islands exhibit a tropical maritime climate, influenced by the prevailing northeast trade winds, elevated sea surface temperatures, and their proximity to the equatorial region. In Trinidad, average daily temperatures are typically observed to range between 24°C and 31°C throughout the year, exhibiting minimal annual variation. The local climate

is further characterized by generally high humidity levels, fluctuating within the 70% to 85% range [25]. Trinidad experiences two distinct climatic seasons: a wet season spanning from June to December and a dry season from January to May. The wet season aligns with the Atlantic hurricane season and is influenced by the movement of the Intertropical Convergence Zone and regional convective patterns, resulting in high rainfall intensity and frequency. The mean annual precipitation is approximately 2,200 mm, with over 65% of this rainfall occurring during the wet season [26]. Conversely, the dry season is characterized by significantly lower precipitation, increased evaporation, and elevated water demand, particularly in the residential and agricultural sectors [27].

The distinct seasonal fluctuations in rainfall and humidity significantly impact water consumption trends in the region. Water management agencies frequently confront supply disruptions during prolonged dry periods and infrastructure strain during the wet season due to flooding and surface runoff. Consequently, modelling these seasonal climate-consumption dynamics is essential for accurate demand forecasting and effective resource planning in the Trinidadian context [16, 28]. This work utilizes water consumption and weather data for Trinidad (Trinidad and Tobago). Historical monthly water consumption data were provided by the Water Resources Agency (WRA), a division of the Water and Sewerage Authority (WASA) while weather data were obtained from the Trinidad and Tobago Meteorological Service (TTMS). The collected data include the following variables: water consumption (m^3), precipitation (Precip) (mm), monthly average of maximum temperature (TempMaxAvg), monthly average of minimum temperature (TempMinAvg), maximum monthly humidity (HumidMax) (%) and monthly average humidity (HumidAvg) (%) from December 2003 to July 2024. This 21-year dataset captures long-term trends, inter-annual variability, and both wet and dry season demand cycles, which is crucial for developing generalizable forecasting. A workflow diagram outlining the steps in model development are shown in Figure 1.

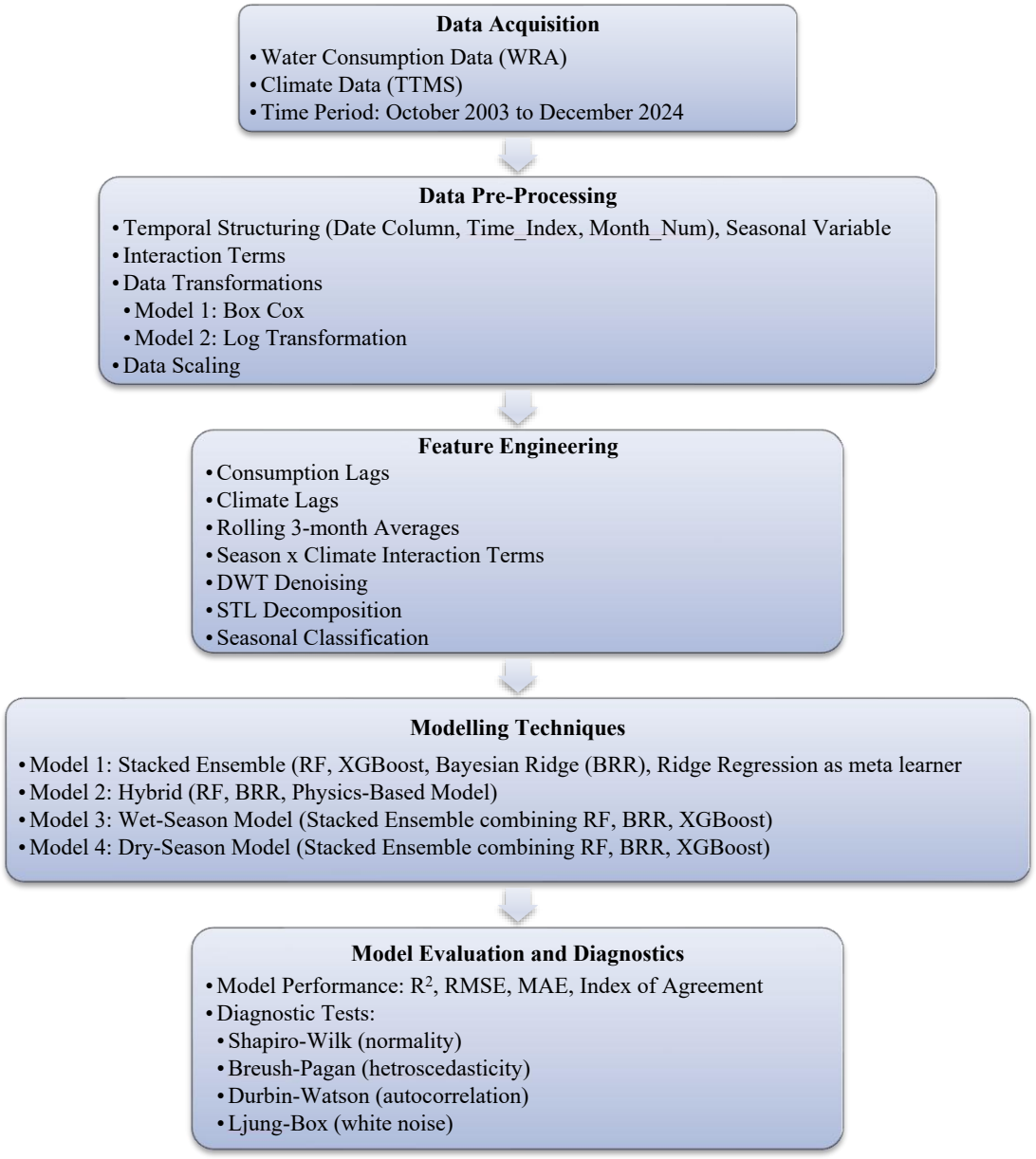


Figure 1. Workflow diagram

2.2. Temporal Structuring

A comprehensive feature engineering and pre-processing pipeline was established to boost the predictive performance of water consumption models influenced by climatic factors. While the overall approach remained consistent, specific steps were tailored to align with the modelling objectives of each individual model.

To preserve the temporal structure of the time series and enable the application of time-aware modelling techniques, such as lag generation, STL decomposition, and trend-season-residual forecasting, the dataset was chronologically indexed using a unified Date column. This column was constructed by combining the Year and Month variables, with the date field assigned to the first day of each month and sorted in ascending order.

A numeric variable representing the month number (Month_Num) and a time index (Time_Index) were created to capture the number of months since the start of the data series. These variables enabled the construction of lagged predictors, time-dependent interactions, and the modelling of long-term trends.

To account for the pronounced seasonal patterns in water consumption, a binary classification variable ('season_Wet') was created. This variable distinguished between the wet season, defined as June through December, and the dry season, spanning January to May, in alignment with the regional climatology of Trinidad and Tobago (McSweeney et al. 2010). This seasonal categorization facilitated the development of models tailored to each period, as well as the incorporation of interaction terms between meteorological variables and the seasonality factor (e.g., TempMaxAvg × season_Wet).

This temporal structure also enabled the use of time-aware validation approaches, such as TimeSeriesSplit, which ensured that training data chronologically preceded test data—a crucial practice for temporal forecasting [29]. The list of water consumption variables and their brief descriptions are presented in Table 1.

Table 1. List of variables with description

Variable	Description
Consumption	Monthly water consumption (m^3)
Precip	Monthly average precipitation (mm)
TempMaxAvg	Monthly average of maximum temperature (℃)
TempMinAvg	Monthly average of minimum temperature (℃)
HumidMax	Monthly maximum relative humidity (%)
HumidAvg	Monthly average relative humidity (%)
Precip_Squared	Squared term for precipitation (mm^2)
TempMaxAvg_Squared	Squared term for maximum temperature $(℃)^2$
TempMaxAvg_season	Interaction between TempMaxAvg and season (℃)
HumidAvg_season	Interaction between HumidAvg and season (%)
Consumption_Lag1	1-month lagged water consumption (m^3)
Consumption_Lag2	2-month lagged water consumption (m^3)
Consumption_MA3	3-month moving average of consumption (m^3)
Precip_Lag1	1-month lagged precipitation (mm)
TempMaxAvg_Lag1	1-month lagged monthly average of maximum temperature (℃)
HumidAvg_Lag1	1-month lagged monthly average humidity (%)
season_Wet	Binary variable for season (1 for wet; 0 for dry)

2.3. Target Variable Transformations

Target transformations were applied to stabilize variance, improve normality and ensure that model assumptions were satisfied.

Model 1: A power transformation [30] was applied with exponent $p = 0.15$ as follows:

$$y' = y^{0.15} \tag{1}$$

The Box-Cox is a family of transformations designed to address skewness and heteroscedasticity. Its use is motivated by the fact that residuals from complex ensembles deviate from a normal distribution or exhibit non-constant variance [31].

Model 2: Log P transformation [32] was applied to the data as follows:

$$y' = log(1 + y) \tag{2}$$

A log transformation was applied to convert the multiplicative relationships between water consumption and its physical predictors into a simpler linear form. This was important in the combined machine learning and physics-based framework, where proportional effects were expected. This transformation ensured interpretability and consistency between the data-driven and physics-based components [33].

2.4. Data Scaling

All continuous variables (precipitation, temperature, humidity, wind speed) were scaled using z-score normalization before the model was fitted by using the following formula to account for stability across models:

$$z = \frac{x-\mu}{\sigma} \tag{3}$$

where μ is $mean$ of variable and σ is standard deviation of variable.

2.5. Decomposition and Smoothing Methods

Models 1 and 2 utilized Discrete Wavelet Transform (DWT) to denoise the data. DWT decomposes a time series $y(t)$ into a multi-scale representation consisting of approximation and detail coefficients as follows:

$$y(t) = \sum_k a_{J,k}\, \phi_{J,k}(t) + \sum_{j=1}^{J} \sum_k d_{j,k}\, \psi_{j,k}(t) \tag{4}$$

where $\phi_{J,k}(t)$ and $\psi_{j,k}(t)$ are scaling and wavelet functions, and $a_{J,k}$, $d_{j,k}$ are the coefficients.

Detail coefficients were soft-thresholded using:

$$\hat{d}_{j,k} = \text{sign}\left(d_{j,k}\right) \cdot \max\left(\left|d_{j,k}\right| - \lambda, 0\right) \tag{5}$$

With λ is $\sigma\sqrt{2\log n}$ and σ estimated from the finest scale [34, 35].

2.6. Advanced Feature Engineering

Multiple feature engineering techniques were applied to construct derivative variables that better captured the complex relationships between climatic factors and water consumption. The specific application of these steps varied somewhat across the different models, as shown in Table 2.

Table 2. Modelling Techniques

Feature Engineering Step	Model 1	Model 2	Model 3	Model 4	Description
Consumption Lags (Lag1, Lag2)	✗	✗	✓	✓	Capture short-term autocorrelation
Precipitation and Temperature Lags	✗	✗	✓	✓	Capture delayed climate effects
Rolling 3-Month Averages	✗	✗	✓	✓	Capture a persistent pattern in water use
Season × Climate Interaction Terms	✓	✓	✗	✗	Allow season-modulated climate effects
DWT Denoising	✓	✓	✗	✗	Suppress noise in the consumption series
STL Decomposition	✓	✗	✗	✗	Separate trend, seasonality, and residual
Seasonal Classification (Wet/Dry)	✓	✓	✓	✓	Account for seasonal regime shifts

Seasonal-Trend Decomposition (STL) [36] was used in Model 1 to extract trend, seasonal and residual components:

$$y_t = T_t + S_t + R_t \tag{6}$$

where y_t is the observed time series at time t, T_t is the long-term trend component, S_t is the seasonal component, R_t is the remainder (or residual) component.

In this study, STL was applied with a seasonal period of 12 (monthly data), and the trend and seasonal components were extracted before modelling the residuals in Model 1.

2.7. Modelling Techniques

The forecasting models in this investigation utilized a combination of ensemble machine learning approaches and statistical regression techniques. These foundational models, including Random Forest, Bayesian Ridge Regression, XGBoost, and Physics-Based model, were consolidated through stacking and hybrid techniques to enhance overall predictive accuracy. The following models were introduced in this study for modelling water demand.

2.7.1. Random Forest Model (RF)

Random forest, which was first proposed by Breiman [37] is an ensemble learning algorithm based on decision trees. Individual trees are trained on bootstrapped samples of the dataset, with node splits determined by a random subset of features. The final prediction is calculated by averaging the predictions from the individual trees as follows:

$$\hat{y} = \frac{1}{T}\sum_{t=1}^{T} f_t\,(x) \tag{7}$$

where $f_t(x)$ is the prediction from the t-th tree, and T is the number of trees in the forest. Random Forests demonstrate strong resistance to overfitting and are effective for datasets with complex relationships and correlated variables.

2.7.2. Bayesian Ridge Regression (BRR)

Bayesian Ridge Regression represents a probabilistic expansion of linear regression, incorporating Gaussian priors for coefficient estimation. In contrast to conventional least squares methods, which generate point estimates for regression coefficients, BRR assumes the parameters adhere to a multivariate normal prior as follows:

$$\beta \sim \mathcal{N}(0, \lambda^{-1}\mathrm{I}) \tag{8}$$

The likelihood of the target variable y is also assumed to be Gaussian:

$$y \mid X, \beta, \alpha \sim \mathcal{N}(X\beta, \alpha^{-1}\mathrm{I}) \tag{9}$$

where α and λ are precision parameters for the noise and prior, respectively. These are treated as hyperparameters and estimated using Type II Maximum Likelihood or Empirical Bayes methods [38]. The predictive distribution for a new observation x_* is given by:

$$p(y_* \mid x_*, X, y) = \mathcal{N}\left(x_*^{\mathrm{T}}\bar{\beta},\ \sigma^2(x_*)\right) \tag{10}$$

where $\bar{\beta}$ is the posterior mean and $\sigma^2(x_*)$ incorporates both data noise and parameter uncertainty [39]. Bayesian Ridge Regression is beneficial in water demand forecasting due to its ability to minimize sensitivity to multicollinearity, effectively manage small or imperfect datasets, and provide predictive uncertainty assessments.

2.7.3. Extreme Gradient Boosting (XGBOOST)

Extreme Gradient Boosting is a scalable machine learning algorithm that utilizes gradient boosting decision trees. It works by sequentially building new regression trees to correct the errors of previous trees, with the goal of minimizing a regularized loss function that considers both predictive accuracy and model complexity. Given a dataset $\{(x_i, y_i)\}_{i=1}^{n}$, the prediction from XGBoost is an additive model of the form:

$$\hat{y}_i = \sum_{t=1}^{T} f_t\,(x_i), \quad f_t \in \mathcal{F} \tag{11}$$

where $\mathcal{F}$ denotes the space of regression trees and T is the total number of trees. Each function f_t corresponds to a decision tree trained at iteration t [40].

During training, the algorithm aims to minimize an objective function. This function includes two parts: a loss function (like squared error) that measures prediction errors, and a regularization term to prevent the model from becoming too complex.

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) \tag{12}$$

The regularization term is defined as:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda\sum_{j=1}^{T} w_j^2. \tag{13}$$

where T is the number of leaf nodes in the tree, w_j is the output score on the leaf j, γ penalizes the number of leaves, and λ applies ℓ_2 regularization to the leaf weights.

XGBoost also provides functionality for handling missing data and early stopping, thereby demonstrating its robustness when applied to high-dimensional, collinear datasets.

2.7.4. Stacking Regressor

Stacking is a meta-learning technique that combines multiple base models by training a second-level model on their predictions [41]. Formally:

$$\hat{y}_{\text{stack}} = g\big(\hat{y}_1(x), \hat{y}_2(x), \ldots, \hat{y}_M(x)\big). \tag{14}$$

where $\hat{y}_m(x)$ are predictions from base learners and $g(\cdot)$ is the meta-learner, typically a linear regressor, $\hat{y}_{\text{stack}}$ is a hybrid ensemble forecast and the actual observed values, enhancing accuracy through residual correction. Ensemble methods, such as stacking, diminish model bias and variance by exploiting the unique advantages of individual base models.

2.7.5. Differential Evolution for Weight Optimization

Differential Evolution is a global optimization algorithm that uses mutation, crossover, and selection to improve potential solutions iteratively [42]. Its goal is to find the optimal weights that minimize the mean squared error (MSE) between the hybrid model's predictions and the actual test data.

2.7.6. Physics-Based Model

To incorporate domain knowledge into the forecasting approach, a basic physics-informed linear model was developed. This model aims to capture established associations between climate variables and water demand, assuming that monthly water usage correlates positively with precipitation and temperature, while being inversely related to humidity due to decreased evapotranspiration. It takes the form:

$$\hat{y}_{\text{phys}} = \beta_1 \cdot \text{Precip} + \beta_2 \cdot \text{TempMaxAvg} + \beta_3 \cdot \text{HumidAvg}\,. \tag{15}$$

This approach reflects broad climate-consumption connections and acts as an additional signal within the combined model. The contribution of this component was combined with machine learning predictions using optimized weights via differential evolution. Simple physics-based models have been shown to improve interpretability and correct for overfitting of pure data-driven models [43].

2.8. Model Performance

The performance of water consumption models was assessed using four common performance statistical measures: coefficient of determination (R^2), root mean square error (RMSE) (ASCE 2020) and mean absolute error (MAE). These metrics are often employed to assess model accuracy and bias in climatic and hydrological research [44] and given as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \tag{16}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{17}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}. \tag{18}$$

where y_i and $\hat{y}_i$ are the actual and estimated values, respectively, $\bar{y}$ is the mean of the observed values, and n is the total number of predictions. The models having the lowest error values are selected as the most accurate.

3. Results and Discussion

Trinidad and Tobago has a tropical climate that is hot and humid year-round. Trinidad has two distinct seasons: the dry season, characterized by reduced rainfall and higher temperatures, and the wet season, marked by increased precipitation and humidity. During the dry season, diurnal temperatures can reach a peak of 36 degrees Celsius, while nocturnal temperatures average around 28 degrees Celsius. Marginally cooler temperatures typically characterize the wet season.

Over the period from October 2003 to July 2024, mean monthly precipitation in Trinidad showed clear seasonal variation, with the dry season averaging 66.56 mm and the wet season averaging 191.51 mm. This reflects the region's characteristic bimodal rainfall pattern. The stark difference in rainfall confirms the well-known climatic

division between the dry and wet seasons. In contrast to this notable seasonal difference in precipitation, average monthly water consumption showed only slight variation, with 4,564,548.22 m³ during the dry season and 4,660,273.92 m³ during the wet season. To determine whether there was any difference in water use between seasons, the Mann-Whitney U test was applied to both the observed data and model residuals for Models 1 and 2. The p-values were 0.56 and 0.39 for Model 1, and 0.51 and 0.72 for Model 2. These results highlight the consistency of water consumption across seasons. Nonetheless, season-specific models were retained to identify any patterns or relationships specific to each season.

Figure 2 shows the plot of water consumption time series data from 2003 to 2024 before data transformation. The series exhibits a gradual upward trend from 2003 to 2015. However, between 2016 and 2017 and again after 2020, the data shows increased variability with extreme spikes and dips. Despite these variations, a seasonal pattern remains apparent, consistent with climatic influences during the wet and dry seasons in Trinidad.

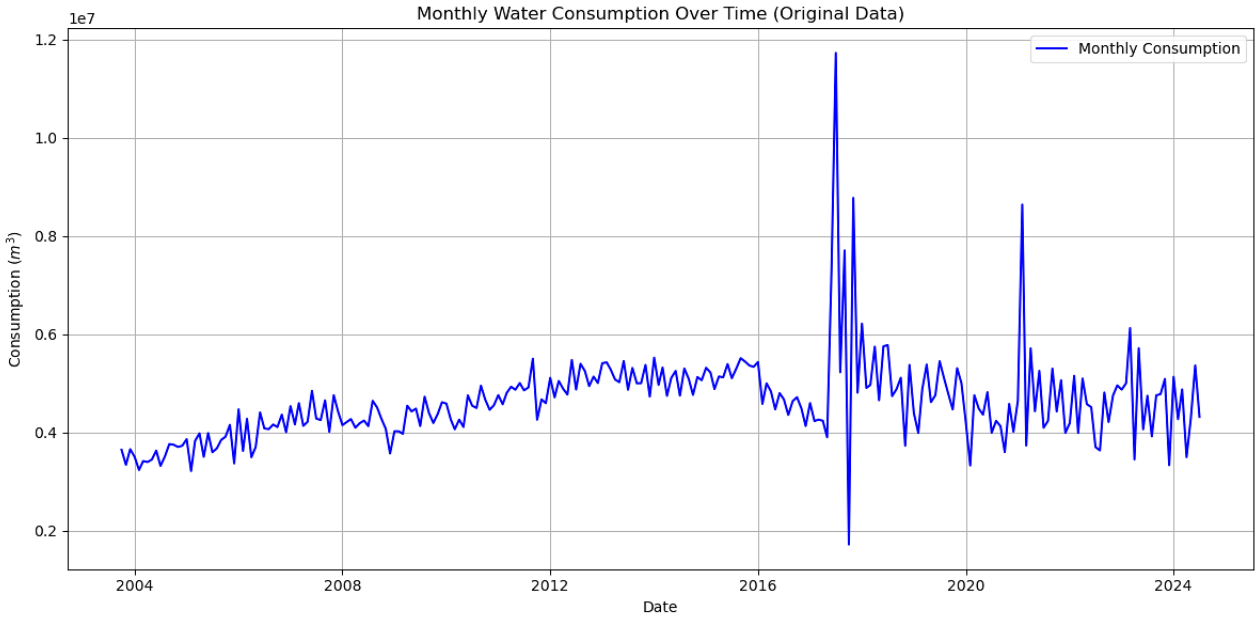


Figure 2. Plot of water consumption time series data from 2003 to 2024

Model 1 uses a design that separates long-term trends from short-term fluctuations, which helps improve forecasting accuracy. To evaluate the presence of temporal autocorrelation in the consumption series, statistical diagnostics were performed on the data. The Durbin-Watson test [45] produced a statistic of 3.31, indicating strong negative autocorrelation (close to 4). Additionally, the Ljung-Box test at lag 10 resulted in a test statistic of 189.22 with a p-value less than 2.85 times 10 to the minus 35, signifying that the null hypothesis of no autocorrelation is rejected. These autocorrelation diagnostics support the use of a layered modelling approach, as applied in the STL + Residual ensemble framework, because STL decomposition aims to capture remaining temporal irregularities not explained by the primary model [36, 46]. Squared terms like Precip_Squared and TempMaxAvg_squared were added to account for potential nonlinearities. Interaction terms such as TempMaxAvg, season_Wet, and HumidAvg times season_Wet were included to address seasonal dependencies. A temperature variable was created to measure the effect of temperatures above the median, following methods used in climate response analysis [47].

The STL trend component was estimated using a Bayesian Ridge Regression model with an intercept ($beta_0$ = -17.424). The model coefficients for the Bayesian Ridge Regression are shown in Table 3**.** The negative coefficients for the humidity-related variables (HumidAvg −0.006801; HumidMax −0.004208) imply that higher humidity is linearly associated with lower water consumption. These results agree with the findings of a study published by Alsulaili et al. [48], which stated that water consumption decreased with humidity. The positive coefficients for the temperature variables (TempMaxAvg 0.005070; TempMinAvg 0.002273) reflect higher water demand during warmer conditions. These findings are supported by a study carried out by Dimkic [49], which showed that as temperature increases, water consumption increases. The small coefficient for the season variable suggests that when climatic variables are included, the seasonal indicator variable contributes minimally to the Bayesian Ridge Regression layer. Precip is positive (0.003034), but Precip_Squared is negative (−0.001450). This indicates a diminishing relationship between rainfall and water consumption, where light to moderate rainfall is associated with higher recorded consumption, but very heavy rainfall reduces demand.

Table 3. Bayesian Ridge Regression coefficients for Model 1

Variable (x_i)	Coefficient (β_i)
Year	0.013605
HumidAvg	-0.006801
TempMaxAvg	0.005070
HumidMax	-0.004208
Precip	0.003034
TempMinAvg	0.002273
TempMaxAvg_season	0.001966
Precip_Squared	-0.001450
TempMaxAvg_Squared	0.001611
Month	0.000397
season_Wet	-0.000095
HumidAvg_season	0.000050

The Bayesian Ridge Equation of the STL trend component is given by:

$$\hat{y}_{trend}(Bayesian\ Ridge) = -17.424 + \sum_{i=1}^{12} \beta_i x_i \quad (19)$$

The relative contributions of each model to the stacking ensemble are shown in Table 4. The Random Forest model's predictions greatly influence (60.43%) the final stacked prediction of the trend layer, capturing most of the key data signals that help the meta-model in minimizing errors. The XGBoost model contributes 39.23% to the final stacked model, most likely capturing different non-linear patterns that could not be explained by the Random Forest Model. The Bayesian Ridge contributed a mere 0.35% to the final stacked ensemble model but was still retained for regularization.

Table 4. Relative contributions of base models to Model 1

Model	Contribution (%)
Random Forest	60.43
Bayesian Ridge	0.35
XGBoost	39.23

The Random Forest model used in the ensemble comprised 100 decision trees, each with an average depth of 13.86 and approximately 126.07 leaf nodes per tree. These trees were constructed using 12 input features. The XGBoost component incorporated 300 boosted trees, resulting in a total of 7974 leaves across all trees, averaging approximately 26.58 leaves per tree, allowing the model to capture complex nonlinear patterns in the data. A Ridge Regression model served as the meta-learner, linearly combining the predictions of the base models. The Stacking Ensemble trend equation is:

$$\hat{y}_{trend} = -0.5797 + 0.4465\hat{y}_{RF} + 0.5984\hat{y}_{XGB} + 0.01314\hat{y}_{BR} \quad (20)$$

The residual component ($\hat{y}_{resid}$)was modelled separately using a Random Forest Regressor, which can capture nonlinear effects and complex interactions that cannot be handled by linear models [37]. The Breusch–Pagan test [50] was conducted to evaluate heteroscedasticity, producing a statistic of 18.67 with a p-value of 0.097, suggesting a consistent variance of the residuals across the fitted value range [51]. In addition, the Shapiro–Wilk test [52] for normality yielded a statistic of 0.801 and a p-value < 0.001. These results indicate a significant deviation from normality, a common occurrence in ensemble models, especially those that employ nonlinear learners like Random Forests [37]. Autocorrelation diagnostics were used to assess the independence assumption on the residuals of the final ensemble model, and are shown in Table 5. The Durbin–Watson statistic was 2.33, which falls within the acceptable range of 1.5 to 2.5, indicating no strong evidence of positive or negative autocorrelation. Furthermore, the Ljung–Box test, conducted at a lag of 10, produced a test statistic of 4.56 and a p-value of 0.92, providing additional evidence that autocorrelation is not present in the residuals ($p > 0.05$). The findings indicate that the model effectively captured the temporal structure of the data, as evidenced by the absence of substantial autocorrelation in the residual component.

Table 5. Autocorrelation Diagnostics for Model 1

	Durbin-Watson Statistic	Ljung-Box (p-value) at lag 10
Before Modelling	3.31	2.85×10^{-35}
After Modelling	2.33	0.92

Model 2 utilized a hybrid ensemble method, integrating Random Forest, Bayesian Ridge, and a physics-informed model with optimized weighting to forecast denoised and log-transformed water consumption. The Durbin-Watson test yielded a statistic of 3.32, indicating strong negative autocorrelation in the data. Additionally, the Ljung-Box test at lag 10 produced a test statistic of 1022.92 with a p-value of 2.16×10^{-213}, also confirming the presence of strong autocorrelation. These findings indicate that water consumption data has complex time-based patterns, thereby justifying the use of denoising, decomposition, and hybrid modelling techniques for analyzing long-term trends and short-term fluctuations [34, 36, 53]. Squared terms, such as Precip_Squared and TempMaxAvg_Squared, were introduced to capture non-linear effects of temperature and precipitation. The denoised water consumption model was estimated using a Bayesian Ridge Regression model with an intercept ($\beta_0 = -2.26$). The model coefficients for the Bayesian Ridge Regression are presented in Table 6. Similar to Model 1, The positive coefficients for the temperature variables TempMaxAvg ($\beta_3 = 0.002047$) and TempMinAvg ($\beta_6 = 0.000968$) indicate higher water use in warmer conditions. This matches prior studies that state that water consumption linearly increases with temperature [49]. Similar to Model 1, the coefficient for Precip was positive, which is inconsistent with previous studies that found rainfall negatively affects water consumption [54]. Meanwhile, the variable Precip_Squared had a negative coefficient in the Bayesian Ridge layer, consistent with prior literature. These findings indicate a non-linear relationship between precipitation and water consumption. TempMaxAvg_season is positive ($\beta_7 = 0.000864$), while HumidAvg_season is negative ($\beta_{12} = -0.000365$), indicating season-dependent sensitivities. This implies that high temperatures intensify water demand during the dry season, whereas elevated humidity levels mitigate demand during the wet season [55-57].

Table 6. Bayesian Ridge Regression coefficients for Model 2

Variable (x_i)	Coefficient (β_i)
Year	0.008737
HumidAvg	-0.002549
TempMaxAvg	0.002047
HumidMax	-0.001562
Precip	0.000650
TempMinAvg	0.000968
TempMaxAvg_season	0.000864
Precip_Squared	-0.000486
TempMaxAvg_Squared	0.001739
Month	0.000380
season_Wet	0.000368
HumidAvg_season	-0.000365

The Bayesian Ridge Regression equation of Denoised water consumption (y_t) is

$$y_t\,(BR) = -2.2656 + \sum_{i=1}^{12} \beta_i x_i. \quad (21)$$

The physics-based model equation at time t is

$$y_t(Phys) = 0.6Precip + 0.3TempMaxAvg - 0.2HumidAvg. \quad (22)$$

The physics-based component of the hybrid model consists of a linear combination of three climatic variables: precipitation, temperature, and humidity. This model suggests that precipitation and temperature are associated with higher water consumption, while higher humidity usually reduces the need for water, especially for cooling or irrigation. This aligns with research by Timotewos et al. [58], which found that temperature had a significant positive effect on water consumption, while humidity had a negative effect.

The relative contributions of each model to the Hybrid Ensemble are shown in Table 7. The final hybrid ensemble assigned relative contributions of 91.36% to the Random Forest model, 8.62% to the Bayesian Ridge regression model, and only 0.01% to the physics-based component. The weights were optimized using differential evolution to minimize prediction error in the denoised and transformed consumption series. The Random Forest model's dominance highlights its robust ability to capture nonlinear patterns and interactions among climatic variables. In contrast, the Bayesian Ridge model contributed modestly to any linear patterns in the data. Despite the minimal weighting of the physics-based model, it was included in the ensemble to preserve the underlying structural relationships, which are expected to evolve under the presence of climate change [59, 60].

Table 7. Relative contributions of base models to Model 2

Model	Contribution (%)
Random Forest	91.36%
Bayesian Ridge	8.62%
Physics-Based	0.01%

The Equation of the hybrid ensemble model is

$$\hat{y}_{hybrid} = 0.9136\,\hat{y}_{RF} + 0.0862\hat{y}_{BR} + 0.0001\hat{y}_{Phys}\,. \quad (23)$$

The Breusch–Pagan test was conducted to evaluate heteroscedasticity, yielding a statistic of 15.71 and a p-value of 0.204, suggesting that the residuals have consistent variance across the fitted value range. In addition, the Shapiro–Wilk test for normality yielded a statistic of 0.928 and a p-value of 0.005. Similar to the stacked model, this result also indicates a slight deviation from normality. Autocorrelation diagnostics were employed to assess the independence assumption of the residuals of the final ensemble model, and are presented in Table 8. The Durbin–Watson statistic was 1.90, indicating no strong evidence of positive or negative autocorrelation. Furthermore, the Ljung–Box test at lag 10 produced a test statistic of 5.59 and a p-value of 0.85, providing additional evidence that autocorrelation is not present in the residuals.

Table 8. Autocorrelation diagnostics for Model 2

	Durbin-Watson Statistic	Ljung-Box (p-value) at lag 10
Before Modelling	3.32	2.16×10^{-213}
After Modelling	1.90	0.85

Table 9 summarizes the impact of transformations on variance stability and overall information criteria for Models 1 and 2. The Box-Cox transformation for Model 1 reduced heteroscedasticity (p-value > 0.05) and substantially lowered Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values, indicating improved model likelihood [61]. Model 2 exhibited constant variance before transformation; however, a log transformation was retained as model likelihood values were improved.

Table 9. Effect of transformations on Model 1 and Model 2

Model	Transformation	Breush-Pagan (p-value)	AIC	BIC
Model 1	Without	0.0094	7550.25	7596.03
Model 1	Box Cox	0.097	-420.40	-397.46
Model 2	Without	0.081	1237.52	1260.46
Model 2	Log	0.204	-332.67	-309.73

Model 3 was specifically designed to forecast water consumption during the wet season, effectively capturing the unique seasonal dynamics inherent to this period. It employs a stacked ensemble framework combining Random Forest, XGBoost, and Bayesian Ridge regressors, using past consumption, weather variables, and their relationships to improve prediction accuracy. The Durbin-Watson test yielded a statistic of 3.32, indicating strong negative autocorrelation in the data. Additionally, the Ljung-Box test at lag 10 produced a test statistic of 71.75 with a p-value of 2.04×10^{-11}, also confirming strong autocorrelation. The wet season model incorporated several lagged features to effectively capture temporal dependencies in water consumption and climatic patterns. Lagged values of consumption (one-month and two-month lags, along with a three-month moving average) were used to account for usage patterns in consumer behavior. Additionally, one-month lags of precipitation, temperature, and humidity were included to reflect delayed environmental influences on demand. These variables helped the model address autocorrelation, allowing it to explicitly learn from past patterns, thereby improving forecasting accuracy and capturing temporal dependencies in water consumption [62, 63]. Bayesian Ridge Regression contributed significantly to the final predictions with an intercept of $\beta_0 = 2.043 \times 10^{-10}$. The coefficients for this Bayesian Ridge Regression is presented in Table 10. Negative coefficients for Consumption_Lag1 and Consumption_Lag2 show that consumption quickly reverts to its average levels, suggesting that higher consumption in previous periods usually leads to subsequent decreases. Farah & Shahrour [57] implemented a multi-scale Artificial Neural Network (ANN) that integrated lagged features to predict daily and hourly water consumption in four areas in France. The 1-hour lagged model consistently outperformed other models, achieving an R^2 up to 0.944. The study concluded that short-term autocorrelation is a primary factor influencing demand variability and that ANN models with lag features offer a reliable and valuable tool. The small lag coefficients for the climate variables

indicate that the wet season water use responds primarily to current climatic conditions rather than those of the preceding month. The Bayesian Ridge model reduces any false lag effects toward zero to maintain model stability. This is consistent with the findings of a monthly econometric model reported by the San Diego Water Authority [64], which confirm small coefficients for the lagged precipitation variable.

Table 10. Bayesian Ridge Regression coefficients for Model 3

Variable(x_i)	Coefficients (β_i)
Precip	2.74×10^{-10}
TempMaxAvg	-5.07×10^{-10}
TempMinAvg	-4.72×10^{-11}
HumidMax	-8.31×10^{-10}
HumidAvg	-4.99×10^{-11}
Consumption_Lag1	-1.005
Consumption_Lag2	-1.007
Consumption_MA3	2.312
Precip_Lag1	3.33×10^{-10}
TempMaxAvg_Lag1	1.004×10^{-11}
HumidAvg_Lag1	-5.92×10^{-11}

The Bayesian Ridge Regression equation for standardized water consumption (y_t) is:

$$\hat{y}(BR) = 2.043 \times 10^{-10} + \sum_{i=1}^{11} \beta_i x_i \, . \quad (24)$$

The relative contributions of each model to the stacked model are presented Table 11. Although the Bayesian Ridge model produces small variable coefficients, its strong cross-validation accuracy resulted in a dominant meta-model weighting of 99.81%. This suggests that the Bayesian Ridge model effectively captured nearly all explainable variance in the wet season model. Random Forest and XGBoost were also incorporated as base learners to capture complex nonlinear patterns that the linear Bayesian Ridge model could not fully address. Despite their low weightings, these models enhanced robustness during cross-validation by offering additional learning approaches.

Table 11. Relative contributions of base models to model 3

Model	Contribution (%)
Bayesian Ridge	99.82
XGBoost	0.16
Random Forest	0.0002

The Random Forest model used in the ensemble comprised 300 decision trees, each with an average depth of 8.0 and approximately 46 leaf nodes per tree. These trees were constructed using 11 input features, including lagged variables and climate predictors. The model's deep structure guarded against overfitting while capturing nonlinear interactions in the wet-season data. Within the wet season ensemble, the XGBoost component incorporated 300 boosted trees, culminating in a total of 4527 leaf nodes, averaging approximately 15.09 leaves per tree. Trained using 11 input features, this model effectively captured both lagged climate and consumption variables. These values demonstrate XGBoost's capability to learn complex non-linear interactions, which is beneficial in modelling the erratic patterns that can be characteristic of the wet season. The final ensemble output for the wet season model was computed via a Ridge regression meta-learner that combined the predictions of the three base learners: Random Forest, XGBoost, and Bayesian Ridge regression. The coefficients of this Meta learner are 0.000172, 0.00163 and 0.99347 for the Random Forest, XGBoost and Bayesian Ridge respectively. This yields the following stacked ensemble Equation (25):

$$\hat{y}_{stacked} = 2.91 \times 10^{-4} + 0.000172\hat{y}_{RF} + 0.00163\hat{y}_{XGB} + 0.993470\hat{y}_{BR} \quad (25)$$

The Breusch–Pagan test was conducted to evaluate heteroscedasticity, producing a statistic of 18.01 and a p-value of 0.08, suggesting a constant variance of the residuals across the fitted value range. In addition, the Shapiro–Wilk test for normality yielded a statistic of 0.9815 with a p-value of 0.8649. This suggests that the model residuals for the wet season model are approximately normally distributed. This result can be attributed to the substantial influence of the Bayesian Ridge learner (99.82%) in the final stacking model. Bayesian Ridge Regression is a linear model that works

by finding the best distribution of coefficients, assuming errors are normally distributed [65]. After fitting the model to the wet season data, autocorrelation checks were conducted. These values are presented in Table 12. The Durbin–Watson statistic was 2.19, indicating no strong evidence of positive or negative autocorrelation. Furthermore, the Ljung–Box test at lag 10 produced a test statistic of 7.53 and a p-value of 0.67, providing additional evidence that autocorrelation is not present in the residuals.

Table 12. Autocorrelation diagnostics for Model 3

	Durbin-Watson Statistic	Ljung-Box (p-value) at lag 10
Before Modelling	3.32	2.04×10^{-11}
After Modelling	2.19	0.67

Model 4 was specifically designed to predict water consumption during the dry season, accounting for the unique climatic and behavioral trends that dominate during this period. It employs a stacked ensemble framework that integrates Random Forest, XGBoost, and Bayesian Ridge regressors, trained on lagged climate variables and historical consumption features. The Durbin-Watson test yielded a statistic of 0.695, indicating strong positive autocorrelation in the data. Additionally, the Ljung-Box test at lag 10 produced a test statistic of 204.23 with a p-value of $2.11\times\ 10^{-38}$, also confirming strong autocorrelation, thereby justifying the incorporation of lagged variables to mitigate this effect. Bayesian Ridge Regression contributed significantly to the final predictions with an intercept of $\beta_0 = -6.27 \times 10^{-15}$. The coefficients for this Bayesian Ridge Regression are presented in Table 13. The dry season model produced strong negative one-month and two-month lag coefficients for consumption, as well as a positive coefficient for the three-month moving average consumption variable. The negative lagged values imply that larger past values are associated with smaller current consumption responses, holding all other variables constant. For the dry season, these effects can reflect delayed responses in conservation or adaptive behaviors to past consumption [66, 67]. This interpretation aligns with Bayesian lag models that use priors to shrink and smooth coefficients, with the signs and sizes driven by the data under those priors. A positive three-month moving average consumption coefficient indicates that higher three-month average water use is associated with higher current water use, after accounting for the other predictors in the model [68]. This aligns with the integration of time series features to capture autocorrelation in consumption data within a regression framework [69].

Table 13. Bayesian Ridge Regression coefficients for Model 4

Variable(x_i)	Coefficients (β_i)
Precip	3.83×10^{-10}
TempMaxAvg	1.47×10^{-10}
TempMinAvg	-2.77×10^{-11}
HumidAvg	-7.3×10^{-10}
Consumption_Lag1	-1.22
Consumption_Lag2	-1.28
Consumption_MA3	2.82
Precip_Lag1	-1.71×10^{-10}
HumidAvg_Lag1	4.85×10^{-10}

The Bayesian Ridge Regression Equation for standardized water consumption (y_t) is

$$\hat{y}(BR) = -6.27 \times 10^{-15} + \sum_{i=1}^{9} \beta_i x_i \quad (26)$$

The relative contributions of each model to the stacked model are presented in Table 14. Although the contributions of both XGBoost and Random Forest were minimal, their inclusion helped capture any nonlinear patterns that the Bayesian Ridge model might have missed.

Table 14. Relative contributions of all models to Model 4

Model	Contribution (%)
Bayesian Ridge	92.71
XGBoost	0.9
Random Forest	6.38

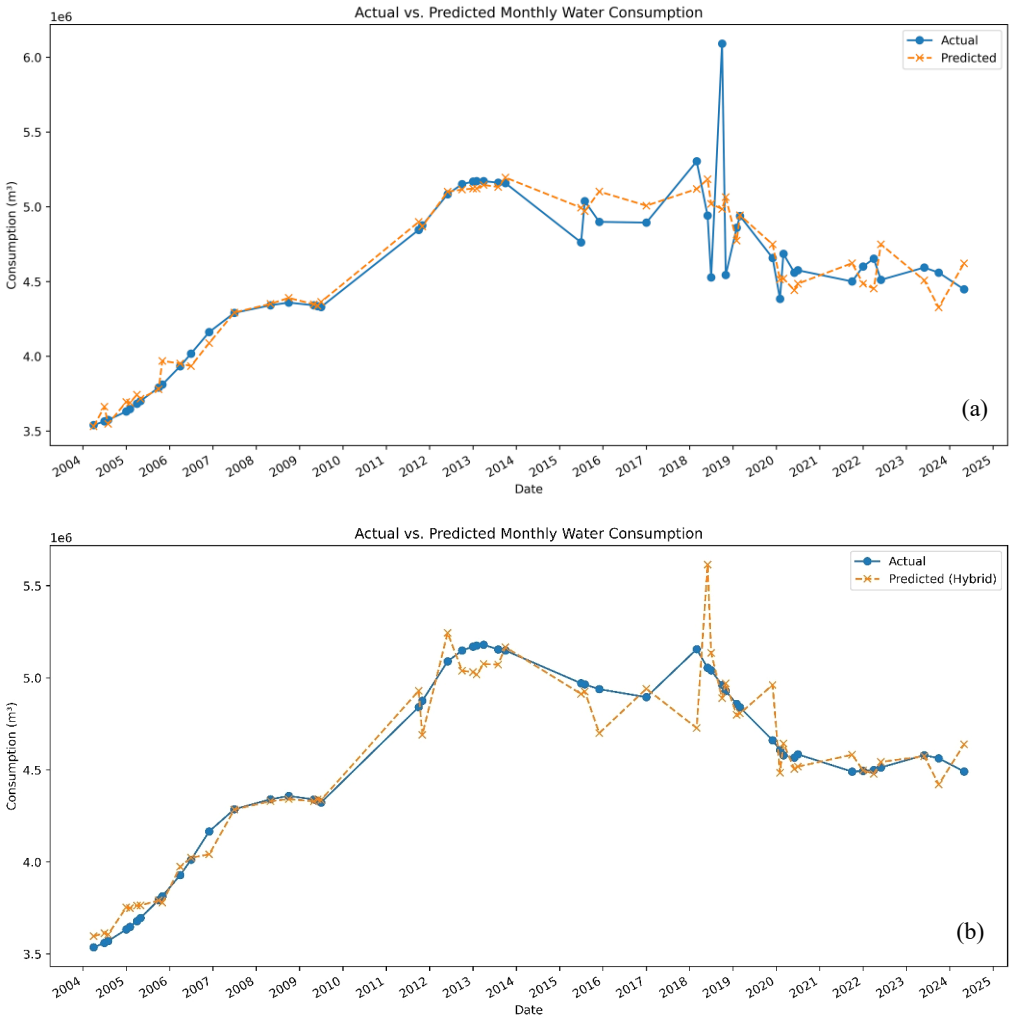
The Random Forest model used in the ensemble comprised 300 decision trees, each with an average depth of 8.0 and approximately 54 leaves per tree. These trees were constructed using nine input features, including lagged variables and climate predictors. Within the wet season ensemble, the XGBoost component incorporated 300 boosted trees, culminating in a total of 5278 leaf nodes, averaging approximately 17.59 leaves per tree. This model effectively captured both lagged climate and consumption variables. The final ensemble output for the wet season model was computed via a Ridge regression meta-learner that combined the predictions of the three base learners: Random Forest, XGBoost, and Bayesian Ridge regression. The coefficients of this meta learner are 0.0655, -0.00926 and 0.9515 for the Random Forest, XGBoost and Bayesian Ridge respectively. This yields the following stacked ensemble Equation:

$$\hat{y}_{stacked} = -0.0058 + 0.0655\hat{y}_{RF} - 0.00926\hat{y}_{XGB} + 0.9515\hat{y}_{BR} \quad (27)$$

The Breusch–Pagan test produced a statistic of 16.14 with a p-value of 0.06, suggesting that the residuals have a constant variance across the fitted value range. In addition, the Shapiro–Wilk test for normality yielded a statistic of 0.8445 with a p-value of 0.043. While the Bayesian Ridge model dominated both the wet and dry season models, yielding weights above 90%, the residuals for the dry season exhibited statistically significant deviation from normality (Shapiro–Wilk p = 0.0043). This non-normality has been supported by findings from a study conducted by Tucker et al. [63], which showed that water consumption models in dry periods often deviate from Gaussian error assumptions due to infrastructural and behavioral adaptations. Autocorrelation diagnostics before and after model fitting are presented in Table 15**.** After model fitting, Durbin–Watson test yielded a value of 2.48, indicating no strong evidence of positive or negative autocorrelation. Furthermore, the Ljung–Box test produced a test statistic of 5.57 and a p-value of 0.85, which indicates that there is not sufficient evidence to reject the null hypothesis that there is no autocorrelation up to lag 10. This further confirms that the residuals behave like white noise which satisfies a key assumption of time series models. Figure 3 shows the actual vs predicted plots for all models.

Table 15. Autocorrelation diagnostics for Model 4

	Durbin-Watson Statistic	Ljung-Box (p-value) at lag 10
Before Modelling	0.695	$2.11\times\ 10^{-38}$
After Modelling	2.48	0.85



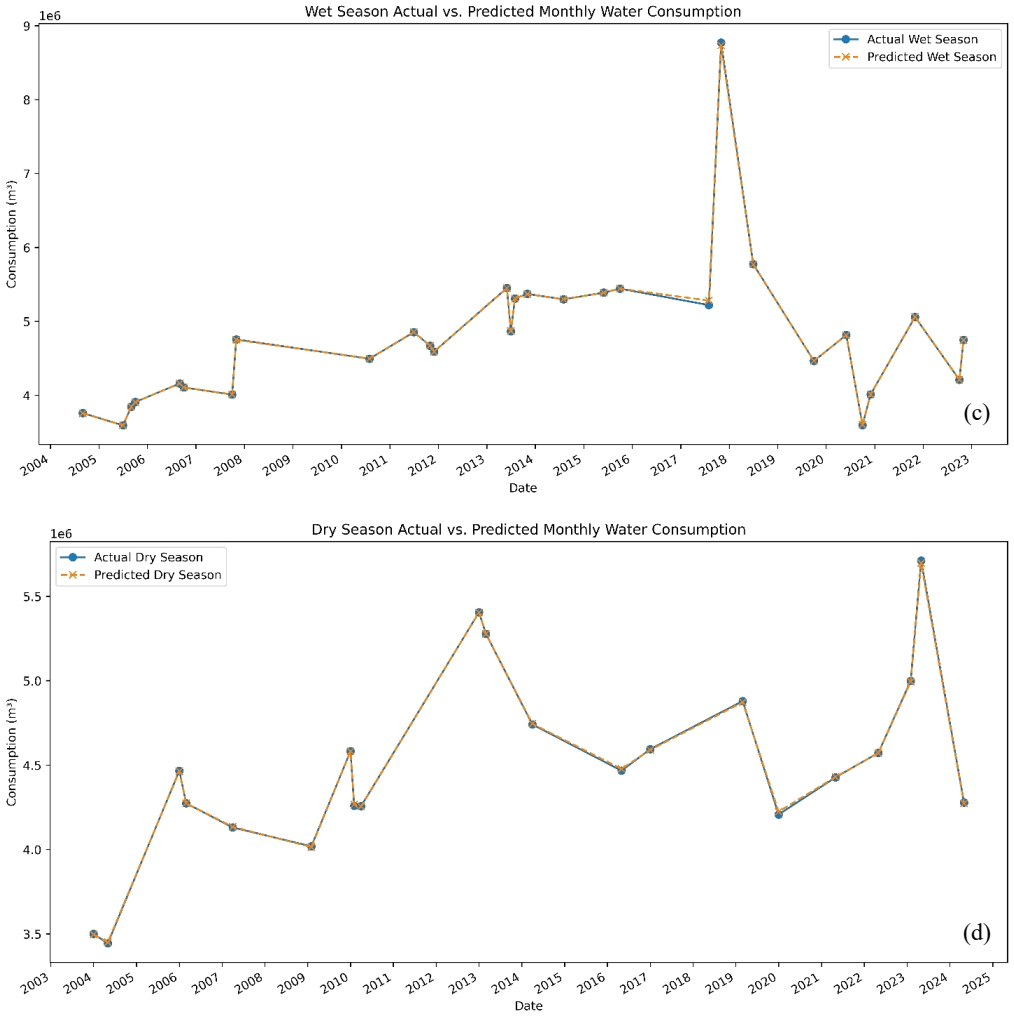


Figure 3. Actual vs Predicted plots for water consumption: (a) Model 1 – stacking ensemble (STL decomposition and residual model); (b) Model 2 – Hybrid model; (c) Model 3 – Stacking ensemble for wet season; (d) Model 4 – Stacking ensemble model for dry season.

Figure 3-a shows the actual versus predicted plot for model 1. The predicted values closely match the observed consumption patterns, accurately reflecting both rising and falling trends and capturing significant variations in consumption behavior. The model effectively tracked the long-term upward trend until 2014 and the subsequent decline from 2015 onward. The strong alignment between the actual and predicted lines from 2007 to 2018 indicates high long-term prediction accuracy with minimal errors. This stacked model also responded to fluctuations observed from 2018 to 2020; however, it did not effectively capture the extreme value in 2019. After 2020, the predictions remained aligned with the actual values, despite consumption levelling off at a slower rate. Some minor discrepancies between actual and predicted values appeared between 2021 and 2024, but overall, the predictions closely matched actual consumption, demonstrating the model's ability to generalize effectively. This performance is further supported by an index of agreement of 0.9571, indicating a high correlation between observed and predicted values in both magnitude and pattern. In the transformed space, Model 1 showed strong predictive accuracy with a mean absolute error (MAE) of 0.0378 and a root mean square error (RMSE) of 0.0643, indicating low average deviations from the true values. The coefficient of determination (R^2) was 0.8761, showing that about 87.6% of the variance in the denoised-transformed consumption signal was explained by the model.

Figure 3-b illustrates the actual vs predicted plots for model 2. The model captures a steady rise from 2004 to 2014, followed by a gradual decline and stabilization post-2015. However, instances of under- and overestimation are apparent, particularly in 2012 and during the 2018 to 2020 period, where deviations between predicted and observed values are noted. The 2019 spike observed in Model 1 is absent here due to smoothing effects from preprocessing. Despite this, the model maintains strong alignment between actual and predicted values as evidenced by the high index of agreement of 0.9808. Performance metrics also confirm the hybrid model's accuracy with mean absolute error (MAE) of 0.0195, root mean square error (RMSE) of 0.0282 and R^2 value of 0.9423.

Figure 3-c displays the actual vs predicted plots for model 3, that is, the wet season model. This model achieves near perfect alignment between observed and predicted values. It accurately captures the steady increase in consumption from 2005 to 2014 and successfully replicates the pronounced spike observed in 2018 which highlights the model's ability to respond to long term trends and sudden anomalies. Minimal deviations are observed post 2019. This wet season model is supported by an R^2 of 0.998, a mean absolute error (MAE) of 0.0029 and root mean square error (RMSE) of 0.0035. In addition, the index of agreement was 0.999, indicating perfect agreement between actual and predicted values. These results show that model 3 is very effective at predicting water consumption during the wet season, as it offers both precision and sensitivity to fluctuating consumption patterns.

Figure 3-d illustrates the actual vs predicted plot for the dry season model. The model demonstrates excellent alignment with the observed consumption across the period 2003 to 2024. The model accurately represents the sharp spikes in consumption between 2005 and 2013, and the subsequent decline in consumption from 2014 to 2016, followed by steady fluctuations thereafter. The model, therefore, reflects sensitivity to long-term trends and variability in the data. Model diagnostics achieve an R^2 of 0.991, with error metrics on the test set as follows: mean absolute error (MAE) of 0.0365 and root mean square error (RMSE) of 0.0541. The index of agreement was close to 1, indicating perfect agreement between actual and predicted values. The performance of the proposed models compares favourably with previous studies. Previous works using hybrid techniques reported R^2 values below 0.9 [21, 70], whereas the present study reported higher values. The seasonal sub-models (Models 3 and 4) yielded very high coefficients for R^2 ($R^2 >$ 0.99), and hence steps were taken to reduce the potential for overfitting.. Each model was trained and validated using an 80/20 time series split by date and 5-fold cross-validation within the training set. A 3-fold cross-validation was also used within the stacking ensemble to improve generalization and reduce overfitting. The difference between training and testing R^2 values was negligible ($\triangle R^2 < 0.01$), demonstrating stable generalization and therefore strongly reducing the likelihood for overfitting [71].

4. Conclusion

This study presents a comprehensive evaluation of advanced ensemble and hybrid modelling techniques for forecasting monthly water consumption patterns in Trinidad, with a specific focus on accounting for seasonality and climatic variability. Random Forest and XGBoost consistently delivered substantial predictive performance across all models, especially in ensemble and hybrid modelling approaches. In contrast, Bayesian Ridge Regression dominated as the primary predictor in the ensembles tailored explicitly for both wet and dry seasons. Analysis of the Bayesian Ridge models indicated positive associations between year and maximum temperature with increased water consumption, while humidity had a negative association. This pattern highlights the sensitivity of water demand to increases in temperature and decreases in humidity, which is consistent with prior studies on climate–demand relationships. Moderate levels of precipitation were associated with higher consumption, whereas heavy rainfall led to a reduction in demand, thereby highlighting its non-linear influence on water use.

Despite nonparametric tests (Mann-Whitney U) showing no statistically significant differences in consumption distributions or residuals between wet and dry periods, season-specific models improved their capacity to capture distinct seasonal dynamics and time-dependent structures. The ensemble methodologies exhibited high predictive performance, as evidenced by R^2 values reaching up to 0.998 during the wet season and 0.991 during the dry season. In addition, while Model 2 showed superior predictive performance over Model 1, Model 1 provided statistically robust residual behavior and fully met residual assumptions. The error metrics indicated consistently low mean absolute errors and root mean square errors across all models, which suggests excellent forecasting capabilities. Residual analysis largely supported assumptions, with minor non-normality during the dry season potentially reflecting behavioral adaptations.

Overall, this study underscores the value of combining linear and non-linear modelling techniques and season-specific models to improve the accuracy of water consumption modelling in tropical climates. One limitation of this research relates to the potential inaccuracy of historical and climate data. Despite being obtained from reputable sources, measurement errors over the years may have introduced uncertainties into the model. Verified data from national agencies were used, and preprocessing steps such as standardization, transformation, and wavelet/STL decomposition were applied to reduce noise and stabilize variance, thus minimizing the impact of data quality limitations common in developing countries. To address potential overfitting, future work can explore larger datasets, apply methods such as early stopping and test transferability of the model across different climatic contexts. Incorporating additional predictors, such as socioeconomic or land-use data, can further improve the model's applicability for water resources management in tropical regions.

5. Declarations

5.1. Author Contributions

Conceptualization, A.R. and V.T.; methodology, A.R. and V.T.; software, A.R.; validation, A.R., V.T., S.T., and A.C.; formal analysis, A.R. and S.T.; investigation, A.R.; resources, A.R., V.T., S.T., and A.C.; data curation, A.R.; writing—original draft preparation, A.R.; writing—review and editing, A.R., V.T., S.T., and A.C.; visualization, A.R.; supervision, V.T.; project administration, A.R.; funding acquisition, A.R. All authors have read and agreed to the published version of the manuscript.

5.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

5.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

5.4. Conflicts of Interest

The authors declare no conflict of interest.

6. References

[1] Beharry, S. L., & Clarke, R. M. (2023). Estimations of future reservoir volumes under different climate scenarios for a tropical reservoir in a small Caribbean Island, Trinidad. Environmental Monitoring and Assessment, 195(5), 590. doi:10.1007/s10661-023-11207-8.

[2] Hemati, A., Rippy, M. A., Grant, S. B., Davis, K., & Feldman, D. (2016). Deconstructing demand: The anthropogenic and climatic drivers of urban water consumption. Environmental Science & Technology, 50(23), 12557–12566. doi:10.1021/acs.est.6b02938.

[3] Mycoo, M. A., & Roopnarine, R. R. (2024). Water resource sustainability: Challenges, opportunities and research gaps in the English-speaking Caribbean Small Island Developing States. PLOS Water, 3(1), e0000222. doi:10.1371/journal.pwat.0000222.

[4] Saeed, F. H., Al-Khafaji, M. S., Al-Faraj, F. A. M., & Uzomah, V. (2024). Sustainable Adaptation Plan in Response to Climate Change and Population Growth in the Iraqi Part of Tigris River Basin. Sustainability (Switzerland) , 16(7), 2676. doi:10.3390/su16072676.

[5] Alikhani, M. R., & Moeini, R. (2025). Predicting the urban water demand by equipping intelligent-based methods with discrete wavelet transform function. Applied Water Science, 15(2), 38. doi:10.1007/s13201-025-02368-7.

[6] 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), 35005. doi:10.1088/2634-4505/ad5e1d.

[7] Skanupong, N., Xu, Y., Yu, L., Wan, Z., & Wang, S. (2024). The convolutional neural network for Pacific decadal oscillation forecast. Environmental Research Letters, 19(12), 124022. doi:10.1088/1748-9326/ad8be2.

[8] Shu, J., Xia, X., Han, S., He, Z., Pan, K., & Liu, B. (2024). Long-term water demand forecasting using artificial intelligence models in the Tuojiang River basin, China. PLOS ONE, 19(5), e0302558. doi:10.1371/journal.pone.0302558.

[9] Kulaczkowski, A., & Lee, J. (2024). Harnessing the Power of Random Forest for Precise Short-Term Water Demand Forecasting in Italian Water Districts †. Engineering Proceedings, 69(1), 81. doi:10.3390/engproc2024069081.

[10] Li, X., Wu, X., Sun, M., Yang, S., & Song, W. (2022). A Novel Intelligent Leakage Monitoring-Warning System for Sustainable Rural Drinking Water Supply. Sustainability (Switzerland), 14(10), 6079. doi:10.3390/su14106079.

[11] Liu, J., Zhou, X. L., Zhang, L. Q., & Xu, Y. P. (2023). Forecasting Short-term Water Demands with an Ensemble Deep Learning Model for a Water Supply System. Water Resources Management, 37(8), 2991–3012. doi:10.1007/s11269-023-03471-7.

[12] Ji, F. (2024). Time Series Prediction Method for Meteorological Data Based on the ARIMA-LSTM Model. Academic Journal of Science and Technology, 13(1), 193–196. doi:10.54097/rx3d5d70.

[13] Rajballie, A., Tripathi, V., & Chinchamee, A. (2022). Water consumption forecasting models - a case study in Trinidad (Trinidad and Tobago). Water Supply, 22(5), 5434–5447. doi:10.2166/ws.2022.147.

[14] Bakhshipour, A. E., Namdari, H., Koochali, A., Dittmer, U., & Haghighi, A. (2024). An Ensemble Data-Driven Approach for Enhanced Short-Term Water Demand Forecasting in Urban Areas. Engineering Proceedings, 69(1), 69. doi:10.3390/engproc2024069069.

[15] Herrera, M., Torgo, L., Izquierdo, J., & Pérez-García, R. (2010). Predictive models for forecasting hourly urban water demand. Journal of Hydrology, 387(1–2), 141–150. doi:10.1016/j.jhydrol.2010.04.005.

[16] Johnson, R. C., Burian, S. J., Halgren, J., Irons, T., Baur, E., Aziz, D., Hassan, D., Li, J., Kirkham, T., Stewart, J., & Briefer, L. (2024). Preparing municipal water system planning for a changing climate: Integrating climate-sensitive demand estimates. Journal of the American Water Resources Association, 60(1), 211–225. doi:10.1111/1752-1688.13165.

[17] Wu, W., & Kang, Y. (2024). Ensemble Empirical Mode Decomposition Granger Causality Test Dynamic Graph Attention Transformer Network: Integrating Transformer and Graph Neural Network Models for Multi-Sensor Cross-Temporal Granularity Water Demand Forecasting. Applied Sciences (Switzerland), 14(8), 3428. doi:10.3390/app14083428.

[18] Banda, P., Bhuiyan, M., Zhang, K., & Song, A. (2021). Multivariate monthly water demand prediction using ensemble and gradient boosting machine learning techniques. Proceedings of the International Conference on Evolving Cities, 14. doi:10.55066/proc-icec.2021.14.

[19] Zubaidi, S. L., Al-Bdairi, N. S. S., Ortega-Martorell, S., Ridha, H. M., Al-Ansari, N., Al-Bugharbee, H., Hashim, K., & Gharghan, S. K. (2023). Assessing the Benefits of Nature-Inspired Algorithms for the Parameterization of ANN in the Prediction of Water Demand. Journal of Water Resources Planning and Management, 149(1), 04022075. doi:10.1061/(asce)wr.1943-5452.0001602.

[20] Hekimoğlu, M., Çetin, A. İ., & Kaya, B. E. (2023). Evaluation of Various Machine Learning Methods to Predict Istanbul's Freshwater Consumption. International Journal of Environment and Geoinformatics, 10(2), 1–11. doi:10.30897/ijegeo.1270228.

[21] Tiwari, M. K., & Adamowski, J. F. (2015). Medium-Term Urban Water Demand Forecasting with Limited Data Using an Ensemble Wavelet–Bootstrap Machine-Learning Approach. Journal of Water Resources Planning and Management, 141(2), 4014053. doi:10.1061/(asce)wr.1943-5452.0000454.

[22] Bachour, R., Maslova, I., Ticlavilca, A. M., Walker, W. R., & McKee, M. (2016). Wavelet-multivariate relevance vector machine hybrid model for forecasting daily evapotranspiration. Stochastic Environmental Research and Risk Assessment, 30(1), 103–117. doi:10.1007/s00477-015-1039-z.

[23] Xu, Y., Zhang, J., Long, Z., & Chen, Y. (2018). A novel dual-scale deep belief network method for daily urban water demand forecasting. Energies, 11(5), 1068. doi:10.3390/en11051068.

[24] GOBTT. (2021). National Water Resources Management Strategy. Ministry of Public Utilities, Government of the Republic of Trinidad and Tobago, Port of Spain, Trinidad & Tobago.

[25] WRA. (2020). Water Resource Assessment for Trinidad and Tobago. Water Resources Agency, Water and Sewerage Authority (WASA), Port of Spain, Trinidad & Tobago.

[26] McSweeney, C., New, M. & G. Lizcano, G. (2010). UNDP Climate Change Country Profiles: Trinidad and Tobago. United Nations Development Programme, New York, United States.

[27] GCCA+. (2019). Trinidad and Tobago National Adaptation Strategy: Climate Resilience in the Water Sector. Global Climate Change Alliance Plus, United Nations Development Programme, New York, United States.

[28] SIDS. (2017). Impacts of Climate Change on Settlements and Infrastructure in the Coastal and Marine Environments of Caribbean Small Island Developing States (SIDS). Science Review 2017, Caribbean Marine Climate Change Report Card. Available online: https://www.gov.uk/government/publications/commonwealth-marine-economies-cme-programme-caribbean-marine-climate-change-report-card-scientific-reviews (accessed on January 2026).

[29] Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: principles and practice. OTexts, Melbourne, Australia.

[30] Box, G. E. P., & Cox, D. R. (1964). An Analysis of Transformations. Journal of the Royal Statistical Society Series B: Statistical Methodology, 26(2), 211–243. doi:10.1111/j.2517-6161.1964.tb00553.x.

[31] Gong, Q., Li, Q., Gavrielides, M. A., & Petrick, N. (2020). Data transformations for statistical assessment of quantitative imaging biomarkers: Application to lung nodule volumetry. Statistical Methods in Medical Research, 29(9), 2749–2763. doi:10.1177/0962280220908619.

[32] Venables, W. N., & Ripley, B. D. (2002). Modern Applied Statistics with S. Statistics and Computing. Springer, New York, United States. doi:10.1007/978-0-387-21706-2.

[33] Yao, D., Zhang, B., Li, X., Zhan, X., Zhan, X., & Zhang, B. (2023). Applying negative sample denoising and multi-view feature for lncRNA-disease association prediction. Frontiers in Genetics, 14. doi:10.3389/fgene.2023.1332273.

[34] Donoho, D. L., & Johnstone, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. Journal of the American Statistical Association, 90(432), 1200–1224. doi:10.1080/01621459.1995.10476626.

[35] Percival, D. B., & Walden, A. T. (2000). Wavelet Methods for Time Series Analysis. Cambridge University Press, Cambridge, United Kingdom. doi:10.1017/cbo9780511841040.

[36] Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). STL: A seasonal-trend decomposition. Journal of Official Statistics, 6(1), 3-73.

[37] Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5–32. doi:10.1023/a:1010933404324.

[38] MacKay, D. J. C. (1992). Bayesian Interpolation. Neural Computation, 4(3), 415–447. doi:10.1162/neco.1992.4.3.415.

[39] Christopher, M. B. (2006). Pattern recognition and machine learning. Springer, New York, United States.

[40] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13-17-August-2016, 785–794. doi:10.1145/2939672.2939785.

[41] Wolpert, D. H. (1992). Stacked generalization. Neural Networks, 5(2), 241–259. doi:10.1016/S0893-6080(05)80023-1.

[42] Storn, R., & Price, K. (1997). Differential Evolution - A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces. Journal of Global Optimization, 11(4), 341–359. doi:10.1023/A:1008202821328.

[43] Daw, A., Karpatne, A., Watkins, W. D., Read, J. S., & Kumar, V. (2022). Physics-guided neural networks (pgnn): An application in lake temperature modeling. Knowledge guided machine learning, Chapman and Hall/CRC, London, United Kingdom. doi:10.1137/1.9781611974973.33.

[44] Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. Transactions of the ASABE, 50(3), 885–900. doi:10.13031/2013.23153.

[45] Durbin, J., & Watson, G. S. (1950). Testing for serial correlation in least squares regression. Biometrika, 37(3–4), 409–428. doi:10.1093/biomet/37.3-4.409.

[46] Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2008). Time Series Analysis. Wiley Series in Probability and Statistics, John Wiley & Sons, Hoboken, United States. doi:10.1002/9781118619193.

[47] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning. In Springer Texts in Statistics. Springer, New York, United States. doi:10.1007/978-1-4614-7138-7.

[48] Alsulaili, A., Alkandari, M., & Buqammaz, A. (2022). Assessing the impacts of meteorological factors on freshwater consumption in arid regions and forecasting the freshwater demand. Environmental Technology & Innovation, 25, 102099. doi:10.1016/j.eti.2021.102099.

[49] Dimkić, D. (2020). Temperature Impact on Drinking Water Consumption. Environmental Sciences Proceedings, 2(1), 31. doi:10.3390/environsciproc2020002031.

[50] Breusch, T. S., & Pagan, A. R. (1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation. Econometrica, 47(5), 1287. doi:10.2307/1911963.

[51] Koenker, R. (1981). A note on studentizing a test for heteroscedasticity. Journal of Econometrics, 17(1), 107–112. doi:10.1016/0304-4076(81)90062-2.

[52] SHAPIRO, S. S., & WILK, M. B. (1965). An analysis of variance test for normality (complete samples). Biometrika, 52(3–4), 591–611. doi:10.1093/biomet/52.3-4.591.

[53] Khashei, M., & Bijari, M. (2011). A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. Applied Soft Computing, 11(2), 2664–2675. doi:10.1016/j.asoc.2010.10.015.

[54] Moosavi, S. F., Salehnia, N., Seifi, A., AsgharpourMasouleh, A., & Salehnia, N. (2023). Designing and calibrating an agent-based platform to evaluate the effect of climate variables on residential water demand. Water and Environment Journal, 37(3), 604–615. doi:10.1111/wej.12864.

[55] Makpiboon, C., Pornprommin, A., & Lipiwattanakarn, S. (2020). Impacts of weather variables on urban water demand at multiple temporal scales. International Journal of GEOMATE, 18(67), 71–77. doi:10.21660/2020.67.5758.

[56] Sadok, W., Lopez, J. R., & Smith, K. P. (2021). Transpiration increases under high-temperature stress: Potential mechanisms, trade-offs and prospects for crop resilience in a warming world. Plant Cell and Environment, 44(7), 2102–2116. doi:10.1111/pce.13970.

[57] Farah, E., & Shahrour, I. (2025). Forecasting Urban Water Demand Using Multi-Scale Artificial Neural Networks with Temporal Lag Optimization. Water (Switzerland), 17(19), 2886. doi:10.3390/w17192886.

[58] Timotewos, M. T., Barjenbruch, M., & Behailu, B. M. (2022). The Assessment of Climate Variables and Geographical Distribution on Residential Drinking Water Demand in Ethiopia. Water (Switzerland), 14(11), 1722. doi:10.3390/w14111722.

[59] Willard, J., Jia, X., Xu, S., Steinbach, M., & Kumar, V. (2020). Integrating physics-based modeling with machine learning: A survey. arXiv preprint, arXiv:2003.04919, 1-34. doi:10.1145/1122445.1122456.

[60] Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. Nature, 566(7743), 195–204. doi:10.1038/s41586-019-0912-1.

[61] DeGuzman, K., Knappenberger, T., Brantley, E., & Olshansky, Y. (2023). Estimating runoff probability from precipitation data: A binomial regression analysis. Hydrological Processes, 37(11), e15029. doi:10.1002/hyp.15029.

[62] Donkor, E. A., Mazzuchi, T. A., Soyer, R., & Alan Roberson, J. (2014). Urban Water Demand Forecasting: Review of Methods and Models. Journal of Water Resources Planning and Management, 140(2), 146–159. doi:10.1061/(asce)wr.1943-5452.0000314.

[63] Tucker, J., MacDonald, A., Coulter, L., & Calow, R. C. (2014). Household water use, poverty and seasonality: Wealth effects, labour constraints, and minimal consumption in Ethiopia. Water Resources and Rural Development, 3, 27-47. doi:10.1016/j.wrr.2014.04.001.

[64] San Diego County Water Authority. (2021). Water Demand Forecasting Model Update 2020. County Water Authority, San Diego, United States.

[65] Tipping, M. E. (2000). Sparse Bayesian Learning and the Relevance Vector Machine. Journal of Machine Learning Research, 1, 211–244. doi:10.1162/15324430152748236.

[66] Zhu, X., Miao, P., Qin, J., Li, W., Wang, L., Chen, Z., & Zhou, J. (2023). Spatio-temporal variations of nitrate pollution of groundwater in the intensive agricultural region: Hotspots and driving forces. Journal of Hydrology, 623, 129864. doi:10.1016/j.jhydrol.2023.129864.

[67] Sollmann, R. (2024). Estimating the temporal scale of lagged responses in species abundance and occurrence. Ecosphere, 15(1), e4704. doi:10.1002/ecs2.4704.

[68] Sazib, N., Bolten, J., & Mladenova, I. (2020). Exploring spatiotemporal relations between soil moisture, precipitation, and streamflow for a large set of watersheds using google earth engine. Water (Switzerland), 12(5), 1371. doi:10.3390/w12051371.

[69] Msigwa, A., Chawanda, C. J., Komakech, H. C., Nkwasa, A., & Van Griensven, A. (2022). Representation of seasonal land use dynamics in SWAT+ for improved assessment of blue and green water consumption. Hydrology and Earth System Sciences, 26(16), 4447–4468. doi:10.5194/hess-26-4447-2022.

[70] Xenochristou, M., & Kapelan, Z. (2020). An ensemble stacked model with bias correction for improved water demand forecasting. Urban Water Journal, 17(3), 212–223. doi:10.1080/1573062X.2020.1758164.

[71] Lovaglio, P. G. (2025). Cross-Learning With Panel Data Modeling for Stacking and Forecast Time Series Employment in Europe. Journal of Forecasting, 44(2), 753–780. doi:10.1002/for.3224.