ASEAN
Journal of Scientific and Technological Reports
Online ISSN:2773-8752

Research article

Forecasting Rainfall in the Ping, Wang, Yom, and Nan River Basins of Thailand using Decomposition and Holt-Winters Methods Enhanced by GRG Nonlinear Optimization

Pradthana Minsan[1*], and Watha Minsan[2]

[1] Faculty of Science and Technology, Chiang Mai Rajabhat University, Chiang Mai, 50300, Thailand
[2] Faculty of Science, Chiang Mai University, Chiang Mai, 50200, Thailand
* Correspondence: pradthana_min@g.cmru.ac.th

Citation:
Minsan, P.; Minsan, W.; Forecasting rainfall in the Ping, Wang, Yom, and Nan River Basins of Thailand using decomposition and holt-winters methods enhanced by GRG nonlinear optimization. *ASEAN J. Sci. Tech. Report.* **2025**, *28*(5), e259297. https://doi.org/10.55164/ajstr.v28i5.259297.

Article history:
Received: May 13, 2025
Revised: August 20, 2025
Accepted: September 1, 2025
Available online: September 14, 2025

Abstract: The Ping, Wang, Yom, and Nan rivers in Northern Thailand are vital for national sustainability, serving as key sources of energy and public utilities. This study utilizes secondary data from monitoring stations operated by the Upper Northern Region Irrigation Hydrology Center under the Royal Irrigation Department (RID), Thailand. The primary objective is to develop optimal parameter estimation methods for rainfall forecasting in the upper Ping, Wang, Yom, and Nan river basins. Time series forecasting techniques examined include decomposition with the Whale Optimization Algorithm (WOA-D), Holt-Winters smoothing enhanced by WOA (WOA-HW), and Generalized Reduced Gradient (GRG) nonlinear optimization methods (GRG-D and GRG-HW). These are compared with traditional decomposition and Holt-Winters models developed using Minitab (Minitab-D) and Excel (ForecastSheet-HW). The secondary objective is to forecast rainfall over a 24-month horizon to support trend analysis and improve water resource management. Seasonal ARIMA (SARIMA) is also employed for comparative analysis. Results show that traditional tools, such as Minitab-D and ForecastSheet-HW, are accessible and effective in minimizing forecasting errors. Metaheuristic models such as WOA offer improved accuracy but require programming expertise. GRG solvers provide a practical balance, offering near-comparable accuracy without the need for coding. GRG-D and GRG-HW produced forecasts closely matching actual rainfall across all basins. GRG-HW and WOA-HW achieved the lowest sMAPE values for the Yom and Nan Rivers, GRG-D outperformed Minitab-D for the Wang River, and ForecastSheet-HW remained most effective for the Ping River.

Keywords: Generalized reduced gradient; whale optimization algorithm; rorecasting; time series

1. Introduction

Accurate rainfall forecasting in Thailand's major river basins is crucial for effective water resource management, informed agricultural planning, flood prevention, and the development of infrastructure related to water management. Thailand comprises several major river basins, among which the Ping, Wang, Yom, and Nan rivers form significant tributaries contributing to the Chao Phraya River, the country's primary waterway. Analyzing rainfall patterns in these basins is therefore essential for developing precise forecasting models.

Over the past decade, Thailand has faced significant challenges due to climate change, which has considerably affected rainfall patterns [1]. Notable examples include the severe flooding of 2011, which caused extensive economic damage, and the recent unprecedented floods in northern Thailand in 2024, primarily driven by the La Niña phenomenon and intensified by typhoons Yagi and Soulik. These conditions caused heavy and persistent rainfall, leading to substantial water accumulation, flash floods, and river overflow, severely impacting over 43,000 households across 19 provinces, resulting in at least 45 fatalities. Chiang Mai, situated on the Ping River, experienced two significant flood events within a year, underscoring the severe limitations of existing flood-warning systems. These events underscore the critical need for improved water management and enhanced early warning systems to mitigate future disasters. Conversely, periods of severe drought have also impacted agricultural productivity and water availability, as evidenced during the 2015–2016 period. Consequently, the increased variability in rainfall makes accurate forecasting particularly vital for disaster preparedness and sustainable water resource management.

The application of statistical and data science methodologies for rainfall forecasting has garnered significant attention globally [2]. Mishra and Desai applied the Box-Jenkins method to forecast droughts using the standardized precipitation index (SPI) series in India's Kansabati river basin, achieving forecasts consistent with observed data 1–2 months ahead, with accuracy decreasing for longer lead times. Wang et al. employed an integrated Ensemble Empirical Mode Decomposition-Autoregressive Integrated Moving Average (EEMD-ARIMA) approach to forecast annual runoff data for Biuliuhe, Dahuofang, and Mopanshan reservoirs in China, demonstrating effective forecasting capabilities [3]. Within the context of Thailand, Jiroge et al. developed forecasting models for inflow to reservoirs managed by the Electricity Generating Authority of Thailand (EGAT), employing decomposition, Holt-Winters, Box-Jenkins methods, and combined forecasting methods, utilizing Minitab and Excel Office 365 for data analysis [4].

Recent years have seen growing interest in combining metaheuristic algorithms with traditional time series forecasting methods. An early contribution in this direction was made by Jiang et al., who improved the Holt–Winters smoothing method by incorporating the Fruit Fly Optimization Algorithm (FOA) to forecast monthly electricity consumption [5]. Inspired by this approach, more recent work in 2024 has focused on applying similar techniques to hydrological forecasting. For instance, Minsan and Minsan integrated the Whale Optimization Algorithm (WOA) with both decomposition and Holt–Winters methods to predict reservoir inflows in southern Thailand [6]. They also applied Cuckoo Search (CS) in combination with these methods for forecasting monthly inflows into reservoirs in eastern Thailand [7]. Beyond hydrology, these integrated metaheuristic approaches have proven effective in other domains, including PM2.5 air quality forecasting and government revenue prediction [8–10].

Therefore, this research is structured around two primary objectives. First, to develop optimal parameter estimation models using Root Mean Square Error (RMSE) as a performance metric, employing time series analysis methods to forecast rainfall in the upper Ping, Wang, Yom, and Nan river basins. The approaches include decomposition combined with WOA (WOA-D), Holt-Winters smoothing enhanced by WOA (WOA-HW), and methods integrating Generalized Reduced Gradient (GRG) nonlinear optimization, namely GRG-D and GRG-HW. Additionally, these methods are compared with decomposition models created by Minitab and Holt-Winters models developed using Excel during the training data phase. The second objective is to forecast rainfall for a 24-month horizon, utilizing these models for practical rainfall trend analysis and prediction, to enhance efficient water management in northern Thailand. Seasonal ARIMA (SARIMA) methodology is also employed for comprehensive analysis. Model performance comparisons and selections are made based on the Symmetric Mean Absolute Percentage Error (sMAPE) criterion, serving as the primary performance indicator for selecting the most suitable forecasting model for each of the four river basins.

2. Materials and Methods

2.1 Data Preparation

This research aims to identify suitable forecasting methods for rainfall in the upper northern region of Thailand. Monthly rainfall data were obtained from the Upper Northern Region Irrigation Hydrology Center, Royal Irrigation Department (RID), Thailand, covering four major river basins: Ping, Wang, Yom, and

Nan. The secondary data [11] were recorded monthly in millimeters at monitoring stations managed by RID. Specific data sources for each river basin included:

Ping River: Stations 071511 (Ban Huai Luek, Chiang Dao district) and 07132 (Chiang Dao district), Chiang Mai Province.

Wang River: Stations W.16A (Ban Hai, Chae Hom district) and Kiew Lom Dam (Mueang district), Lampang Province.

Yom River: Stations Y.20 (Ban Huai Sak, Song district) and Y.1C (Ban Nam Khong, Mueang district), Phrae Province.

Nan River: Stations 28053 (Thung Chang district), 28073 (Tha Wang Pha district), and Mueang district, Nan Province.

The dataset encompasses 120 months from April 2014 to March 2024. The data were partitioned into two subsets:

Training Dataset (80% of the data, April 2014 to March 2022, totaling 96 months) for developing forecasting models.

Testing Dataset (20% of the data, April 2022 to March 2024, totaling 24 months) for evaluating model performance.

Initial examination of Figure 1 indicated that the time series exhibited both seasonal and trend components. Deseasonalized data analyzed through the Autocorrelation Function (ACF) (Figure 2) showed significant correlations at lag 1, which gradually decreased with increasing lag, indicating a trend component. Detrended data (Figure 3) revealed prominent negative autocorrelations at lags 6 and 18, and positive autocorrelations at lags 12 and 24, suggesting the presence of seasonal fluctuations. Thus, the data contain both seasonal and trend characteristics. Variance stability was assessed using Levene's Test to evaluate homoscedasticity. Prior to the test, the time series data were transformed to reduce autocorrelation, ensuring independence among observations. The results (Table 1) showed p-values greater than the significance level of 0.05 for all river basins, indicating constant variance across all river basins. Consequently, an additive model was deemed appropriate for subsequent analyses and forecasting.

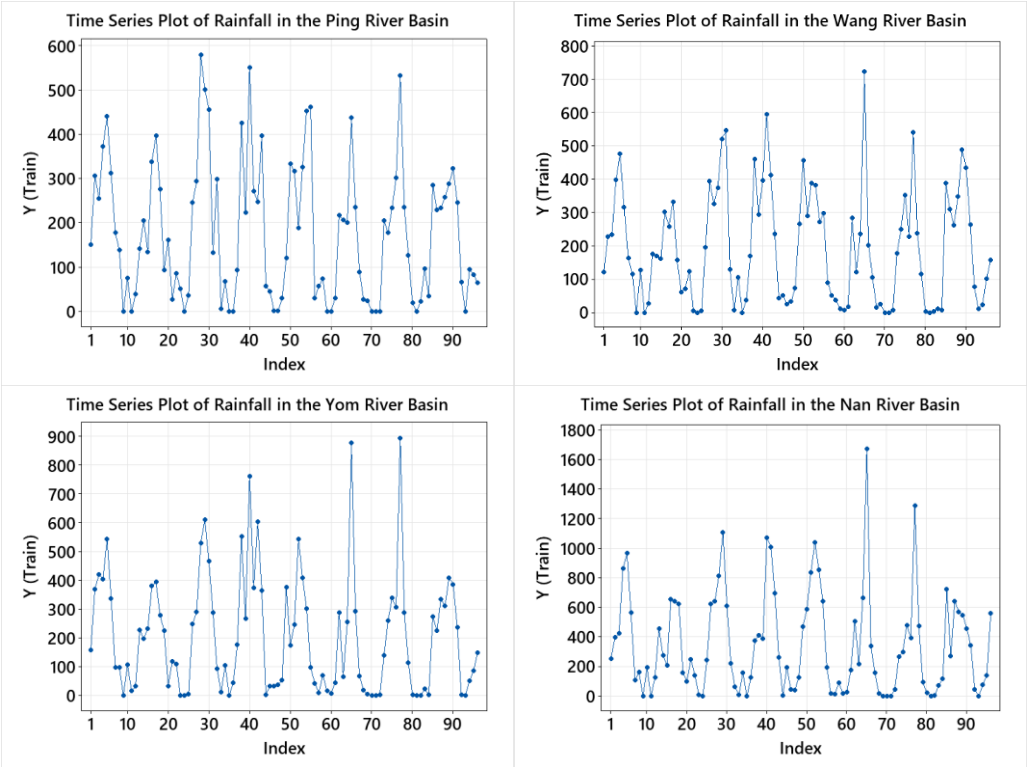

Figure 1. Monthly precipitation in the Ping, Wang, Yom, and Nan Basins.

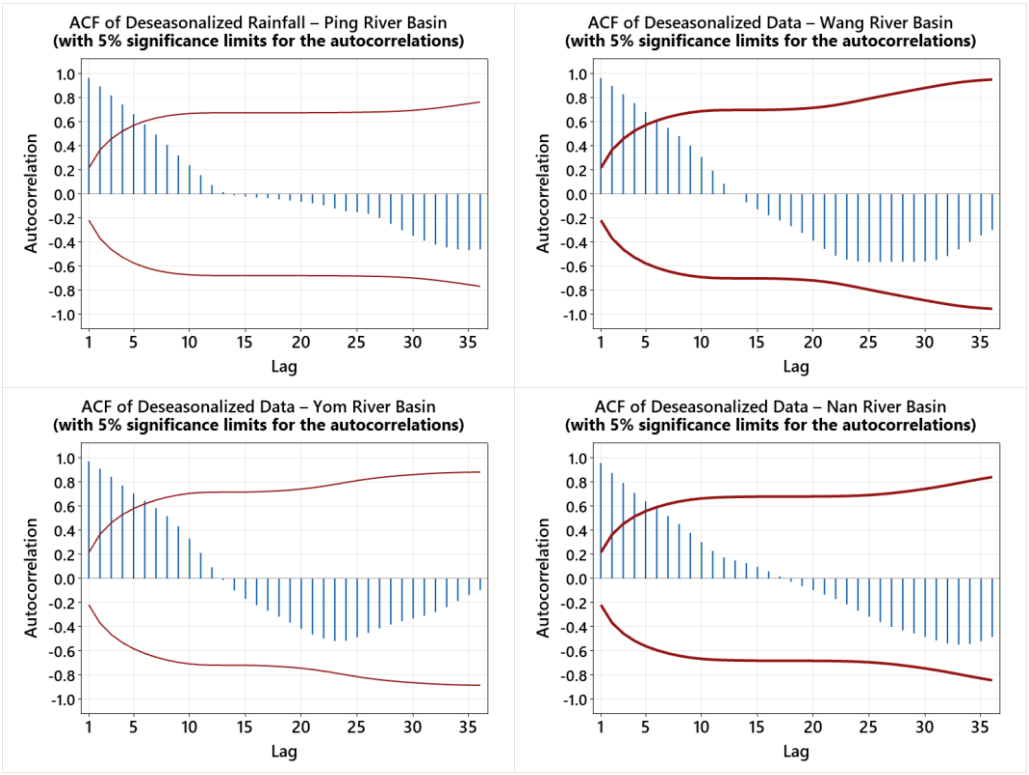

Figure 2. ACF of the deseasonalized time series data of the four river basins.

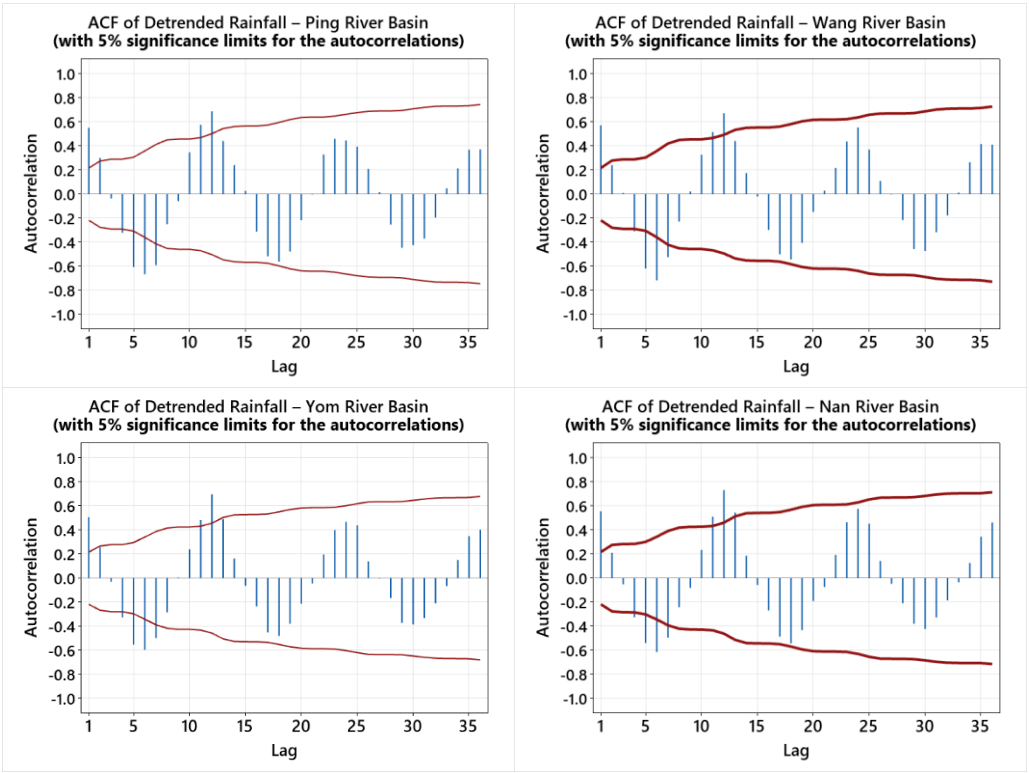

Figure 3. ACF of the detrended time series data of the four river basins.

Table 1. Results of Levene's test for homogeneity of variances across the four river basins.

Levene's Test	River Basin			
	Ping	**Wang**	**Yom**	**Nan**
Test Statistic	2.31	0.22	0.10	0.69
P-Value	0.132	0.638	0.749	0.408

2.2 Holt-Winters Method

The Holt-Winters method is widely recognized for its ability to accommodate changes in level, trend, and seasonality within time series data. It can effectively accommodate both additive and multiplicative forms, depending upon the underlying characteristics of the time series data. Notably, the Holt-Winters method remains robust even with limited training data. For the additive model, the forecasting process follows equations (1) through (4) sequentially, as detailed below:
Additive Forecasting:

$$\hat{Y}_{t+p} = \hat{T}_t + p\hat{\beta}_t + \hat{S}_{t-s+1+((p-1) \bmod s)} \text{ for } p = 1, 2, ..., \quad (1)$$

$$\hat{T}_t = \alpha(Y_t - \hat{S}_{t-s}) + (1-\alpha)(\hat{T}_{t-1} + \hat{\beta}_{t-1}), \quad (2)$$

$$\hat{\beta}_t = \gamma(\hat{T}_t - \hat{T}_{t-1}) + (1-\gamma)\hat{\beta}_{t-1}, \quad (3)$$

$$\hat{S}_t = \delta(Y_t - \hat{T}_t) + (1-\delta)\hat{S}_{t-s}, \quad (4)$$

In these equations, $\hat{Y}_{t+p}$ denotes the forecasted data at a future time $t+p$, and p denotes the number of periods ahead to forecast. The parameter s is set to 12 months in this study. Additionally, $\hat{T}_t$ denotes the estimated level at time t, $\hat{\beta}_t$ denotes the trend component, and $\hat{S}_t$ denotes the seasonal component.

2.2.1 Integration of Holt-Winters with Whale Optimization Algorithm (WOA-HW)

Following the approach of Minsan and Minsan [6, 8], the Holt-Winters method requires the estimation of three key smoothing parameters: level (α), trend (γ), and seasonality (δ). Each parameter lies within the range [0, 1] and plays a crucial role in adjusting the model components to best reflect the underlying time series behavior, including long-term trends and seasonal fluctuations. To determine the optimal values for these parameters, this study employs the Whale Optimization Algorithm (WOA), a nature-inspired metaheuristic algorithm. The goal is to minimize forecasting error by tuning the parameters specifically for the additive Holt-Winters model. The optimization process, including a detailed explanation of its implementation, is illustrated in the pseudo-code provided in Figure 4. The performance of the WOA-HW model is evaluated using the RMSE, which serves as the objective function for optimization. The objective is to minimize the RMSE, as defined in Equation (5), calculated over the training dataset.

The number of whales: $N = 100$, the number of parameters: $m = 3$, maximum iterations: $T_{\max} = 300$, time limit: $MaxTime = 30$ sec., the fitness value fails to improve after a specified number of iterations $T_{improve} = 50$, the bound of search area: Range Initialize $X_i = (x_i^j, x_i^2, x_i^3), X^*$
While (t < $T_{\max}$ or time < $MaxTime$ or the fitness value fails to improve after a specified $T_{improve}$)
For i = 1 to N
    Check if any search agent goes beyond the search space and amend it
    For j = 1 to m
        $p = rand[0,1]$
        Update $a, r, A, C, D, D', b, l, X_{rand}$
        If $p \geq 0.5$ then
            Update x_i^j the position of the current search agent #Exploitation Phase
        Elseif $p < 0.5$ and $|A| < 1$
            Update x_i^j the position of the current search agent #Encircling Prey
        Elseif $p < 0.5$ and $|A| \geq 1$
            Update x_i^j the position of the current search agent #Exploration Phase
        Endif
    End for
End for
Calculate $fitness(X_i)$ using WOA-HW by Equation (5)
Update X^* if there is a better solution
$t = t + 1$
End while
Return $X^* = (x^{1*}, x^{2*}, x^{3*})$ # Objective Minimize RMSE($\alpha^*, \gamma^*, \delta^*$) where $\alpha^* = x^{1*}, \gamma^* = x^{2*}, \delta^* = x^{3*}$

Figure 4. Pseudo-code of the WOA-HW [6, 12].

$$\text{Objective Minimize RMSE}(\alpha,\gamma,\delta),$$
$$\text{Variable range } \begin{cases} 0 \leq \alpha \leq 1 \\ 0 \leq \gamma \leq 1, \\ 0 \leq \delta \leq 1 \end{cases}$$
$$\text{RMSE} = \frac{1}{n_1}\sum_{t=1}^{n_1}\left(Y_t - \hat{Y}_t\right)^2, \tag{5}$$

where n_1 denotes the number of observations in the training dataset (96 months in this study),
Y_t denotes the observed rainfall, and
$\hat{Y}_t$ denotes the corresponding forecasted value from the WOA-HW model.

2.2.2 Integration of Holt-Winters with GRG Nonlinear Optimization (GRG-HW)

This method builds upon the WOA-HW model by incorporating GRG Nonlinear optimization, available through Microsoft Excel's Solver add-in. Based on the findings of Minsan [13], GRG Nonlinear, when used in Excel 2019, is capable of efficiently identifying optimal parameters for forecasting models. Therefore, in this study, GRG Nonlinear was employed as an alternative optimization technique to WOA, to determine the optimal values of the three smoothing parameters (α, γ, and δ) for the Holt-Winters model.

The GRG-HW model follows the same forecasting equations (1) to (4) as in WOA-HW, with GRG Nonlinear used to replace the WOA metaheuristic for parameter tuning. The optimization setup was configured in Excel Solver with detailed parameter settings to ensure consistency and replicability. These settings are summarized as follows:

Constraint precision: 0.000001
Solving limits: Maximum time = 30 seconds
Use automatic scaling: True
Convergence: 0.0001
Population size: 100
Random seed: 0
Require bounds on variables: True
Derivatives: Forward; and use multistart: True

This configuration enables Solver to efficiently explore the solution space and determine optimal values for the Holt-Winters parameters. The objective function minimized in the GRG-HW model is the same RMSE as defined in Equation (5). The final forecast values, $\hat{Y}_t$, are generated using the parameter set obtained from the GRG-HW model, evaluated on the same training dataset of 96 monthly rainfall observations used in WOA-HW. To facilitate replication, the Excel file containing the GRG-HW Solver setup, parameter constraints, and example data has been provided as supplementary material [14].

2.2.3 Forecasting in Excel using Forecast Sheet (ForecastSheet-HW)

Microsoft Excel 365 includes a built-in tool called Forecast Sheet, which enables users to create time series forecasting models through a guided interface. This tool employs the exponential smoothing method and enables the automatic or manual configuration of the three smoothing parameters (level, trend, and seasonality) used in Holt-Winters forecasting. In this study, the ForecastSheet-HW approach is incorporated as a baseline method and compared with the GRG-HW and WOA-HW models discussed in earlier sections. The comparison aims to evaluate whether the Forecast Sheet can estimate the three smoothing parameters and generate forecasts with comparable accuracy and model adaptability. For consistency, the following settings were manually specified when using the Forecast Sheet in Excel:

Seasonality: Set manually to 12
Include forecast statistics: True

These configurations ensure a controlled environment for evaluating the forecasting performance of ForecastSheet-HW in relation to other optimization-based Holt-Winters implementations.

2.3 Decomposition

This study explores three decomposition-based forecasting approaches. The first approach involves traditional decomposition using Minitab software, referred to in this study as Minitab-D. The other two approaches, described in Sections 2.3.1 and 2.3.2, incorporate optimization algorithms.

2.3.1 Integration of Decomposition with Whale Optimization Algorithm (WOA-D)

Following the approach proposed by Minsan and Minsan [6, 8], the decomposition method is applied to forecast rainfall time series by estimating 14 parameters: $\hat{\beta}_0, \hat{\beta}_1, \hat{S}_1, \hat{S}_2, \hat{S}_3, \hat{S}_4, \hat{S}_5, \hat{S}_6, \hat{S}_7, \hat{S}_8, \hat{S}_9, \hat{S}_{10}, \hat{S}_{11}, \hat{S}_{12}$. These parameters are optimized within the range [0, 1] and are rescaled based on a scaling method proposed in Minsan and Minsan [6, 8]. The effectiveness of the WOA-D model is evaluated by minimizing RMSE, as defined by the objective function below:

$$\text{Objective Minimize RMSE}(\hat{\beta}_0, \hat{\beta}_1, \hat{S}_1, \hat{S}_2, ..., \hat{S}_{12})$$

$$\text{Variable range} \begin{cases} \hat{\beta}_0' - 0.2\left|\hat{\beta}_0'\right| \le \hat{\beta}_0 \le \hat{\beta}_0' + 0.2\left|\hat{\beta}_0'\right| \\ \hat{\beta}_1' - 0.2\left|\hat{\beta}_1'\right| \le \hat{\beta}_1 \le \hat{\beta}_1' + 0.2\left|\hat{\beta}_1'\right| \\ \text{LB}_i \le \hat{S}_i \le \text{UB}_i) \text{ for } i = 1, 2, ..., 12 \end{cases}$$

$$\text{RMSE} = \frac{1}{n_1}\sum_{t=1}^{n_1}\left(Y_t - \hat{Y}_t\right)^2, \tag{6}$$

where $\hat{Y}_t$ denotes the corresponding forecasted value from the WOA-D model.

The seasonal component bounds for the additive model UB_i LB_i are derived as follows: $\mathrm{UB}_i = \frac{|\Delta Y_t|}{2}$, $\mathrm{LB}_i = -\frac{|\Delta Y_t|}{2}$ where $\Delta Y_t = Y_t - Y_{t-1}$. This range constraint is adapted from [6, 8] to enhance the efficiency of the optimization search space, particularly when implemented using spreadsheet-based solvers such as Excel. The trend component is represented by a simple linear equation and estimated using the ordinary least squares (OLS) method as follows: $\hat{Y}_t' = \hat{\beta}_0' + \hat{\beta}_1' t$. The optimization process, including a detailed explanation of its implementation, is illustrated in the pseudo-code provided in Figure 5.

The number of whales: $N = 100$, the number of parameters: $m = 14$, maximum iterations: $T_{\max} = 300$, time limit: $MaxTime = 30$ sec., the fitness value fails to improve after a specified: $T_{improve} = 50$, the bound of search area: Range Initialize $X_i = (x_i^j, x_i^2, ..., x_i^{14}), X^*$
While (t < $T_{\max}$ or time < $MaxTime$ or the fitness value fails to improve after a specified $T_{improve}$)
For i = 1 to N
    Check if any search agent goes beyond the search space, and amend it
  For j = 1 to m
    $p = rand[0,1]$
    Update $a, r, A, C, D, D', b, l, X_{rand}$
    If $p \geq 0.5$ then
      Update x_i^j the position of the current search agent #Exploitation Phase
    Elseif $p < 0.5$ and $|A| < 1$
      Update x_i^j the position of the current search agent #Encircling Prey
    Elseif $p < 0.5$ and $|A| \geq 1$
      Update x_i^j the position of the current search agent #Exploration Phase
    Endif
  End for
End for
Scaling Parameters
Calculate $fitness(X_i)$ using WOA-D by Equation (6)
Update X^* if there is a better solution
$t = t + 1$
End while
Return $X^* = (x^{1*}, x^{2*}, ..., x^{14*})$
Objective Minimize RMSE($\hat{\beta}_0^*, \hat{\beta}_1^*, \hat{S}_1^*, \hat{S}_2^*, \hat{S}_3^*, \hat{S}_4^*, \hat{S}_5^*, \hat{S}_6^*, \hat{S}_7^*, \hat{S}_8^*, \hat{S}_9^*, \hat{S}_{10}^*, \hat{S}_{11}^*, \hat{S}_{12}^*$) where $\hat{\beta}_0^* = x^{1*}, \hat{\beta}_1^* = x^{2*}, ..., \hat{S}_{12}{}^* = x^{14*}$

Figure 5. Pseudo-code of the WOA-D [6, 12].

2.3.2 Integration of Decomposition with GRG Nonlinear Optimization (GRG-D)

This method builds upon the WOA-D approach described in Section 2.3.1 by substituting the metaheuristic algorithm with the GRG Nonlinear optimization available in Microsoft Excel. The GRG Nonlinear algorithm was applied to determine 14 parameters for the decomposition model, including the trend coefficients ($\hat{\beta}_0$, $\hat{\beta}_1$) and 12 seasonal components ($\hat{S}_1$, $\hat{S}_2$, $\hat{S}_3$, $\hat{S}_4$, $\hat{S}_5$, $\hat{S}_6$, $\hat{S}_7$, $\hat{S}_8$, $\hat{S}_9$, $\hat{S}_{10}$, $\hat{S}_{11}$, $\hat{S}_{12}$). The structure of the objective function and forecasting equations remains consistent with those used in the WOA-D model. Following the procedure of Minsan and Minsan [6, 8], the GRG Solver in Excel was configured with the setting "Use automatic scaling" enabled, as it improves the effectiveness of parameter estimation, particularly when the data scale is small. This adjustment aligns with the scaling approach applied in WOA-D. To facilitate replication, the Excel file containing the GRG-D Solver setup, parameter constraints, and example data has been provided as supplementary material [14]. Users can test the model directly using the default Solver

configuration. Model performance is evaluated using RMSE, as defined in Equation (6). Forecast values $\hat{Y}_t$ are calculated using the GRG-D model on the training dataset, with the default Solver tolerance parameters retained, similar to the configuration used in GRG-HW.

2.4 Box-Jenkins Method

The SARIMA model, or Seasonal Autoregressive Integrated Moving Average, extends the classical ARIMA model by incorporating seasonal components. It is commonly expressed as SARIMA(p,d,q)(P, D, Q)$_s$, where s denotes the seasonal period (e.g., 12 for monthly data). This model accounts for both non-seasonal and seasonal autoregressive (AR), moving average (MA), and differencing components. Equation (7) provides the general form of the SARIMA model:

$$\phi_p(B)\Phi_P(B^L)(1-B)^d(1-B^L)^D Y_t = \delta + \theta_q(B)\Theta_Q(B^L)\varepsilon_t \,. \tag{7}$$

Where $\delta = \phi_p(B)\Phi_P(B^L)\mu$, with μ being the mean of the stationary series

B denotes the backward shift operator, such that $B^p y_t = y_{t-p}$

d denotes the degree of non-seasonal differencing

D denotes the degree of seasonal differencing

L denotes the length of the seasonal cycle (e.g., $L = 12$ for monthly data)

$\phi_p(B) = (1-\phi_1 B - \phi_2 B^2 - ... - \phi_p B^p)$ denotes a non-seasonal autoregressive operator of order p

$\theta_q(B) = (1-\theta_1 B - \theta_2 B^2 - ... - \theta_q B^q)$ denotes a non-seasonal moving average operator of order q

$\Phi_P(B^L) = (1-\Phi_1 B^L - \Phi_2 B^{2L} - ... - \Phi_P B^{PL})$ denotes a seasonal autoregressive operator of order P

$\Theta_Q(B^L) = (1-\Theta_1 B^L - \Theta_2 B^{2L} - ... - \Theta_Q B^{QL})$ denotes the seasonal moving average operator of order Q

The SARIMA model combines both non-seasonal and seasonal components into a unified framework, allowing it to effectively capture both short-term fluctuations and recurring seasonal patterns within time series data.

The Box-Jenkins approach involves a systematic sequence of steps to identify, estimate, validate, and forecast using SARIMA models. The key steps are as follows:

2.4.1 Stationarity Check

The time series is analyzed using the autocorrelation function (ACF) and partial autocorrelation function (PACF) to determine whether the series is stationary. If the series is found to be non-stationary, differencing must be applied to achieve stationarity.

2.4.2 Regular Differencing

If the ACF and PACF suggest a non-stationary trend component, a non-seasonal differencing of order $d = 1$ or $d = 2$ may be applied to stabilize the mean of the series.

2.4.3 Seasonal Differencing

If the ACF and PACF exhibit a clear seasonal pattern, seasonal differencing must be applied using an appropriate seasonal order D, typically 1 or 2, depending on the periodicity of the seasonal structure.

2.4.4 Data Transformation

If the variance of the series is non-constant, a variance-stabilizing transformation, such as the Box-Cox transformation, may be applied. This transformation helps stabilize the variance across time and supports the use of modeling techniques that assume homoscedasticity, thereby improving the reliability of parameter estimation and residual diagnostics.

2.4.5 Model Identification

Once a stationary series is obtained, the initial model structure is identified by analyzing the ACF and PACF plots to determine the appropriate values for the non-seasonal and seasonal orders (p, d, q) and (P, D, Q), respectively.

2.4.6 Parameter Estimation

Model parameters are estimated using an appropriate method, typically the Ordinary Least Squares (OLS) method or the Maximum Likelihood Estimation (MLE) method.

2.4.7 Model Diagnostics

Model adequacy is assessed through the following diagnostic procedures:

1. Test the significance of each model parameter using the t-test.
2. Examine residual autocorrelation using the Ljung-Box Q-test at appropriate lags (e.g., lag 12) to check for randomness of residuals.
3. Assess residual normality using the Kolmogorov–Smirnov (KS) test.
4. Evaluate mean equality of residuals over time using the t-test to determine temporal stability.
5. Assess the variance equality of residuals using Levene's test to verify homoscedasticity across forecast horizons.
6. Use the Autoregressive Conditional Heteroscedasticity–Lagrange Multiplier (ARCH-LM) test to detect conditional heteroscedasticity, indicating the presence of time-varying volatility in residuals.

If any diagnostic test fails, steps 2.4.5 through 2.4.7 should be revisited to refine the model until all assumptions are reasonably satisfied.

2.4.8 Forecasting

Once a suitable model passes all diagnostics, it is used to generate forecasts for the desired period. The resulting forecasts are expected to provide reliable estimates for future values in the time series.

2.5 Evaluation Criteria

The evaluation in this study comprises two main aspects. The first focuses on in-sample accuracy, which evaluates the parameter estimation process using the training dataset. The objective is to minimize the RMSE, calculated as:

$$\text{RMSE} = \frac{1}{n_1}\sum_{t=1}^{n_1}\left(Y_t - \hat{Y}_t\right)^2$$

Where n_1 is the number of observations in the training dataset (96 months), and Y_t and $\hat{Y}_t$ denote the actual and forecasted rainfall values, respectively.

The RMSE values were computed from six forecasting approaches:

1. Minitab-D: Parameter estimation using Minitab.
2. WOA-D: Parameter estimation using Python and Whale Optimization Algorithm.
3. GRG-D Solver: Parameter estimation in Excel using GRG Nonlinear Solver.
4. ForecastSheet-HW: Holt-Winters parameters estimated using Excel's Forecast Sheet.
5. WOA-HW: Holt-Winters parameters estimated using Whale Optimization Algorithm.
6. GRG-HW: Parameters estimated using the GRG Solver in Excel for the Holt-Winters method.

The RMSE results from methods 1 through 3 (Minitab-D, WOA-D, GRG-D) and 4 through 6 (ForecastSheet-HW, WOA-HW, GRG-HW) are compared within their respective groups to identify the most efficient parameter estimation strategy.

The second evaluation aspect focuses on forecasting accuracy, assessed using the testing dataset. This stage aims to evaluate each model's ability to forecast rainfall 24 months ahead (i.e., two years into the future). The performance metric employed is the Symmetric Mean Absolute Percentage Error (sMAPE). The adjusted MAPE was first introduced by Flores in 1986 [15]. It has been widely adopted in hydrological forecasting studies and has also been implemented in the HydroErr open-source library by Elise et al. [16], which supports both Python and MATLAB environments.

One of the main advantages of sMAPE is that the forecast error is expressed as a percentage bounded between 0% and 100%, making it easier to interpret and compare across different models and time periods. The formula is defined as:

$$\text{sMAPE} = \frac{100}{24}\sum_{t=97}^{n}\frac{\left|Y_t - \hat{Y}_t\right|}{\left|Y_t\right| + \left|\hat{Y}_t\right|}$$

Where Y_t and $\hat{Y}_t$ denote the actual and forecasted rainfall values for the testing period. Here, $n = 120$ refers to the total number of months in the full dataset, with testing data covering months 97 through 120. Although $n = 120$ the sMAPE is computed using the last 24 months (i.e., $t = 97$ to 120)

Seven forecasting approaches were compared based on their sMAPE values: Minitab-D, WOA-D, GRG-D, ForecastSheet-HW, WOA-HW, GRG-HW, and Box-Jenkins. The approach with the lowest sMAPE is considered the most effective in terms of long-term forecasting accuracy.

3. Results and Discussion

3.1 Analysis of Training Dataset Performance

This study compares the performance of parameter estimation for decomposition-based forecasting using three methods: Minitab-D, WOA-D, and GRG-D. Additionally, it evaluates the performance of parameter estimation for Holt-Winters models using three methods: ForecastSheet-HW, WOA-HW, and GRG-HW. All models were applied to monthly rainfall data from the upper northern region of Thailand, covering the Ping, Wang, Yom, and Nan river basins. The rainfall data were obtained from hydrological monitoring stations under the Upper Northern Region Irrigation Hydrology Center, Royal Irrigation Department, and span 10 years. The RMSE values used to evaluate the forecasting performance on the training dataset are presented in Table 2.

Table 2. RMSE of the training dataset for each river basin.

Model		River basin			
		Ping	**Wang**	**Yom**	**Nan**
Decomposition	Minitab-D	84.2	89.5	110.0	165.1
	WOA-D	81.1	87.7*	104.1*	162.9*
	GRG-D	80.9*	87.8	104.7	163.9
Holt-Winters	ForecastSheet-HW	121.6	125.3	138.2	218.7
	WOA-HW	91.2*	102.2*	124.8*	187.5*
	GRG-HW	91.2*	102.2*	124.8*	187.5*

Note: The lowest RMSE value in group decomposition, and group Holt-Winters is *

For the WOA-HW and WOA-D methods, the number of iterations was set to 300, with execution performed using Python via Google Colab. Each iteration required approximately 20 seconds to converge. GRG-D and GRG-HW models were implemented in Excel using the GRG Nonlinear Solver. These optimization approaches aimed to enhance forecast accuracy by effectively tuning parameters. Compared to manual parameter estimation using Minitab-D and Excel's ForecastSheet-HW, the optimization-based approaches generally yielded better performance. The results presented in Table 2 show the following key findings. WOA-D and GRG-D both produced lower RMSE values than Minitab-D across all river basins. Among them, WOA-D achieved the lowest RMSE values in three out of four basins (Wang, Yom, and Nan), while GRG-D slightly outperformed WOA-D only in the Ping River basin. For Holt-Winters models, both WOA-HW and GRG-HW yielded the same RMSE values and clearly outperformed ForecastSheet-HW across all river basins. The best-performing models in each group (decomposition and Holt-Winters) are marked with an asterisk (*) in Table 2.

3.2 Analysis of Forecasting Performance on the Testing Dataset

3.2.1 Box-Jenkins

In the testing phase, the Box-Jenkins method was employed to construct seasonal ARIMA (SARIMA) models, following the methodology proposed by Minsan and Minsan [7]. They emphasized that forecasting future observations often involves higher uncertainty due to the accumulation of model and estimation errors. The testing dataset (i.e., the last 24 months) was used to assess how well the SARIMA models could predict future rainfall values across the four river basins. The goal was to ensure that the Box-Jenkins models not only fit the training data but also perform well in out-of-sample forecasting. The best SARIMA model for each basin was selected based on its ability to minimize residual autocorrelation and pass all relevant diagnostic tests, including the t-test for coefficient significance, Levene's test for homoscedasticity, the ARCH–LM test for conditional heteroscedasticity, the Kolmogorov–Smirnov (KS) test for residual normality, and the Ljung–Box test for independence. These models, along with their equations and transformation techniques (if applied), are summarized below. Ping River Basin: The selected model was SARIMA$(0,1,1)(1,0,0)_{12}$, with no transformation applied. The forecasting equation is

$$Y_t = Y_{t-1} + 0.8185Y_{t-12} - 0.8185Y_{t-13} - 0.9760e_{t-1}.$$

Wang River Basin: The model SARIMA$(0,1,2)(0,1,1)_{12}$ was used, with a Box-Cox transformation applied to the data $Z = Y^{\frac{1}{1.1}}$. The forecasting equation is

$$Z_t = Z_{t-1} + Z_{t-12} - Z_{t-13} - 0.7870e_{t-1} - 0.2041e_{t-2} - 0.8608e_{t-12} + 0.6774e_{t-13} + 0.1757e_{t-14}.$$

Yom River Basin: The SARIMA$(0,1,2)(0,1,3)_{12}$ model was selected, along with a Box-Cox transformation with $Z = Y^{\frac{1}{1.1}}$. The equation is

$$\begin{aligned} Z_t = {} & Z_{t-1} + Z_{t-12} - Z_{t-13} - 1.129e_{t-1} + 0.219e_{t-2} - 0.681e_{t-12} + 0.7689e_{t-13} \\ & - 0.1490e_{t-14} - 0.549e_{t-24} + 0.6197e_{t-25} - 0.1203e_{t-26} + 0.522e_{t-36} - 0.5891e_{t-37} + 0.1143e_{t-38}. \end{aligned}$$

Nan River Basin: The chosen model was SARIMA$(0,1,1)(1,1,0)_{12}$, using a Box-Cox transformation with $Z = Y^{\frac{1}{1.448}}$. The equation is

$$Z_t = Z_{t-1} + 0.446Z_{t-12} - 0.446Z_{t-13} + 0.554Z_{t-24} - 0.554Z_{t-25} - 0.9192e_{t-1}.$$

Table 3 summarizes the selected Box–Jenkins SARIMA models for the Ping, Wang, Yom, and Nan river basins, together with adequacy diagnostics including the t-test, Levene's test, ARCH-LM test, Kolmogorov–Smirnov (KS) test, and Ljung–Box (LB) test at lag 12. All models satisfied the diagnostic criteria with no evidence of residual autocorrelation, variance heterogeneity, non-normality, or conditional heteroskedasticity, confirming their suitability for forecasting rainfall in each river basin.

Table 3. Box-Jenkins models and model adequacy diagnostics by river basin.

River basin	Box-Jenkins SARIMA(p,d,q)(P,D,Q)$_{12}$	t-test (P-value)	Levene (P-value)	Arch-LM (P-value)	KS (P-value)	LB Lag 12 (P-value)
Ping	SARIMA$(0,1,1)(1,0,0)_{12}$	0.05 (0.963)	0.00 (0.963)	11.372 (0.251)	0.090 (0.054)	14.15 (0.166)
Wang	SARIMA$(0,1,2)(0,1,1)_{12}$	0.15 (0.878)	0.33 (0.565)	4.672 (0.862)	0.085 (0.145)	13.28 (0.151)
Yom	SARIMA$(0,1,2)(0,1,3)_{12}$	-0.71 (0.480)	0.05 (0.832)	10.411 (0.318)	0.074 (>0.150)	5.23 (0.632)
Nan	SARIMA$(0,1,1)(1,1,0)_{12}$	0.11 (0.916)	3.2 (0.078)	14.312 (0.112)	0.095 (0.066)	6.59 (0.763)

3.2.2 Summary of Forecasting Performance Based on sMAPE

To identify the most effective forecasting models for monthly rainfall across each river basin, the symmetric Mean Absolute Percentage Error (sMAPE) values from the testing dataset were compared, as shown in Table 4. The model with the lowest sMAPE value in each basin was selected as the most appropriate for 24-month-ahead forecasting. The results can be summarized as follows. For the Ping River Basin, the ForecastSheet-HW method achieved the lowest sMAPE of 37.0, outperforming all other models. Thus, ForecastSheet-HW is identified as the most suitable model for forecasting rainfall in this basin. In the Wang River Basin, the GRG-D method yielded the lowest sMAPE value of 32.1, indicating its superiority over other decomposition-based and Holt-Winters models. Therefore, GRG-D was selected as the best model for this basin. For the Yom and Nan River Basins, both GRG-HW and WOA-HW shared the lowest sMAPE values of 40.6 and 32.5, respectively. This suggests that the optimization-enhanced Holt-Winters models provided the most accurate forecasts for these two basins. Overall, when evaluating all models across the four basins, GRG-D or GRG-HW optimization methods produced the lowest sMAPE values in three out of four river basins. In contrast, the ForecastSheet-HW method was the most suitable model only for the Ping River Basin.

Table 4. sMAPE of the testing dataset for monthly rainfall in each river basin.

Model		River basin			
		Ping	**Wang**	**Yom**	**Nan**
Decomposition	Minitab-D	40.8	34.8	42.0	33.0
	WOA-D	46.3	34.4	50.0	34.9
	GRG-D	45.3	32.1*	45.1	33.5
Holt-Winters	ForecastSheet-HW	37.0*	47.0	43.7	57.3
	WOA-HW	42.2	39.0	40.6*	32.5*
	GRG-HW	42.2	39.0	40.6*	32.5*
Box-Jenkins		40.6	35.5	43.1	35.5

Note: * indicates the lowest sMAPE for each river basin.

Table 5 and Figure 6 present the 24-month-ahead rainfall forecasts for each river basin from April 2024 to March 2026. The results indicate that the highest inflow volumes are expected during the rainy season, typically between August and September for the Ping, Wang, and Yom Rivers, and July to August for the Nan River. This seasonal trend aligns closely with Thailand's climatological patterns, where monsoon rainfall peaks during these months. The forecasted rainfall volumes clearly demonstrate strong seasonal variation, with particularly high rainfall predicted between July and September, reflecting the annual rainy season. These peaks suggest a significant increase in water inflows to the river systems during this period. Such forecasts are highly valuable for water resource planning and management. They support proactive strategies in maintaining river levels, preparing for flood risks during the rainy season, and allocating water for downstream use during the dry season. Overall, the 24-month forecasts provide essential insights for sustainable water management across the upper northern region of Thailand.

Table 5. Forecasted Monthly Rainfall for Each River Basin (in millimeters), April 2024 – March 2026.

Month/Year	River basin			
	Ping	**Wang**	**Yom**	**Nan**
Apr-24	123	160	148	296
May-24	326	289	315	409
Jun-24	209	258	272	458
Jul-24	305	322	401	799*
Aug-24	411*	445*	524*	983*
Sep-24	397*	336*	434*	586
Oct-24	255	230	185	176
Nov-24	19	62	36	104
Dec-24	15	22	9	24
Jan-25	11	50	48	124
Feb-25	37	15	19	18
Mar-25	15	34	38	128
Apr-25	126	158	148	296
May-25	329	288	315	409
Jun-25	212	257	272	458
Jul-25	308	321	401	799*
Aug-25	414*	444*	524*	983*
Sep-25	400*	335*	434*	586
Oct-25	258	229	185	176
Nov-25	22	60	36	104
Dec-25	18	20	9	24
Jan-26	14	49	48	124
Feb-26	40	14	19	18
Mar-26	18	33	38	128

Note: * indicates the first and second highest rainfall volumes within each 12 months.

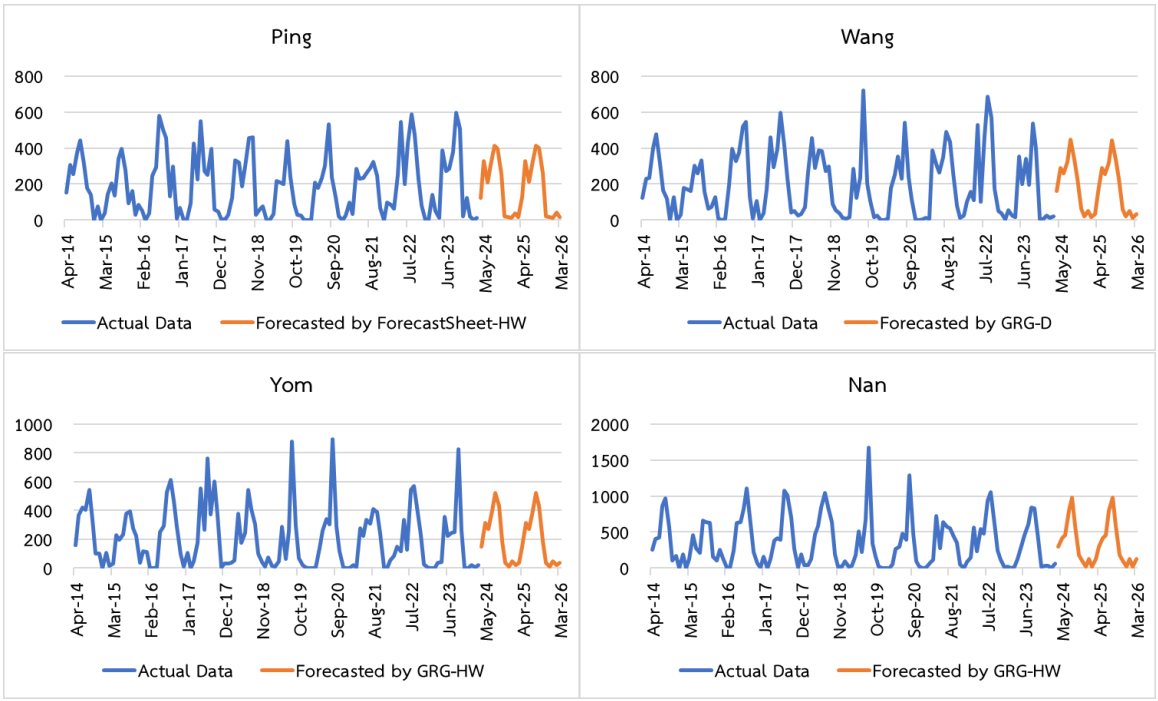

Figure 6. Actual and Forecasted Monthly Rainfall for Each River Basin Using the Most Appropriate Forecasting Methods.

4. Conclusions

This study examined the use of the GRG Nonlinear Solver, a built-in optimization tool in Microsoft Excel, to estimate parameters for time series forecasting models with trend and seasonality, specifically the decomposition (GRG-D) and Holt-Winters (GRG-HW) models. Based on the 24-month ahead forecasts, the Holt-Winters models proved most effective for the Ping, Yom, and Nan Rivers, reflecting rainfall patterns where recent observations should be given greater importance and older data progressively less weight through exponential smoothing. In contrast, the Wang River forecasts were best captured by the decomposition model, which treats all historical data points equally in the prediction process. These results indicate that the suitability of each method is closely tied to the way it weights past data, aligning with the inherent temporal characteristics of rainfall in each basin. While traditional estimation tools such as Minitab-D and ForecastSheet-HW are user-friendly and widely accessible, they often fall short in minimizing forecasting error as defined by the RMSE criterion. In particular, it is known to be inefficient in exploring the entire solution space and often fails to reach the global optimum. In contrast, metaheuristic algorithms such as the Whale Optimization Algorithm (WOA) can significantly improve forecasting accuracy, but they require programming expertise, which limits their practical use in non-technical settings. The GRG Nonlinear Solver offers a balanced alternative. It eliminates the need for coding while achieving RMSE values that are competitive with WOA, with differences typically within one decimal place. In this study, GRG-D and GRG-HW produced forecasts closely aligned with actual rainfall across the four basins. During the 24-month testing period, GRG-HW and WOA-HW delivered the lowest sMAPE values in the Yom and Nan Rivers, while GRG-D consistently provided the most accurate forecasts for the Wang River and showed robust performance overall. ForecastSheet-HW, however, remained the best-performing model for the Ping River. These findings suggest that the GRG Nonlinear Solver is a practical and effective alternative for rainfall forecasting, particularly when ease of use is a key consideration. To further enhance its utility, future research may explore the tuning of GRG parameters, such as solver bounds and convergence settings. Additionally, integrating the forecasting models with physical and climatic variables, such as reservoir inflows, watershed characteristics, and regional weather patterns, may support more robust decision-making in long-term water resource planning.

5. Acknowledgement

We would like to thank the Department of Mathematics and Statistics, Faculty of Science and Technology, Chiang Mai Rajabhat University, for providing us with research space and equipment. We also gratefully acknowledge the Data Science Research Center, Department of Statistics, Faculty of Science, Chiang Mai University, for providing access to Minitab software, which was used in the data analysis.

Author Contributions: For research articles with multiple authors, a brief paragraph outlining their individual contributions must be included. The following statements should be used "Conceptualization, Pradthana Minsan; methodology, P.M.; software and coding, P.M. and Watha Minsan.; analysis and visualization, P.M.; validation, P.M., and W.M.; writing—original draft preparation, P.M.; writing—review and editing, P.M., and W.M. All authors have read and agreed to the published version of the manuscript.

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

References

[1] Limsakul, A.; Singhruck, P. Long-term trends and variability of total and extreme precipitation in Thailand. *Atmospheric Research* **2016**, *169*, 301–317. https://doi.org/10.1016/j.atmosres.2015.10.015

[2] Mishra, A. K.; Desai, V. R. Drought forecasting using stochastic models. *Stoch Environ Res Risk Assess* **2005**, *19*, 326–339. https://doi.org/10.1007/s00477-005-0238-4

[3] Wang, W. C.; Chau, K. W.; Xu, D. M.; Chen, X. Y. Improving forecasting accuracy of annual runoff time series using ARIMA based on EEMD decomposition. *Water Resour Manage* **2015**, *29*(8), 2655–2675. https://doi.org/10.1007/s11269-015-0962-6

[4] Saeying, J.; Minsan, W.; Taninpong, P. Forecasting model for the amount of water flowing into the reservoirs of the electricity generating authority of Thailand (EGAT). *Recent Science and Technology* **2023**, *15*(2), 494–510. (In Thai)

[5] Jiang, W.; Wu, X.; Gong, Y.; Yu, W.; Zhong, X. Holt–Winters smoothing enhanced by fruit fly optimization algorithm to forecast monthly electricity consumption. *Energy* **2020**, *193*, 116779. https://doi.org/10.1016/j.energy.2019.116779

[6] Minsan, W.; Minsan, P. Decomposition and Holt-Winters enhanced by the whale optimization algorithm for forecasting the amount of water inflow into the large dam reservoirs in southern Thailand. *Journal of Current Science and Technology* **2024**, *14*(2), article 38.

[7] Minsan, P.; Minsan, W. Monthly volumes of water inflow into the large dam reservoirs in eastern Thailand forecasting by the cuckoo search optimization enhanced decomposition and Holt-Winters techniques. *Thai Journal of Operations Research* **2024a**, *12*(2), 69–89. (In Thai)

[8] Minsan, P.; Minsan, W. Decomposition and Holt-Winters techniques enhanced by whale optimization algorithm: Case study of pm2.5 forecasting in 8 northern provinces of Thailand. *Thai Science and Technology Journal* **2024b**, *32*(6), 12–34. (In Thai)

[9] Minsan, W.; Minsan, P. Incorporating decomposition and the Holt-Winters method into the whale optimization algorithm for forecasting monthly government revenue in Thailand. *Science & Technology Asia* **2023**, *28*(4), 38–53.

[10] Minsan, W.; Minsan, P.; Panichkitkosolkul, W. Enhancing decomposition and Holt-Winters weekly forecasting of pm2.5 concentrations in Thailand's eight northern provinces using the cuckoo search algorithm. *Thailand Statistician* **2024**, *22*(4), 963–985.

[11] The upper northern region irrigation hydrology center, the Royal Irrigation Department (RID), Thailand. https://www.hydro-1.net/ (accessed 2025-01-16).

[12] Mirjalili, S.; Lewis, A. The whale optimization algorithm. *Advances in Engineering Software* **2016**, *95*, 51–67. https://doi.org/10.1016/j.advengsoft.2016.01.008

[13] Minsan, P. Comparing methods of optimization in solver of Microsoft excel 2007 and 2019: A case study of statistical models. *The Journal of King Mongkut's University of Technology North Bangkok* **2021**, *31*(3), 496–511. https://doi.org/10.14416/j.kmutnb.2021.05.013

[14] Minsan, W.; Minsan, P. GRG-HW and GRG-D optimization models for rainfall forecasting in the Yom river basin. *Figshare* **2025**. https://doi.org/10.6084/m9.figshare.28953389
[15] Flores, B. E. A pragmatic view of accuracy measurement in forecasting. *Omega* **1986**, *14*(2), 93–98.
[16] Jackson, E. K.; Roberts, W.; Nelsen, B.; Williams, G. P.; Nelson, E. J.; Ames, D. P. Introductory overview: Error metrics for hydrologic modelling – A review of common practices and an open source library to facilitate use and adoption. *Environmental Modelling & Software* **2019**, *119*, 32–48. https://doi.org/10.1016/j.envsoft.2019.05.001.