E-ISSN 2586-9027
Homepage : https://tci-thaijo.org/index.php/SciTechAsia Science & Technology Asia
Vol.31 No.1 January - March 2026 Page: [1-17]
Original research article

Optimizing Combined Holt–Winters and Decomposition Forecasting Models with the Whale Optimization Algorithm for Monthly Maximum of Daily 24-hour Average PM2.5 Concentrations in the Bangkok Metropolitan Region

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

[1]*Department of Mathematics and Statistics, Faculty of Science and Technology, Chiang Mai Rajabhat University, Chiang Mai 50300, Thailand*
[2]*Data Science Research Center, Department of Statistics, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand*

Received 22 June 2025; Received in revised form 29 October 2025
Accepted 1 December 2025; Available online 27 March 2026

ABSTRACT

This study proposes a novel hybrid forecasting model, Combined Holt–Winters and Decomposition optimized by the Whale Optimization Algorithm (CHD–WOA), to predict the monthly maximum of daily 24-hour average PM2.5 concentrations in the Bangkok Metropolitan Region. The model integrates decomposition and Holt–Winters exponential smoothing forecasts through weighted averaging, with 18 parameters simultaneously optimized using the Whale Optimization Algorithm. Ten methods are compared, including WOA-optimized variants, classical combination models, and benchmark approaches such as Box–Jenkins and Long Short-Term Memory (LSTM). Performance is evaluated using RMSE for in-sample fit and MAPE for out-of-sample accuracy. The proposed CHD–WOA model achieves the lowest MAPE at four of eight stations and consistently outperforms benchmarks. Forecasts for 2025 accurately capture seasonal pollution peaks during the dry season, demonstrating the robustness and effectiveness of combining classical statistical techniques with metaheuristic optimization for complex urban air quality forecasting.

Keywords: Bangkok metropolitan region; Combined decomposition and holt–winters; PM2.5; Time series forecasting; Whale optimization algorithm

***Corresponding author:** pradthana_min@g.cmru.ac.th **doi:** xx.xxxxx/scitechasia.xxxx.xx

1. Introduction

Air pollution has become a persistent and increasingly critical environmental concern in urbanized areas worldwide. In Thailand, the Bangkok Metropolitan Region (BMR), which includes Bangkok, Nonthaburi, Pathum Thani, Samut Prakan, Samut Sakhon, and Nakhon Pathom, frequently suffers from elevated concentrations of fine particulate matter (PM2.5). These concentrations are largely driven by transportation emissions, industrial activities, biomass burning, and construction. In addition to local sources, recent studies have highlighted the role of transboundary transport, especially during the dry season. A modeling study using the Community Multiscale Air Quality (CMAQ) framework [1] indicated that biomass burning in neighboring countries such as Myanmar, Laos, Cambodia, and Vietnam contributes to elevated PM2.5 concentrations in Thailand. While the direct impact on the BMR was found to be minor overall, the study revealed that smoke plumes from Myanmar can significantly raise PM2.5 levels in the region during the peak burning period between March and April. These findings underscore the importance of considering both domestic and regional emission sources when developing forecasting systems and air quality management strategies. The resulting pollution levels pose significant public health risks, including respiratory and cardiovascular diseases, and call for effective monitoring and forecasting systems to support timely decision-making and mitigation measures (see Fig. 1 for the geographic scope of the BMR).

Forecasting PM2.5 concentrations presents several methodological challenges due to the data's inherent seasonality, long-term trends, and high variability. Classical time series forecasting techniques, particularly decomposition and Holt–Winters exponential smoothing (HW), have long been employed to capture such characteristics. Decomposition separates a time series into its trend, seasonal, and residual components, enabling interpretable and component-wise forecasting. Meanwhile, the HW method, originally formulated by Holt in 1957 [2] and Winters in 1960 [3], uses smoothing equations to estimate level, trend, and seasonality parameters, making it effective for structured seasonal data. However, both techniques rely on optimal parameter settings, which are difficult to determine manually, especially in complex datasets.

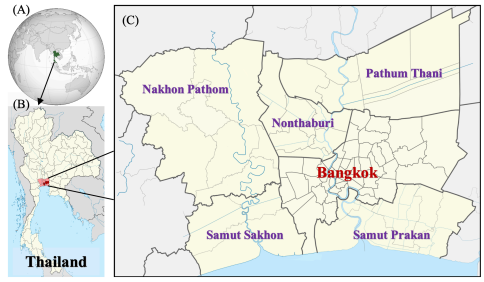


Fig. 1. Study area: (A) Modeling domain. (B) Target source Thailand. (C) Bangkok Metropolitan Region.

Recent research has demonstrated the effectiveness of using metaheuristic algorithms to optimize such parameters automatically. In particular, the Whale Optimization Algorithm (WOA), proposed by Mirjalili and Lewis in 2016 [4], has proven to be a robust method for global optimization in forecasting applications [5-11]. Previous studies conducted by the authors have explored the use of WOA and other metaheuristics in optimizing both HW and decomposition models. For example, WOA was employed to improve monthly government revenue forecasting and monthly dam inflow prediction in Southern Thai-

land by optimizing 3 and 14 parameters for HW and decomposition, respectively [12, 13]. Building on these, we applied Cuckoo Search Optimization (CSO) to both models in a Thai-language study forecasting monthly inflows into Eastern Thailand's dam reservoirs [14]. A separate study published in Thai extended the WOA framework to a more complex decomposition model with 54 parameters and applied it to weekly PM2.5 forecasting in northern Thailand [15].

In parallel, our most recent study introduced a hybrid approach integrating CSO-optimized HW with CSO-optimized decomposition to forecast weekly PM2.5 levels in eight provinces of northern Thailand [16], highlighting the complementary strengths of different algorithms and forecasting frameworks. Additionally, Vimolsutjarit et al. [17] employed the Fruit Fly Optimization Algorithm (FOA) to enhance parameter estimation in classical decomposition and HW models for dam reservoir forecasting in Northern Thailand, further supporting the effectiveness of metaheuristic-enhanced approaches in hydrological time series modeling.

These cumulative findings have demonstrated that forecasting models enhanced by metaheuristic optimization significantly outperform classical methods when applied to high-variability environmental time series data.

Building on this research trajectory, the present study introduces a novel hybrid forecasting model: the Combined HW and Decomposition (CHD) method. This model integrates the mathematical formulations of both the HW and decomposition methods into a unified forecasting framework. A weighting parameter is introduced to combine their structural contributions, and all parameters—including the smoothing coefficients, decomposition components, and the combination weight—are simultaneously optimized using the WOA. Thus, the CHD model comprises 18 parameters in total, with 3 from HW, 14 from decomposition, and 1 weight for combining the two methods. All parameters are estimated using the WOA, leveraging its efficiency and convergence properties. The resulting integrated model is referred to as CHD–WOA, which denotes the combined HW and decomposition method optimized via the Whale Optimization Algorithm.

To evaluate the proposed CHD–WOA model, we apply it to the forecasting of monthly maximum values of daily 24-hour average PM2.5 concentrations in the BMR. This dataset poses substantial forecasting challenges due to its irregular temporal patterns, seasonal fluctuations, and the influence of both local and transboundary emission sources. The BMR also serves as a representative case study for urban air quality forecasting, given its population density and environmental significance. The performance of the CHD–WOA model is compared with several benchmark methods, including standalone decomposition, HW, Box–Jenkins, Long Short-Term Memory (LSTM) networks, and a classical combination of HW and decomposition (CHD–Classical). This study aims to advance time series forecasting methodologies by combining the interpretability of classical statistical techniques with the optimization strength of metaheuristics, thereby supporting more informed environmental health policy and management in Thailand's most urbanized region.

2. Research Methods

2.1 Dataset and preliminary analysis

This study aims to develop an effective forecasting approach for the monthly

maximum values of daily 24-hour average PM2.5 concentrations in the BMR. This metric refers to the highest recorded daily mean PM2.5 concentration within each month, derived from 24-hour average data collected at air quality monitoring stations located across the BMR. These values reflect the most severe air pollution episodes experienced monthly, offering insight into population-level peak exposure.

Secondary data were obtained from the Pollution Control Department of Thailand [18], comprising measurements from eight air quality monitoring stations. Three stations are located within Bangkok:

- Kheha Chumchon Huai Khwang Stadium (Station Code: 11T)
- Nonsi Witthaya School (Station Code: 12T)
- The Government Public Relations Department (Station Code: 59T)

The remaining five stations are situated in adjacent provinces:

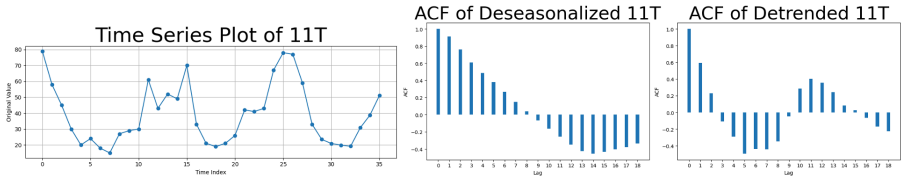
- Nonthaburi (Station Code: 13T)
- Samut Prakan (Station Code: 16T)
- Pathum Thani (Station Code: 20T)
- Samut Sakhon (Station Code: 27T)
- Nakhon Pathom (Station Code: 81T)

The monthly maximum for each station was computed as the highest daily mean within each calendar month. Only two missing observations were found in the entire dataset—at station 11T (October 2024) and station 16T (January 2023)—and these were estimated using linear interpolation to maintain continuity. The resulting monthly dataset was complete and uniformly applied to all forecasting models to ensure fair and consistent comparison.

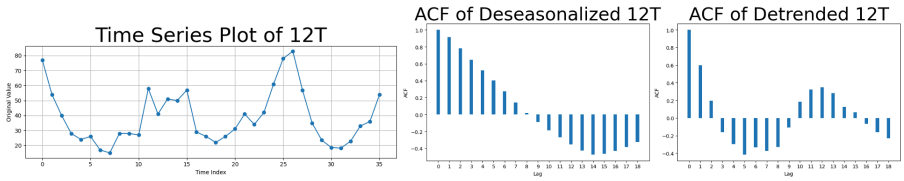
The data span 48 months, from January 2021 to December 2024, recorded on a monthly basis with units in micrograms per cubic meter ($\mu g/m^3$). The full dataset was divided into two subsets:

- The training set (75%) which covers January 2021 to December 2023 (36 months), used for model development.
- The test set (25%) which covers January 2024 to December 2024 (12 months), used to assess forecasting accuracy. The forecasting objective is to predict the values for January 2025 to December 2025, and the chosen test period aligns with this future horizon.

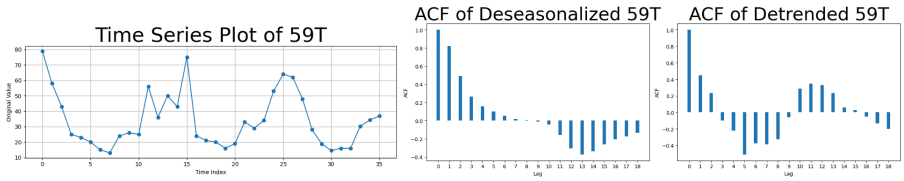
A preliminary time series analysis was conducted to explore the data structure. Figs. 2-3 display the time series plots and autocorrelation functions (ACF) of the deseasonalized and detrended series for each station. Visual inspection of the original series indicates clear seasonal patterns across all locations. Upon removing the seasonal component, the ACF plots of the deseasonalized series reveal a strong positive autocorrelation at lag 1, gradually declining with increasing lags, indicating a trend component. After detrending, the ACF shows a strong negative peak around lag 6 and a strong positive peak near lag 12, confirming the presence of annual seasonality consistent with the environmental behavior of PM2.5. These characteristics support the inclusion of both trend and seasonal components in the proposed forecasting models across all eight stations. All data analyses, optimization procedures, and forecasting experiments in this study were conducted in the Google Colab environment using the Python programming language.



A (11T) Kheha Chumchon Huai Khwang Stadium, Bangkok

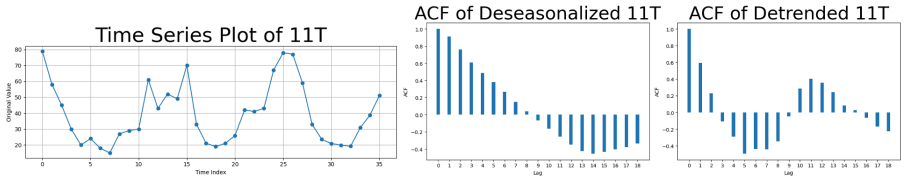


B (12T) Nonsi Witthaya School, Bangkok

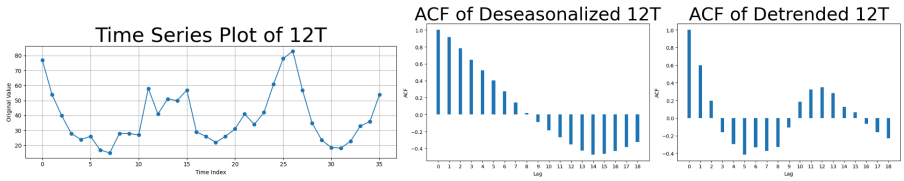


C (59T) The Government Public Relations Department, Bangkok

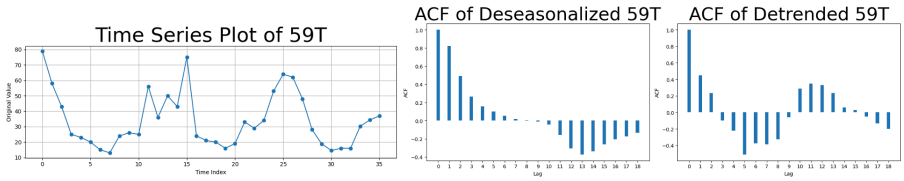
Fig. 2. A-C: Time series plot and ACF of the deseasonalized and detrended of the three stations in the Bangkok.



A (11T) Kheha Chumchon Huai Khwang Stadium, Bangkok



B (12T) Nonsi Witthaya School, Bangkok



C (59T) The Government Public Relations Department, Bangkok

Fig. 3. A-E: Time series plot and ACF of the deseasonalized and detrended of the five stations in metropolitan area.

2.2 The combined Holt–Winters and decomposition method optimized via the Whale Optimization Algorithm (CHD–WOA)

This section presents the formulation of the CHD model and its parameter optimization using the WOA. The CHD model integrates two classical time series forecasting techniques—decomposition and HW—within a unified hybrid framework. Forecasts from both models are aggregated using a weighted averaging approach, in which the weight is treated as an additional parameter. All model parameters are estimated simultaneously using the WOA to enhance forecasting accuracy and ensure global convergence. The resulting hybrid model is referred to as CHD–WOA, which denotes the Combined HW and decomposition method optimized via the WOA.

The time series models applied in this study—namely, decomposition and Holt–Winters—were implemented in both additive and multiplicative forms. The choice between these formulations depends on the characteristics of the seasonal variation relative to the overall level of the series. The additive form is suitable when the magnitude of seasonal fluctuations remains approximately constant over time, independent of the mean level of PM2.5 concentrations. Conversely, the multiplicative form is more appropriate when the seasonal variation changes proportionally with the series level—typically observed when pollution peaks become more pronounced during periods of higher overall concentrations. To account for these differences, both additive and multiplicative versions of each model were developed and compared in this study.

2.2.1 Decomposition model

The decomposition model separates the original time series into two main components: trend and seasonality. These components are modeled independently and then recombined to produce the final forecast. The trend is represented as a simple linear function of time, while the seasonal component captures recurring patterns across a 12-month period. The model can be expressed in either additive or multiplicative form, as shown in Eqs. (2.1)-(2.2), respectively.

Additive Forecasting:

$$\hat{D}_{t+p} = \hat{\beta}_0 + \hat{\beta}_1(t+p) + \hat{S}_{t+p}. \quad (2.1)$$

Multiplicative Forecasting:

$$\hat{D}_{t+p} = (\hat{\beta}_0 + \hat{\beta}_1(t+p)) \times \hat{S}_{t+p}. \quad (2.2)$$

In these equations, $\hat{D}_{t+p}$ is the forecasted value at future time point $t+p$, where t is the index of the most recent observed time period and p is the number of steps ahead to be forecasted. The term $\hat{\beta}_0$ is the estimated intercept, and $\hat{\beta}_1$ is the estimated slope of the linear trend component. The seasonal component $\hat{S}_{t+p}$ denotes the estimated seasonal index corresponding to month $t + p$, which falls within a specific season $i (i = 1, ..., s)$. We define s as a 12-month cycle. Each time point t is thus associated with one of these 12 distinct seasons.

In the WOA–optimized decomposition model (WOA–D), as applied in [13, 16, 19], 14 parameters are estimated simultaneously: two for the trend component and twelve for the seasonal component. The WOA is employed to search for the optimal parameter values within predefined bounds, which are determined based on the seasonal amplitude observed in the training dataset. The WOA–D model is included as a baseline method in the comparative analysis.

2.2.2 Holt–Winters model

The HW exponential smoothing method is used to model the level, trend, and seasonality of the time series. Separate

formulations are applied based on the nature of the seasonal component—additive or multiplicative. The recursive updating equations for level, trend, and seasonality, as well as the forecasting function, are presented in Eqs. (2.3)-(2.10).

Additive Forecasting:

$$\hat{H}_{t+p} = \hat{T}_t + p\hat{\beta}_t + \hat{S}_{t-s+1+((p-1) \mod s)} \quad \text{for } p = 1, 2, ..., \tag{2.3}$$

$$\hat{T}_t = \alpha(Y_t - \hat{S}_{t-s}) + (1-\alpha)(\hat{T}_{t-1} + \hat{\beta}_{t-1}), \tag{2.4}$$

$$\hat{\beta}_t = \gamma(\hat{T}_t - \hat{T}_{t-1}) + (1-\gamma)\hat{\beta}_{t-1}, \tag{2.5}$$

$$\hat{S}_t = \delta(Y_t - \hat{T}_t) + (1-\delta)\hat{S}_{t-s}. \tag{2.6}$$

Multiplicative Forecasting:

$$\hat{H}_{t+p} = (\hat{T}_t + p\hat{\beta}_t) \times \hat{S}_{t-s+1+((p-1) \mod s)} \quad \text{for } p = 1, 2, ..., \tag{2.7}$$

$$\hat{T}_t = \alpha(Y_t / \hat{S}_{t-s}) + (1-\alpha)(\hat{T}_{t-1} + \hat{\beta}_{t-1}), \tag{2.8}$$

$$\hat{\beta}_t = \gamma(\hat{T}_t - \hat{T}_{t-1}) + (1-\gamma)\hat{\beta}_{t-1}, \tag{2.9}$$

$$\hat{S}_t = \delta(Y_t / \hat{T}_t) + (1-\delta)\hat{S}_{t-s}. \tag{2.10}$$

In these equations, $\hat{T}_t, \hat{\beta}_t$ and $\hat{S}_t$ represent the estimated level, trend, and seasonal components at time t. The parameters α, γ and δ are smoothing constants for level, trend, and seasonality, each constrained to the interval [0,1]. The seasonal cycle length is denoted by $s = 12$ for monthly data, and the forecast horizon is represented by p.

In the WOA–optimized HW model (WOA–HW), the WOA is applied to estimate the three smoothing parameters and . The WOA–HW model, as proposed in [13, 16, 19], is included as a baseline method in the comparative analysis.

2.2.3 The combined Holt–Winters and decomposition method (CHD)

The Combined HW and Decomposition (CHD) model linearly integrates the forecasts generated by the decomposition and Holt–Winters models using a weight parameter $w \in [0, 1]$. The final forecasted value is defined in Eq. (2.11):

$$\hat{Y}_{t+p} = w\hat{D}_{t+p} + (1+w)\hat{H}_{t+p}, \tag{2.11}$$

where $\hat{D}_{t+p}$ and $\hat{H}_{t+p}$ are the forecasts from the decomposition and HW models, respectively. The parameter w determines the relative contribution of each model to the final forecast and is optimized concurrently with the model-specific parameters. In total, the CHD model requires the estimation of 18 parameters. These include 14 parameters from the decomposition model—comprising two trend coefficients ($\hat{\beta}_0, \hat{\beta}_1$) and twelve monthly seasonal indices ($\hat{S}_1, \hat{S}_2, \hat{S}_3, ..., \hat{S}_{12}$) as well as three smoothing parameters from the HW model (α, γ, δ) and one combination weight w. The complete parameter set is optimized using the WOA.

The CHD–WOA+ model refers to the configuration where both $\hat{D}_{t+p}$ and $\hat{H}_{t+p}$ are derived from their respective additive models (i.e., Eq. (2.1) for decomposition and Eq. (2.3) for HW). Similarly, the CHD–WOA× model refers to the configuration where both components follow the multiplicative form (i.e., Eq. (2.2) and Eq. (2.7)).

The objective function used to evaluate and optimize the CHD–WOA model is the Root Mean Square Error (RMSE), calculated over the training dataset as shown in Eq. (2.12):

$$RMSE = \frac{1}{n_1}\sum_{t=1}^{n_1}(Y_t - \hat{Y}_t)^2, \tag{2.12}$$

where Y_t is the actual observed value, $\hat{Y}_t$ is the forecasted value from the CHD–WOA model, and n_1 is the number of observations in the training set.

The optimization problem is formulated as follows: minimize RMSE with respect to all 18 parameters, subject to the constraints:

$$\begin{aligned}
&\hat{\beta}_0 \in \left[\hat{\beta}'_0 - 0.2|\hat{\beta}'_0|, \hat{\beta}'_0 + 0.2|\hat{\beta}'_0|\right], \\
&\hat{\beta}_1 \in \left[\hat{\beta}'_1 - 0.2|\hat{\beta}'_1|, \hat{\beta}'_1 + 0.2|\hat{\beta}'_1|\right], \\
&\hat{S}_i \in [LB, UB], \quad \text{for } i = 1, 2, ..., s, \\
&\alpha, \gamma, \delta \in [0, 1]
\end{aligned} \tag{2.13}$$

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 seasonal component bounds, UB_i and LB_i, are determined differently depending on the chosen model structure.

For the additive decomposition model, the bounds are derived from the first-order difference of the original series:

$$UB_i = \Delta Y_t, \; LB_i = -\Delta Y_t,$$

where $\Delta Y_t = Y_t - Y_{t-1}$.

For the multiplicative decomposition model, the seasonal index bounds are defined as:

$$UB_i = s - 0.01, LB_i = 0.01.$$

To ensure identifiability of seasonal components, a normalization constraint is applied to the seasonal indices. For the additive model, the seasonal indices are adjusted so that their sum equals zero, i.e., $\sum_{i=1}^{s} \hat{S}_t = 0$. For the multiplicative model, the seasonal indices are normalized such that their average equals one (or equivalently, their sum equals 12 for monthly data), i.e., $\sum_{i=1}^{s} \hat{S}_t = s$. These constraints prevent overlapping effects between the level and seasonal components and maintain consistency in the decomposition structure. where s represents the seasonal cycle, which equals 12 for monthly data. These bounds ensure that seasonal variation remains within a reasonable range based on the observed data behavior.

The HW method requires the estimation of three smoothing parameters: level (α), trend (γ), and seasonality (δ). Each parameter lies within the interval [0,1] and plays a critical role in adjusting the model components to reflect the underlying structure of the time series, including both long-term trend and seasonal fluctuations.

The optimization routine used in this study is implemented according to the WOA, whose pseudocode is summarized in Fig. 4. The configuration follows the guidelines proposed in [4] and is adapted from our previous implementations in [13, 16, 19] for optimizing the decomposition and HW models, with necessary modifications for the CHD–WOA model.

The configuration of the WOA is summarized in Fig. 4, with key parameters set as follows: number of whales = 100, number of parameters = 18, maximum iterations = 300, and a time limit of 30 seconds. The search process terminates if the global fitness fails to improve for 50 iterations ($T_{improve}$= 50). Parameter bounds are defined according to empirical ranges of each model component. Random initialization is used without fixing a seed to preserve stochastic exploration.

In contrast, the CHD–Classical model applies a two-step procedure in which the decomposition and HW models are independently fitted to the training dataset. Each model is calibrated separately using classical methods to obtain fitted values. After both fitted series are obtained, the combination weight w in

the number of whales: $N = 100$, the number of parameters: $m = 18$, 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 $X_i = (x_i^j, x_i^2, ..., x_i^{18}), X^*$
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 CHD–WOA by the Eq. (12)
Update X^* if there is a better solution
$t = t + 1$
End while
Return $X^* = (x^{1*}, x^{2*}, ..., x^{18*})$
Objective Minimize RMSE($\hat{\beta}_0^*, \hat{\beta}_1^*, \hat{S}_1^*, \hat{S}_2^*, \hat{S}_3^*, \hat{S}_4^*, \hat{S}_5^*, \hat{S}_6^*, \hat{S}_7^*, \hat{S}_8^*, \hat{S}_9^*, \hat{S}_{10}^*, \hat{S}_{11}^*, \hat{S}_{12}^*, \alpha^*, \gamma^*, \delta^*, w^*$) where $\hat{\beta}_0^* = x^{1*}, \hat{\beta}_1^* = x^{2*}, ..., \hat{S}_{12}^* = x^{14*}$ and $\alpha^* = x^{15*}, \gamma^* = x^{16*}, \delta^* = x^{17*}$ and $w^* = x^{18*}$

Fig. 4. Pseudo-code of the CHD–WOA.

Eq. (2.11) is determined by minimizing the RMSE on the training set. Since the model parameters and the weight are not optimized simultaneously, this approach results in a local optimum for the combined forecast. The CHD–Classical model is included as a baseline method in this study to compare with the global optimization strategy of CHD–WOA.

Therefore, the WOA is employed to achieve robust global optimization in non-convex error spaces typical of environmental time series data, ensuring stable convergence and reducing the risk of local minima.

2.3 Box–Jenkins method

The Box–Jenkins methodology is employed in this study through the application of Seasonal Autoregressive Integrated Moving Average (SARIMA) models to provide a statistical benchmark for forecasting. This approach is well-established for its ability to capture both short-term fluctuations and recurring seasonal patterns in time series data. The SARIMA model is specified as SARIMA$(p, d, q)(P, D, Q)_s$, where p, d, q represent the non-seasonal autoregressive, differencing, and moving average orders, and P, D, Q represent their seasonal counterparts, with s indicating the seasonal length (e.g., $s = 12$ for monthly data).

The Box–Jenkins procedure involves a systematic sequence of steps, beginning with a stationarity check using the autocorrelation function (ACF) and partial autocorrelation function (PACF). If necessary, dif-

ferencing is applied to stabilize the series, including both regular and seasonal differencing. Variance stabilization is also considered through data transformation techniques such as logarithmic transformation when required.

Once the series is rendered stationary, candidate models are identified based on ACF and PACF patterns, and model parameters are estimated using maximum likelihood estimation (MLE). Model selection was further guided by the Akaike Information Criterion (AIC), which balances model fit and parsimony by penalizing excessive parameterization. Diagnostic checking is then conducted to validate model adequacy. This includes:

- Assessing the significance of estimated parameters using t-tests.

- Using the t-test for testing whether the mean of residuals is statistically different from zero.

- Evaluating the randomness of residuals with the Ljung–Box Q test.

- Checking residual normality via the Kolmogorov–Smirnov (KS) test.

- Employing the ARCH LM test to detect heteroscedasticity in the model residuals, particularly to identify time-varying volatility.

If any diagnostic assumption is violated, model refinement is conducted by revisiting the identification and estimation steps. Upon successful validation, the selected SARIMA model is used to generate forecasts for the test period. The Box–Jenkins model serves as one of the comparative baselines to evaluate the effectiveness of the proposed CHD–WOA approach [20].

2.4 Long-short term memory (LSTM) method

The LSTM network is a specialized form of recurrent neural network (RNN) designed to model sequential data with long-term dependencies. Leveraging its gated memory cell architecture, the LSTM is particularly effective in capturing temporal dynamics and nonlinear patterns in time series data, making it suitable for forecasting tasks such as PM2.5 concentrations.

In this study, the LSTM model is configured using a sequential structure comprising one hidden LSTM layer followed by a dense output layer. The input data—monthly maximum values of daily 24-hour average PM2.5 concentrations in the BMR—is first normalized using Min-Max scaling to ensure numerical stability. A sliding window approach is used to construct input-output pairs, where each input consists of a specified number of lagged observations (look-back window), and the output corresponds to the PM2.5 value of the following month.

Two look-back windows, 6 and 12 months, are employed to capture short- and medium-term temporal dependencies. For each setting, the model is trained on the first 36 months (January 2021 to December 2023) and used to generate forecasts for the subsequent 12 months (January to December 2024). The network is trained for 100 epochs using the Adam optimizer with a RMSE loss function and a batch size of 1. The number of epochs was set to 100 based on preliminary convergence analysis, which showed that the model's loss function stabilized after approximately 80 epochs, and further training beyond 100 epochs led to overfitting without improving forecast accuracy.

To evaluate the model's internal forecasting performance, the predicted values were inverse-transformed to the original scale and compared with the observed values using the RMSE. This metric, consistent with the model's training loss function,

quantifies the average magnitude of prediction errors and provides a straightforward measure of the model's accuracy.

The LSTM-based approach is included as a benchmark method in the comparative evaluation to assess the relative accuracy of the proposed CHD–WOA framework. This model helps demonstrate the applicability of deep learning techniques in forecasting high-variability environmental time series data.

2.5 Evaluation criteria

The forecasting performance of all models is evaluated through both in-sample and out-of-sample comparisons. The in-sample evaluation measures how well each model fits the training data, while the out-of-sample evaluation assesses the forecasting accuracy on the test data.

2.5.1 In-sample Evaluation (Training Phase)

The in-sample accuracy is assessed by comparing the fitted values against the actual observations within the training dataset. The RMSE is employed as the primary metric and is computed using:

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

where n_1 is the number of observations in the training dataset (36 months), and Y_t and $\hat{Y}_t$ denote the actual and fitted values, respectively.

The comparison focuses on the proposed hybrid models. Specifically, the CHD–WOA+ model is compared with CHD–Classical+, and CHD–WOA× is compared with CHD–Classical×, to examine the impact of WOA–based optimization on model fitting performance.

2.5.2 Out-of-sample evaluation (Forecasting Phase)

The forecasting accuracy on the test dataset is evaluated using the mean absolute percentage error (MAPE), defined as:

$$MAPE = \frac{100}{12}\sum_{t=37}^{n}\left|\frac{Y_t - \hat{Y}_t}{Y_t}\right|,$$

where Y_t and $\hat{Y}_t$ are the actual and forecasted values for the testing period. Here, $n = 48$ refers to the total number of months in the full dataset, with testing data covering months 37 through 48. The MAPE measures the average relative error between the predicted and observed values, expressed as a percentage.

In total, ten forecasting methods were compared in this study, covering both statistical and machine learning approaches. All models are included in the MAPE-based comparison:

- WOA–HW+ and WOA–HW×
- WOA–D+ and WOA–D×
- CHD–WOA+ and CHD–WOA×
- CHD–Classical+ and CHD–Classical×
- Box–Jenkins
- LSTM

This two-stage evaluation framework provides a comprehensive assessment of each model's performance across both historical fitting and future prediction tasks.

3. Results and Discussion

3.1 In-sample evaluation (Training Phase)

To evaluate the model fitting performance during the training phase, the RMSE was computed between the actual observed values and the fitted values generated by each model. The comparison was conducted between the CHD–WOA and CHD–Classical models for both additive and mul-

tiplicative forms, across all eight air quality monitoring stations in the BMR.

Table 1 presents the RMSE results for the additive models. Across all stations, the CHD–WOA+ model consistently achieved lower RMSE values compared to the CHD–Classical+ model. Notably, the CHD–WOA+ model outperformed its classical counterpart at every station, with the most substantial improvements observed at stations 16T (4.7 vs. 8.7), 12T (7.2 vs. 9.6), and 27T (11.8 vs. 14.2). These results indicate that the integration of the WOA in the additive CHD model leads to more accurate parameter estimation and better in-sample fit.

Table 2 summarizes the RMSE results for the multiplicative models. Similar to the additive case, the CHD–WOA× model outperformed the CHD–Classical× model at all stations. The lowest RMSE values in each column are highlighted to emphasize the advantage of using the WOA–based optimization. Significant improvements were again observed at stations 16T (4.8 vs. 8.7), 59T (5.5 vs. 9.3), and 11T (5.9 vs. 9.2). These results reinforce the robustness of the CHD–WOA framework in both seasonal forms and across diverse data conditions.

3.1.1 Out-of-sample evaluation (Testing Phase)

The out-of-sample forecasting performance of all models is evaluated over a 12-month prediction horizon (January to December 2024) using the MAPE. This metric reflects the average percentage deviation between forecasted and actual values and is widely used to assess forecasting accuracy, particularly when comparing different model frameworks across multiple monitoring locations.

Prior to comparative evaluation, the Box–Jenkins models were established for each station based on the SARIMA specification that best fit the training data. Table 3 summarizes the selected SARIMA structures, along with any applied transformations and results from model adequacy diagnostics. Half of the monitoring stations required data transformation prior to model fitting. Logarithmic transformation was applied at stations 12T, 20T, and 81T, while a square root transformation was used for 16T. The remaining four stations were modeled using the original untransformed data. The model identification process followed standard Box–Jenkins procedures, including visual inspection of ACF/PACF plots and selection based on AIC.

The selected models passed all diagnostic tests at the 5% significance level. Additionally, although not shown in Table 3 to maintain clarity and conciseness, all model parameters in the final SARIMA specifications were statistically significant based on individual t-tests.

As shown in Table 4, the CHD–WOA+ model produced the lowest MAPE at two stations: 11T (16.8%) and 59T (16.4%). The CHD–WOA× model also achieved top performance at 12T (18.9%) and 16T (15.7%). In addition, the WOA–D+ model yielded the most accurate forecasts at 13T (12.7%) and 20T (13.1%), while WOA–D× provided the lowest error at 81T (17.3%). Interestingly, the LSTM model delivered the best result only at 27T (15.0%), though its performance at other stations was notably poor.

Overall, the CHD–WOA framework—comprising both additive (CHD–WOA+) and multiplicative (CHD–WOA×) seasonal formulations—demonstrated superior out-of-sample forecasting performance across multiple monitoring stations. It achieved the lowest

MAPE at four out of eight stations (11T, 12T, 16T, and 59T), representing 50% coverage in terms of best-performing locations and highlighting its robustness under diverse air quality conditions within the BMR. This superiority can be attributed to the model's integrated structure and global optimization strategy. By combining the strengths of the Holt–Winters and decomposition methods, CHD–WOA effectively captures both long-term trend–seasonal patterns and localized fluctuations in PM2.5 concentrations. The simultaneous parameter estimation via the WOA enables global exploration of the solution space, avoiding local minima typical of conventional approaches, while the weighted averaging mechanism allows adaptive adjustment to the temporal characteristics of each monitoring site. These features collectively make the CHD–WOA model particularly suitable for complex urban environments such as Bangkok, where air quality dynamics are shaped by interacting factors including traffic emissions, meteorological variability, and transboundary influences.

In the final stage of this study, the most appropriate forecasting model for each air quality monitoring station—determined from the lowest MAPE values in the out-of-sample evaluation—was retrained using the entire dataset (January 2021 to December 2024) in order to maximize model calibration. These fully trained models were then used to forecast the monthly maximum of daily 24-hour average PM2.5 concentrations for a 12-month horizon, covering January to December 2025.

The resulting forecasts are summarized in Table 5, while the corresponding time series plots are illustrated in Fig. 5. Each subplot shows the actual data, the fitted values over the training period, and the 12-month forecasted values. The visual comparison demonstrates that the chosen models not only fit the historical data well but also capture the seasonal structure of PM2.5 concentrations with reasonable forecast continuity.

The forecasts in Table 5 further emphasize typical seasonal patterns, with the highest predicted PM2.5 values generally occurring between January and April, aligning with the dry season and peak pollution period in the BMR. For example, the highest forecasted values occur in February 2025 at several stations—76.2 μg/m^3 at 20T, 73.5 μg/m^3 at 81T, and 72.9 μg/m^3 at 27T—indicating a strong seasonal peak. In contrast, values drop considerably during the rainy season (June–September), which is consistent with the natural dispersion and removal of airborne particles due to precipitation.

These results validate the reliability of the selected models not only in fitting past data but also in generating plausible forecasts for future air quality. This capability is particularly valuable for environmental policy planning, early warning systems, and public health preparedness in the region.

The differences in the best-performing forecasting methods across the monitoring stations can be attributed to the heterogeneous characteristics of PM2.5 time series in the BMR. Each station is influenced by distinct emission sources, local meteorological conditions, and surrounding land use. For instance, urban stations such as 11T and 59T, located in the inner city of Bangkok, exhibit more dynamic and irregular PM2.5 patterns driven by traffic and meteorological variability. These complex fluctuations favor the flexibility of the CHD–WOA models, which integrate both structural and adaptive components.

Table 1. RMSE of additive model of train dataset for each air quality monitoring station.

Additive Model	Air Quality Monitoring Station							
	11T	**12T**	**59T**	**13T**	**16T**	**20T**	**27T**	**81T**
CHD–WOA+	**7.0**	**7.2**	**7.5**	**8.3**	**4.7**	**10.1**	**11.8**	**9.1**
CHD–Classical+	9.2	9.6	9.1	10.3	8.7	11.9	14.2	10.7

Note: The lowest RMSE value between CHD–WOA and CHD–Classical is bold.

Table 2. RMSE of multiplicative model of train dataset for each air quality monitoring station.

Additive Model	Air Quality Monitoring Station							
	11T	**12T**	**59T**	**13T**	**16T**	**20T**	**27T**	**81T**
CHD–WOA×	**5.9**	**6.0**	**5.5**	**7.2**	**4.8**	**9.4**	**11.8**	**7.7**
CHD–Classical×	9.2	8.7	9.3	10.2	8.7	11.7	14.4	10.7

Note: The lowest RMSE value between CHD–WOA and CHD–Classical is bold.

Table 3. Box–Jenkins models and model adequacy diagnostics by air quality monitoring station.

Air Quality Monitoring Station	Transfor-mation	Box–Jenkins SARIMA(p,d,q)(P,D,Q)$_{12}$	t-test (P-value)	Arch LM (P-value)	KS (P-value)	LB Lag 12 (P-value)
11T	-	SARIMA(0,1,1)(0,1,1)$_{12}$	-0.10 (0.918)	1.460 (0.834)	0.168 (0.090)	16.06 (0.098)
12T	Ln	SARIMA(0,1,1)(0,1,0)$_{12}$	0.66 (0.515)	2.433 (0.657)	0.116 (>0.150)	7.32 (0.773)
59T	-	SARIMA(0,1,0)(1,1,0)$_{12}$	0.57 (0.576)	3.845 (0.427)	0.176 (0.065)	6.52 (0.836)
13T	-	SARIMA(2,1,2)(1,1,0)$_{12}$	-0.89 (0.381)	8.465 (0.076)	0.096 (>0.150)	9.46 (0.221)
16T	Square Root	SARIMA(0,1,1)(0,1,0)$_{12}$	1.19 0.246	0.624 (0.960)	0.180 (0.050)	17.83 (0.086)
20T	Ln	SARIMA(0,1,1)(0,1,0)$_{12}$	0.85 0.406	2.996 (0.559)	0.152 (>0.150)	18.10 (0.079)
27T	-	SARIMA(0,1,0)(1,1,0)$_{12}$	0.76 (0.455)	3.435 (0.488)	0.176 (0.063)	9.69 (0.558)
81T	Ln	SARIMA(0,1,1)(0,1,0)$_{12}$	-0.25 (0.806)	4.156 (0.385)	0.174 (0.072)	16.71 (0.117)

Table 4. MAPE of test dataset for each air quality monitoring station.

Model	Air Quality Monitoring Station							
	11T	12T	59T	13T	16T	20T	27T	81T
WOA–HW+	19.5	22.7	25.1	18.7	20.7	14.9	15.9	18.4
WOA–HW×	20.9	23.0	25.1	17.8	20.7	17.4	17.5	21.8
WOA–D+	17.6	29.0	17.2	**12.7**	18.9	**13.1**	17.0	20.8
WOA–D×	18.5	24.9	18.8	13.3	16.9	13.6	17.5	**17.3**
CHD–WOA+	**16.8**	33.7	**16.4**	21.2	23.3	19.3	18.5	17.4
CHD–Classical+	17.7	27.8	16.4	13.2	18.4	13.3	16.3	20.1
CHD–WOA×	24.6	**18.9**	21.3	20.0	**15.7**	15.7	19.6	21.1
CHD–Classical×	18.6	24.7	18.9	13.5	17.1	14.0	16.7	17.4
Box–Jenkins	37.2	40.7	63.7	36.8	20.9	22.1	29.4	23.7
LSTM	38.9	34.4	31.6	23.5	31.2	17.9	**15.0**	26.6

Note: The lowest MAPE value for each province is bold.

In contrast, suburban and peri-urban stations such as 13T and 20T tend to show more regular seasonal cycles and stable long-term trends, making the WOA–D models—based on explicit decomposition of trend and seasonality—more suitable.

The LSTM model achieved its best performance at station 27T, likely due to its

Table 5. Forecasted monthly maximum of daily 24-hour average PM2.5 for each station, January 2025 – December 2025.

Month/Year	Air Quality Monitoring Station							
	11T	12T	59T	13T	16T	20T	27T	81T
Jan-25	51.7	47.1	40.8	**58.6**	53.1	**65.8**	**72.5**	**65.4**
Feb-25	**65.3**	**63.4**	**52.7**	**64.8**	**67.0**	**76.2**	**72.9**	**73.5**
Mar-25	**57.2**	**55.0**	46.1	52.0	53.5	60.1	66.0	53.0
Apr-25	48.1	49.8	**55.5**	48.8	**56.4**	52.5	52.9	52.1
May-25	41.6	29.0	25.8	24.7	28.1	30.2	40.0	28.3
Jun-25	19.1	18.5	16.4	14.5	22.8	19.2	28.4	21.5
Jul-25	13.5	15.2	14.9	11.0	14.0	17.8	21.7	14.3
Aug-25	14.4	16.9	13.5	15.6	14.7	21.9	23.0	18.6
Sep-25	19.4	17.6	22.7	17.3	22.8	23.5	31.2	21.4
Oct-25	30.5	29.5	22.8	30.9	35.7	35.6	44.0	36.1
Nov-25	35.0	30.9	23.1	31.9	34.3	41.3	56.9	38.4
Dec-25	39.9	35.0	30.1	46.7	40.4	52.9	66.1	52.5

Note: The first and second highest values within each 12-month period.

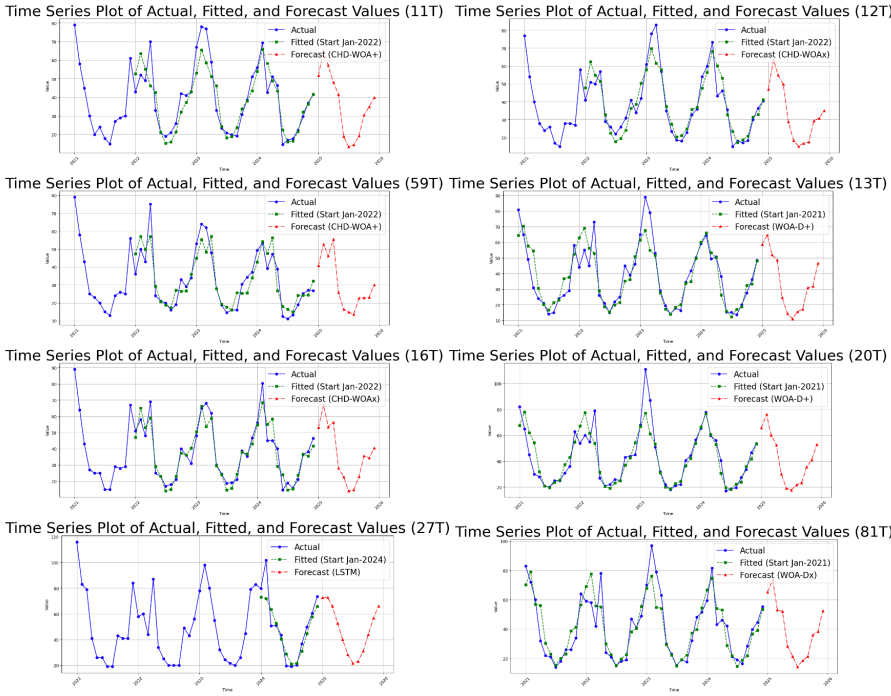


Fig. 5. Actual, fitted and forecasted monthly maximum of daily 24-hour average pm2.5 for each stations using the most appropriate forecasting methods.

ability to capture nonlinear variations associated with dynamic emission patterns in industrial and coastal areas. These variations indicate that no single forecasting method universally dominates across all locations, as model performance depends strongly on the temporal dynamics and local pollution regimes of each station.

4. Conclusion

This study introduced a novel hybrid time series forecasting model, referred to as CHD–WOA, which combines the decomposition and Holt–Winters methods through a weighted averaging scheme and optimizes all parameters simultaneously using the WOA. The proposed model was evaluated using monthly maximum values of daily 24-hour average PM2.5 concentrations across eight air quality monitoring stations in the BMR. Comparative evaluations against classical methods, standalone models, Box–Jenkins, and LSTM demonstrated the superior forecasting capability of the CHD–WOA framework.

The in-sample evaluation demonstrated that WOA–based parameter estimation significantly improved model fit in both additive and multiplicative formulations, while the out-of-sample forecasting results showed that CHD–WOA achieved the lowest MAPE at 50% of the monitoring stations. Forecasts for the year 2025, generated using fully retrained models, successfully captured seasonal pollution patterns consistent with historical trends and meteorological conditions.

These findings highlight the practical value of integrating classical time series structures with metaheuristic optimization, particularly for forecasting complex environmental phenomena with seasonal and non-linear dynamics. The CHD–WOA model offers a flexible and accurate tool that can support proactive air quality management, health risk mitigation, and decision-making processes. Future work may explore extending this framework to multivariate forecasting scenarios or incorporating additional optimization strategies for real-time adaptive applications.

References

[1] Chantaraprachoom N, Shimadera H, Uranishi K, Mui LV, Matsuo T, Kondo A. A nation-by-nation assessment of the contribution of Southeast Asian open biomass burning to $PM_{2.5}$ in Thailand using the community multiscale air quality-integrated source apportionment method model. Atmosphere. 2024;15(11):1358. doi:10.3390/atmos15111358.

[2] Holt CC. Forecasting seasonals and trends by exponentially weighted moving averages. Report to the Office of Naval Research (ONR 52); 1957.

[3] Winters PR. Forecasting sales by exponentially weighted moving averages. Management Science. 1960;6(3):324-42.

[4] Mirjalili S, Lewis A. The whale optimization algorithm. Adv Eng Softw. 2016;95:51-67.

[5] Liu L, Wu L. Predicting housing prices in China based on modified Holt's exponential smoothing incorporating whale optimization algorithm. Socioecon Plann Sci. 2020;72:1-8.

[6] Sapnken FE, Tazehkandgheshlagh AK, Diboma BS, Hamaidi M, Noumo PG, Wang Y, et al. A whale optimization algorithm-based multivariate exponential smoothing grey-holt model for electricity price forecasting. Expert Syst Appl. 2024;255(B):124663.

[7] Du P, Ye Y, Wu H, Wang J. Study on deterministic and interval forecasting of electricity load based on multi-objective whale optimization algorithm and transformer model. Expert Syst Appl. 2025;268:126361.

[8] Hasan MW. Building an IoT temperature and humidity forecasting model based on long short-term memory (LSTM) with improved whale optimization algorithm. Memories Mater Devices Circuits Syst. 2023;6:100086.

[9] Diboma BS, Sapnken FE, Hamaidi M, Wang Y, Noumo PG, Tamba JG. Improved exponential smoothing grey-holt models for electricity price forecasting using whale optimization. MethodsX. 2024;13:102926.

[10] Alameer Z, Elaziz MA, Ewees AA, Ye H, Jianhua Z. Forecasting gold price fluctuations using improved multilayer perceptron neural network and whale optimization algorithm. Resour Policy. 2019;61:250-60.

[11] Guo W, Liu T, Dai F, Xu P. An improved whale optimization algorithm for forecasting water resources demand. Appl Soft Comput. 2020;86:105925.

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

[13] 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. J Curr Sci Technol. 2024;14(2):Article 38.

[14] 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 J Oper Res. 2024;12(2):69-89. Thai.

[15] Minsan W, Minsan P, Panichkitkosolkul W. Enhancing decomposition and Holt–Winters weekly forecasting of $PM_{2.5}$ concentrations in Thailand's eight northern provinces using the cuckoo search algorithm. Thailand Statistician. 2024;22(4):963-85.

[16] Minsan P, Minsan W. Decomposition and Holt–Winters techniques enhanced by whale optimization algorithm: case study of $PM_{2.5}$ forecasting in eight northern provinces of Thailand. Thai Sci Technol J. 2024;32(6):12-34. Thai.

[17] Vimolsutjarit N, Minsan W, Taninpong P, Thumronglaohapun S. Forecasting water volume of dam reservoirs in northern Thailand using classical decomposition and Holt–Winters enhanced by fruit fly optimization algorithm. KKU Sci J. 2025;53(2):205-18.

[18] Pollution Control Department of Thailand. Air4Thai [Internet]. [cited 2025 Mar 1]. Available from: http://air4thai.pcd.go.th/webV3//History

[19] Minsan W. Decomposition and Holt–Winters enhanced by the whale optimization algorithm. figshare [Internet]. 2025 [cited 2025 Jun 1]. Available from: https://doi.org/10.6084/m9.figshare.29177183.v1

[20] Box GEP, Jenkins GM. Time series analysis: forecasting and control. 2nd ed. San Francisco (CA): Holden-Day; 1976.