STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **x**, Month 202x, pp 0–15.
Published online in International Academic Press (www.IAPress.org)

Forecasting PM10 concentration in Iraq based on improving v-support vector regression

Mohammed Ibrahim Mardini, Zakariya Yahya Algamal*

Department of Statistics and Informatics, University of Mosul, Mosul, Iraq

Abstract This manuscript develops and evaluates an improved v-support vector regression (v-SVR) framework, optimized via the Bamboo Forest Growth Optimization (BFGO) algorithm, to forecast daily PM10 concentrations in Baghdad, Iraq. Daily PM10 data from a ground-based monitoring station at the Baghdad U.S. Embassy, covering the period 1 March 2019 to 1 June 2023, are combined with relevant meteorological and environmental predictors to construct a data-driven forecasting system. The proposed method simultaneously performs v-SVR hyperparameter tuning and feature selection, addressing limitations of conventional v-SVR schemes and existing nature-inspired approaches that typically optimize only model parameters. BFGO is tailored to guide the search in the v-SVR hypothesis space through meme-based grouping, whip extension (exploitation), and shoot growth (exploration), thereby enhancing convergence toward a parsimonious and high-performing regression model. Model performance is assessed under multiple forecast horizons using standard accuracy indices such as RMSE, MAE, and R^2, and is benchmarked against baseline v-SVR, Random Forest, and other machine-learning forecasters commonly used in air quality prediction. The empirical results show that BFGO-v-SVR consistently achieves lower prediction errors and higher explanatory power than competing models, particularly in capturing sharp peaks and rapid fluctuations in PM10 driven by dust storms and intense anthropogenic emissions. These gains are attributed to the joint optimization of kernel parameters, regularization strength, ϵ-insensitive loss width, and an informative subset of predictors, which together improve generalization and reduce overfitting. The proposed BFGO-v-SVR framework offers a robust and flexible tool for real-time PM10 forecasting in highly polluted urban environments and can support more effective air quality management, early-warning systems, and evidence-based environmental policy in Iraq and similar semi-arid regions.

Keywords PM10 forecasting,-support vector regression, Bamboo Forest Growth Optimization, Hyperparameter tuning, Time-series prediction

DOI: 10.19139/soic-2310-5070-3735

1. Introduction

Air pollution constitutes a critical environmental challenge that drives several global issues such as climate change, ozone layer depletion, and acid rain formation. The deterioration of air quality, particularly in densely populated urban areas, primarily stems from rapid industrialization, expanding infrastructure, and urban development. Although these developments are often pursued to accommodate the accelerating pace of population growth, they simultaneously impose severe adverse effects on both human health and ecological systems [36]. These detrimental outcomes encompass an increased risk of mortality associated with serious illnesses, including respiratory infections and cardiovascular diseases, while also posing substantial threats to animal welfare and contributing significantly to climate change [43].

High population density is typically correlated with greater air quality degradation, driven by elevated energy consumption from residential, commercial, and industrial activities. A primary source of this issue lies in the

*Correspondence to: Zakariya Yahya Algamal (Email: zakariya.algamal@uomosul.edu.iq). Department of Statistics and Informatics, University of Mosul, Mosul, Iraq.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

combustion of fossil fuels for power generation and heating, which releases large quantities of greenhouse gases, nitrogen oxides, sulfur oxides, and particulate matter into the atmosphere [34, 45]. Furthermore, the escalating number of vehicles in cities—often resulting in severe traffic congestion exacerbates air quality deterioration and directly affects urban residents [22, 46, 47].

According to the World Health Organization, nine out of ten people globally reside in areas where air pollution levels surpass the limits recommended by WHO [27]. Exposure to outdoor air pollution, primarily caused by PM10, has been associated with an estimated 3.3 million premature deaths annually worldwide, with projections indicating a potential doubling of this figure by 2050 [18]. Hazardous fine particles such as PM10 are pervasive and pose a widespread environmental problem affecting multiple regions across the globe. Despite many countries reporting reductions in air pollution levels, PM10 continues to have a considerable and persistent impact on global public health [23, 48, 49, 50].

PM10 represents one of the most significant airborne environmental contaminants due to its small particle size, high toxicity, extensive dispersibility, and prolonged atmospheric suspension [39]. These properties render it particularly harmful to environmental stability, human health, and socio-economic systems. PM10 not only carries toxic compounds but also travels over long distances, substantially affecting air quality and visibility [32, 51]. Prolonged exposure to PM10 has been shown to increase the risk of cancer, respiratory illnesses, and cardiovascular diseases major threats to public health. Empirical evidence demonstrates that chronic exposure to PM10 significantly elevates the risk of premature mortality from various causes [3].

Accurate and efficient prediction of future PM10 concentrations plays an essential role in environmental management and policy formulation [33]. Human activities related to fossil fuel consumption especially those involving transportation, power generation, industrial operations, and heating/cooling systems—are primary anthropogenic contributors to PM10 emissions. Additionally, episodic events such as crop residue burning, human-induced forest fires, and dust storms linked to deforestation markedly contribute to elevated PM10 levels in certain regions [30, 36]. The intricate interplay of meteorological factors, including temperature inversions and thermal updrafts, further influences the accumulation and dispersion of atmospheric pollutants, often beyond the detection capabilities of traditional meteorological observations [24, 52, 53, 54].

Mitigating PM10 concentrations cannot be achieved instantaneously; hence, there is a pressing need for accurate, real-time forecasting models that can inform effective mitigation strategies. Such predictive capabilities not only facilitate timely public health advisories but also assist policymakers in developing targeted environmental interventions. However, forecasting PM10 concentrations remains challenging due to their pronounced spatial and temporal variability, influenced by meteorological, geographic, and anthropogenic factors [21].

Recently, PM10 forecasting models have attracted increasing attention and continuous improvement, driven by advances in computational modeling and environmental data collection. Reliable, real-time monitoring and forecasting of PM10 concentrations are indispensable for environmental protection and air quality management [37]. Addressing the complex spatiotemporal dynamics of air pollution requires sophisticated algorithms capable of capturing non-linear interactions among multiple influencing variables [25]. Researchers have highlighted the necessity of robust estimation frameworks capable of generating actionable insights to inform environmental policy and decision-making [29]. Accordingly, adopting data-driven modeling techniques, particularly those leveraging machine learning, has proven instrumental in improving both the spatial and temporal accuracy of air quality predictions.

In recent years, machine learning and deep learning approaches have emerged as transformative tools in the fields of earth and environmental sciences [43]. They have been successfully employed in various complex environmental applications, such as urban flood prediction [15], water quality estimation [26], and regional air pollution forecasting (Wong et al., 2021). These technologies have substantially enhanced the capacity to forecast PM10 concentrations and to issue timely pollution alerts. Numerous algorithms—ranging from relatively simple models like Random Forests (RF), Support Vector Machines (SVM), and Backpropagation Neural Networks (BPNN), to more complex architectures such as Gated Recurrent Units (GRU), Long Short-Term Memory (LSTM), and Convolutional Neural Networks (CNN)—have been developed to improve predictive performance [38].

Machine learning algorithms, particularly SVM and RF, have demonstrated significant potential for PM10 forecasting due to their strong capability to model complex, non-linear relationships within environmental data.

Their ability to capture intricate temporal dynamics in PM10 concentration series offers a clear advantage over traditional, inventory-based atmospheric models, which are often constrained by limited data availability [8]. In this research, the Random Forest algorithm is adopted due to its flexibility, robustness against overfitting, and high adaptability in handling complex environmental datasets [33].

Consequently, this study aims to overcome the limitations of short-term and coarse temporal resolution in existing prediction frameworks and to develop a robust, data-driven model capable of reliably forecasting daily variations in PM10 concentrations. The research focuses on the development and enhancement of accurate forecasting models using an RF-based approach that integrates both environmental and meteorological variables. To this end, daily PM10 data collected from ground-level monitoring sensors located in one of the most congested urban areas of Baghdad, Iraq, are employed. This dataset serves as the foundation for model training, testing, and validation and comprises daily PM10 measurements spanning a four-year period.

2. Data and Methods

2.1. Study Area

This study utilizes a daily average PM10 concentration dataset collected between 1 March 2019 and 1 June 2023 for Baghdad, Iraq. The data were obtained from the air quality monitoring station situated within the Baghdad U.S. Embassy, which provides reliable ground-level measurements of urban air quality. Although data from a single monitoring site may primarily represent the atmospheric conditions of the central urban zone rather than the entire metropolitan area, it nonetheless offers valuable insights into overall urban air pollution trends.

Baghdad, as Iraq's largest metropolitan center (Figure 1), hosts approximately 22% of the national population, resulting in intense anthropogenic pressure on the environment. The city faces substantial air pollution challenges arising from multiple emission sources, including vehicular exhaust, industrial processes, power generation, open waste burning, and unregulated disposal activities. These factors collectively contribute to a consistent upward trajectory in the city's annual average PM10 concentrations over the past decade.

Recent assessments have identified Baghdad among the most polluted cities globally, based on PM10 concentration levels reported in 2022. The combination of rapid urban expansion, population growth, and limited environmental regulation has amplified the city's vulnerability to particulate pollution. Moreover, frequent dust storms, a characteristic feature of Iraq's semi-arid climate, exacerbate airborne particulate matter concentrations, thereby intensifying both environmental and public health risks.

Given these conditions, Baghdad serves as a critical case study for assessing urban air pollution dynamics and developing accurate forecasting models for PM10 concentrations. The city's atmospheric environment provides an ideal context for evaluating the performance of data-driven predictive approaches that integrate meteorological, environmental, and anthropogenic variables.

2.2. v-Support Vector Regression

Support Vector Machines (SVMs) have effectively addressed numerous classification challenges. By incorporating the ϵ-insensitive loss function proposed by Vapnik, SVMs were extended to handle nonlinear regression tasks, leading to the development of Support Vector Regression (SVR). The primary goal of the v-SVR method is to model quantitative variables in diverse domains such as chemical activity prediction [44], spectral analysis [31], and stock price forecasting.

A training dataset of size n $\{(x_i, y_i)\}_{i=1}^{*}$ consists of samples $x_i = (x_{i,1}, x_{i,2}, \ldots, x_{i,p}) \in R^p$ that represent feature vectors corresponding to i^{th} " observations, along with $y_i \in R$ target values representing quantitative variables. The ϵ-insensitive loss function allows the resolution of an optimization problem to determine the SVR model:

$$\min_{M,d} \left\{ \frac{1}{2} M^r M + H \sum_{i=1}^{n} (\alpha_i + \tilde{\alpha_i}) \right\} \tag{1}$$

$$S.T. \{ y_i - (M \cdot \pi(x_i) + d) \leq \varepsilon + \tilde{\alpha_i} \;\; (M \cdot \pi(x_i) + d) - y_i \leq \varepsilon + \alpha_i \;\; \alpha_i, \tilde{\alpha_i} \geq 0$$

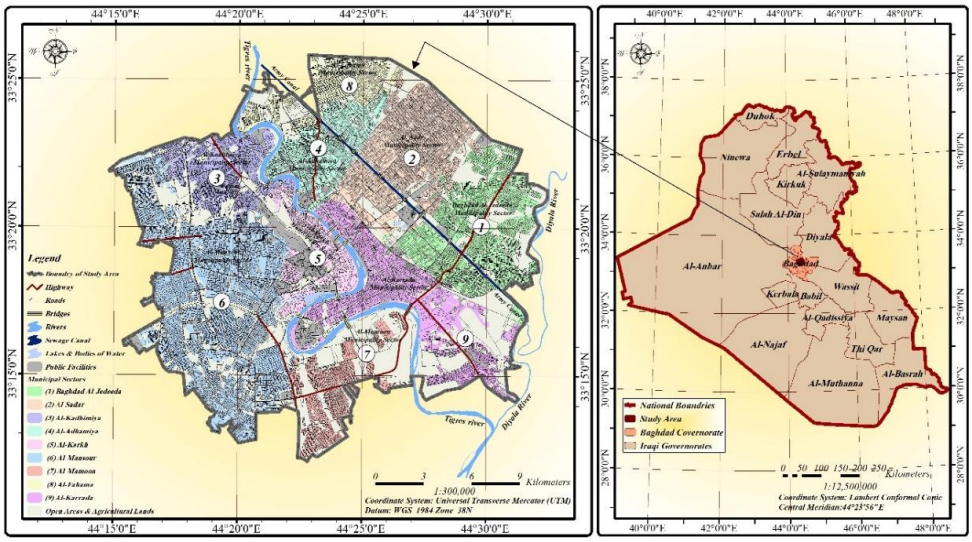

Figure 1. Map of Baghdad, highlighting metropolitan areas and municipalities.

This algorithm includes a single positive parameter H that balances the trade-off between training error and model complexity. It introduces slack variables α_i and $\alpha^{\tilde{}}_i$ along with a nonlinear mapping $\pi(x_i)$ using a kernel function, a weight vector M, and a bias term d.

After reformulating Equation 1 into its dual problem, the optimization is solved using Lagrange multipliers:

$$\underset{\tilde{\beta}/\beta}{min}\ \frac{1}{2}\sum_{i,j=1}^{n}\left(\beta^{\tilde{}}_i-\beta_i\right)\left(\beta^{\tilde{}}_j-\beta_j\right)K\left(x_i,x_j\right)\ +\varepsilon\sum_{i=1}^{n}\left(\beta^{\tilde{}}_i-\beta_i\right)\ -\sum_{i=1}^{n}y_i\left(\beta^{\tilde{}}_i-\beta_i\right) \tag{2}$$

Such that:

$$\sum_{i=1}^{n}\left(\beta_i-\beta^{\tilde{}}_i\right)\ =0\ 0\le\beta_i,\beta^{\tilde{}}_i\le H$$

Here, the kernel function $K(x_i,x_j)$ represents the nonlinear mapping, and the variables serve $\beta_i,\beta^{\tilde{}}_i$ as Lagrange multipliers. The regression hyperplane of the underlying regression problem is obtained as follows:

$$y_i=g(x_i)=\sum_{x_i=sv}\left(\beta^{\tilde{}}_i+\beta_i\right)K\left(x_i,x_j\right)+d \tag{3}$$

The set of support vectors is denoted as SV.

A newer class of nonlinear kernels, v-SVR, seeks to minimize structural risk in a high-dimensional feature space by identifying the most efficient regression hyperplane. However, in ϵ-SVR, the number of support vectors cannot be directly controlled (Chou, 2017). [28] introduced an improved version, v-SVR, which enhances the computational efficiency of SVR by regulating both the number of support vectors and training errors, while simultaneously estimating ϵ automatically. A schematic representation of support vectors and errors is shown below [17].

In v-SVR, the parameter v determines the proportion of support vectors, allowing for an automatic and optimal computation of ϵ during optimization. Conversely, in ϵ-SVR, the number of selected support vectors from the dataset cannot be predetermined and may vary significantly. The key distinction between ϵ-SVR and v-SVR lies in the parameterization during training. Through the v parameter, users can directly control the number of support vectors produced by the SVM [19].

v-SVR reformulates the original SVR optimization problem into a convex quadratic programming problem with inequality constraints [5, 19]:

$$\underset{Md}{min}\left\{\frac{1}{2}M^{T}M+H\left[v\varepsilon+\frac{1}{n}\sum_{i=1}^{n}\left(\alpha_{i}+\alpha\tilde{}_{i}\right)\right]\right\} \tag{4}$$

Such that:

$$\{y_{i}-\left(M\cdot\pi\left(x_{i}\right)+d\right)\leq\varepsilon+\alpha\tilde{}_{i}\ \left(M\cdot\pi\left(x_{i}\right)+d\right)-y_{i}\leq\varepsilon+\alpha_{i}\ \alpha_{i},\alpha\tilde{}_{i}\geq0,\varepsilon\geq0$$

According to [6, 10], the v-SVR optimization framework results in a convex quadratic programming problem with inequality constraints.

where $0\leq v\leq1$. [28] demonstrated that

$$v$$

provides an effective constraint by linking it to margin errors and the fraction of support vectors. The inclusion of

$$v$$

within Equation (4) introduces a new level of flexibility, enabling automatic determination of $\varepsilon\geq0$. Moreover,

$$v$$

is easier to estimate than the regularization constant C [12, 13, 55, 56, 57].

After converting Equation (4) into its dual form, the optimization problem is solved using the Lagrange multiplier technique [12, 13]:

$$\begin{aligned}L\left(M,d,\varepsilon,\alpha_{i},\alpha\tilde{}_{i}\right)=\frac{1}{2}M^{T}M+H\left(v\varepsilon+\frac{1}{n}\sum_{i=1}^{n}\left(\alpha_{i}+\alpha\tilde{}_{i}\right)\right)-\sum_{i=1}^{n}\tau_{i}\alpha_{i}\ -\sum_{i=1}^{n}\tau\tilde{}_{i}\alpha\tilde{}_{i}\ -\rho\varepsilon\\+\sum_{i=1}^{n}\beta_{i}\left(M^{T}\pi\left(x_{i}\right)+d-y_{i}-\varepsilon-\alpha_{i}\right)\ +\\\sum_{i=1}^{n}\widetilde{\beta}_{i}\left(M^{T}\pi\left(x_{i}\right)+d-y_{i}-\varepsilon-\alpha\tilde{}_{i}\right)\end{aligned} \tag{5}$$

The optimization involves non-negative multiplier variables $\beta_{i},\beta\tilde{}_{i},\tau_{i},\tau\tilde{}_{i}$,and $\rho\geq0$. The solution to Equation (5) is obtained by partially differentiating the Lagrangian function $\alpha_{i},M,d,\varepsilon$, with respect to $\alpha\tilde{}_{i}$ the model parameters:

$$\begin{aligned}\frac{\partial L}{\partial M}=M+\sum_{i=1}^{n}\beta_{i}x_{i}\ -\sum_{i=1}^{n}\beta\tilde{}_{i}x_{i}\ =0\ \frac{\partial L}{\partial d}=\sum_{i=1}^{n}\beta_{i}\ -\sum_{i=1}^{n}\beta\tilde{}_{i}\ =0\ \frac{\partial L}{\partial\delta}\\=\frac{H}{n}\sum_{i=1}^{n}v\ -\rho-\sum_{i=1}^{n}\left(\beta_{i}+\beta\tilde{}_{i}\right)\ =0\Rightarrow\ \frac{\partial L}{\partial\alpha}=\sum_{i=1}^{n}\frac{H}{n}\ -\sum_{i=1}^{n}\beta_{i}\ -\sum_{i=1}^{n}\tau_{i}\ =0\ \frac{\partial L}{\partial\alpha\tilde{}}\\=\sum_{i=1}^{n}\frac{H}{n}\ -\sum_{i=1}^{n}\beta\tilde{}_{i}\ -\sum_{i=1}^{n}\widetilde{\tau}_{i}\ =0\left\{M=\sum_{i=1}^{n}\left(\beta\tilde{}_{i}-\beta_{i}\right)x_{i}\ \ \sum_{i=1}^{n}\left(\beta\tilde{}_{i}-\beta_{i}\right)\right.\\=0\ \sum_{i=1}^{n}\left(\beta\tilde{}_{i}-\beta_{i}\right)\ =Hv-\rho\leq Hv\ \beta_{i}=\frac{H}{n}-\tau_{i}\leq\frac{H}{n}\ \beta\tilde{}_{i}=\frac{H}{n}-\tau\tilde{}_{i}\leq\frac{H}{n}\end{aligned} \tag{6}$$

Substituting Equation (6) into Equation (5) yields a new form of the Lagrangian function:

$$L = -\frac{1}{2}\sum_{i,j=1}^{n} \left(\beta^{\sim}{}_i - \beta_i\right)\left(\beta^{\sim}{}_j - \beta_j\right) K\left(x_i, x_j\right) + \sum_{i=1}^{n} \left(\beta^{\sim}{}_i - \beta_i\right) y_i, \tag{7}$$

According to the Karush–Kuhn–Tucker (KKT) conditions, finding the optimal solution to Equation (7) corresponds to achieving the optimal configuration of the v-SVR model. Consequently, the final decision function of v-SVR is derived from the optimal dual formulation, providing an efficient and theoretically grounded regression model [12, 13].

$$y_i = g\left(x_i\right) = \sum_{i=1}^{n} \left(\beta^{\sim}{}_i + \beta_i\right) K\left(x_i, x_j\right) + d. \tag{8}$$

2.3. Proposed Method

The performance of the Support Vector Regression (v-SVR) model is highly dependent on the appropriate selection of its key hyperparameters. Conventional mathematical techniques are insufficient for accurately determining these parameters [31]. Consequently, the optimal tuning of v-SVR hyperparameters has become a central research topic. Prior studies have introduced various techniques and methodological approaches aimed at improving the selection of these hyperparameters [31]. Additionally, numerous researchers have employed nature-inspired optimization algorithms to identify optimal v-SVR hyperparameter values [2, 40, 41].

Support Vector Regression requires the adjustment of several critical hyperparameters, including the penalty parameter H, the ϵ-insensitive loss function, and the kernel parameter. Since there is no universally reliable mathematical procedure for selecting these values, proper hyperparameter tuning plays a fundamental role in achieving high predictive performance. Several studies have demonstrated significant performance improvements in v-SVR through the adoption of suitable hyperparameter-selection strategies [42, 10, 11]. Nature-inspired optimization methods have also been widely adopted for v-SVR parameter tuning [5, 6, 1, 9, 4]. However, these approaches typically focus solely on parameter tuning without addressing feature selection simultaneously.

The Bamboo Forest Growth Optimization (BFGO) algorithm. Bamboo is a fast-growing herbaceous plant and is considered one of the fastest-growing plants globally. It possesses an underground rhizome system—referred to as bamboo whips—that expands horizontally and develops roots at the nodes, known as whip roots. Each node contains a bud that may develop into either a new whip or a bamboo shoot. New whips extend underground, while bamboo shoots emerge above the soil and gradually mature into bamboo culms, eventually forming a bamboo forest. Figure 1 illustrates the structural components of bamboo.

The bamboo whip plays a pivotal role in the development of bamboo forests, as it both expands the territory of the bamboo population and supplies nutrients necessary for growth. According to [14], bamboo exhibits unique growth characteristics compared to other grasses: its tall stems grow rapidly within 2–3 months, following an intensive yet rhythmic growth pattern. This biological trait likely enhances bamboo's adaptability and competitiveness for survival. The development of a bamboo forest can generally be divided into two main stages: (a) the extension of the bamboo whip and (b) the growth of bamboo shoots. Additionally, a bamboo forest may contain several whips, although each whip is associated with only one bamboo plant.

Recently, Chu, Feng, and colleagues introduced a novel optimization method inspired by bamboo forest growth dynamics, termed Bamboo Forest Growth Optimization (BFGO). The global extension of bamboo whips corresponds to the exploitation phase of the algorithm, whereas the emergence of bamboo shoots represents the exploration phase. Bamboo shoots that break through the soil surface have a relatively small probability of developing into mature bamboo culms.

2.3.1. Extension of the Bamboo Whip : Based on the biological relationship between bamboo forests and bamboo whips, the BFGO algorithm incorporates a clustering mechanism. During optimization, individuals are grouped according to memes, with dynamic transitions between uniform and random grouping. In uniform grouping, individuals are sorted in descending order based on fitness values, forming a sequence from which meme groups are

evenly constructed. Random grouping is applied when improvement in the best individuals of each group stagnates, redistributing individuals randomly. The meme-grouping concept is illustrated in Figure 2.

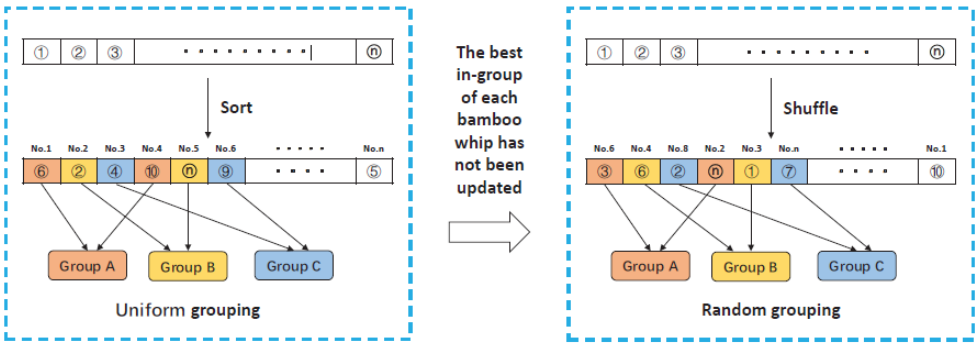

Figure 2. The idea of meme grouping.

The extension direction of an underground bamboo whip is influenced by three primary factors: group cognition, whip memory, and the bamboo forest center. These correspond respectively to the global optimum, the within-group optimum, and the forest center. The equation for computing the forest center is provided in Eq. (9), while the extension direction is defined in Equations (9-12):

$$\overrightarrow{C\left(k\right)} = \frac{1}{n}\sum_{i=1}^{n}\overrightarrow{X\left(k_i\right)} \tag{9}$$

$$coscos\ \left(\alpha\right) = \frac{\overrightarrow{X_t}\ .\ \overrightarrow{X_G}}{\left|\overrightarrow{X_t}\right| \times \left|\overrightarrow{X_G}\right|} \tag{10}$$

$$coscos\ \left(\beta\right) = \frac{\overrightarrow{X_t}\ .\ \overrightarrow{X_{P(k)}}}{\left|\overrightarrow{X_t}\right| \times \left|\overrightarrow{X_{P(k)}}\right|} \tag{11}$$

$$coscos\ \left(\gamma\right) = \frac{\overrightarrow{X_t}\ .\ \overrightarrow{C(k)}}{\left|\overrightarrow{X_t}\right| \times \left|\overrightarrow{C(k)}\right|} \tag{12}$$

where $X(k_i)$ represents the ith bamboo shoot position on the k bamboo whip, $\overrightarrow{X_t}$ is the current bamboo shoot position, $\overrightarrow{X_G}$ is the global optimal bamboo shoot position, $\overrightarrow{X_{P(k)}}$ is the optimal bamboo shoot position on the k bamboo whip and $C(k)$ is the central position of the bamboo forest. Moreover, $coscos\ (\alpha)$, $coscos\ (\beta)$ and $coscos\ (\gamma)$ represent the degree of extension of the current bamboo shoot position to $\overrightarrow{X_G}$, $\overrightarrow{X_{P(k)}}$, and $C(k)$, respectively. The formula for the update is shown in Equation (13).

$$\begin{aligned} X_{t+1} = \{ & X_G + Q \times (c_1 \times X_G - X_t) \times coscos\ (\alpha)\ ,\ rand \leq 0.4\ X_{p(k)} + \\ & Q \times \left(c_1 \times X_{p(k)} - X_t\right) \times coscos\ (\beta)\ ,\ 0.4 < rand \leq 0.8\ C(k) + Q \times \\ & (c_1 \times C(k) - X_t) \times coscos\ (\gamma)\ ,\ else \end{aligned} \tag{13}$$

$$Q = 2 \times \left(1 - \frac{t}{T}\right) \tag{14}$$

where Q is a crucial parameter impacting the step size of the algorithm development and steadily reduces from 2 to 0 as the number of iterations grows, t is the current iteration, and T indicates the maximum number of iterations; c_1 is a random number from 0 to 1. By comparing a random number to 0.4 and 0.8, the algorithm determines the extension direction for the next generation, thereby preserving population diversity and improving the capacity for global optimization.

2.3.2. Shoot Growth of the Bamboo : It is well established that tree growth is influenced by numerous stochastic factors. Throughout the growth cycle, these factors—whether individually or collectively—exert varying degrees of influence. Because their precise interactions are difficult to measure or predict, tree growth is generally modeled as a stochastic process. Due to environmental variability and random disturbances, the cumulative growth of different bamboo culms at a given time *t* fluctuates randomly, illustrating the inherent randomness of bamboo growth.

Based on these characteristics, Shi et al. developed a stochastic process model for bamboo shoot growth using stochastic process theory and the Sloboda growth equation [44]. The bamboo shoot growth process consists of two stages: a slow-growth phase and a rapid-growth phase. This behavior is modeled mathematically as shown in Eq. (15).

$$X\left(t\right) = X_G \times e^{\frac{b}{\omega \times e^{\omega}}} \tag{15}$$

The growth increment pattern is depicted in Figure 3. The rapid-growth period lasts approximately 55 days, and the process can be divided into two sub-stages around day 25:

1. Days 1–25: slow and steady growth
2. Days 25–55: sudden and accelerated growth

Moreover, X_G represents the maximum bamboo height in a particular growth environment, varying with the growth environment; b is the bamboo measurement factor, a random variable; and is the shape parameter of ω the model, independent of the environmental conditions.

Changes in incremental growth over time are computed using Equation (16), which extends to vector form in multi-dimensional cases.

$$\Delta H = \frac{X\left(t\right) - X(t-1)}{X_G - C\left(k\right) + 1} \tag{16}$$

where ΔH represents the relationship between the increment between the two generations and X_G and $C\left(k\right)$, the denominator represents the distance between X_G and $C\left(k\right)$, and $X(t)$ indicates the cumulative length of the bamboo growth within the t-th generation.

The individual renewal of bamboo shoots at this stage is shown in Equations (17) and (18). In the multi-dimensional case, the result of Equation (18) is a vector.

$$X_{temp} = \{X_t + X_D \times \Delta H, rand < 0.5\, X_t + X_D \times \Delta H, else \tag{17}$$

$$X_D = 1 - \left| \frac{X_t - C\left(k\right)}{X_G - C\left(K\right) + 1} \right| \tag{18}$$

where X_D represents the ratio relationship between $C(k)$ and the distance between X_t and X_G, which varies to a large extent in the early stages of exploration and stabilizes or even

remains constant in the later stages of exploration as X_t converges extremely closely to X_G and $C(k)$. This results in a more extensive exploration in the early stages and a slow growth towards convergence in the later stages.

In Equation (17) '+' means the distance increases and '-'means the distance decreases, increasing the capacity to search for the optimal solution by expanding the search range and balancing global exploitation and local exploration.

In contrast, the method proposed in this study incorporates the use of Gaussian kernel functions with parameter $\sigma > 0$.

Each beluga whale in the search population is defined by three numerical values representing the hyperparameters H, $_{\sigma}$, and $_{v}$.

The outline of the proposed optimization algorithm is summarized as follows:

Step 1: Initialize a population of 30 beluga whales and define the maximum number of iterations $S_{max} = 500$.

Step 2: Assign the three hyperparameters H, $_{\sigma}$, and $_{v}$to each whale using a uniform distribution, ensuring that all values remain within predefined bounds. As $H \sim U(0, 4)$, $_{\sigma \sim U(0,2.5)}$, and $_{v \sim U(0,1)}$.

Step 3: Evaluate each whale using the fitness function defined in Equation (19).

$$fitness = \left[\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_{i(H,v,\sigma)})^2 \right] \tag{19}$$

Using the calculated fitness values, determine each whale's personal best position and the global best position.

Step 4: Update the positions of the beluga whales using the algorithm's mathematical updating rules.

Step 5: Repeat Steps 3 and 4 until the maximum number of iterations is reached s_{max} .

2.3.3. Forecast Accuracy Criteria Several evaluation metrics are commonly used to assess prediction accuracy in forecasting applications. In this study, six performance criteria were employed to evaluate the proposed optimization approach and the predictive capability of the model, consistent with prior research [2, 20, 7] The assessment metrics included mean absolute error (MAE) and mean squared error (MSE)and root mean square error (RMSE) together with trend accuracy (DA) and coefficient of determination (R2) (IA) Index of Agreement. Table 1 presents the mathematical expressions of these evaluation metrics along with their decision rules.

Table 1. Forecast accuracy criteria

	Evaluation criterion	Mathematical formula	Decision
1	MAE	$\frac{1}{n}\sum_{s=1}^{n} \lvert y_t - \hat{y}_t \rvert$	Lower value is better
2	RMSE	$\sqrt{\frac{1}{n}\sum_{s=1}^{n} (y_t - \hat{y}_t)^2}$	Lower value is better
3	MSE	$\frac{1}{n}\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$	Lower value is better
4	DA	$\frac{1}{n}\sum_{t=1}^{n} b_s, \quad b_s = \{1, \quad if(y_{s+1} - y_t)(\hat{y}_{s+1} - \hat{y}_t) \geq 0\ 0, \ otherwise$	Higher value is better
5	R^2	$1 - \left[\frac{\sum_{s=1}^{n} (y_s - \hat{y}_s)^2}{\sum_{s=1}^{n} (\hat{y}_s - \hat{y}_s)^2} \right]$	Larger value near to 1 is better
6	IA	$\frac{^2(sA_s - C)_{s=1}^{n}}{^2(\lvert\_(\lvert\ _S A_S - C\rvert + \lvert C\rvert)_{S=1}^{n}} - 1$	Equal to 1

3. Results Analysis

The primary objective of employing the proposed BFGO optimization framework is to demonstrate that an optimal selection of hyperparameters can substantially enhance the predictive performance of daily PM10 concentration forecasting. To rigorously evaluate the effectiveness of the proposed approach, an extensive comparative analysis was conducted against several widely used hyperparameter optimization techniques, namely Random Search (RS), Bayesian Optimization (BO), Cross-Validation (CV), and Grid Search (GS). These benchmark methods were utilized to assess and compare the forecasting capability of the proposed BFGO-based model. The dataset was divided into two subsets to ensure a robust evaluation of the model's predictive performance. The training dataset consists of 1270 daily observations of PM10 concentrations, covering the period from 1 March 2019 to 31 December 2022, whereas the testing dataset includes 151 daily observations spanning from 1 January 2023 to 1 June 2023. Figures 3 and 4 illustrate the time-series behavior of daily PM10 concentrations for the training and testing datasets, respectively. A visual inspection of these figures clearly reveals that the temporal evolution

of PM10 concentrations exhibits pronounced nonlinear and non-stationary characteristics, which are commonly observed in environmental and atmospheric pollution datasets. To further investigate the statistical characteristics of the data, Table 2 presents the descriptive statistics of the daily PM10 concentrations for both datasets. These statistics provide valuable insights into the underlying distributional properties of the data. The results indicate the presence of substantial variability, asymmetric distribution, intermittent fluctuations, and potential trend components, highlighting the complex stochastic structure of PM10 concentration dynamics.

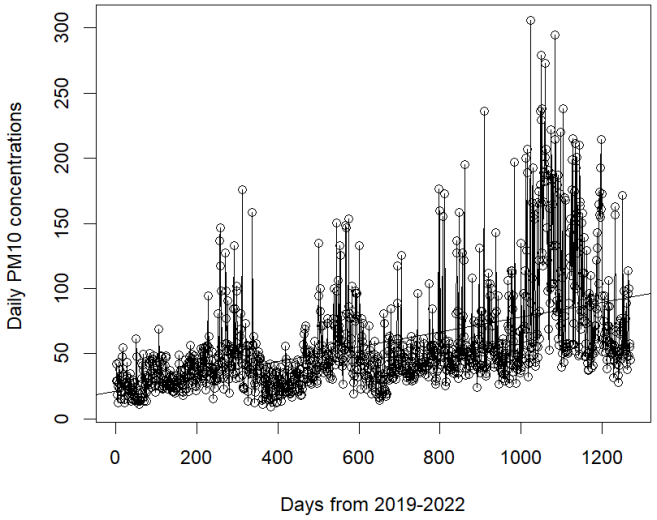

Figure 3. Time series of the daily PM10 concentrations for the training data set.

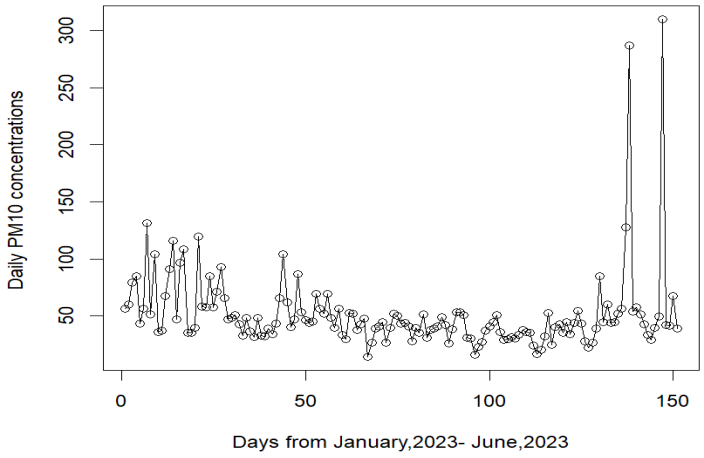

Figure 4. Time series of the daily PM10 concentrations for the testing data set.

The forecasting performance of the **BFGO**, **RS**, **BO**, **CV**, and **GS** models, evaluated using multiple performance metrics, is summarized in **Tables 3 and 4** for the training and testing datasets, respectively. The reported results clearly demonstrate that the proposed **BFGO optimization strategy significantly enhances both forecasting accuracy and model generalization ability**. This improvement is evidenced by **lower values of MAE and RMSE**, along with **higher values of DA and R^2**, compared with the benchmark approaches.

Table 2. The statistical descriptive of daily PM10 concentrations

	Training data set	Testing data set
Mean	57.53971	51.55414
Standard deviation	42.4731	35.83342
Skewness	2.272025	4.735841
Kurtosis	5.774809	29.5129
Minimum	9.471	14.361
Maximum	306.0128	310.3792

A careful examination of **Table 3** reveals that the **BFGO-based model consistently outperforms the competing optimization techniques** across all evaluation metrics. In particular, the BFGO algorithm achieves substantial error reductions relative to the conventional optimization methods. Specifically, when compared with **RS**, **BO**, **CV**, and **GS**, the BFGO model reduces **MAE** by **87.26%**, **85.53%**, **86.01%**, and **86.55%**, respectively. Similarly, the corresponding reductions in **RMSE** are **87.22%**, **86.75%**, **86.99%**, and **87.16%**, respectively. These findings highlight the effectiveness of the BFGO algorithm in identifying optimal hyperparameter configurations for the forecasting model.

A comparable pattern is observed for the **testing dataset**, as shown in **Table 4**. The BFGO model achieves remarkable improvements in predictive accuracy relative to the benchmark approaches. Specifically, the reductions in **MAE** compared with **RS**, **BO**, **CV**, and **GS** are **87.61%**, **86.67%**, **86.91%**, and **87.21%**, respectively. Furthermore, the reductions in **RMSE** are **83.61%**, **83.05%**, **83.34%**, and **83.54%**, respectively. These results confirm the **robust generalization capability** of the proposed BFGO-based forecasting framework when applied to unseen data.

It is worth noting that **meta-heuristic optimization algorithms generally demonstrate strong predictive capabilities**, even in the absence of extensive data preprocessing procedures. In contrast, the comparatively weaker performance of traditional hyperparameter tuning techniques such as **RS**, **BO**, **CV**, and **GS** may largely be attributed to **suboptimal or randomly selected hyperparameter configurations**, which may fail to adequately capture the complex nonlinear relationships present in **PM10 concentration data**.

Furthermore, when comparing the performance of **CV**, **RS**, and **GS** in estimating the optimal hyperparameters of the **Random Forest (RF)** model, the evaluation results based on the training dataset indicate that the **CV approach performs better than both RS and GS**. This suggests that systematic validation strategies may provide more reliable hyperparameter estimates than purely random or grid-based search procedures. In contrast, the **RS method yielded relatively unsatisfactory results**, indicating its limited ability to consistently identify effective hyperparameter combinations.

Table 3. The forecasting results of the BFGO and the benchmark approaches RS, BO, CV, and GS for the training set

	BFGO	RS	BO	CV	GS
MAE	0.148	0.998	0.881	0.91	0.946
RMSE	1.332	10.261	9.901	10.081	10.213
DA	0.971	0.638	0.681	0.668	0.65
R^2	0.975	0.601	0.698	0.691	0.612

Table 4. The forecasting results of the BFGO and the benchmark approaches RS, BO, CV, and GS for the testing set

	BFGO	RS	BO	CV	GS
MAE	0.229	1.679	1.562	1.591	1.627
RMSE	1.813	10.942	10.582	10.762	10.894
DA	0.951	0.622	0.665	0.652	0.634
R^2	0.959	0.585	0.682	0.675	0.596

The time-series plots of the forecasting results further highlight the superior predictive stability of the RF model optimized using the BFGO algorithm for both the training and testing datasets. As illustrated in Figures 5 and 6,

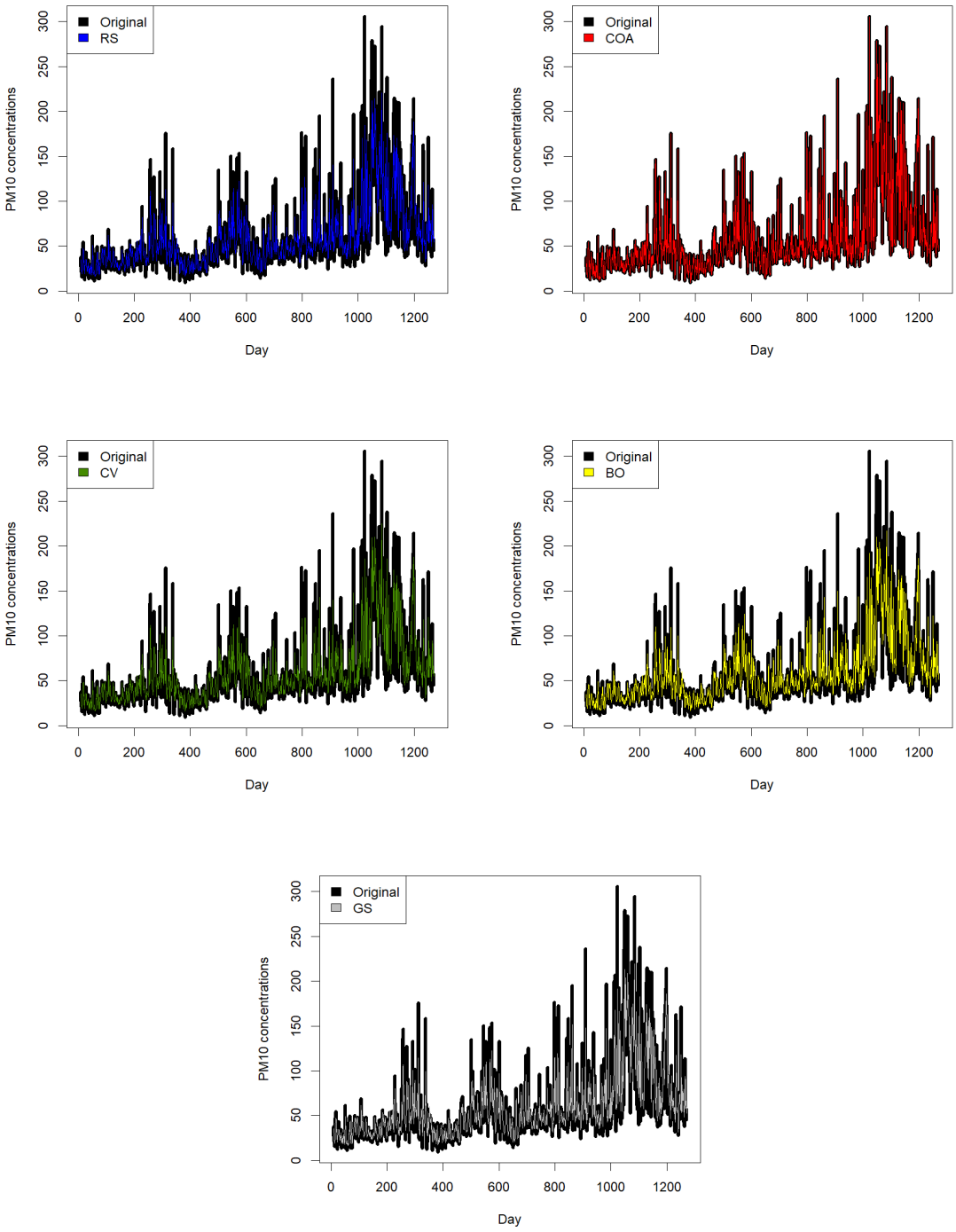

Figure 5. Forecasting results in training data set based on methods used.

the predicted daily PM10 concentrations generated by the BFGO-based model closely follow the actual observed values, demonstrating the high reliability and predictive accuracy of the proposed framework. Moreover, the forecasting trajectories produced by the BFGO model exhibit smooth and consistent temporal behavior, which

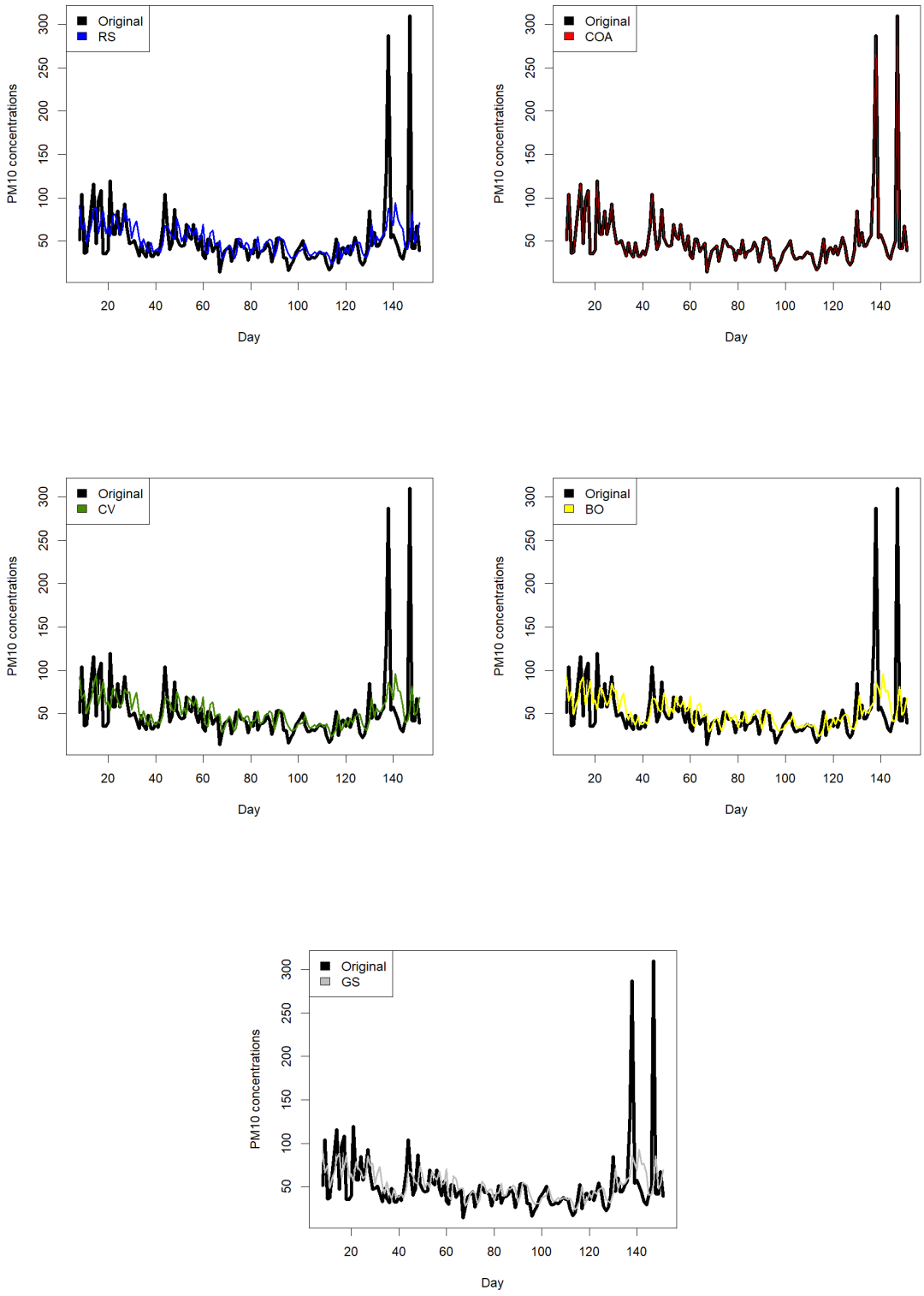

Figure 6. Forecasting results in testing data set based on methods used.

reflects the model's ability to effectively capture the underlying dynamics of PM10 concentration variability. Consequently, the BFGO algorithm enables the RF model to deliver highly accurate and stable forecasts of daily PM10 concentrations. Additionally, as shown in Figures 5 and 6, the forecasting results obtained using the CV-based optimization approach appear slightly closer to the actual observations compared to those obtained using the BO method. In contrast, the predictions generated by the RS method display comparatively weaker performance over time, indicating its limited effectiveness in identifying optimal hyperparameter configurations for the forecasting model.

4. Conclusion

This study developed an enhanced v-support vector regression framework, integrated with a Bamboo Forest Growth Optimization strategy, to provide accurate daily forecasts of PM10 concentrations in Baghdad, Iraq. By jointly optimizing v-SVR hyperparameters and performing implicit feature selection, the proposed approach effectively captured the complex nonlinear and non-stationary behavior of PM10, which is driven by intertwined meteorological, environmental, and anthropogenic factors. The empirical analysis demonstrated that the BFGO-based v-SVR model substantially outperforms conventional v-SVR configurations and widely used hyperparameter tuning strategies such as random search, grid search, cross-validation, and Bayesian optimization, across both training and testing periods. In particular, the proposed framework consistently achieved marked reductions in MAE and RMSE alongside notable gains in R2 and directional accuracy, indicating not only improved point prediction performance but also a stronger ability to track turning points and extreme pollution episodes. These improvements were especially evident during high-pollution conditions associated with dust storms and intensive emission events, where many benchmark models typically deteriorate. From an applied perspective, the BFGO–v-SVR model offers a robust, flexible, and computationally efficient tool for real-time PM10 forecasting in data-constrained yet highly polluted urban environments. Its capacity to exploit rich environmental information while avoiding overfitting makes it well suited for supporting early-warning systems, short-term exposure management, and targeted regulatory interventions in Baghdad and similar semi-arid megacities. More broadly, the methodological framework is generic and can be readily extended to other pollutants, geographic regions, and forecasting horizons, as well as hybridized with additional machine learning or deep learning architectures. Future research could explore multi-station and spatially explicit extensions, probabilistic forecast formulations, and integration with health impact models to better quantify and mitigate the public health burden of particulate pollution.

REFERENCES

1. M. N. Amar, and N. Zeraibi, *Application of hybrid support vector regression artificial bee colony for prediction of MMP in CO2-EOR process*, Petroleum, vol. 6, no. 4, pp. 415–422, 2020.
2. G. Cao, and L. Wu, *Support vector regression with fruit fly optimization algorithm for seasonal electricity consumption forecasting*, Energy, vol. 115, pp. 734–745, 2016.
3. Q. Chen, W. Chen, and G. Pan, *An improved picture-based prediction method of PM2.5 concentration*, IET Image Processing, vol. 16, no. 11, 2022.
4. J. Cheng, J. Qian, and Y. N. Guo, *Adaptive chaotic cultural algorithm for hyperparameters selection of support vector regression*, In International Conference on Intelligent Computing, pp. 286–293, 2009.
5. V. Cherkassky, and Y. Ma, *Practical selection of SVM parameters and noise estimation for SVM regression*, Neural Networks, vol. 17, no. 1, pp. 113–126, 2004.
6. J. S. Chou, and A. D. Pham, *Nature-inspired metaheuristic optimization in least squares support vector regression for obtaining bridge scour information*, Information Sciences, vol. 399, pp. 64–80, 2017.
7. F. X. Diebold, and R. S. Mariano, *Comparing predictive accuracy*, Journal of Business & Economic Statistics, vol. 20, no. 1, pp. 134–144, 2002.
8. R. Feng, H. Gao, K. Luo, and J. Ren Fan, *Analysis and accurate prediction of ambient PM2.5 in China using Multi-layer Perceptron*, Atmospheric Environment, vol. 232, 2020.
9. W. C. Hong, Y. Dong, L. Y. Chen, and S. Y. Wei, *SVR with hybrid chaotic genetic algorithms for tourism demand forecasting*, Applied Soft Computing, vol. 11, no. 2, pp. 1881–1890, 2011.
10. C. F. Huang, *A hybrid stock selection model using genetic algorithms and support vector regression*, Applied Soft Computing, vol. 12, no. 2, pp. 807–818, 2012.

11. J. Huang, and H. Hu, *Hybrid beluga whale optimization algorithm with multi-strategy for functions and engineering optimization problems*, Journal of Big Data, vol. 11, no. 1, p. 3, 2024.
12. O. M. Ismael, O. S. Qasim, and Z. Y. Algamal, *Improving Harris hawks optimization algorithm for hyperparameters estimation and feature selection in v-support vector regression*, Journal of Chemometrics, vol. 34, no. 11, e3311, 2020.
13. O. M. Ismael, O. S. Qasim, and Z. Y. Algamal, *A new adaptive algorithm for v-support vector regression with feature selection using Harris hawks optimization algorithm*, Journal of Physics: Conference Series, vol. 1897, no. 1, p. 012057, 2021.
14. G. Jin, P. F. Ma, X. Wu, L. Gu, M. Long, C. Zhang, and D. Z. Li, *New genes interacted with recent whole-genome duplicates in the fast stem growth of bamboos*, Molecular Biology and Evolution, vol. 38, pp. 5752–5768, 2021.
15. I. F. Kao, J. Y. Liou, M. H. Lee, and F. J. Chang, *Fusing stacked autoencoder and long short-term memory for regional multistep-ahead flood inundation forecasts*, Journal of Hydrology, vol. 598, 2021.
16. A. Kazem, E. Sharifi, F. K. Hussain, M. Saberi, and O. K. Hussain, *Support vector regression with chaos-based firefly algorithm for stock market price forecasting*, Applied Soft Computing, vol. 13, no. 2, pp. 947–958, 2013.
17. R. Laref, E. Losson, A. Sava, and M. Siadat, *On the optimization of the support vector machine regression hyperparameters setting for gas sensors array applications*, Chemometrics and Intelligent Laboratory Systems, vol. 184, pp. 22–27, 2019.
18. J. Lelieveld, J. S. Evans, M. Fnais, D. Giannadaki, and A. Pozzer, *The contribution of outdoor air pollution sources to premature mortality on a global scale*, Nature, vol. 525, no. 7569, 2015.
19. S. Li, H. Fang, and X. Liu, *Parameter optimization of support vector regression based on sine cosine algorithm*, Expert Systems with Applications, vol. 91, pp. 63–77, 2018.
20. X. Li, T. Sengupta, K. S. Mohammed, and F. Jamaani, *Forecasting the lithium mineral resources prices in China: Evidence with Facebook Prophet and artificial neural networks methods*, Resources Policy, vol. 82, p. 103580, 2023.
21. N. C. Liou, C. H. Luo, S. Mahajan, and L. J. Chen, *Why is short-time PM2.5 forecast difficult? The effects of sudden events*, IEEE Access, vol. 8, 2020.
22. S. H. Mohd Shafie, M. Mahmud, S. Mohamad, N. L. F. Rameli, R. Abdullah, and A. F. Mohamed, *Influence of urban air pollution on the population in the Klang Valley, Malaysia: A spatial approach*, Ecological Processes, vol. 11, no. 1, 2022.
23. C. Murray et al., *Global burden of 87 risk factors in 204 countries and territories, 1990–2019*, The Lancet, vol. 396, no. 10258, 2020.
24. B. S. Murthy, R. Latha, A. Tiwari, A. Rathod, S. Singh, and G. Beig, *Impact of mixing layer height on air quality in winter*, Journal of Atmospheric and Solar-Terrestrial Physics, vol. 197, 2020.
25. P. Muthukumar et al., *Predicting PM2.5 atmospheric air pollution using deep learning*, Air Quality, Atmosphere and Health, vol. 15, no. 7, 2022.
26. A. Najah Ahmed et al., *Machine learning methods for better water quality prediction*, Journal of Hydrology, vol. 578, 2019.
27. World Health Organization, *WHO global air quality guidelines: PM2.5 and PM10*, 2021.
28. Schölkopf, B., Smola, A. J., Williamson, R. C., & Bartlett, *New support vector algorithms*, Neural Computation, vol. 12, no. 5, pp. 1207–1245, 2000.
29. V. A. Southerland et al., *Global urban temporal trends in fine particulate matter*, The Lancet Planetary Health, vol. 6, no. 2, 2022.
30. M. Tian et al., *Effects of dust emissions on ambient air quality*, Atmospheric Pollution Research, vol. 12, no. 7, 2021.
31. P. Tsirikoglou et al., *A hyperparameters selection technique for support vector regression models*, Applied Soft Computing, vol. 61, pp. 139–148, 2017.
32. T. Tsurumi, and S. Managi, *Effects of PM2.5 on life satisfaction*, Economic Analysis and Policy, vol. 67, 2020.
33. J. Wang et al., *A forecasting framework on fusion of spatiotemporal features for PM2.5*, Expert Systems with Applications, vol. 238, p. 121951, 2024.
34. W. Wang et al., *Estimation of PM2.5 concentrations in China using neural networks*, Scientific Reports, vol. 9, no. 1, 2019.
35. P. Y. Wong et al., *Using land use regression with machine learning to estimate PM2.5*, Environmental Pollution, vol. 277, 2021.
36. D. A. Wood, *Trend decomposition aids forecasts of PM2.5*, Atmospheric Pollution Research, vol. 13, no. 3, 2022.
37. A. Wu et al., *Machine learning methods for predicting particulate matter levels*, Concurrency and Computation: Practice and Experience, vol. 34, no. 19, 2022.
38. K. Y. Wu et al., *High-spatiotemporal-resolution PM2.5 forecasting by hybrid models*, Journal of Cleaner Production, vol. 433, p. 139825, 2023.
39. Z. Xu et al., *Impacts of land supply on PM2.5 concentration*, Journal of Cleaner Production, vol. 347, 2022.
40. C. H. Zhang, G. H. Tzeng, and R. H. Lin, *Hybrid genetic algorithm for SVR optimization*, Expert Systems with Applications, vol. 36, no. 3, pp. 4725–4735, 2009.
41. J. Zhang et al., *Optimization enhanced GA-SVR for prediction*, Neurocomputing, vol. 240, pp. 183–190, 2017.
42. C. Zhong, G. Li, and Z. Meng, *Beluga whale optimization: A novel nature-inspired metaheuristic algorithm*, Knowledge-Based Systems, vol. 251, p. 109215, 2022.
43. S. Zhong et al., *Machine learning in environmental science and engineering*, Environmental Science and Technology, vol. 55, no. 19, pp. 12741–12754, 2021.
44. B. Zhu et al., *Forecasting carbon price using decomposition and SVR*, Applied Energy, vol. 191, pp. 521–530, 2017.
45. Saleem, W. W. *Using The Biorthogonal Wavelet to Address the Problem Heterogeneity of Variance* , Statistics, Optimization Information Computing, 15(4), p. 3085-3100, 2026.
46. Qasim, M. K., Algamal, Z. Y., Ali, H. M. *A binary QSAR model for classifying neuraminidase inhibitors of influenza A viruses (H1N1) using the combined minimum redundancy maximum relevancy criterion with the sparse support vector machine*, SAR and QSAR in Environmental Research, vol. 29, no.(7), p.517-527, 2018
47. Al-Taweel, Y., Algamal, Z. *Some almost unbiased ridge regression estimators for the zero-inflated negative binomial regression model*, Periodicals of Engineering and Natural Sciences, vol. 8, no.(1), p.248-255, 2020
48. Ewees, A. A., Algamal, Z. Y., Abualigah, L., Al-Qaness, M. A., Yousri, D., Ghoniem, R. M., Abd Elaziz, M. *A cox proportional-hazards model based on an improved aquila optimizer with whale optimization algorithm operators*, Mathematics, vol. 10, no.(8), p.1273, 2022

49. Algamal, Z. Y., Lee, M. H. *Applying penalized binary logistic regression with correlation based elastic net for variables selection*, Journal of Modern Applied Statistical Methods, vol. 14, no.(1), p.15, 2015
50. Algamal, Z. Y., Lukman, A. F., Abonazel, M. R., & Awwad, F. A. *Performance of the ridge and Liu estimators in the zero-inflated Bell regression model* Scientifica, vol. 1, p. 9503460, 2022.
51. Algamal, Z. Y., Lee, M. H., Al-Fakih, A. M., & Aziz, M. *High-dimensional QSAR modelling using penalized linear regression model with L 1/2-norm*, SAR and QSAR in Environmental Research, vol. 27, no.(9), p.703-719, 2016
52. Shamany, R., Alobaidi, N. N., & Algamal, Z. Y. *A new two-parameter estimator for the inverse Gaussian regression model with application in chemometrics*, Electronic Journal of Applied Statistical Analysis, vol. 12, no.(2), p.453-464, 2019
53. Algamal, Z. Y., Qasim, M. K., Lee, M. H., & Ali, H. T. M. *QSAR model for predicting neuraminidase inhibitors of influenza A viruses (H1N1) based on adaptive grasshopper optimization algorithm*, SAR and QSAR in Environmental Research, vol.31, no.11, p.803-814, 2020.
54. Al-Fakih, A. M., Algamal, Z. Y., Lee, M. H., Aziz, M., & Ali, H. T. M. *QSAR classification model for diverse series of antifungal agents based on improved binary differential search algorithm*, SAR and QSAR in Environmental Research, vol.30, no.2, p.131-143, 2019.
55. Alkhateeb, A. N., Algboory, A. M., Hadied, Z. A., Al-Saqal, O. E., & Algamal, Z. Y. *Enhancing Surface Water Potability Assessment through a v-SVM Hybrid Statistical Learning Model*, Statistics, Optimization Information Computing, 15(6), p.5481-5494, 2026.
56. Hawa, N. S., Hadied, Z. A., Al-Saqal, O. E., & Algamal, Z. Y. *). Linearized Ridge-Type Estimator for the Poisson Modification of the Quasi-Lindley Regression Model*, Statistics, Optimization & Information Computing, 16(2), p.1169-1181, 2026.
57. Hadied, Z. A., Al-Saqal, O. E., & Algamal, Z. Y. *Liu-type estimator in inverse gaussian regression model based on (r-(kd)) class estimator*, Iraqi Journal for Computer Science and Mathematics, vol. 6, no. 1, p. 51–56, 2025.