Carbon Emission Prediction and Emission Reduction Analysis of Wastewater Treatment Plant Based on Machine Learning and Deep Learning

Fangqin Liu
China Jiliang University
Ning Ding
tinin@hdu.edu.cn
Hangzhou Dianzi University
Guanghua Zheng
Hangzhou Dianzi University
Jiangrong Xu
China Jiliang University

Research Article

Keywords: wastewater treatment, carbon emission, carbon accounting, carbon emission prediction, carbon neutrality

Posted Date: June 27th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-4561438/v1

Additional Declarations: No competing interests reported.

Carbon Emission Prediction and Emission Reduction Analysis of Wastewater Treatment Plant Based on Machine Learning and Deep Learning

Fangqin Liu[a], Ning Ding[b,*], Guanghua Zheng[b], Jiangrong Xu[a,*]
[a]College of Metrology and Measurement Instrument, China Jiliang University, Hangzhou, 310018, Zhejiang Province, China
[b]School of sciences, Hangzhou Dianzi University, Hangzhou, 310018, Zhejiang Province, China
***Corresponding author: Ning Ding，Jiangrong Xu**
Abstract: Accurate accounting and prediction of carbon emissions from sewage treatment plants is the basis for exploring low-carbon sewage treatment plants and measures to reduce pollution and carbon emissions. Although carbon emission prediction models have been widely used in construction, transportation and other fields, research in the field of wastewater treatment is still lacking, and the existing research is mostly limited to the prediction of carbon emissions from a single link or energy consumption, which makes it difficult to control the carbon emissions of the whole plant as a whole in order to realize the carbon emission reduction of the whole plant. This study proposes a hybrid prediction framework based on machine learning and deep learning, which integrates multiple algorithms and has strong adaptability and generalization ability. The pre-framework uses Pearson correlation coefficient to select feature values, constructs a combined prediction model based on the selected features using support vector machine (SVR) and artificial neural network (ANN), and optimizes the model parameters and structure using Gray Wolf Optimization (GWO) algorithm. The results show that the model has stronger prediction performance compared with other prediction models, with a mean absolute percentage error (MAPE) of 0.49% and an R2 of 0.9926. In addition, this study establishes six future development scenarios based on the historical data trends and policy outlines, which provide recommendations for the development of carbon emission reduction measures for wastewater treatment plants. This study can provide a reference for exploring efficient carbon management and achieving carbon neutrality in wastewater treatment plants.
Keywords: wastewater treatment; carbon emission; carbon accounting; carbon emission prediction; carbon neutrality.

1Introduction

The wastewater treatment sector is one of the ten highest emitters of greenhouse gases worldwide, accounting for around 2 to 3% of total carbon emissions. Moreover, its overall carbon footprint continues to rise annually. Furthermore, its carbon footprint is steadily increasing annually. Thus, the industry must adopt carbon reduction measures to align with "dual carbon" objectives and foster synergies between pollution mitigation and carbon abatement[1]. While wastewater treatment significantly improves water ecosystem health by diminishing influent pollutants and achieving high-quality effluent, it also generates greenhouse gases and solid wastes, notably sludge, alongside considerable energy and resource consumption, all negatively impacting the environment[2]. Therefore, accurate accounting and prediction of carbon emissions from wastewater treatment processes is essential to optimize pollution and decarbonization of the industry. These efforts can provide guidance on future emission trends and assist in the development of relevant energy efficiency and emission reduction strategies.

Analyzing and predicting GHG emissions from wastewater treatment plants based on accounting results can accelerate the decarbonization process of the wastewater treatment industry. Carbon emission prediction models are mainly divided into two categories: machine learning models and deep learning models. The machine learning emphasizes solving complex tasks by constructing mathematical models, while deep learning mainly simulates the composition of human neural networks and constructs hierarchical structures to automatically complete complex tasks. Both types of predictive models have been widely used in the fields of wastewater treatment and carbon emission prediction. Specific methods including regression analysis, random forests, and neural networks have been widely used in real-world scenarios such as wastewater quality prediction, carbon emission estimation, and greenhouse gas emission impact assessment [3-6]. For example, Azeez et al. developed a support vector regression (SVR) model to predict carbon emissions from automobiles [7], Zhu et al. investigated and predicted the peak carbon emissions from the transportation sector in China using an SVR model and a scenario analysis model [8] ; furthermore, Chu and Zhao used an enhanced PSO-SVR model to predict carbon emissions from buildings [9]. Safa et al. employed MLR model and ANN model to anticipate on-farm wheat production, with the ANN model demonstrating superior predictive capability [10]. Meanwhile, Vasilaki et al. developed an SVR model to forecast N2O concentration within both the anaerobic and aeration stages [11]. Additionally, Alali et al. conducted a comparative analysis of prediction accuracies among diverse machine learning models including ANN, SVR, KNN, etc., to estimate energy consumption in wastewater treatment [12].

However, model prediction performance is influenced by historical carbon emissions, demanding greater accuracy in carbon accounting models. Currently, carbon emission accounting methods for wastewater treatment systems mainly include the field monitoring method, emission factor methods, modeling methods and life cycle assessment methods. The field monitoring method closely reflects real emissions under specific conditions, especially direct greenhouse gas emissions. Although it provides valuable insights into emission reduction at treatment plant levels, it demands substantial manpower and resources. While the modeling method can estimate overall GHG emissions using partial monitoring data, their validation is hindered by the complexity of parameters and state variables. Additionally, constant parameter updates are required to adapt to changing environmental conditions, necessitating lengthy experimentation [13]. Moreover, modeling struggles to encompass all relevant factors, demanding high accuracy in data inputs. The life cycle assessment involves greenhouse gas quantification and environmental impact assessment, albeit with a more intricate accounting process.

The emission factor method is presently the most prevalent accounting approach, widely utilized at national and city scales due to its operational simplicity [14]. X. Zhao et al. estimated methane emissions from prefectural-level municipal wastewater treatment plants in China using this method [15], while Dan Wang et al. compiled an emission inventory for China's enterprise-level wastewater treatment plants spanning 2006-2019 [16]. Jiahui Han et al. employed a combination of the emission factor method and mass balance method to model on-site and off-site GHG accounting for paper mill WWTPs [17]. However, due to the lack of reliable emission factors, this method yields results with significant uncertainty. Presently, it primarily relies on emission factors provided by the IPCC. Nonetheless, research indicates that China's wastewater treatment plant emission factors are lower than those measured in European countries and by the IPCC [18], and variations exist in carbon emission factors among different provinces and treatment processes [19]. Hence, employing local carbon emission factors can effectively mitigate the accounting error associated with the emission factor method.

Despite the aforementioned progress, significant shortcomings persist in both accounting for and projecting carbon emissions from wastewater treatment. The localization of emission factors poses a substantial challenge, particularly in developing countries. While carbon emission prediction models have found widespread application in various domains like construction and transportation, research in the wastewater treatment field remains limited. Existing studies predominantly focus on predicting carbon emissions from individual processes or energy consumption within wastewater treatment systems, neglecting comprehensive greenhouse gas emission predictions for entire plants. This gap hampers efforts to effectively control plant-wide carbon emissions and achieve overall emission reductions. To address this gap, this study develops a carbon accounting model using a literature review and local emission factor databases to assess carbon emissions from entire sewage treatment plants. Additionally, it utilizes treated water volume, effluent data, pharmaceutical consumption, and energy consumption as input parameters to construct a hybrid prediction model integrating machine learning and deep learning techniques for predicting carbon emissions from sewage treatment plants. Finally, the prediction model is employed for scenario analysis to devise a rational emission reduction strategy based on the prediction outcomes.

2 Materials and methods

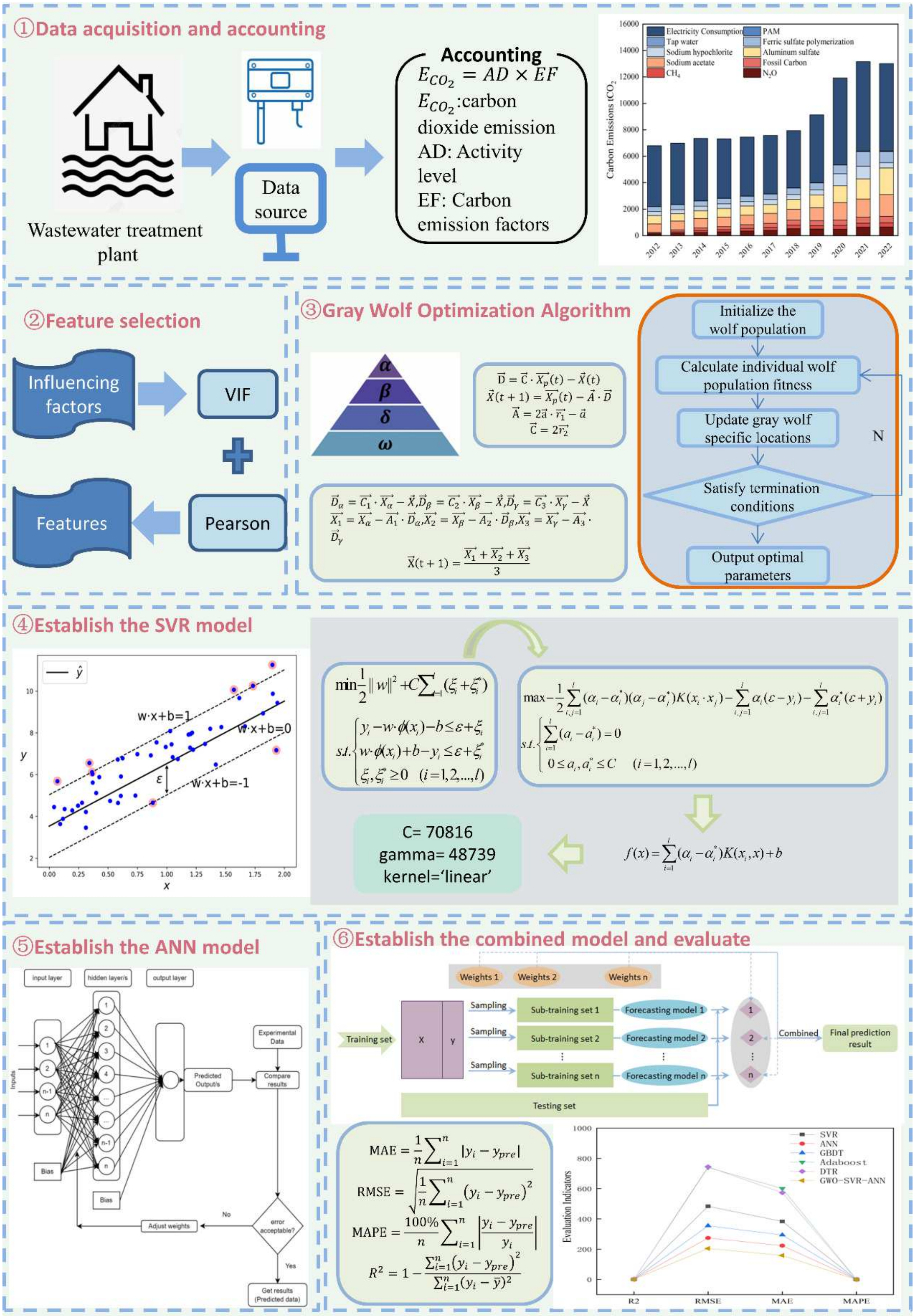


Fig. 1 Flowchart of the combined prediction model

Fig.1 illustrates the forecasting framework proposed in this study. The main steps are outlined as follows:

Step 1. Data acquisition and accounting

Acquire fundamental data from wastewater treatment plants via sensors and internet transmission. Establish an enhanced carbon accounting model based on this data.

Step 2. Feature selection

Utilizing the carbon accounting model and data exploration to narrow down influential factors, along with covariance and correlation analyses to ascertain model eigenvalues.

Step 3. Gray wolf optimization algorithm

The Gray Wolf Optimization algorithm enhances the traditional SVR model, seeking optimal kernel functions and parameter combinations to enhance prediction accuracy.

Step 4. Establish SVR model

Based on the accounting results in Step 1 and the feature selection in Step 2, construct the main part of the SVR model and optimize the model using the parameters from Step 3.

Step 5. Establish ANN model

Construct an ANN model based on the accounting results of Step 1 and the feature selection of Step 2, set the specific parameters of the model to debug the model, and evaluate the prediction results of the model.

Step 6. Establish combined model

Merge the models from Step 4 and Step 5, employ a model fusion technique to determine optimal weight combinations, and assess prediction performance against other models.

The advantages of the prediction model are as follows:

(1) The prediction model can realize the accurate prediction of plant-wide carbon emissions from a single wastewater treatment plant, which bridges the gap in the prediction of carbon emissions from wastewater treatment, and provides valuable insights for the emission prediction and carbon reduction strategies of other wastewater treatment plants.

(2) Integration of the improved SVR model and ANN model based on the Gray Wolf optimization algorithm makes full use of the nonlinear feature extraction capability of deep learning as well as the generalization capability and interpretability of machine learning. This integration can capture complex nonlinear data relationships more accurately and improve prediction accuracy, while reducing the risk of overfitting associated with individual models, thus enhancing the robustness of the overall model.

(3) Based on this prediction model and related policy outlines, specific quantification and analysis of the peak carbon attainment of individual wastewater treatment plants can be achieved, and the industry-wide peak carbon attainment and carbon neutral process can be promoted through individual carbon reduction.

2.1 Carbon emission accounting method

This study focuses on accounting for carbon emissions during the operation and maintenance phase of the wastewater treatment plant, drawing primarily from the IPCC National Greenhouse Gas Guidelines and relevant literature on wastewater treatment plant carbon emission accounting[1,20,21]. The accounting boundary, illustrated in Fig.2, encompasses direct emissions primarily from fossil sources, CH_4 emissions from anaerobic digestion, and N_2O emissions from denitrification. Only fossil carbon emissions are considered for CO_2 emissions in the biological treatment stage. Indirect emissions mainly account for carbon emissions from electricity and chemical consumption, purchased water sources, etc. Since the sludge is treated by thickening and transportation, only carbon emissions from thickening and dewatering are considered in the accounting. The main power-consuming equipment of sewage treatment plants are lifting pumps, grating machines, aeration equipment, sludge treatment equipment, etc. The chemicals are mainly flocculants, coagulants, phosphorus removers, disinfectants and so on. The specific calculation formula is as follows:

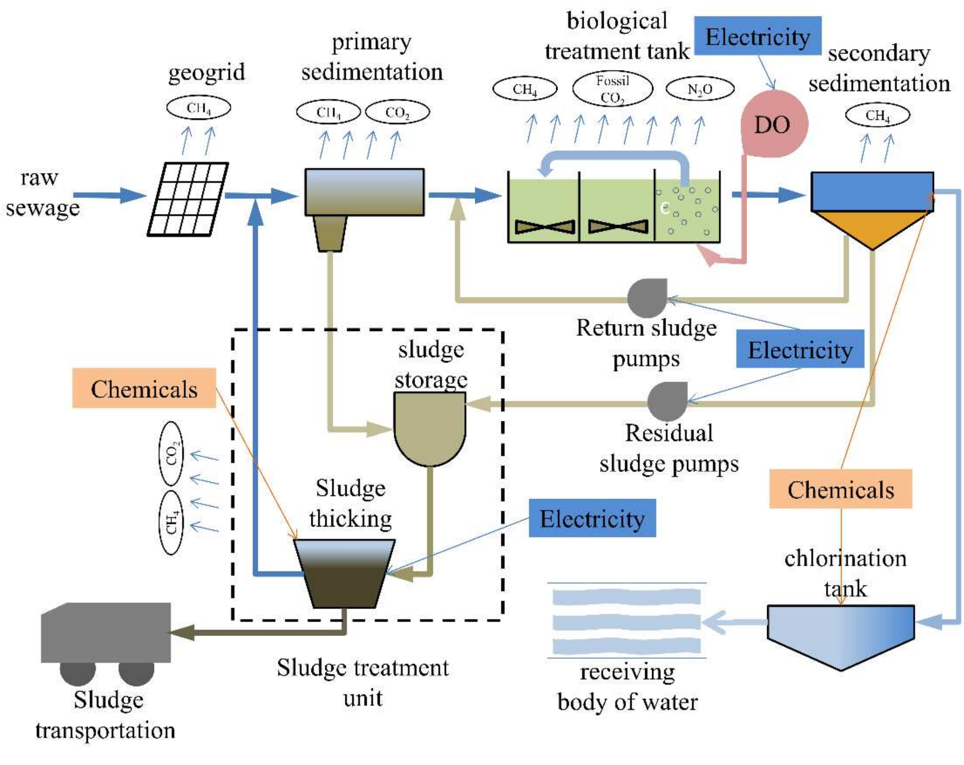


Fig. 2 GHG accounting boundary for wastewater treatment plants

(1) The formula for calculating for fossil carbon emissions in wastewater treatment plants is as follows:

$$m_{CO_2} = Q \times MFCF \cdot \Bigg\{ \Bigg[1.1\left(B_{in} + B_{ex} - B_{eff}\right) \times \left(1.47 - 1.42 \times \left(\frac{0.67}{1+K_d \cdot SRT}\right)\right) \Bigg] +$$

$$(1.947HRT \cdot MLVSS \cdot K_d) - 4.49 \times \Big[\left(TN_{in} - TN_{eff}\right) - \left(B_{in} + B_{ex} - B_{eff}\right) \times$$

$\left(\frac{0.67}{1+K_d \cdot SRT}\right) \times 0.124\right]\Bigg\} \times 10^{-3}$

$$\mathrm{MFCF} = \frac{FCF \times B_{in} + B_{ex}}{B_{in} + B_{ex}}$$
$$K_d = 0.05 \times 1.047^{T_b - 20}$$

Where: m_{CO_2} is the fossil source emissions from wastewater treatment, in terms of emission equivalents produced by treating each cubic meter of wastewater,$kgCO_2/m^3$; ; Q is the daily volume of water treated in the wastewater treatment plant,m^3; $MFCF$ is the proportion of fossil sources discharged; SRT is the mean biosolids retention time, d; 1.947 is the yield of microorganisms corresponding to the consumption of synthetic cellular material per 1 kg of endogenous respiration, $kgCO_2/kg$[22]; and HRT is the hydraulic retention time of the bioreactor, d; MLVSS is the mean mass concentration of volatile suspended solids in the mixed liquor of the bioreactor, mg/L;4.49 is the mass fixed per unit mass of ammonia nitrogen nitrification, $kgCO_2/kg$;TN_{in} is the mass concentration of total nitrogen in the influent of the wastewater treatment plant, mg/L，TN_{eff} is the mass concentration of total nitrogen in the effluent of the wastewater treatment plant, mg/L，0.124 is the mass ratio of the microorganisms ($C_5H_7O_2N$) body with N mass ratio; FCF is the proportion of fossil source organic matter in the influent water of the sewage treatment plant, taking the value of 10%[23];1.1 is the yield of microbial degradation of organic matter; $B_{in}$is the mass concentration of BOD in the influent water of the wastewater treatment plant, mg/l; $B_{ex}$is the additional carbon source artificially added during the operation, mg/l; $B_{eff}$is the effluent BOD of the sewage treatment plant, mg/l;1.47is the ratio of total BOD to BOD_5 of the influent water of the sewage treatment plant; 1.42 is the BOD_5 equivalence of the microbial cells, $kgCO_2/kg$; 0.67is the absolute yield coefficient, $kgCO_2/kg$; $K_d$is the decay coefficient, d; 0.05 is the reaction rate constant,d^{-1};1.047 is the temperature coefficient.

(2) The formula for calculating for CH4 at a wastewater treatment plant is as follows:

$$m_{CH_4} = \frac{Q \times (B_{in} - B_{eff})}{1000} \times EF_{CH_4}$$

m_{CH_4} is the discharge of CH_4 in the sewage treatment plant, kgCH_4/m^3;Q is the daily volume of water treated in the sewage treatment plant, m^3; B_{in} is the sewage treatment plant influent BOD concentration, mg/l; B_{eff} is the sewage treatment plant effluent BOD concentration, mg/l; EF_{CH_4} is the methane emission factor, kg$CH_4/kgBOD$.

(3) The formula for calculating for N_2O at a wastewater treatment plant is as follows:

$$m_{N_2O} = \frac{Q \times (T_{in} - T_{eff}) \times EF_{N_2O}}{1000} \times C_{N_2O/N_2}$$

m_{N_2O} is the discharge of N_2O in the sewage treatment plant, kgN_2O/m^3;Q is the daily volume of water treated in the sewage treatment plant, m^3; TN_{in} is the mass concentration of total nitrogen in the influent water of the sewage treatment plant, mg/l;TN_{eff} is the mass concentration of total nitrogen in the effluent water of the sewage treatment plant, mg/l; EF_{N_2O} is the emission factor of N_2O , kg$N_2O/kgCOD$; C_{N_2O/N_2} is the ratio of the molecular mass of N_2O to that of N_2.

(4) Electricity consumption is the purchased electricity during the production and operation of the wastewater treatment plant, excluding the electricity consumption of office and living areas. The formula for calculating for the electricity consumption of the wastewater treatment plant is as follows:

$$m_e = Q \times fe \times W$$

m_e is the indirect emissions from sewage treatment power consumption, $kgCO_2/m^3$; Q is the daily volume of water treated in the sewage treatment plant, m^3; fe is the carbon emission factor of power consumption, $kgCO_2/kwh$;W is the purchased electricity used for production and operation, kwh.

(4) Physical consumption is the coagulant, flocculant, carbon source, disinfectant and cleaning agent and other chemicals consumed during the production and operation of sewage treatment plants. The formula for calculating for carbon emissions from material consumption is as follows:

$$M_c = \sum_{g=1}^{m} \left(f_{c,g} \times M_{cg}\right)$$

M_c is the emission equivalent of the consumption, $kgCO_2/kg$; $f_{c,g}$ is the emission factor for chemical g, $kgCO_2/kg$; M_{cg} is the amount of chemical g used, kg.

(6) The formula for calculating the total carbon emissions from a wastewater treatment plant is as follows:

$$M_t = m_{CO_2} + m_{CH_4} \times 28 + m_{N_2O} \times 265 + m_e + M_c$$

M_t is the total carbon emission from the operation and maintenance of the wastewater treatment plant, $kgCO_2$; 28 is the global warming potential (GWP) of CH_4, constant, $kgCO_2/kgCH_4$; 265 is the global warming potential (GWP) of N_2O, constant, $kgCO_2/kgN_2O$.

2.2 Data sources

2.2.1 Basic data for operation and maintenance of wastewater treatment plants

The wastewater treatment plant examined in this study is situated in the northeastern region of the Cixi Binhai Economic Development Zone, designed to process a daily treatment capacity of 10×10^4 m^3. Data utilized include the daily wastewater treatment volume, average daily pollutant concentrations entering and exiting the facility, daily electricity consumption, and daily chemical usage from 2012 to 2022. These data are sourced from the wastewater treatment plant's routine measurements and statistical records. Table 1 presents specific operation and maintenance (O&M) parameters for the WWTP.

Table 1 Basic O&M data for wastewater treatment plants

Year	Q	CODin	TNin	CODeff	TNeff
2012	38259	169	21.51	37	13.19
2013	33861	242	31.71	45	16.00
2014	34879	335	33.51	49	16.01
2015	37600	353	36.24	50	14.57
2016	39246	417	38.42	48	12.53
2017	38332	445	45.29	51	16.28
2018	42432	486	48.19	41	13.47
2019	52361	374	37.99	34	10.93
2020	68759	327	30.05	28	9.95
2021	73284	309	34.59	29	9.63
2022	82672	300	30.01	26	7.75

2.2.2 Emission factor data

The N_2O and CH_4 emission factors of different treatment processes are summarized and summarized by checking relevant literature, as shown in Fig. 3, and the carbon emission factors of pharmaceuticals are shown in Table 2, and the emission factors of China's East China Regional Power Grid for the year 2012-2022 are shown in Fig. 4.

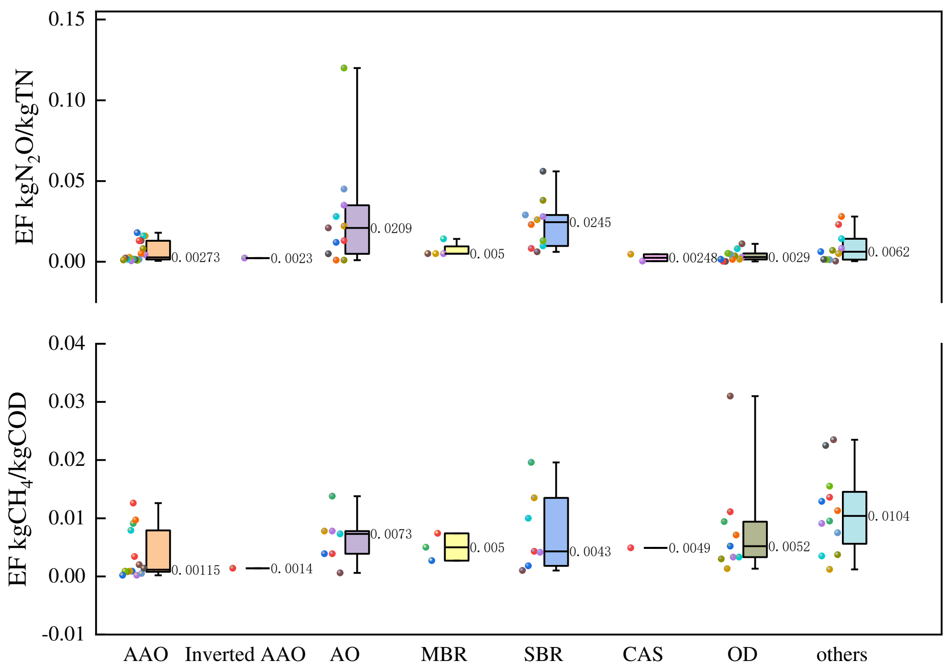


Fig. 3 Carbon emission factors for different wastewater treatment processes

Table 2 Pharmaceutical carbon emission factors

	Emission Factors	Source
PAM	0.851 $kgCO_2/kg$	[24]
Aluminum sulfate	0.5 $kgCO_2/kg$	Technical specification for low-carbon operation evaluation of sewage treatment plant
Sodium hypochlorite	0.92 $kgCO_2/kg$	Technical specification for low-carbon operation evaluation of sewage treatment plant
Sodium acetate	0.623 $kgCO_2/kg$	[25]
Tap water	0.168 $kgCO_2/t$	CPCD, China Products Carbon Footprint Factors Database
Other pharmaceuticals	1.6 $kgCO_2/kg$	[26]

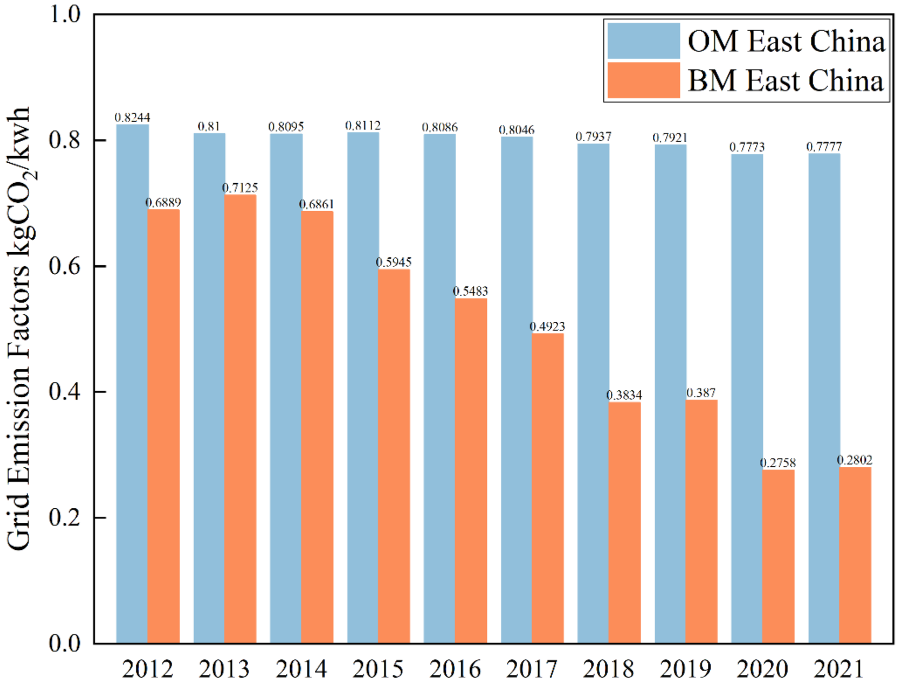


Fig. 4 Carbon emission factor of power grid in East China region

2.3 Combined forecasting model

2.3.1 GWO Grey Wolf Optimization Algorithm

Grey wolf optimization algorithm (GWO) is one of the meta-inspired optimization algorithms, which is based on the hunting behavior of grey wolves in Canidae[27]. As shown in Fig. 5, the gray wolf pack consists of α wolves、β wolves、γ wolves and ω wolves, α wolves dominate decision making about hunting location, sleeping time, etc., β wolves assist α wolves in decision making, β wolves take over the dominant position of α wolves when α wolves are in danger and unable to make decisions, γ wolves and ω wolves are located in the lowest rank of the pack, γ wolves provide services to α wolves and β wolves, ω wolves exist as scapegoats in the pack in order to succumb to the other wolves that are dominant. The GWO optimization algorithm firstly hierarchizes the wolves into a social hierarchy and then stalks and surrounds its prey, the mathematical model of this behavior is represented as:

$$\vec{\mathrm{D}} = \vec{\mathrm{C}} \cdot \overrightarrow{X_p}(t) - \vec{X}(t)$$

$$\vec{X}(t+1) = \overrightarrow{X_p}(t) - \vec{A} \cdot \vec{D}$$

$$\vec{\mathrm{A}} = 2\vec{\mathrm{a}} \cdot \overrightarrow{r_1} - \vec{a}$$

$$\vec{\mathrm{C}} = 2\overrightarrow{r_2}$$

Where t is the current iteration number, $\vec{A}$ and $\vec{\mathrm{C}}$ are the coefficient vectors, $\overrightarrow{X_p}$ denotes the position vector of the prey, $\vec{X}(t)$ denotes the current position vector of the gray wolf, and a decreases linearly from 2 to 0 throughout the iteration process and $\overrightarrow{r_1}$、$\overrightarrow{r_2}$ are random vectors between [0,1].

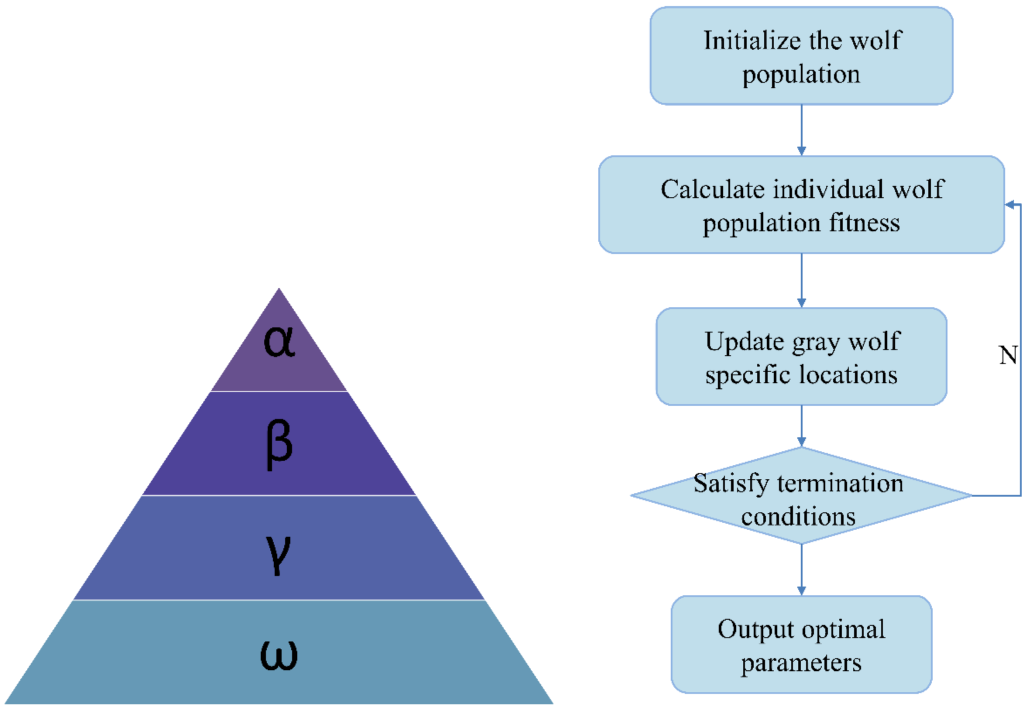


Fig. 5 Top-to-bottom hierarchy of gray wolves(a)Schematic diagram of the gray wolf optimization algorithm(b)

Gray wolves have the ability to identify the location of potential prey (optimal solution), and the search process is mainly accomplished by the guidance of α wolves、β wolves and γ wolves. However, the solution space characteristics of many problems are unknown, and the gray wolf is unable to determine the precise location of the prey (optimal solution). In order to simulate the search behavior of the gray wolf (candidate solution), it is assumed that the α wolves、β wolves、γ wolves has a strong ability to identify the location of the potential prey. Therefore, during each iteration, the best three gray wolves (α wolves、β wolves、γ wolves) in the current population are retained, and then the locations of the other search agents are updated based on their location information. The mathematical model of this behavior can be expressed as follows: $\vec{D}_\alpha = \overrightarrow{C_1} \cdot \overrightarrow{X_\alpha} - \vec{X}, \vec{D}_\beta = \overrightarrow{C_2} \cdot \overrightarrow{X_\beta} - \vec{X}, \vec{D}_\gamma = \overrightarrow{C_3} \cdot \overrightarrow{X_\gamma} - \vec{X}$

$$\overrightarrow{X_1} = \overrightarrow{X_\alpha} - \overrightarrow{A_1} \cdot \vec{D}_\alpha, \overrightarrow{X_2} = \overrightarrow{X_\beta} - \overrightarrow{A_2} \cdot \vec{D}_\beta, \overrightarrow{X_3} = \overrightarrow{X_\gamma} - \overrightarrow{A_3} \cdot \vec{D}_\gamma$$

$$\vec{\mathrm{X}}(\mathrm{t}+1) = \frac{\overrightarrow{X_1} + \overrightarrow{X_2} + \overrightarrow{X_3}}{3}$$

2.3.2 **Support Vector Machine (SVR)**

SVM（Support Vector Machine）was initially used as a classification machine learning algorithm, and was later applied to regression prediction problems. While the purpose of regression is to find the function f(x) between the training input samples and the training output samples, so as to accurately predict the expected output of the test samples, SVR (Support Vector Regression) is mainly designed to solve the nonlinear regression problem, to obtain the global optimum through regression prediction, and to have a good generalization ability to unknown samples[28,29] . Different from the general regression model, the SVR algorithm is set with the acceptable maximum error ε ， w 、 b as the parameters to be determined. If the output of the prediction error value $\varepsilon_0 > \varepsilon$, the loss is calculated. At this time, it is equivalent to constructing a prediction band with a width of 2ε centered on f(x), and if the predicted quantity falls into this prediction circle, it is considered to be predicted correctly.

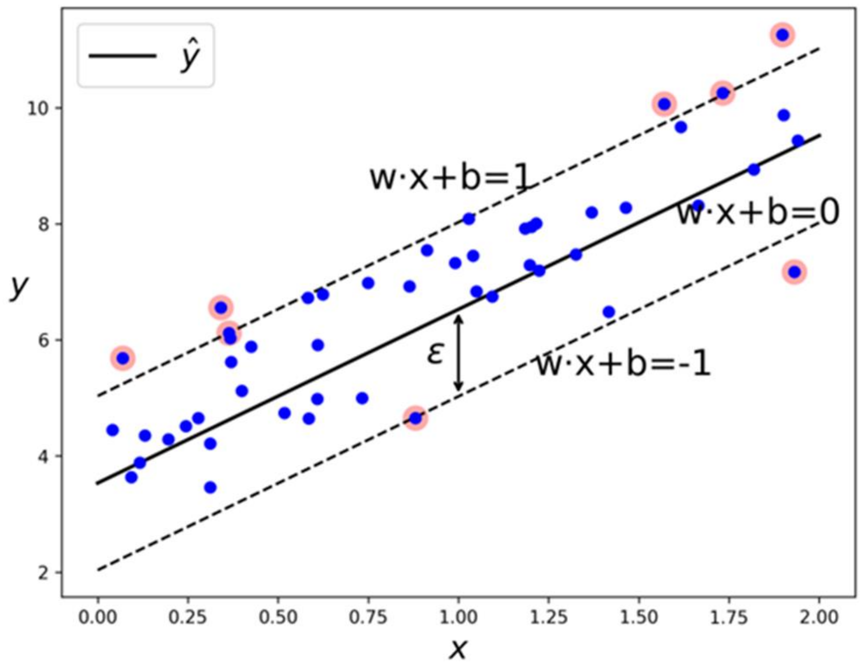


Fig. 6 Schematic diagram of SVR principle

As shown in Fig. 6, $w \cdot x + b = 0$ is the separating hyperplane, and 1 and -1 in $w \cdot x + b = 1$ and $w \cdot x + b = -1$ are the binary labeling values, in order to ensure that $y_i(wx_i + b)$ is always greater than 0. SVR mainly solves the nonlinear regression problem by mapping the nonlinearly mapped data x to the high-dimensional feature space.

Let the known training set be $T = \{(x_1, y_1), (x_2, y_2), ..., (x_l, y_l)\}$, The function is obtained from SVR training regression:

$$f(x) = w^T \phi(x) + b$$

Where w and b are the weight coefficients and constant coefficients, respectively. $\phi(x)$ is the nonlinear mapping function in the feature space. An insensitive loss function ε is introduced, errors smaller than ε will not be included in the loss function. Regularize the risk function by minimizing the following. Introduce the regularization parameter C to optimize the regression problem as:

$$\min \frac{1}{2} \| w \|^2 + C \frac{1}{l} \sum_{i=1}^{l} L_\varepsilon (y_i, f(x_i))$$

Where C is the penalty factor and L is the loss function.

Introducing slack variables ξ_i and ξ_i^* can express the optimization objective as:

$$\min \frac{1}{2} \| w \|^2 + C \sum_{i=1}^{l} (\xi_i + \xi_i^*)$$

$$s.t.\begin{cases} y_i - w\cdot\phi(x_i) - b \le \varepsilon + \xi_i \\ w\cdot\phi(x_i) + b - y_i \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0 \quad (i = 1,2,...,l) \end{cases}$$

The dyadic problem of the issue:

$$\max -\frac{1}{2}\sum_{i,j=1}^{l}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)K(x_i \cdot x_j) - \sum_{i,j=1}^{l}\alpha_i(\varepsilon - y_i) - \sum_{i,j=1}^{l}\alpha_i^*(\varepsilon + y_i)$$

$$s.t.\begin{cases} \sum_{i=1}^{l}(a_i - a_i^*) = 0 \\ 0 \le a_i, a_i^* \le C \qquad (i = 1,2,...,l) \end{cases}$$

By solving the above problem, the decision function for support vector regression can be obtained:

$$f(x) = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)K(x_i, x) + b$$

Where $K(x_i, x_j) = \phi(x_i)\phi(x_j)$ denotes the kernel function, it is crucial to choose a suitable kernel function in the prediction model, and the most commonly used is the RBF (Radial Basis Function) kernel function.

2.3.3 Artificial neural network (ANN)

ANN model consists of three parts: input, hidden and output layers, Neural Network (NN) processes the input data and indicates the significance relationship between the input variables and the output mechanism. It can learn from the original dataset and test the generated network with a new dataset, further correct and strengthen the internal logical relationships for the training model, and finally validate the model with a validation dataset.

Compared with traditional prediction models, ANN has superior learning ability in multi-input data processing, and can quickly sort out the hidden internal logic of the input and output layers, and this internal structure enables it to effectively manage and interpret the original time series data. Meanwhile, ANN has superior ability in mining the nonlinear features of total carbon emissions from sewage treatment plants, and this ability significantly improves the accuracy of carbon emission prediction.

2.3.4 Combined prediction model

Different prediction models have their strengths and weaknesses, and to maximize the strengths and avoid the weaknesses of the combination model is introduced. The combined prediction model integrates the prediction results of multiple models according to different weights, and the combined new model combines the prediction advantages of the sub-models to effectively improve the prediction performance of the model. The framework of the method is shown in Fig. 7.

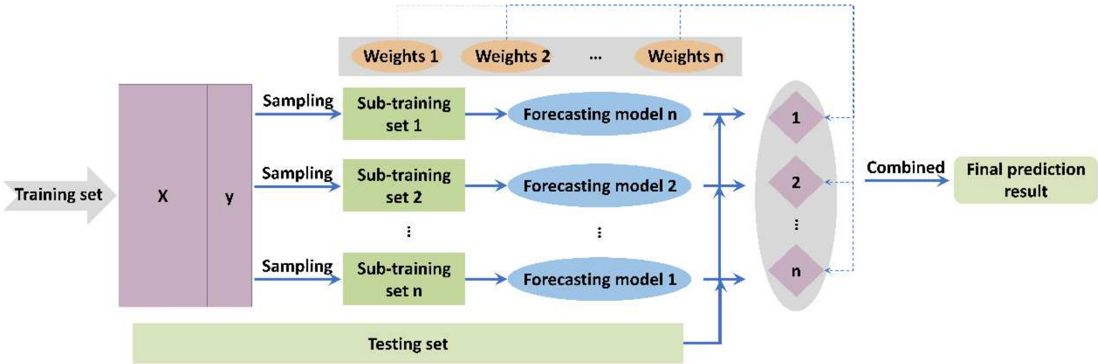


Fig. 7 Combined modelling framework

The calculation method of the combined model is as follows: for a given model j with input X_i, the output Y_{ij} is obtained. Each model output value is assigned a weight α_j. The predicted value of the combined model is the combination of the output results Y_{ij} and their corresponding weights α_j, represented as W_i. This process determines the optimal weight for the model evaluation indices, resulting in the final combined model.

$$W_i = \sum_{j=1}^{k} \alpha_j \times Y_{ij}\,, i = 1,2,\cdots,k$$

$$\sum_{j=1}^{k} \alpha_j = 1$$

$$\mathrm{MinRMSE} = \sqrt{\frac{1}{n}\sum\nolimits_{i=1}^{n}(W_i - y_i)^2}$$

2.3.5 Model evaluation indicators

Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE) and Goodness of Fit (R^2) were chosen as the indicators for judging the final prediction results, and the formulas were calculated as follows:

$$\mathrm{MAE} = \frac{1}{n}\sum\nolimits_{i=1}^{n}|y_i - y_{pre}|$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum\nolimits_{i=1}^{n}\left(y_i - y_{pre}\right)^2}$$

$$\mathrm{MAPE} = \frac{100\%}{n}\sum\nolimits_{i=1}^{n}\left|\frac{y_i - y_{pre}}{y_i}\right|$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - y_{pre}\right)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Where y_i is the actual water quality monitoring value, y_{pre} is the simulated prediction value, $\bar{y}$ is the average value of measured water quality, n is the number of samples. The value range of MAE, RMSE, MAPE is [0,+∞], and the value range of [0,1]. A smaller value for MAE, RMSE, and MAPE, as well as a value closer to 1 for R^2, indicate better prediction performance[30].

3 Results and discussion

3.1 Carbon accounting for sewage treatment

Fig. 8 illustrates the total carbon emissions of the WWTP from 2012 to 2022, computed using the refined carbon accounting method and carbon emission factor data. The overall trend reveals a continuous increase in total carbon emissions over the years, with indirect carbon emissions emerging as the primary contributor to the WWTP's total carbon footprint. Particularly noteworthy is the significant contribution of carbon emissions from electricity consumption. In contrast to the conventional AAO process, the inverted AAO process reverses the anaerobic and anoxic positions. This allows the microorganisms to consume a significant quantity of the primary carbon source in the anaerobic zone. When the nitrifying solution in the aerobic zone flows back to the anoxic zone, the lack of carbon source in the traditional AAO process results in insufficient denitrification and the release of more N_2O[31]. This results in the lowest N_2O and CH_4 emission factors of the inverted AAO process. Furthermore, the direct carbon emissions from this plant will only account for 12% of the total carbon emissions in 2022. Concurrently, the influent of this wastewater treatment plant is a mixture of domestic and industrial wastewater, which contains a high concentration of nitrogen (N), phosphorus (P), and industrial carbon. This results in a higher proportion of N_2O and fossil carbon emissions than CH_4.

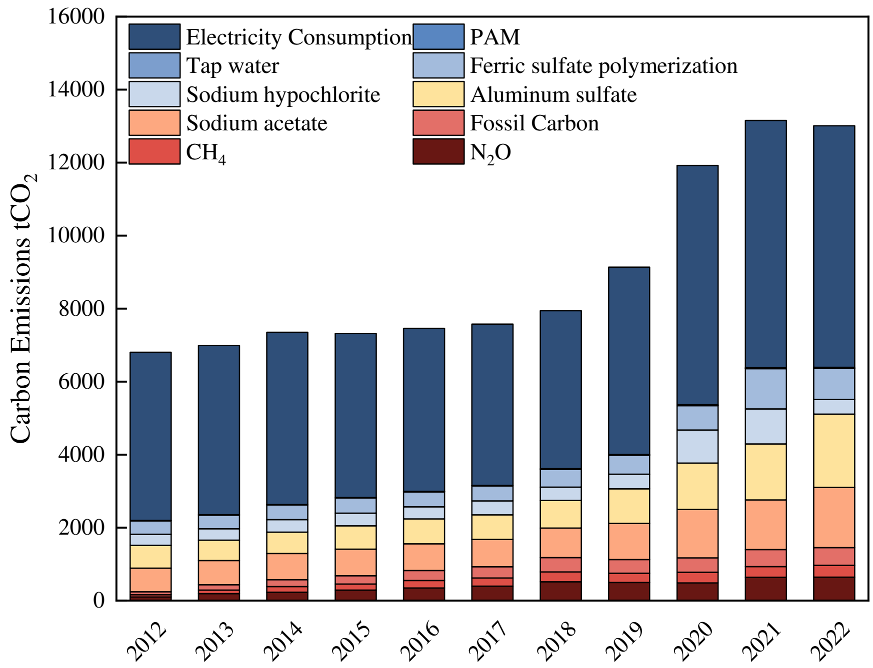


Fig. 8 Total carbon emissions from 2012 to 2022

3.2Analysis of influencing factors

In this study, the presence of multicollinearity among the influencing factors in relation to the available data might lead to feature redundancy or model overfitting. Therefore, the variance inflation factor (VIF) is introduced to assess multicollinearity. Table 3 indicates significant multicollinearity among most influencing factors. Such multicollinearity not only distorts regression model results but also compromises model stability. To address this issue, the Pearson correlation coefficient is employed to evaluate the correlation between parameters and extract model eigenvalues further. eigenvalues further.

The Pearson correlation coefficient, based on covariance analysis, effectively examines feature correlations without directly influencing model analysis[32] . The correlation analysis heatmap depicted in Fig. 9 reveals significant correlations between total carbon emissions and treated water volume, effluent COD concentration, effluent TN concentration, and various types of drug consumption. To enhance model prediction performance and efficiency while mitigating the impact of multiple covariances among features, the model prediction eigenvalues are selected as treated water volume, electricity consumption, effluent COD concentration, effluent TN concentration, and total drug consumption.

Table 3 VIF test results

Influencing factors	significance	covariance statistics	
		tolerance	VIF
CODin	0.8001	0.0156	63.966
TNin	0.7982	0.0102	97.737

CODeff	0.3054	0.0406	24.612
TNeff	0.4977	0.0131	76.436
Aluminium sulfate	0.2075	0.0006	1682.481
Sodium hypochlorite	0.2888	0.0056	178.754
Polymerised iron sulphate	0.3872	0.0021	478.648
Tap water	0.3156	0.0543	18.403
Electricity consumption	0.8669	0.0005	1983.018
Pharmaceutical consumption	0.1968	0.0001	6833.279
Q	-	0.0000	7953.177
Sodium acetate	-	0.0000	4248.949
PAM	-	0.0000	1063973.3

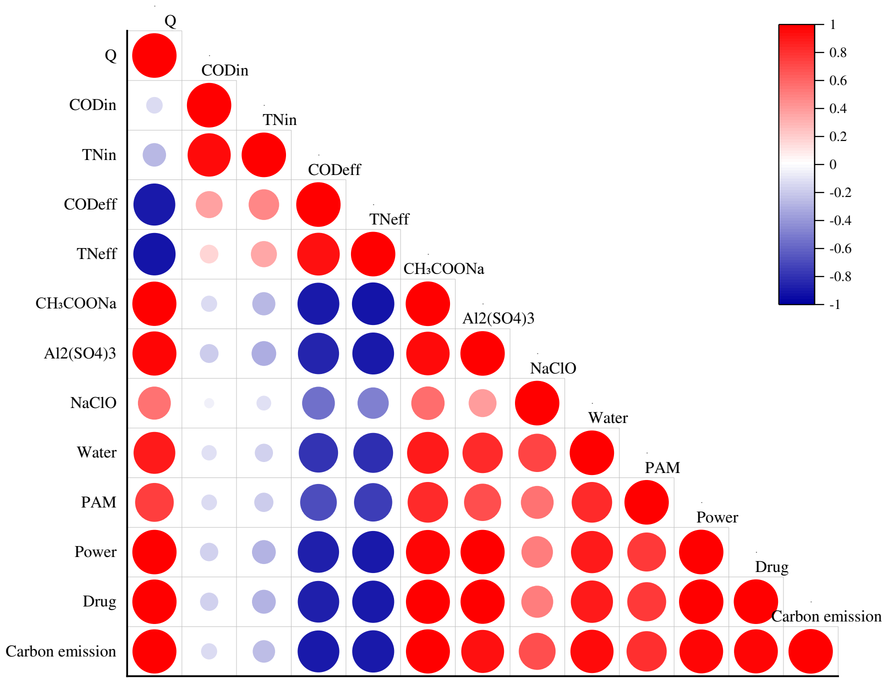


Fig. 9 Heat map for correlation analysis

3.3Model construction and evaluation

Normalizing the data before model construction mitigates the impact of varying data scales on model performance and enhances its accuracy and robustness. Common normalization techniques include Min-Max normalization, Z-Score normalization, decimal scaling normalization, sigmoid normalization, and Robust-Scaler normalization. In this study, Min-Max normalization was employed, with the calculation being performed in accordance with the following formula:

$$B' = \frac{B - B_{Min}}{B_{Max} - B_{Min}}$$

Where B' is the normalised data, B is the original data, B_{Min} is the minimum value of the original data column and B_{Max} is the maximum value of the original data column.

The data is partitioned into a training set comprising the initial 80% and a test set consisting of the remaining 20%. In order to enhance the prediction accuracy of the Support Vector Regression (SVR) model, the Grey Wolf Optimization algorithm (GWO) is employed for parameter optimization, including the selection of the optimal kernel function. Through a systematic comparison of model scores across various parameter combinations, the optimal parameter configuration is determined. Notably, the linear kernel function emerges as the most effective choice, with optimal parameter values identified as C = 70816 and gamma = 48739. On the other hand, the Artificial Neural Network (ANN) model is structured with a two-layer architecture. The first layer incorporates a hidden layer comprising 16 neurons, while the second layer features five hidden layers, each housing 32 neurons. This configuration has been selected in order to facilitate robust learning and prediction within the ANN framework.

To enhance the overall predictive accuracy of the model, a hybrid approach combining the GWO-SVR and ANN models was employed, with weights assigned to each model to optimize performance. The optimal weight ratio was determined by minimizing the Root Mean Square Error (RMSE), thereby leveraging the predictive strengths of both sub-models to improve overall performance. The calculated optimal weight distribution assigns a weight of 0.404 to the ANN model and 0.596 to the SVR model, achieving the highest predictive performance.

To evaluate the effectiveness of the combined model, several comparative models were introduced, and their prediction results are summarized in Table 4. Notably, the combined model yields a remarkable R-squared (R2) value of 0.9926, signifying a substantial improvement in prediction accuracy compared to standalone SVR and ANN models. Additionally, error metrics including RMSE, MAE, and MAPE demonstrate significant reductions, underscoring the superior performance of the combined model.

The GWO algorithm, functioning as a heuristic optimization technique, plays a pivotal role in aiding both SVR and ANN models to attain more optimal parameter configurations during training. Furthermore, compared to individual models, the combined approach exhibits enhanced robustness, mitigating the risk of overfitting to some extent.

Although Gradient Boosting Decision Trees (GBDT) and Decision Tree Regression (DTR) outperform SVR and ANN models within the training dataset, their susceptibility to overfitting poses a limitation due to the constrained number of training samples. Moreover, these tree-based models are unable to simultaneously capture both linear and nonlinear relationships within the data, thus imposing constraints on their predictive capabilities.

Table 4 Values of model assessment indicators

	Model	SVR	ANN	GBDT	AdaBoost	DTR	GWO-SVR-ANN
test	R2	0.9590	0.9868	0.9779	0.9038	0.9027	0.9926
	RMSE	483.2494	274.3218	354.8518	739.8989	744.2982	204.8652

	MAE	383.4922	223.4162	293.9377	601.1521	572.8457	158.6957
	MAPE	1.19%	0.70%	0.91%	1.87%	1.76%	0.49%
train	R2	0.9577	0.9818	0.9997	0.9235	1.0000	0.9934
	RMSE	518.3161	339.9446	43.9357	697.3520	0.0000	204.9784
	MAE	402.7506	237.4547	33.6327	601.1239	0.0000	154.4703
	MAPE	1.23%	0.73%	0.10%	1.85%	0.0000	0.47%

3.4 Scenario prediction

3.4.1 Future rate of change settings for different indicators

The parameters defining three development scenarios (low, medium, and high speed) for the seven development indicators underwent adjustments based on historical data from the wastewater treatment plant, coupled with the "Outline of the Fourteenth Five-Year Plan for the Development of the National Economy and Society of Ningbo Municipality and the Visionary Targets for the Year 2035", alongside the " Synergetic Control of Environmental Pollution and Carbon Emissions Implementation Program ".

The medium-speed development scenario mirrors the average rate of change observed over the preceding three years. Conversely, the high-speed and low-speed development scenarios were formulated and fine-tuned in alignment with the national green low-carbon policy and the objectives outlined in Ningbo's ecological planning and construction initiatives. Detailed scenario change rates are delineated in Table 5.

Table 5 Rate of change of indicators for different scenarios (%)

Indicator name	Scenario	2023-2025	2026-2030	2031-2035	2035-2040
Amount of treated water (A)	L	4.8	3.6	2.4	1.2
	M	5.3	4.1	2.9	1.7
	H	6	4.8	3.6	2.4
Pollutant concentration (B)	L	-2.6	-2.4	-2.2	-2
	M	-3	-2.8	-2.6	-2.4
	H	-3.4	-3.2	-3	-2.8
Energy consumption (C)	L	3.8	2.8	1.9	1.3
	M	4.2	3.2	2.3	1.7
	H	4.6	3.6	2.7	2.1
drug consumption (D)	L	4.8	3.6	2.4	1.2
	M	5.3	4.1	2.9	1.7
	H	6	4.8	3.6	2.4
clean energy substitution rate (E)	L	4	6	8	10
	M	5	8	11	14
	H	8	14	20	26
Methane recovery power generation rate (F)	L	4	6	8	10
	M	5	7	10	12
	H	6	10	14	18
Biological carbon sequestration rate	L	4	6	8	10
	M	6	10	14	18

(G)	H	8	14	20	26

3.3.2 Scenario design

To project the carbon emissions of the wastewater treatment plant (WWTP) over the next 20 years, this study established six distinct scenarios, outlined in Table 6. These scenarios aim to assess the carbon emissions and abatement potential of the WWTP under various developmental contexts, to evaluate the impact of different abatement strategies on total carbon emissions. By integrating considerations such as clean energy substitution, biological carbon sequestration, methane recovery for power generation, and other abatement techniques to fulfil the sustainable development objectives of WWTPs, this research offers valuable insights for other WWTPs in Ningbo. These insights aid in formulating carbon abatement measures and sustainable development strategies.

1) Baseline Scenario: The Baseline Scenario functions as a control group for the other scenarios. It assumes the continuation of the existing operational mode without alterations, with annual rates of change for different indicators based on reasonable extrapolations of historical data, all set to a medium level. Additionally, the projected outcomes of this scenario reflect the anticipated trend of carbon emissions from the wastewater treatment plant in the future.

2) No Emission Reduction Measures Scenario: The No Emission Reduction Measures Scenario functions as a control group for comparing other emission reduction scenarios. It maintains a medium level for the rate of change of treated water volume, pollutant concentration, energy consumption, and pharmaceutical consumption. Moreover, it does not include clean energy substitution, biological carbon sequestration, or methane recovery for power generation. This scenario assumes that the wastewater treatment plant retains its original treatment mode without implementing any emission reduction measures.

3) Technological Innovation Scenario: The scenario sets energy and pharmaceutical consumption at a low growth rate, energy substitution rate and biological carbon sequestration rate at a high growth rate, and other indicators at a medium growth rate. Through the introduction of new management policies and technological paths to reduce energy and drug consumption in the treatment process, keep the growth rate of energy and drug consumption low in order to achieve the efficient use of resources, and at the same time increase the rate of production and proportion of clean energy in the plant, adjust the energy structure of the plant, and reduce the direct carbon emissions of the plant by combining with the advanced biological carbon sequestration technology, so as to promote the simultaneous development of efficient governance and low-carbon construction.

4) Energy Saving and Consumption Reduction Scenario: Energy conservation and consumption reduction serve as proactive measures for source control, fostering synergies between pollution mitigation and carbon reduction. This is achieved through enhancing resource and energy conservation and efficiency, and expediting the transition towards industrial structures, production methods, and lifestyles conducive

to both pollution and carbon reduction. The scenario prioritizes water and electricity conservation, aiming to maintain a low growth rate in treated water volume and energy consumption. However, water resource reuse may elevate pollutant concentrations in water bodies. Hence, the rate of pollutant concentration reduction is set low, while other indicators experience moderate growth rates.

5) Sustainable Development Scenario: The sustainable development scenario is defined as a coordinated development of the economy, society and the environment. In this scenario, the wastewater treatment plant introduces new treatment technologies while achieving energy savings and consumption reduction. Accordingly, the proportion of clean energy substitution, the rate of methane recovery for power generation and the rate of biological carbon sequestration are set to a high-growth level in order to reduce the emission of pollutants to the maximum extent possible.

6) Low Development Scenario: The Low Development Scenario is marked by the plant's conservative approach to future planning and the delayed updating of treatment equipment and technology. Consequently, there is a pronounced upward trend in energy and chemical consumption. Moreover, inadequate capital investment has resulted in a minimal increase in the proportion of clean energy substitution, methane recovery power generation rate, and biological carbon sequestration rate.

Table 6 Scenario settings

Scenario	A	B	C	D	E	F	G
Baseline Scenario	M	M	M	M	M	M	M
No Emission Reduction Measures Scenario	M	M	M	M	-	-	-
Technological Innovation Scenario	M	M	L	L	H	M	H
Energy Saving and Consumption Reduction Scenario	L	L	L	M	M	M	M
Sustainable Development Scenario	L	H	L	L	H	H	H
Low Development Scenario	H	H	H	H	L	L	L

3.3.3 Scenario prediction analysis

The results depicted in Fig. 11 illustrate the scenario analysis projections. Findings indicate that under the no-abatement-measures scenario and low-development scenario, the WWTP will not achieve carbon peak before 2040. Conversely, the baseline scenario, energy-saving scenario, technological innovation scenario, and sustainable development scenario forecast carbon peak occurrences in 2039, 2036, 2035, and 2034, respectively. Their respective peaks are projected at 19,138 t, 18,073 t, 17,484 t, and 16,819 t, approximately 1.37 times that of 2022. Notably, the no abatement measures scenario and low development scenario lack sustainable development potential. Furthermore, neither the baseline scenario nor the energy saving and consumption reduction scenario can attain carbon peak by 2035. To meet the carbon peak target for water pollution prevention and control by 2030, it's imperative to bolster the proportion of clean energy substitution, methane recycling for power generation, and the rate of biological carbon sequestration as per the technological innovation scenario and sustainable development scenario. Additionally, efforts should be intensified to fortify prevention and control measures at the source, restrain high-pollution and high-emission industries and projects, promote water-saving and low-carbon lifestyles, and prioritize the adoption of advanced, low-energy-consuming process technologies and equipment.

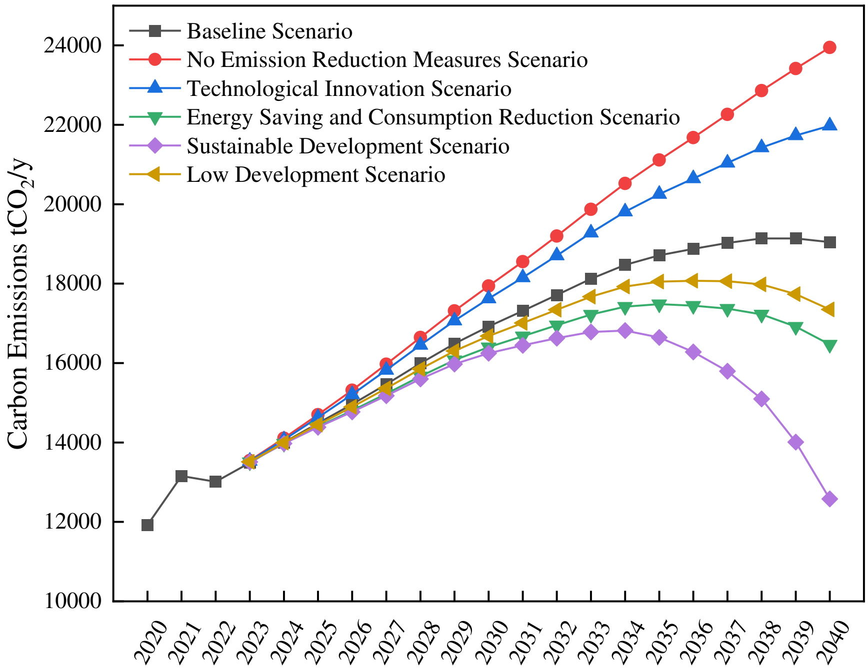


Fig. 10 Scenario analysis prediction results

3.5 Carbon reduction pathway analysis

1) energy conservation and consumption reduction

Carbon emissions from pharmaceutical consumption and electricity consumption are the main sources of total carbon emissions from wastewater treatment plants, with pharmaceutical consumption accounting for 35.4% of the total carbon emissions from wastewater treatment plants, and the total carbon emissions from pharmaceutical con sumption during a half-year period can be up to 2811 tCO_2eq. Sodium acetate and oth er additional carbon sources can be used to reduce the dosage by means of carbon cyc ling and carbon recycling, thus improving the sustainability of wastewater treatment p lants. Additional carbon sources, disinfectants, flocculants, and other pharmaceuticals should be purchased with priority given to those that meet sustainability criteria to red uce indirect greenhouse gas emissions, and recycling to reduce the carbon emissions a ssociated with the pharmaceutical production process. The indirect carbon emissions f rom the power consumption of this wastewater treatment plant accounted for 51% of t he total carbon emissions, and the average power consumption reached 0.36 kWh/m^3. At present, the average power consumption of China's urban wastewater treatment pla nts is 0.29 kWh/m^3 and the power consumption of more than 82% of the wastewater tr eatment plants is not less than 0.440 kWh/m^3 , so there is still a great potential for ene rgy saving in this wastewater treatment plant. Electricity consumption in municipal w astewater treatment plants mainly occurs in the three parts of the sewage lifting syste m, the oxygen supply system of the secondary biochemical treatment and the sludge tr eatment system, and the reduction of electricity consumption can be achieved from th e perspectives of both power saving and clean energy substitution. It has been found t hat electricity consumption can be reduced by 7-9% by adjusting the sludge age and d issolved oxygen concentration in water, while the use of energy-saving equipment can further reduce energy consumption [33]. Some emerging reaction devices such as aer ated membrane bioreactors, which can significantly increase the oxygen transfer rate, and precision aeration systems can reduce the energy consumption of the aeration pro cess. Currently most of the electricity used is coal-fired power generation, increasing t he use of renewable energy such as wind power, nuclear power, photovoltaic power ca n also reduce the indirect carbon emissions of sewage treatment plants, can be in the s and sedimentation tanks, secondary sedimentation tanks and unused land in the plant t o increase the proportion of photovoltaic power to reduce the total carbon emissions o f the whole plant.

2) Energy recycling

There is a large space for energy recovery in the wastewater treatment process, and the use of water source heat pumps to recover part of the secondary wastewater heat energy can reduce the energy consumption of the plant, thus reducing the total carbon emissions of the plant, and the theoretical carbon emission reduction formula is as follows:

$$m_{hp-R} = [M_{hp} \times |T_o - T_i| \times c \times (fe \times 2.778 \times 10^{-4} - EF_{hp})] \times (1 \pm \frac{1}{COP \mp 1})$$

m_{hp-R} is the carbon emission reduction of heat recovery, $kgCO_2/kg$; M_{hp} is the mass of secondary effluent involved in heat exchange, kg; T_o is the water source heat pump discharge temperature, ℃; T_i is the water source heat pump inlet temperature, ℃; c is the specific heat capacity of water, 4.18KJ/(kg · °C); fe is the electric power emission factor, $kgCO_2/kwh$; 2.778×10^{-4} is the conversion coefficient of kilowatt hour and kilojoule, kwh/KJ; EF_{hp} is the water source heat pump emission factor, $kgCO_2/KJ$, which is generally taken 6.208×10^{-4}; COP is the ratio between the heating or cooling capacity of the water source heat pump and the consumed electric power Ratio.

There are four main sludge treatment technologies: mechanical dewatering, thermal drying, anaerobic digestion, aerobic fermentation[34] , a large amount of biogas will be produced in the process of anaerobic digestion of sludge, and upgrading the discharged biogas to bio-methane, and reducing the energy consumption of the plant through biogas cogeneration can also reduce the total carbon emissions of the plant to a certain degree, and the carbon emission reduction formula is as follows:

$$m_{m-R} = M_m \times \alpha \times (k \times fe - \frac{44}{16})$$

m_{m-R} is the carbon emission reduction of biogas cogeneration, $kgCO_2/m^3$; M_m

is the amount of biogas produced, m^3; α is the methane share of the biogas; k is the amount of electricity generated per unit volume of methane, kwh/m^3; fe is the electricity emission factor, $kgCO_2/kwh$; 44/16 is the molecular mass ratio of carbon dioxide to methane.

For the sludge after mechanical dewatering and thermal drying treatment, sludge incineration can be used to generate electricity to supply energy for the plant, with the following carbon reduction formula:

$$m_{i-R} = M_i \times a \times fe - M_i \times (EF_1 \times 28 + EF_2 \times 265 + dm \times CF \times FCF \times OF \times \frac{44}{12})$$

m_{i-R} is the mass of sludge for incineration and power generation, kg; a is the power generation capacity of incineration unit sludge, kwh/kg; fe is the power emission factor, $kgCO_2/kwh$; EF_1 is the sludge N_2O emission factor, kgN_2O/kg; EF_2 is the sludge CH_4 emission factor, kg CH_4/kg; dm is the content of dry matter in the sludge; CF is the proportion of carbon in the dry matter, which is 30% in the IPCC; FCF is the proportion of mineral charcoal in carbon, which is generally between 14.3%-16.3%; OF is the oxidation factor, generally taken as 100%; 44/12 is the molecular mass ratio of CO_2 and C.

3) Optimization of the treatment process

Efficient wastewater treatment process can improve the effluent quality and reduce carbon emission at the same time, for AAO wastewater treatment process, low DO concentration promotes the production of N_2O and high DO concentration inhibits the production of N_2O, meanwhile, the increase of aeration will increase the production of N_2O and vaporization, in addition, the low temperature will reduce the enzyme activity, which affects the nitrification and denitrification reactions, so maintaining complete nitrification without inhibiting the complete heterotrophic denitrification process by controlling DO and temperature at a reasonable level and controlling a certain amount of aeration is an effective way to reduce N_2O production[35] . For direct emission of the greenhouse gas methane, co-digestion techniques can be used to increase biogas and methane production and reduce the operating costs and energy requirements of the plant [36].

4) Enhancing carbon sinks

Biomass sequestration and energy recovery from biogas conversion are the two main strategies for carbon sequestration and sinks[37]. in addition to the use of microalgae-bacterial microbiomes to capture carbon dioxide during biodegradation of organic compounds to reduce direct greenhouse gas emissions[38], and the use of technologies such as microbial electrochemical technology (MET), photocatalytic, and electrochemistry to convert carbon dioxide into utilizable chemicals, such as formic acid, methanol, and methane, to enhance carbon sinks[39].

4 Conclusion

In conclusion, this study proposes a prediction framework that applies to carbon emissions from wastewater treatment plants with a high degree of prediction accuracy. The prediction framework initially located certain parameter values or coefficients provided in the IPCC and then combines them with a literature analysis to collate and update the carbon emission factors. It replaces the previous international data with the baseline emission factors for regional power grids issued by the Ministry of Ecology and Environment of China in different years and updates the GWP values of CH4 and N2O to the latest public data of IPCC 2019. Concurrently, the model incorporates fossil carbon into the accounting process, thereby resolving the issue of underestimation of the proportion of fossil carbon emissions in the traditional emission factor accounting method. Subsequently, the SVR and ANN prediction models were developed by combining covariance analysis and correlation analysis to extract features, and the GWO algorithm was used to further optimize the model parameters to improve the model's prediction performance. Finally, by minimizing the root mean square error (RMSE) and combining the two models according to the optimal weights, the coefficient of determination (R^2) of the model reaches 0.9926, and the mean absolute percentage error (MAPE) reaches 0.49%, which demonstrates superior prediction performance. This combined model effectively integrates the strengths of traditional machine learning and deep learning, capitalizing on the respective advantages of both. The introduction of nonlinear feature extraction capability, derived from deep learning, in conjunction with the generalization and interpretability of machine learning, enables the model to more accurately capture the complex nonlinear relationships between the data, thereby improving the prediction accuracy and robustness of the model.

Concurrently, this study employs scenario analysis, integrating the composite prediction model with pertinent policy frameworks to assess the carbon emissions of the wastewater treatment plant and propose targeted emission reduction strategies. Findings indicate that no emission reduction measures scenario and low development scenario lack sustainable potential, while baseline scenario and energy saving and Consumption Reduction Scenario fall short of achieving carbon neutrality by 2035. realizing the 2030 target for carbon neutrality in water pollution prevention and control necessitates augmenting the share of clean energy substitution, methane-recycling power generation, and biological carbon sequestration rates, as outlined in the technological innovation scenario and sustainable development scenario. Moreover, robust source prevention and control measures are imperative, alongside efforts to curtail high-pollution industries, promote water-saving and low-carbon lifestyles, and prioritize the adoption of advanced, low-energy process technologies and equipment. This study lays groundwork for optimizing wastewater treatment processes towards "zero carbon" and serves as a blueprint for effective carbon management in wastewater treatment plants, facilitating the journey towards carbon neutrality.

Acknowledgments

We gratefully acknowledge supports by the Baima Lake Laboratory Joint Funds of the Zhejiang Provincial Natural Science Foundation of China under Grant No. LBMHY24E060014.

Declaration of interest statement

We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of.

References

[1] Pahunang RR, Buonerba A, Senatore V, et al. Advances in technological control of greenhouse gas emissions from wastewater in the context of circular economy. Science of the total environment. 2021 Oct 20;792:148479.

[2] Zang Y, Li Y, Wang C, et al. Towards more accurate life cycle assessment of biological wastewater treatment plants: a review. Journal of Cleaner Production. 2015;107:676-692.

[3] Adibimanesh B, Polesek-Karczewska S, Bagherzadeh F, et al. Energy consumption optimization in wastewater treatment plants: Machine learning for monitoring incineration of sewage sludge. Sustainable Energy Technologies and Assessments. 2023 Mar;56:103040.

[4] Yang F, Xiong X. Carbon emissions, wastewater treatment and aquatic ecosystems. Science of The Total Environment. 2024 Apr 15;921:171138.

[5] Wei R, Hu Y, Yu K, et al. Assessing the determinants of scale effects on carbon efficiency in China's wastewater treatment plants using causal machine learning. Resources, Conservation and Recycling. 2024 Apr;203:107432.

[6] Jadhav AR, Pathak PD, Raut RY. Water and wastewater quality prediction: current trends and challenges in the implementation of artificial neural network. Environmental Monitoring and Assessment. 2023 Feb;195(2):321.

[7] Azeez OS, Pradhan B, Shafri HZ. Vehicular CO emission prediction using support vector regression model and GIS. Sustainability. 2018 Oct;10(10):3434.

[8] Zhu C, Wang M, Du W. Prediction on peak values of carbon dioxide emissions from the Chinese transportation industry based on the SVR model and scenario analysis. Journal of Advanced Transportation. 2020 Oct 23;2020:1-14.

[9] Chu X, Zhao R. A building carbon emission prediction model by PSO-SVR method under multi-criteria evaluation. Journal of Intelligent & Fuzzy Systems. 2021;41(6):7473-7484.

[10] Safa M, Nejat M, Nuthall P, et al. Predicting CO_2 emissions from farm inputs in wheat production using artificial neural networks and linear regression models-Case study in Canterbury, New Zealand. International Journal of Advanced Computer Science and Applications. 2016:268-274.

[11] Vasilaki V, Conca V, Frison N, et al. A knowledge discovery framework to predict the N2O emissions in the wastewater sector. Water Research. 2020 Jul 1;178:115799.

[12] Alali Y, Harrou F, Sun Y. Unlocking the potential of wastewater treatment: Machine learning

based energy consumption prediction. Water. 2023 Jul;15(13):2349.
[13] Jeppsson U. Modelling aspects of wastewater treatment processes. 1996.
[14] Xi J, Gong H, Zhang Y, et al. The evaluation of GHG emissions from Shanghai municipal wastewater treatment plants based on IPCC and operational data integrated methods (ODIM). Science of the Total Environment. 2021 Nov 25;797:148967.
[15] Zhao X, Jin X, Guo W, et al. China's urban methane emissions from municipal wastewater treatment plant. Earth's Future. 2019 Apr;7(4):480-490.
[16] Wang D, Ye W, Wu G, et al. Greenhouse gas emissions from municipal wastewater treatment facilities in China from 2006 to 2019. Scientific data. 2022 Jun 16;9(1):317.
[17] Han J, Liu Y, Li W, et al. Modeling greenhouse gas emissions from biological wastewater treatment process with experimental verification: A case study of paper mill. Science of The Total Environment. 2024:171637.
[18] Yang W-B, Yuan C-S, Chen W-H, et al. Diurnal variation of greenhouse gas emission from petrochemical wastewater treatment processes using in-situ continuous monitoring system and the associated effect on emission factor estimation. Aerosol and Air Quality Research. 2017 Oct;17(10):2608-2623.
[19] Hua H, Jiang S, Yuan Z, et al. Advancing greenhouse gas emission factors for municipal wastewater treatment plants in China. Environmental Pollution. 2022 Feb 15;295:118648.
[20] IPCC. 2006 IPCC guidelines for national greenhouse gas inventories, prepared by the national greenhouse gas inventories program. Vol. 5. IGES, Japan. Volume 5, Chapter 6; 2006. (Eggleston H, Buendia L, Miwa K, et al., editors.).
[21] IPCC. 2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories. IPCC, Switzerland. Volume 5, Chapter 6.; 2019. (Calvo Buendia E, Tanabe, K., Kranjc, A., Baasansuren, J., Fukuda, M., Ngarize, S., Osako, A., Pyrozhenko, Y., Shermanau, P. and Federici, S., editor.).
[22] Boiocchi R, Viotti P, Lancione D, et al. A study on the carbon footprint contributions from a large wastewater treatment plant. Energy Reports. 2023 Sep;9:274-286.
[23] Law Y, Jacobsen GE, Smith AM, et al. Fossil organic carbon in wastewater and its fate in treatment plants. Water research. 2013 Sep 15;47(14):5270-5281.
[24] Zhang J, Zhou L, Wu X, et al. Application of blockchain technology in carbon footprint tracing and accounting of textile and apparel products. Journal of Silk. 2023;60(2):14-23.
[25] Sun Q, Chen Y. Characteristics of carbon emission from municipal wastewater treatment plants in a south-China province. Chinese Journal of Environmental Engineering. 2023;17(10):3231-3244.
[26] Parravicini V, Svardal K, Krampe J. Greenhouse gas emissions from wastewater treatment plants. Energy Procedia. 2016;97:246-253.
[27] Aboobakar A, Cartmell E, Stephenson T, et al. Nitrous oxide emissions and dissolved oxygen profiling in a full-scale nitrifying activated sludge treatment plant. Water research. 2013 Feb 1;47(2):524-534.
[28] Ceylan Z, Bulkan S, Elevli S. Prediction of medical waste generation using SVR, GM (1, 1) and ARIMA models: a case study for megacity Istanbul. Journal of Environmental Health Science and Engineering. 2020 Dec;18(2):687-697.
[29] Chen X, Xu J, Sun R, editors. Improved GM-SVR combined prediction model of pavement skid resistance condition based on finite data. Journal of Physics: Conference Series; 2021:

IOP Publishing.
[30] Zhang D. A coefficient of determination for generalized linear models. The American Statistician. 2017;71(4):310-316.
[31] Yin Y, Hui Y, Xu G, et al. Enhanced nutrients removal resulting from energy metabolism improvement in the anoxic-anaerobic-oxic process (reversed AAO). Desalination and Water Treatment. 2021 Mar;217:286-293.
[32] Mu Y, Liu X, Wang L. A Pearson's correlation coefficient based decision tree and its parallel implementation. Information Sciences. 2018 Apr;435:40-58.
[33] Daskiran F, Gulhan H, Guven H, et al. Comparative evaluation of different operation scenarios for a full-scale wastewater treatment plant: Modeling coupled with life cycle assessment. Journal of Cleaner Production. 2022 Mar 20;341:130864.
[34] Zhen G, Lu X, Kato H, et al. Overview of pretreatment strategies for enhancing sewage sludge disintegration and subsequent anaerobic digestion: Current advances, full-scale application and future perspectives. Renewable and Sustainable Energy Reviews. 2017 Mar;69:559-577.
[35] Lu H, Wang H, Wu Q, et al. Automatic control and optimal operation for greenhouse gas mitigation in sustainable wastewater treatment plants: A review. Science of the Total Environment. 2023;855:158849.
[36] Lima D, Appleby G, Li L. A Scoping review of options for increasing biogas production from sewage sludge: Challenges and opportunities for enhancing energy self-sufficiency in wastewater treatment plants. Energies. 2023 Mar;16(5):2369.
[37] Rosso D, Stenstrom MK. The carbon-sequestration potential of municipal wastewater treatment. Chemosphere. 2008 Feb;70(8):1468-1475.
[38] Zhu X, Lei C, Qi J, et al. The role of microbiome in carbon sequestration and environment security during wastewater treatment. Science of The Total Environment. 2022 Sep 1;837:155793.
[39] Kumar S, Priyadarshini M, Ahmad A, et al. Advanced biological and non-biological technologies for carbon sequestration, wastewater treatment, and concurrent valuable recovery: A review. Journal of CO2 Utilization. 2023;68:102372.

Tables with captions

Table 1 Basic O&M data for wastewater treatment plants

Year	Q	CODin	TNin	CODeff	TNeff	Sodium acetate	Aluminum sulfate	Sodium hypochlorite	Ferric polymerized sulfate	Tap water	PAM	Power
2012	38259	169	21.51	37	13.19	13.74	9.96	4.19	5.77	24.99	37.18	19512
2013	33861	242	31.71	45	16.00	21.13	12.09	3.44	7.15	22.11	32.91	16930
2014	34879	335	33.51	49	16.01	27.91	12.08	3.26	4.78	22.79	33.90	17091
2015	37600	353	36.24	50	14.57	28.27	11.03	3.22	4.71	24.55	36.54	18048
2016	39246	417	38.42	48	12.53	32.65	9.41	3.04	2.49	25.65	38.14	18446
2017	38332	445	45.29	51	16.28	35.15	12.41	2.67	4.00	24.99	37.25	17633
2018	42432	486	48.19	41	13.47	36.79	9.98	2.61	1.70	27.80	41.24	19095
2019	52361	374	37.99	34	10.93	24.16	8.57	2.80	1.65	33.98	50.89	23039
2020	68759	327	30.05	28	9.95	14.27	6.93	3.13	0.10	45.49	66.82	29948

2021	73284	309	34.59	29	9.63	10.65	8.39	2.87	1.89	64.20	84.57	31778
2022	82672	300	30.01	26	7.75	7.27	10.99	1.21	0.74	83.28	95.46	31505

Table 2 Pharmaceutical carbon emission factors

	Emission Factors	Source
PAM	0.851 $kgCO_2/kg$	[24]
Aluminum sulfate	0.5 $kgCO_2/kg$	Technical specification for low-carbon operation evaluation of sewage treatment plant
Sodium hypochlorite	0.92 $kgCO_2/kg$	Technical specification for low-carbon operation evaluation of sewage treatment plant
Sodium acetate	0.623 $kgCO_2/kg$	[25]
Tap water	0.168 $kgCO_2/t$	CPCD, China Products Carbon Footprint Factors Database
Other pharmaceuticals	1.6 $kgCO_2/kg$	[26]

Table 3 VIF test results

Influencing factors	significance	covariance statistics	
		tolerance	VIF
CODin	0.8001	0.0156	63.966
TNin	0.7982	0.0102	97.737
CODeff	0.3054	0.0406	24.612
TNeff	0.4977	0.0131	76.436
Aluminium sulfate	0.2075	0.0006	1682.481
Sodium hypochlorite	0.2888	0.0056	178.754
Polymerised iron sulphate	0.3872	0.0021	478.648
Tap water	0.3156	0.0543	18.403
Electricity consumption	0.8669	0.0005	1983.018
Pharmaceutical consumption	0.1968	0.0001	6833.279
Q	-	0.0000	7953.177
Sodium acetate	-	0.0000	4248.949
PAM	-	0.0000	1063973.3

Table 4 Values of model assessment indicators

	Model	SVR	ANN	GBDT	AdaBoost	DTR	GWO-SVR-ANN
test	R2	0.9590	0.9868	0.9779	0.9038	0.9027	0.9926
	RMSE	483.2494	274.3218	354.8518	739.8989	744.2982	204.8652
	MAE	383.4922	223.4162	293.9377	601.1521	572.8457	158.6957
	MAPE	1.19%	0.70%	0.91%	1.87%	1.76%	0.49%
train	R2	0.9577	0.9818	0.9997	0.9235	1.0000	0.9934
	RMSE	518.3161	339.9446	43.9357	697.3520	0.0000	204.9784
	MAE	402.7506	237.4547	33.6327	601.1239	0.0000	154.4703
	MAPE	1.23%	0.73%	0.10%	1.85%	0.0000	0.47%

Table 5 Rate of change of indicators for different scenarios (%)

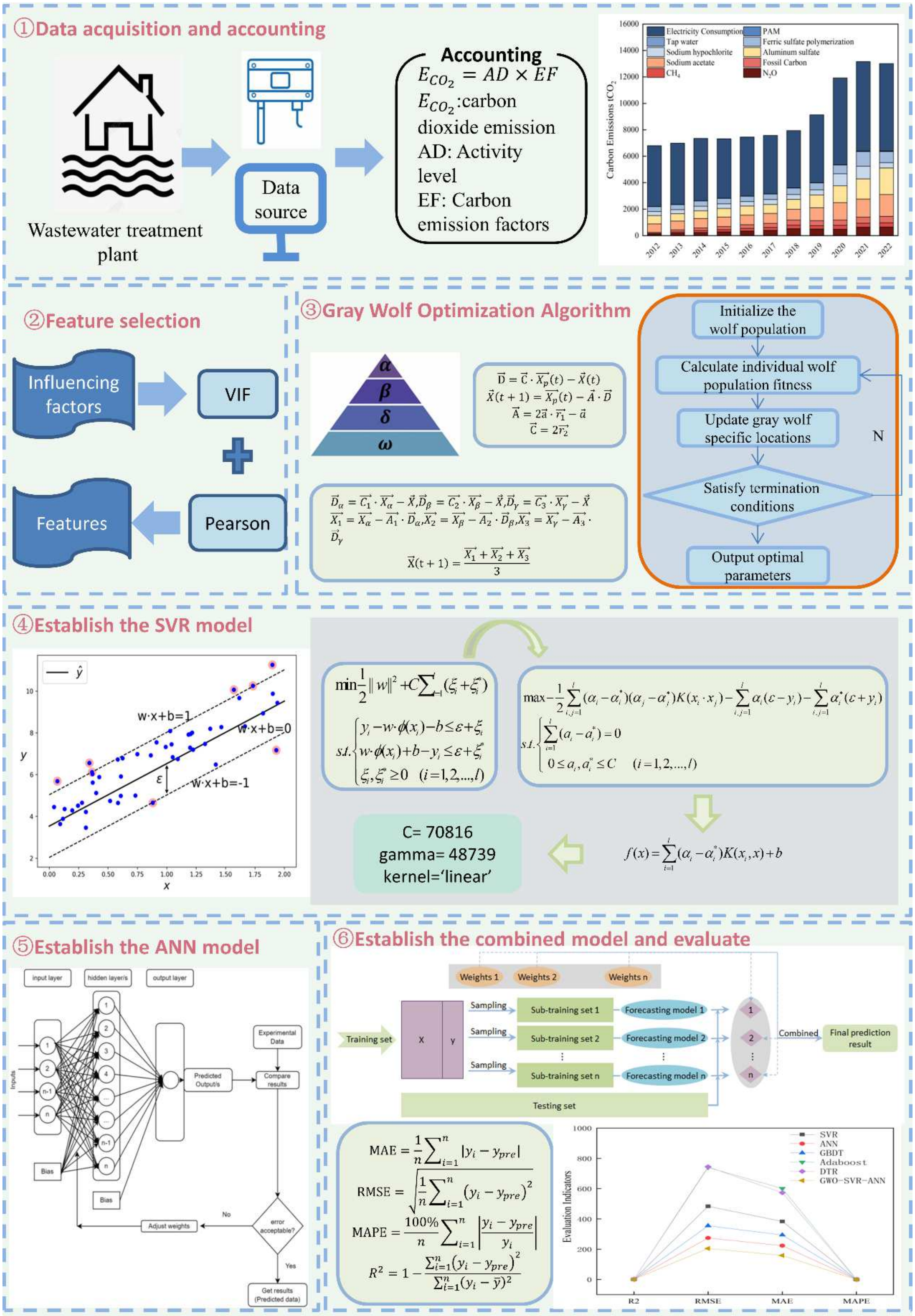
Indicator name	Scenario	2023-2025	2026-2030	2031-2035	2035-2040
Amount of treated water (A)	L	4.8	3.6	2.4	1.2
	M	5.3	4.1	2.9	1.7
	H	6	4.8	3.6	2.4
Pollutant concentration (B)	L	-2.6	-2.4	-2.2	-2
	M	-3	-2.8	-2.6	-2.4
	H	-3.4	-3.2	-3	-2.8
Energy consumption (C)	L	3.8	2.8	1.9	1.3
	M	4.2	3.2	2.3	1.7
	H	4.6	3.6	2.7	2.1
drug consumption (D)	L	4.8	3.6	2.4	1.2
	M	5.3	4.1	2.9	1.7
	H	6	4.8	3.6	2.4
clean energy substitution rate (E)	L	4	6	8	10
	M	5	8	11	14
	H	8	14	20	26
Methane recovery power generation rate (F)	L	4	6	8	10
	M	5	7	10	12
	H	6	10	14	18
Biological carbon sequestration rate (G)	L	4	6	8	10
	M	6	10	14	18
	H	8	14	20	26

Table 6 Scenario settings

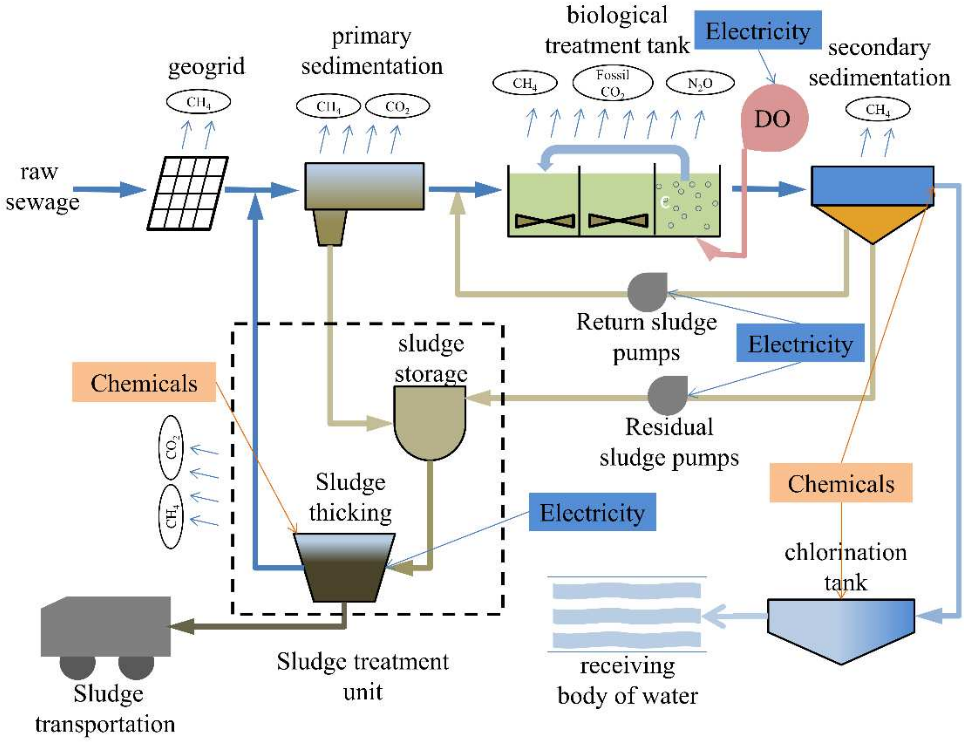
Scenario	A	B	C	D	E	F	G
Baseline Scenario	M	M	M	M	M	M	M
No Emission Reduction Measures Scenario	M	M	M	M	-	-	-
Technological Innovation Scenario	M	M	L	L	H	M	H
Energy Saving and Consumption Reduction Scenario	L	L	L	M	M	M	M
Sustainable Development Scenario	L	H	L	L	H	H	H
Low Development	H	H	H	H	L	L	L

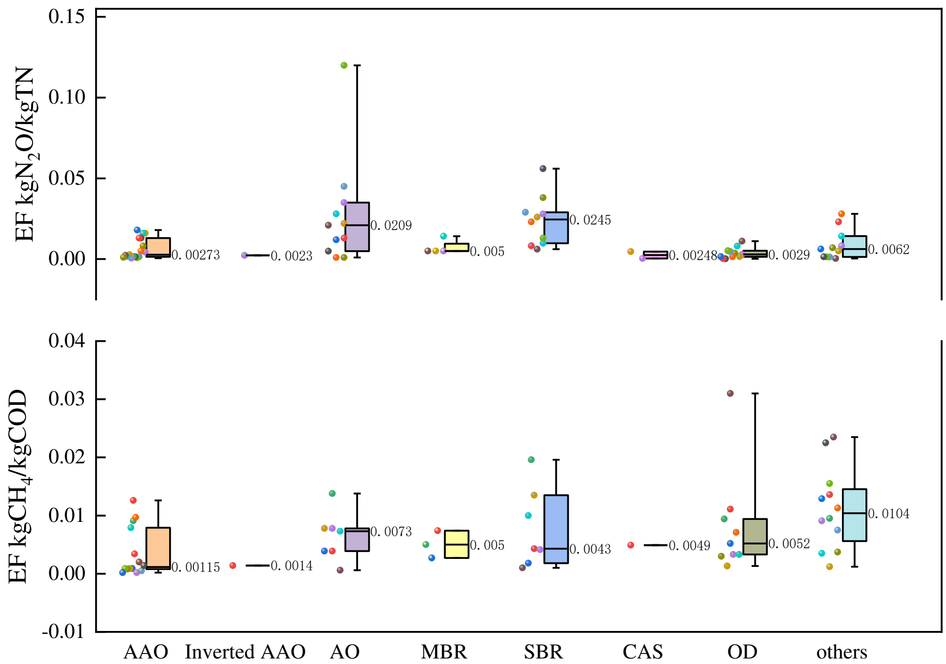
Scenario

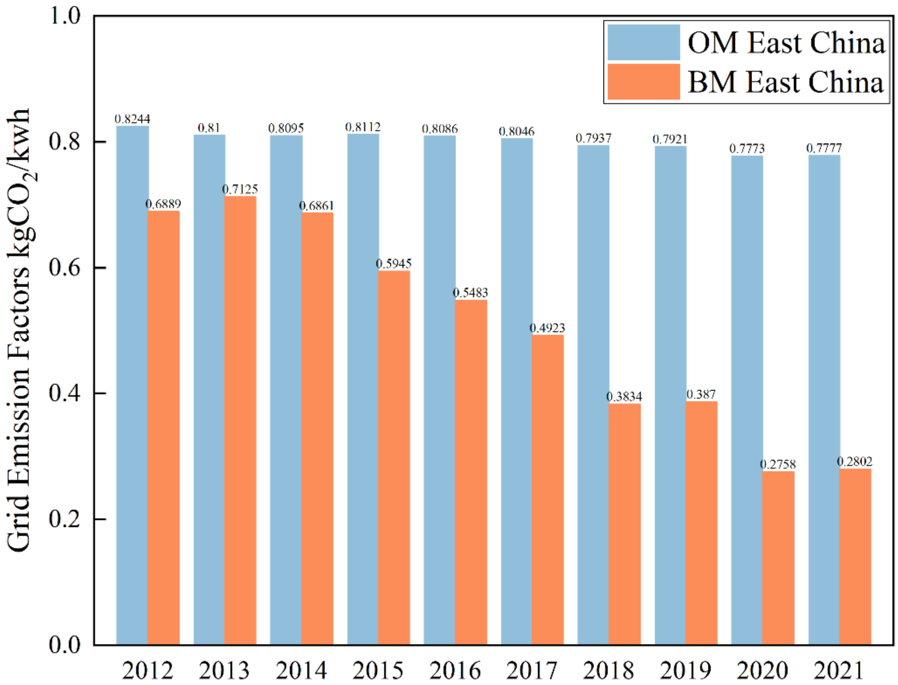
Figures

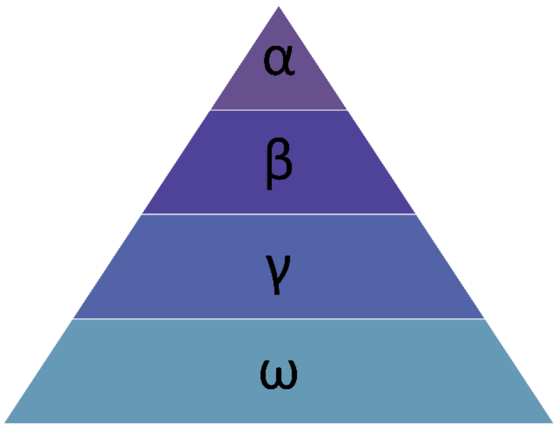


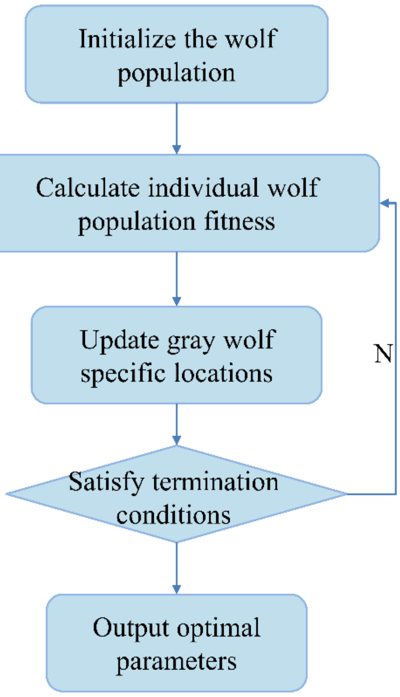
①Data acquisition and accounting
Wastewater treatment plant
Data source
Accounting
$E_{CO_2} = AD \times EF$
E_{CO_2}:carbon dioxide emission
AD: Activity level
EF: Carbon emission factors
Electricity Consumption
Tap water
Sodium hypochlorite
Sodium acetate
CH_4
PAM
Ferric sulfate polymerization
Aluminum sulfate
Fossil Carbon
N_2O
Carbon Emissions tCO_2
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
②Feature selection
Influencing factors
VIF
Pearson
Features
③Gray Wolf Optimization Algorithm
α
β
δ
ω
Initialize the wolf population
Calculate individual wolf population fitness
Update gray wolf specific locations
Satisfy termination conditions
N
Output optimal parameters
④Establish the SVR model
w·x+b=1
w·x+b=0
w·x+b=-1
C= 70816
gamma= 48739
kernel='linear'
⑤Establish the ANN model
input layer
hidden layer/s
output layer
Inputs
Bias
Predicted Output/s
Experimental Data
Compare results
error acceptable?
No
Yes
Adjust weights
Get results (Predicted data)
⑥Establish the combined model and evaluate
Weights 1
Weights 2
Weights n
Training set
Sampling
Sub-training set 1
Sub-training set 2
Sub-training set n
Forecasting model 1
Forecasting model 2
Forecasting model n
Combined
Final prediction result
Testing set
SVR
ANN
GBDT
Adaboost
DTR
GWO-SVR-ANN
Evaluation Indicators
R2
RMSE
MAE
MAPE

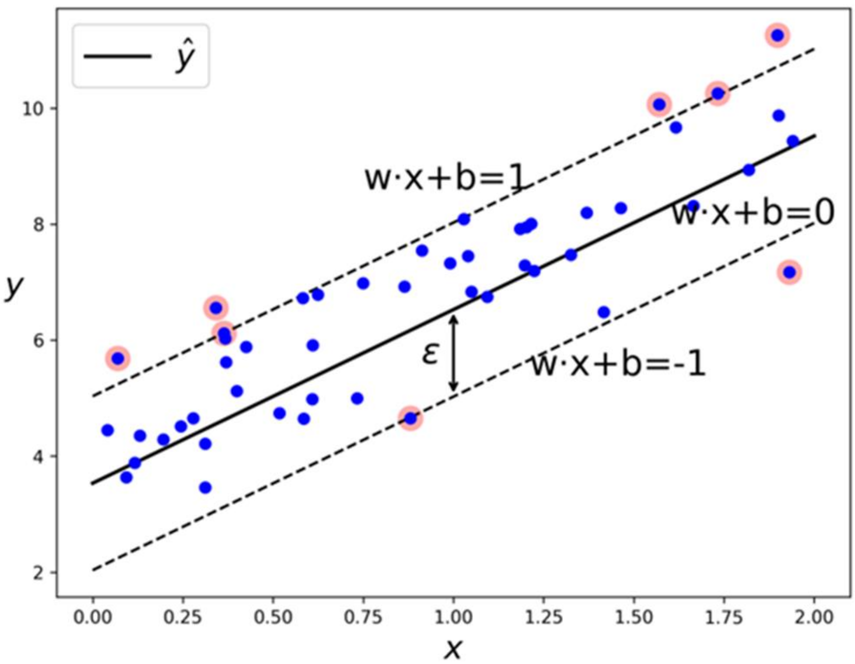

raw sewage
geogrid
primary sedimentation
biological treatment tank
Electricity
secondary sedimentation
DO
Return sludge pumps
Electricity
sludge storage
Chemicals
Residual sludge pumps
Sludge thicking
Electricity
Chemicals
chlorination tank
receiving body of water
Sludge treatment unit
Sludge transportation

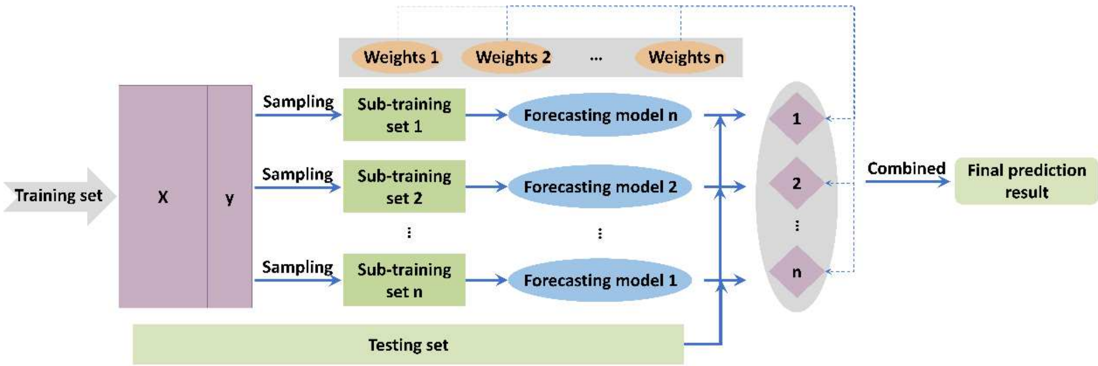

EF kgN2O/kgTN
0.15
0.10
0.05
0.00
0.00273
0.0023
0.0209
0.005
0.0245
0.00248
0.0029
0.0062
EF kgCH4/kgCOD
0.04
0.03
0.02
0.01
0.00
-0.01
0.00115
0.0014
0.0073
0.005
0.0043
0.0049
0.0052
0.0104
AAO
Inverted AAO
AO
MBR
SBR
CAS
OD
others

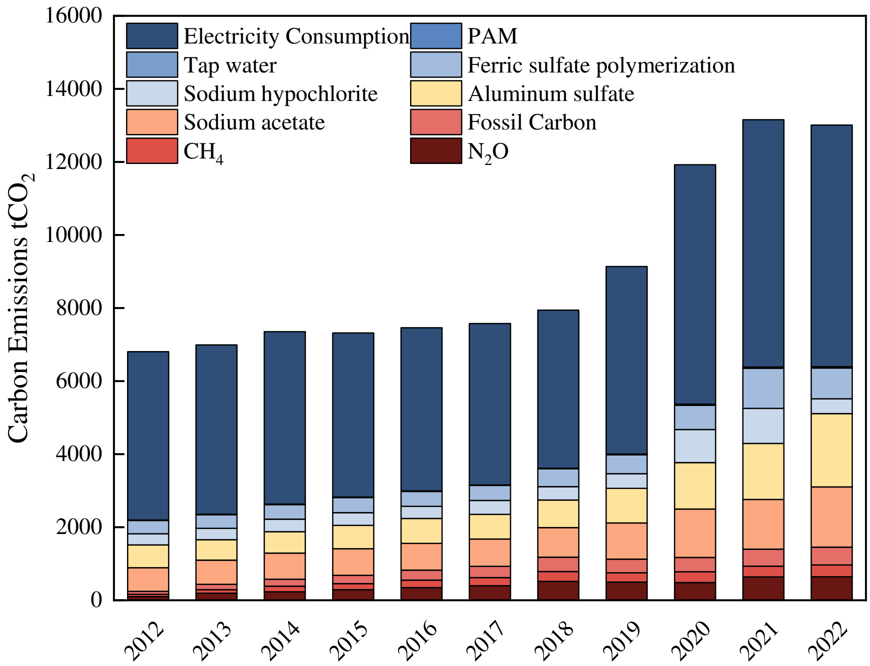

OM East China
BM East China
Grid Emission Factors $kgCO_2/kwh$
1.0
0.8
0.6
0.4
0.2
0.0
0.8244
0.6889
0.81
0.7125
0.8095
0.6861
0.8112
0.5945
0.8086
0.5483
0.8046
0.4923
0.7937
0.3834
0.7921
0.387
0.7773
0.2758
0.7777
0.2802
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021

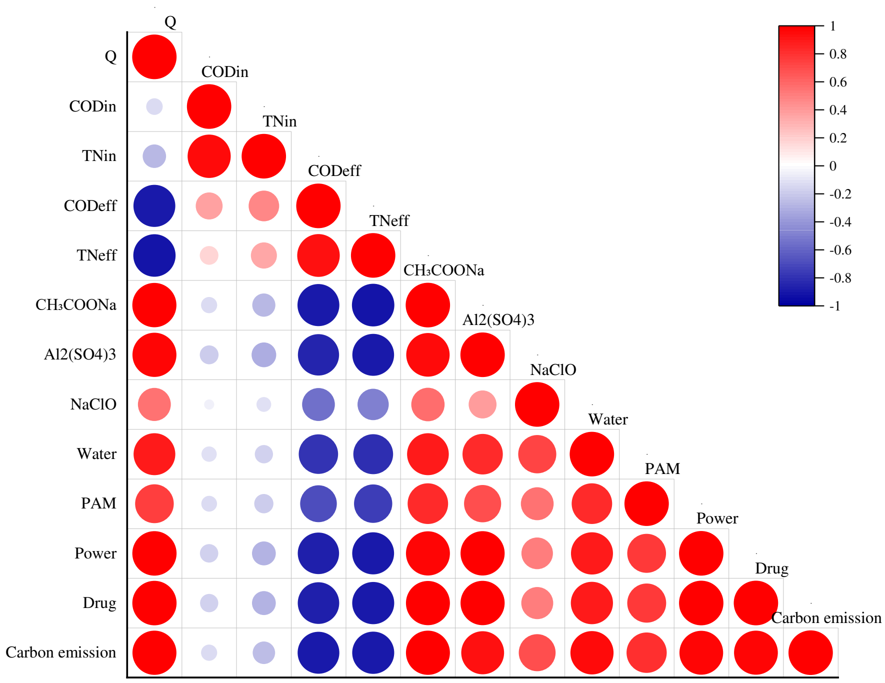

α
β
γ
ω


Initialize the wolf population
Calculate individual wolf population fitness
Update gray wolf specific locations
N
Satisfy termination conditions
Output optimal parameters


$\hat{y}$
w·x+b=1
w·x+b=0
w·x+b=-1
ε
y
x
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
2
4
6
8
10


Weights 1
Weights 2
...
Weights n
Training set
X
y
Sampling
Sub-training set 1
Forecasting model n
Sampling
Sub-training set 2
Forecasting model 2
Sampling
Sub-training set n
Forecasting model 1
Testing set
1
2
n
Combined
Final prediction result


Electricity Consumption
PAM
Tap water
Ferric sulfate polymerization
Sodium hypochlorite
Aluminum sulfate
Sodium acetate
Fossil Carbon
CH_4
N_2O
Carbon Emissions tCO_2
16000
14000
12000
10000
8000
6000
4000
2000
0
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022



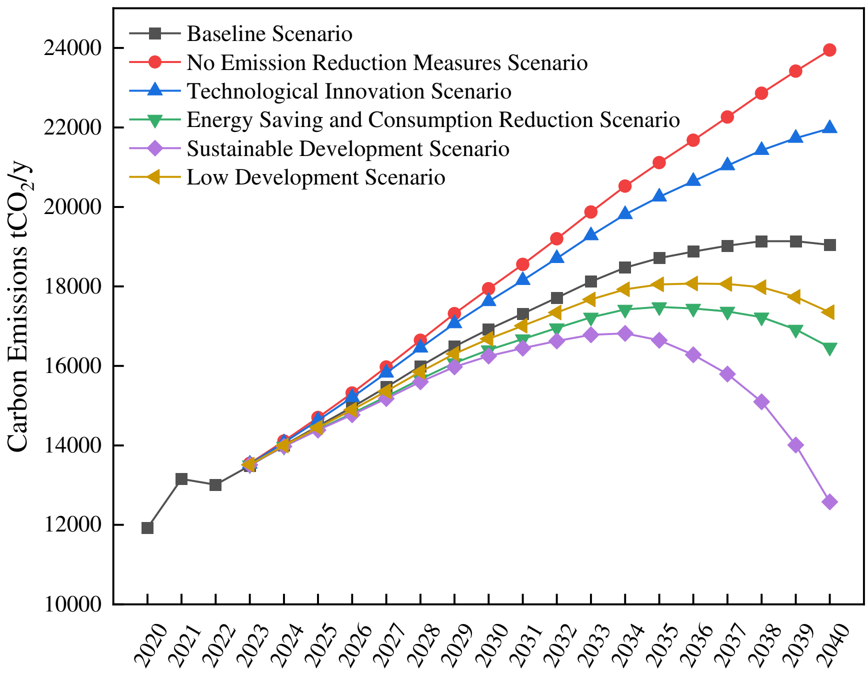


Figure captions

Figure	Figure captions
Fig. 1	Flowchart of the combined prediction model
Fig. 2	GHG accounting boundary for wastewater treatment plants
Fig. 3	Carbon emission factors for different wastewater treatment processes
Fig. 4	Carbon emission factor of power grid in East China region
Fig. 5a	Top-to-bottom hierarchy of gray wolves
Fig. 5b	Schematic diagram of the gray wolf optimization algorithm
Fig. 6	Schematic diagram of SVR principle