



XGB-SEGA coupled energy saving method for wastewater treatment plants

Zijian Wang^{1,2} · Xuwen Zhou^{3,2} · Hao Wang¹ · Zehan Huang⁴ · Jiantao Ji⁵ · Zhaoxu Peng² · Kun Jiang⁶

Received: 19 September 2023 / Accepted: 13 December 2023 / Published online: 23 January 2024
© The Author(s) 2024

Abstract

To reduce the amount of energy consumed in wastewater treatment plants, nine methods were used to select the key operation parameters that affected energy consumption according to daily operation records, and an intelligent operation management system based on a genetic algorithm was constructed by mapping the relationships between energy consumption and the key operation parameters. The results showed that the prediction and management of energy consumption could be achieved by incorporating the strengthened elastic genetic algorithm into the extreme gradient boosting model. The main parameters affecting energy consumption were the influent flow rate, effluent total nitrogen, $\text{NH}_4^+\text{-N}$ loading rate, etc., and the energy consumption could be reduced by 13–27% (with an average of 22%). The parameters were all selected from the daily operation records of the wastewater treatment plant, and no additional complex data acquisition system was needed to collect specific parameters. This study provided a cost-effective strategy to reduce energy consumption in wastewater treatment plants.

Keywords Energy saving · Extreme gradient boosting (XGB) · Genetic algorithm optimization · Carbon neutrality · Wastewater treatment

Introduction

With obvious climate change occurring during the last decade, reducing carbon emissions is an urgent need (Yang et al. 2022). Wastewater treatment is one of the most energy-consuming industries in China; it consumed 18.4 billion kWh in 2020, and its consumption level continues to increase every year (Chang et al. 2021). Large amounts of greenhouse gases (GHGs), such as CO_2 , N_2O , and CH_4 , are generated during

wastewater treatment and have been identified as anthropogenic sources of GHG emissions (Yoshida et al. 2014). Among the total energy consumption of wastewater treatment plants (WWTPs), the electricity consumed during aeration accounts for 70–80% of this amount, followed by pumps and chemicals (Yang et al. 2021). At present, the aeration control process in Chinese WWTPs is slipshod and mostly manual, and excess aeration is always supplied, which generates unnecessary energy consumption (Yang et al. 2021), especially in small-scale WWTPs. In most European WWTPs, the SCADA system has been installed to precisely control the aeration process. However, most Chinese WWTPs are only equipped with facilities to check whether the effluent meets the imposed discharge standards. Currently, as discharge standards become stricter, the requirement for energy savings is increasing. It is necessary to determine the energy savings potential of the existing equipment, as well as meet these standards.

During the last decade, with the rapid development of information technologies such as big data and artificial intelligence (AI), new energy saving solutions have been provided by constructing intelligent management systems (Wang et al. 2022). Machine learning (ML) is an AI technology that is used to recognize specific patterns and provide necessary

✉ Zhaoxu Peng
pzx@zzu.edu.cn

¹ Faculty of Architecture, Civil and Transportation Engineering, Beijing University of Technology, Beijing 100124, China

² School of Water Conservancy and Transportation, Zhengzhou University, Zhengzhou 450001, China

³ School of Civil and Environmental Engineering, Harbin Institute of Technology, Shenzhen 518055, China

⁴ Central Plains Environmental Protection Co., Ltd, Zhengzhou 450001, China

⁵ School of Ecological and Environment, Zhengzhou University, Zhengzhou 450001, China

⁶ CSD Water Service Co., Ltd, Beijing 100124, China

data for prediction or classification. Benefiting from its high precision based only on data relationships, ML is becoming more popular in various fields, such as effluent prediction and process optimization (Picos-Benítez et al. 2020). Since wastewater treatment is a nonlinear process, it is often difficult to construct simple models, while a data-driven approach based on ML is preferable (Bagherzadeh et al. 2021). Many studies have verified the feasibility of ML in wastewater treatment tasks. Nourani et al. (2021) proposed an approach based on black-box models, including a feedforward neural network, support vector regression (SVR) and an adaptive neuro-fuzzy inference system, to predict effluent biological oxygen demand (BOD₅) and chemical oxygen demand (COD). Ly et al. (2022) investigated and compared six ML algorithms for predicting effluent total phosphorus (TP). El-Rawy et al. (2021) compared the performance of different models for predicting the removal efficiencies of total suspended solids, COD, BOD₅, and NH₄⁺-N (ammonia nitrogen). Wan et al. (2022) proposed a model based on a convolutional neural network, weight-sharing long short-term memory and Gaussian process regression for paper-making wastewater treatment, which exhibited a comprehensive forecasting ability. Das et al. (2021) used standard mean absolute error (MAE) and root mean square error (RMSE) metrics as evaluation indices to compare four ML algorithms, and a gated recurrent unit was selected as the best model. Żyłka et al. (2020) evaluated the least-squares linear regression model for predicting electricity consumption and found that the main parameters were the organic loading rate and temperature.

However, to date, research using ML in wastewater treatment scenarios has mainly focused on effluent quality, while few studies has been conducted on energy consumption. In addition, some factors for determining energy consumption have not been considered, and their influence has not yet been fully evaluated; thus, model performance is strongly determined by the accuracy of the input data.

In this study, a mapping relationship between energy consumption and management parameters was established, and an energy-saving strategy for WWTPs was developed based on a genetic algorithm. Daily operation parameters were ranked through regression analysis, and then the top-ranking parameters were selected as inputs to establish an XGB model for predicting energy consumption. Furthermore, an energy-saving control strategy was evaluated, which is expected to offer an instant energy-saving strategy for WWTPs in practical applications.

Materials and methodology

Background of the target WWTP

An urban WWTP (an anaerobic anoxic aerobic process) located in Henan Province (China) had a designed flow rate

of 30,000 m³ day⁻¹, and it followed the Chinese discharge standard of GB18918-2022. The sludge was dewatered by a belt filtering press and then treated by a sanitary landfill (Fig. 1). The operation data were collected for 353 days from 1st January to 31st December 2020, which consisted of the influent flow rate (IFR), influent NH₄⁺-N concentration (IAN), influent COD (ICOD), influent TN (ITN), influent TP (ITP), effluent NH₄⁺-N concentration (EAN), effluent COD (ECOD), effluent TN (ETN), effluent TP (ETP), mixed liquid suspended solids (MLSS) of the aerobic tank, DO at the end of the aerobic tank (DO), ORP at the end of the anoxic tank (ORP), organic loading rate (OLR), NH₄⁺-N loading rate (ANLR) and energy consumption per cubic metre (EC). The units and statistics of each feature are shown in Table 1.

Feature selection

Ordinary least squares

Ordinary least squares (OLS) is a method that can be used for variable selection. For a dataset $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, where $x \in \mathbb{R}^d, y \in \mathbb{R}$, one of its basic expressions is as follows. The fitting criterion is to minimize the sum of squared residuals between y and $f(x)$. The parameters without zero coefficients are selected.

$$f(x) = \omega^T x + \beta \quad (1)$$

$$\min \sum_{i=1}^n (y_i - f(x_i))^2 \quad (2)$$

where ω^T and β are undetermined coefficients.

Least absolute shrinkage and selection operator

To avoid overfitting during OLS, a penalty function with the L₁ norm is added to the objective function to simplify the structure and decrease the empirical risk in the least absolute shrinkage and selection operator (Lasso). Compared with ridge regression, which adapts the L₂ norm, the Lasso more easily obtains sparse solutions and selects features.

$$\min \sum_{i=1}^n (y_i - f(x_i))^2 + \lambda \omega_1 \quad (3)$$

Smoothly clipped absolute deviation

The smoothly clipped absolute deviation (SCAD) approach was proposed by Fan and Li (2001). Compared with the Lasso, this method reduces the bias of parameter

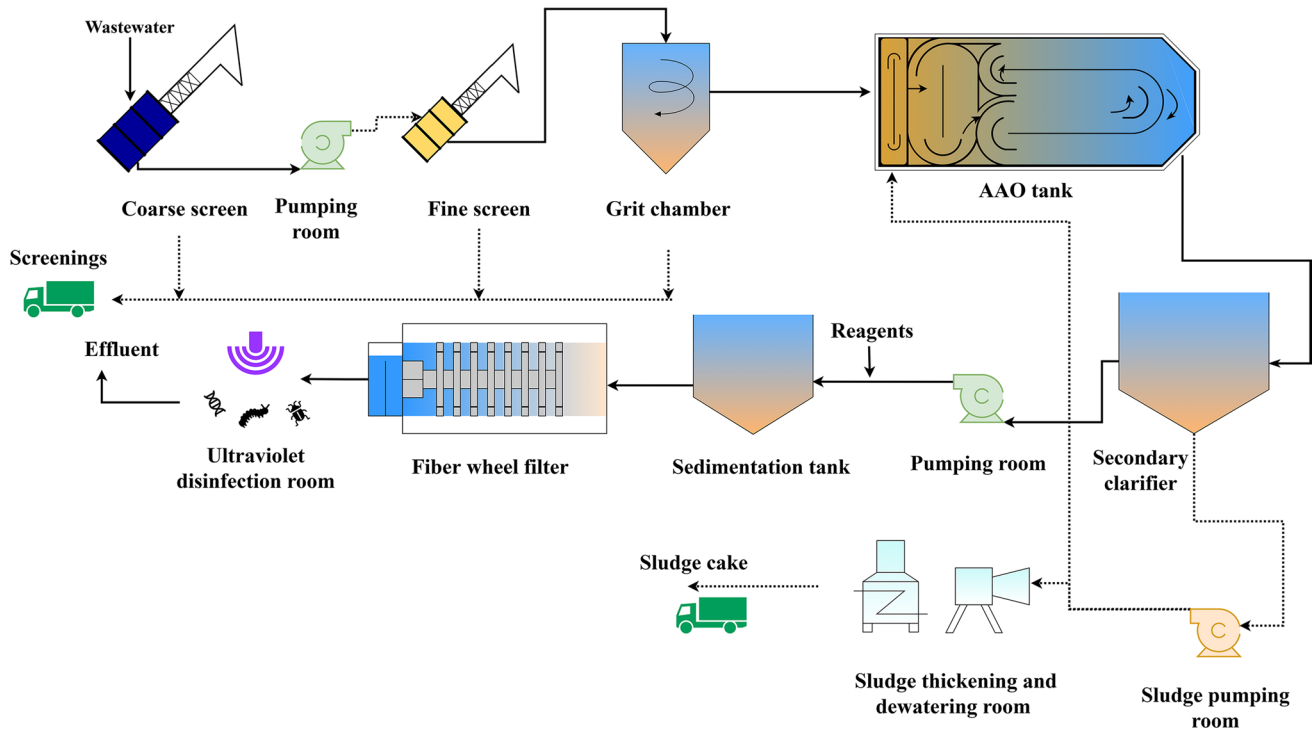


Fig. 1 Treatment process of the WWTP

Table 1 Dataset

| Parameter | Units | Abbreviation | Mean | Std | Min | Max |
|---|------------------------------------|--------------|-----------|---------|---------|--------|
| Influent flow rate | m ³ d ⁻¹ | IFR | 29,269.01 | 4983.09 | 21,238 | 45,672 |
| Influent ammonia nitrogen | mg L ⁻¹ | IAN | 36.68 | 8.93 | 16.62 | 63.85 |
| Influent COD | mg L ⁻¹ | ICOD | 364.34 | 133.05 | 150.5 | 934.8 |
| Influent TN | mg L ⁻¹ | ITN | 48.49 | 9.68 | 26.49 | 85.09 |
| Influent TP | mg L ⁻¹ | ITP | 5.66 | 2.69 | 1.21 | 17.66 |
| Effluent ammonia nitrogen | mg L ⁻¹ | EAN | 0.66 | 0.65 | 0.03 | 3.95 |
| Effluent COD | mg L ⁻¹ | ECOD | 28.06 | 4.80 | 12.1 | 44.74 |
| Effluent TN | mg L ⁻¹ | ETN | 8.24 | 2.37 | 3.53 | 17.98 |
| Effluent TP | mg L ⁻¹ | ETP | 0.18 | 0.08 | 0.04 | 0.52 |
| Mixed liquid suspended solids of the aerobic tank | mg L ⁻¹ | MLSS | 6379.86 | 797.92 | 4472.5 | 7861 |
| DO at the end of the aerobic tank | mg L ⁻¹ | DO | 0.95 | 0.34 | 0.42 | 3.65 |
| ORP at the end of the anoxic tank | Mv | ORP | -40.19 | 30.77 | -178.06 | 46.52 |
| Organic loading rate | kgBOD·(kgMLSS·d) ⁻¹ | OLR | 0.10 | 0.03 | 0.03 | 0.27 |
| Ammonia nitrogen loading rate | Kg m ⁻³ d ⁻¹ | ANLR | 0.01 | 0.00 | 0.00 | 0.02 |
| Energy consumption per cubic metre | kWh m ⁻³ | EC | 0.54 | 0.12 | 0.28 | 0.90 |

estimation. The basic principle behind the imposed penalty is similar to MCP, but the difference is the utilized transition method. The penalty function of SCAD is defined as Eq. 4.

$$p_{\lambda,\gamma}(\theta) = \begin{cases} \lambda\theta, & \theta \leq \lambda \\ -\frac{\theta^2 - 2\gamma\lambda\theta + \lambda^2}{2\theta - 2}, & \lambda < \theta \leq \gamma\lambda \\ \frac{(\gamma+1)\lambda^2}{2}, & \theta > \gamma\lambda \end{cases} \quad (4)$$

where $\lambda \geq 0$ and $\gamma > 2$.

The objective function used by SCAD is as follows:

$$\min \sum_{i=1}^n (y_i - f(x_i))^2 + \lambda \sum_{j=1}^d p_{\lambda, \gamma}(\omega_j) \tag{5}$$

C_p criterion

Mallow's C_p is used for variable selection through the sum of the residual squares obtained from an OLS regression model (James et al. 2021). The independent variable subset that minimizes C_p is selected, and the regression equation corresponding to this independent variable subset is the optimal regression equation.

$$C_p = \frac{SSR_d}{\hat{\sigma}^2} - n + 2d \tag{6}$$

where $\hat{\sigma}^2$ is an estimate of the variance of the residuals, d is the number of parameters, and n is the sample size.

The above methods are based on a linear regression model. By adding a penalty item, the coefficients of the input variables are compressed to varying degrees, and this plays a role in the feature selection process.

Akaike information criterion

The Akaike Information Criterion (AIC) is widely used to evaluate the fitness levels of statistical models, and it works as follows. The first term reflects fitness, and the second term reflects the number of parameters. The best model is the one with the minimum AIC (Ingdal et al. 2019).

$$AIC = -2 \ln \hat{L} + 2d \tag{7}$$

where \hat{L} is the likelihood estimator and d is the number of parameters.

Bayesian information criterion

Similar to the AIC, the Bayesian information criterion (BIC) is also a basic method for decision-making tasks involving statistical models, and it helps to strike a balance between simplicity and mapping ability. The standard expression of the BIC is as follows. Different from the AIC, its second term is concerned with both the number of parameters and the sample size. The model with the minimum value achieves the best balance between simplicity and mapping ability (Liu et al. 2022b).

$$BIC = -2 \ln \hat{L} + d \ln n \tag{8}$$

where \hat{L} is the likelihood estimator, d is the number of parameters, and n is the sample size.

Minimum redundancy–maximum relevance

The minimum redundancy-maximum relevance (mRMR) algorithm is used to solve the problem that due to the existence of redundant variables, the best eigenvalue obtained by maximizing the correlation degree between a feature and the target variable does not necessarily obtain the best prediction accuracy (Ding and Peng 2003). Mutual information is used to measure the correlation between two variables. The mutual information I of two discrete random variables x and y is defined as follows:

$$I(x, y) = \sum_{i,j} p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(x_i)p(y_j)} \tag{9}$$

where $p(x, y)$ is the joint probabilistic distribution of two variables x and y , and $p(x)$ and $p(y)$ are the marginal probabilities of x and y , respectively.

For the mRMR algorithm, the mutual information $I(x, c)$ is used to find the feature subset S among the m features that are most closely related to category c .

$$\max D(S, c), D = \frac{1}{|S|} \sum_{x_i} I(x_i, c) \tag{10}$$

The minimum redundant feature condition is:

$$\min R(S), R = \frac{1}{|S|^2} \sum_{x_i, x_j \in S} I(x_i, x_j) \tag{11}$$

Then, the maximal correlation-minimal redundancy feature set S is:

$$mRMR = \max \left[\frac{1}{|S|} \sum_{x_i} I(x_i, c) - \frac{1}{|S|^2} \sum_{x_i, x_j \in S} I(x_i, x_j) \right] \tag{12}$$

By selecting the subset of variables that minimize or maximize the target value, the redundant variables can be eliminated.

Model construction

XGB

XGB is an optimized version of boosting (Chen and Guestrin 2016). When a tree is added, a new function $f(x)$ can be obtained to fit the residual of the last prediction. Once K trained trees are obtained, every tree falls to a corresponding leaf node, and every node corresponds to a score. The predicted value is the summation of the scores produced by different trees, which is calculated as follows:

$$\hat{y} = \sum_{k=1}^K f_k(x_i) \tag{13}$$

where K is the number of trees and $f_k(x_i)$ is the score of each tree.

The objective function consists of a loss function and a regularization penalty:

$$\text{Obj}(\theta) = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \tag{14}$$

where $l(y_i, \hat{y}_i)$ is the error of the i -th sample and $\Omega(f_k)$ is the regularization penalty term of the k -th tree.

$$\Omega(f_k) = \alpha T + \frac{1}{2} \lambda \sum_{j=1}^T \omega_j^2 \tag{15}$$

where T is the total number of leaf nodes in the t -th tree, ω_j is the weight of the j -th leaf node, and α, λ are scalars. α controls the number of leaf nodes, and γ guarantees that the weights of the leaf nodes are small.

For the t -th iteration, the objective function can be expressed as follows:

$$\text{Obj}^{(t)} = \sum_{i=1}^n l[y_i, \hat{y}_i^{(t-1)} + f_t(x_i)] + \sum_{k=1}^K \Omega(f_k) + C \tag{16}$$

where $f_t(x_i)$ is the newly added t -th tree and C is the complexity of the previous $t-1$ trees, that is, $C = \sum_{i=1}^{k-1} \Omega(f_i)$.

With a descending constant term $l(y_i, \hat{y}_i^{(t-1)})$ and C , the second-order Taylor expansion is employed to approximate the original loss function:

$$\text{Obj}^{(t)} \simeq \sum_{j=1}^T \left[G_j w_j + \frac{1}{2} (H_j + \lambda) w_j^2 \right] + \gamma T \tag{17}$$

where $G_j = \sum_{i \in I_j} \partial_{\hat{y}_{(t-1)}} l(y_i, y^{(t-1)})$, $H_j = \sum_{i \in I_j} \partial_{\hat{y}_{(t-1)}}^2 l(y_i, y^{(t-1)})$

When the partial derivative of the objective function $\text{Obj}^{(t)}$ with respect to ω_j equals 0, the optimal weight value can be obtained as follows:

$$\omega_j^* = -\frac{G_j}{H_j + \lambda} \tag{18}$$

The optimal objective function value is

$$\text{Obj}^{(t)} = -\frac{1}{2} \sum_{j=1}^T \frac{G_j^2}{H_j + \lambda} + \gamma T \tag{19}$$

Multilayer perceptron artificial neural network

As a robust ML technology, the multilayer perceptron artificial neural network (MLPANN) model has been widely used for prediction in various energy systems. The basic MLPANN consists of three layers: an input layer, a hidden layer and an output layer (Faegh et al. 2021). Given a series

of characteristics $X = (x_1, x_2, \dots)$ and a target Y , a multilayer perceptron can learn the relationships between the features and targets for classification or regression purposes.

Light gradient boosting machine

The light gradient boosting machine (LightGBM) is a distributed gradient boosting framework based on a decision tree algorithm. For a given training set, the LightGBM can obtain a strong classifier by combining multiple classification and regression trees (Sun et al. 2022).

Support vector regression

Support vector regression (SVR) is a powerful approach for problems with small sample sizes and high dimensionality (Huang et al. 2022). With the SVR algorithm, a regression plane can be found, and the distance between all data in a set and the plane can be minimized.

Model evaluation

To evaluate different models in terms of energy consumption prediction, the coefficient of determination (R^2), MAE, mean absolute percentage error (MAPE), and RMSE can be calculated as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n \left(y_i - \frac{1}{n} \sum_{i=1}^n y_i \right)^2} \tag{20}$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \tag{21}$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \tag{22}$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \times 100\% \tag{23}$$

where n is the sample size, y_i is the actual value, \hat{y}_i is the predicted value and i is the index.

R^2 is a classic indicator for evaluating the fitness between actual and predicted values. Both the MAE and RMSE are not sensitive to dimensions, but the RMSE magnifies the gap between larger errors. A MAPE of 0% indicates a perfect model, while a value greater than 100% indicates an inferior model.

SEGA The genetic algorithm is a heuristic algorithm originating from the Darwin evolutionary system that simulates natural selection as reproduction, crossover, and mutation in DNA. During the evolutionary process, individuals with low adaptability are eliminated. Through continuous selection and the 3 behaviors of DNA, the optimal result can be obtained. However, it has been proven that the canonical GA, which only uses the selection, crossover and mutation operators with crossover and mutation probabilities between (0,1), cannot converge to the optimal value. Thus, an elitism strategy is adopted to select the best individual and copy it to the next generation without a crossover operator (Fig. 2).

Results and discussion

Parameter selection results

If all the existing parameters are used as the inputs of the constructed model, its complexity will be very high, which may result in a long training time and poor performance. Therefore, it is necessary to select inputs first, which is crucial to model performance (Chu et al. 2009). Both the linear and nonlinear selection methods mentioned above were used

to select the input parameters (Table 2). The eight parameters with the highest frequencies were selected as the inputs, and they could be classified into three categories: water quantity (IFR), water quality (ETN, IAN, ITP and ETP) and management regulation (DO, ANLR and ORP) parameters. Furthermore, a Kendall correlation analysis was carried out between all parameters and the energy consumption level to evaluate the above results (Fig. 3). The absolute values of the correlation coefficients between the selected parameters and the energy consumption level were mostly at the top, which confirmed that the selected results were reasonable. Among the eight selected parameters, ETP and ANLR had limited correlations with energy consumption ($0 < r < 0.2$) (Khamis 2008).

In terms of water quantity, the IFR exhibited a strong negative correlation with EC. This can be attributed to the fact that most equipment cannot operate under effective energy conditions when the IFR deviates from the designed value (Hanna et al. 2018). In addition, poor management and limited regulation may also account for the excessive energy consumption of small-scale WWTPs (Vaccari et al. 2018).

In terms of water quality, the ETN was closely related to nitrogen removal performance. The conventional nitrogen removal process consists of nitrification and denitrification. To achieve sufficient nitrification, a large amount of energy

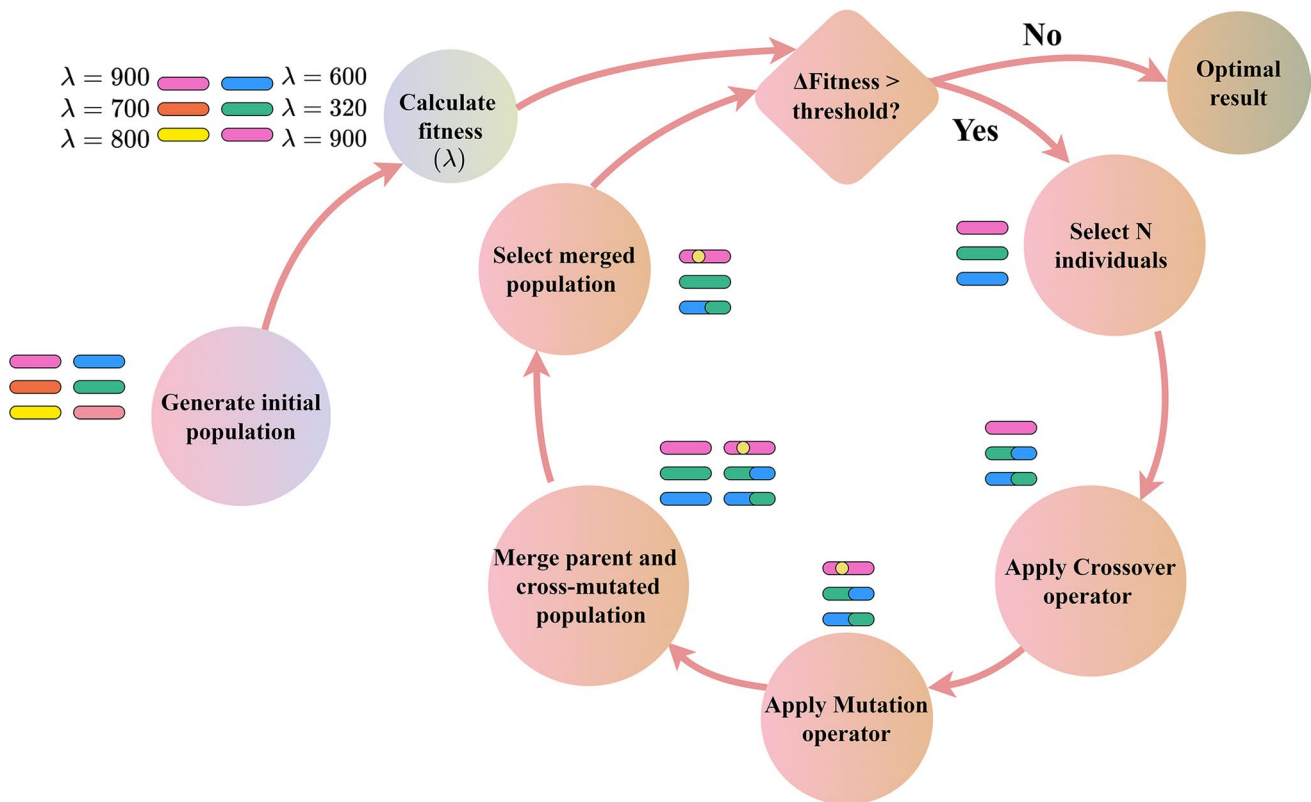


Fig. 2 Flow chart of the strengthened elitist genetic algorithm

Table 2 Parameter selection frequency

| Parameter | Water quantity | Water quality | | | | | | | | | Management regulation | | | |
|--------------------------|----------------|---------------|------|-----|-----|-----|------|-----|-----|------|-----------------------|-----|-----|------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
| | IFR | IAN | ICOD | ITN | ITP | EAN | ECOD | ETN | ETP | MLSS | DO | ORP | OLR | ANLR |
| OLS | * | * | | | * | | | * | * | | * | * | | * |
| Lasso | * | * | | * | * | * | | * | * | * | | * | | * |
| SCAD | * | * | | | * | | | * | * | | * | * | | * |
| C _p criterion | * | * | | | * | | | * | * | | * | * | | * |
| AIC | * | * | | | | | | * | * | | | * | | * |
| BIC | * | * | | | * | | | * | * | | * | * | | * |
| mRMR | * | * | | * | * | * | | | | * | * | * | | * |
| Selected times | 7 | 7 | 0 | 2 | 6 | 2 | 0 | 6 | 6 | 2 | 5 | 7 | 0 | 6 |
| Select results | * | * | | | * | | | * | * | | * | * | | * |

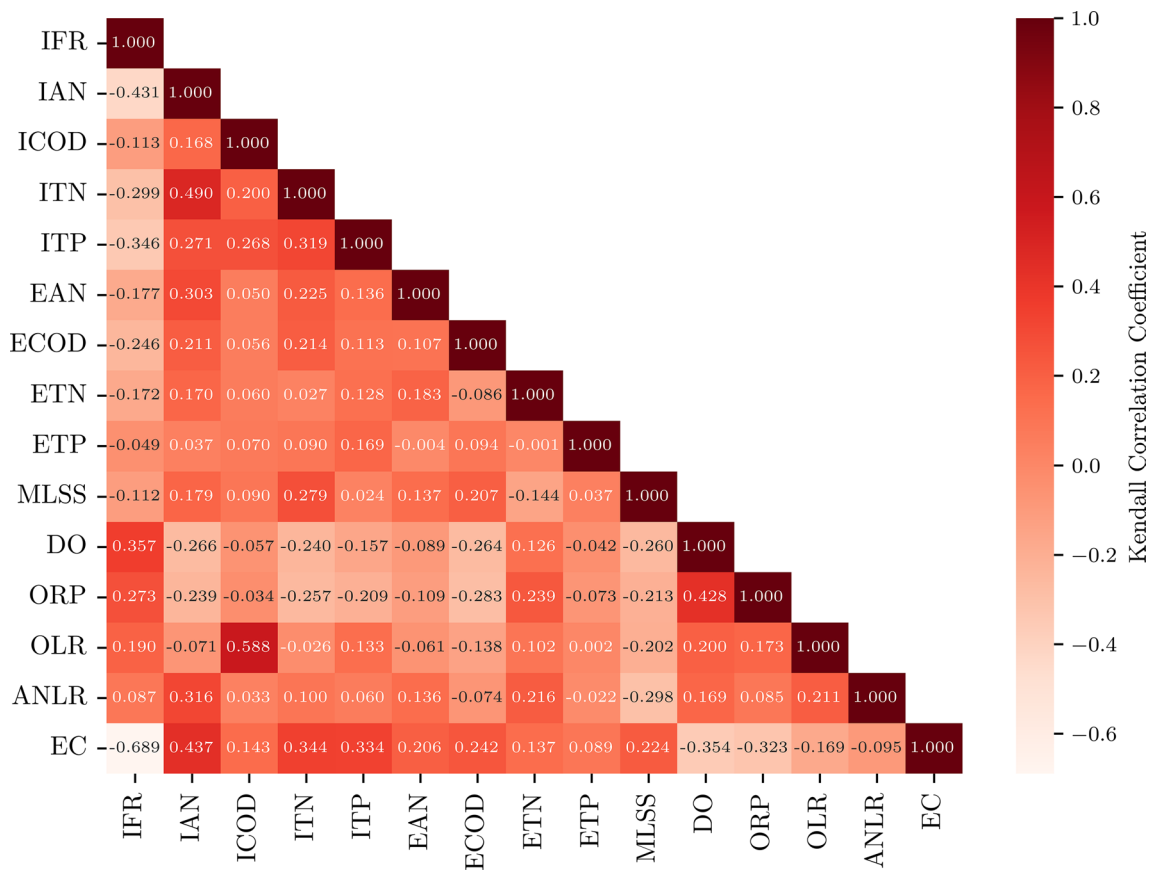


Fig. 3 Heatmap of the Kendall correlation coefficient

is consumed for aeration. Once the needed oxygen is supplied, excessive aeration not only wastes energy but also increases the effluent NO₃⁻-N (Liu et al. 2022a). When the MLSS is low, the phosphorus removal process is mainly dependent on the performance of phosphorus-accumulating

organisms. The biological phosphorus removal process consists of phosphorus release and phosphorus uptake. Electron acceptors such as oxygen are necessary to uptake phosphorus, so more aeration is needed when the ITP is larger.

In terms of management regulation, a lower OLR implied a longer SRT, which increased the MLSS. Hence, less NH₄⁺-N was associated with a lower ANLR, and DO should theoretically have a negative correlation with the OLR. Due to the excess aeration strategy, DO usually changed more violently than the OLR requirement. With the increases in the OLR and influent COD, excessive aeration was supplied to ensure that the effluent COD satisfied the discharge standard, and DO was positively correlated with the OLR. The excessive DO at the end of the aerobic tank reflowed to the anaerobic tank through external reflow, resulting in a higher ORP. Therefore, the correlation between the energy consumption level and the ORP at the end of the anaerobic tank was weak. However, when the influent COD decreased, aeration control fell behind the actual demand, resulting in excessive energy consumption for aeration.

Performance analysis of the XGB model

An XGB model and three other models were established based on the relationships between energy consumption and the selected parameters. The first 70% of the data were used as the training set, and the other 30% were used as the test set. A grid search was used to optimize the hyperparameters of all models in Table 3. The prediction performance is shown in Table 4 and Fig. 4. Compared with the other methods, XGB is the best model. Although a gap remains between the real and predicted values, their variation trends

Table 4 XGB model evaluation indicators

| Model | Dataset | R ² | MAE | MAPE | RMSE |
|----------|---------|----------------|-------|--------|-------|
| XGB | Train | 0.808 | 0.041 | 0.074 | 0.056 |
| | Test | 0.722 | 0.023 | 0.043 | 0.029 |
| MLPANN | Train | 0.978 | 0.009 | 1.662 | 0.019 |
| | Test | 0 | 0.064 | 10.787 | 0.089 |
| LightGBM | Train | 0.79 | 0.044 | 0.083 | 0.058 |
| | Test | 0.364 | 0.037 | 0.077 | 0.051 |
| SVR | Train | 0.754 | 0.053 | 0.101 | 0.063 |
| | Test | 0.145 | 0.047 | 0.094 | 0.059 |

are almost the same, which verifies that the model is feasible for predicting energy consumption.

Parameter impact analysis

The F score was used to evaluate the influence of each input on energy consumption (Fig. 5). The effect of each parameter in descending order was IFR, DO, IAN, ITP, ETN, ANLR, ORP, and ETP. It was found that this order was similar to that obtained by the Kendall analysis. If a significant parameter changes obviously, the energy consumption also changes significantly.

Among the uncontrollable parameters, the IFR, IAN and ITP had large influences. This means that once the treatment process and designed flow rate have been selected, the energy consumption levels has almost been determined.

Table 3 Hyperparameters of each model

| Model | Hyper parameter | Value | Model | Hyper parameter | Value |
|-----------------------|-----------------------|---------|----------|------------------|-------|
| XGB | Learning_rate | 0.06 | LightGBM | boosting_type | gbdt |
| | Max_depth | 6 | | metric | L2 |
| | n_estimators | 22 | | learning_rate | 0.02 |
| | min_child_weight | 1 | | num_leaves | 30 |
| | Gamma | 0 | | max_depth | 4 |
| | Subsample | 0.9 | | subsample | 0.8 |
| | Colsample_bytree | 0.6 | | colsample_bytree | 0.8 |
| | Alpha | 0 | | n_estimators | 60 |
| | Input layer | 8 | | kernel | rbf |
| | Hidden layer 1 (Relu) | 14 | | C | 350 |
| Hidden layer 2 (Relu) | 15 | gamma | 1E-7 | | |
| Hidden layer 3 (Relu) | 30 | epsilon | 0.1 | | |
| Hidden layer 4 (Relu) | 40 | coef0 | 0.01 | | |
| MLPANN | Hidden layer 5 (Relu) | 30 | SVR | max_iter | -1 |
| | Hidden layer 6 (Relu) | 15 | | degree | 5 |
| | Hidden layer 7 (Relu) | 5 | | shrinking | False |
| | Hidden layer 8 (Relu) | 3 | | tol | 1E-7 |
| | Output layer | 1 | | | |
| | Loss | L1 | | | |
| | Optimizer | Adam | | | |

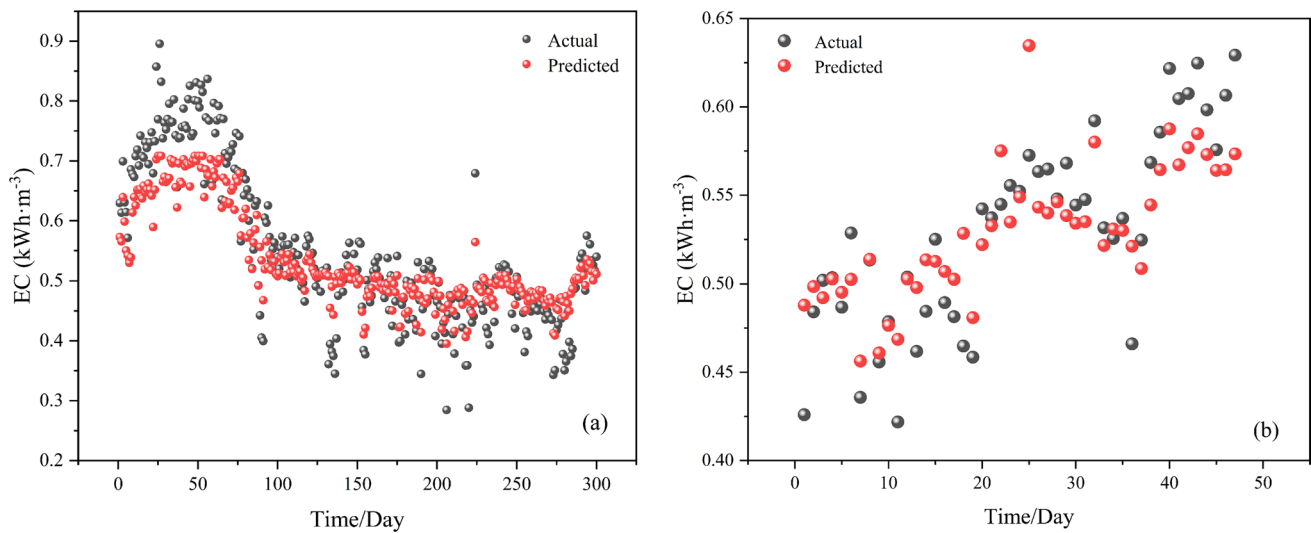
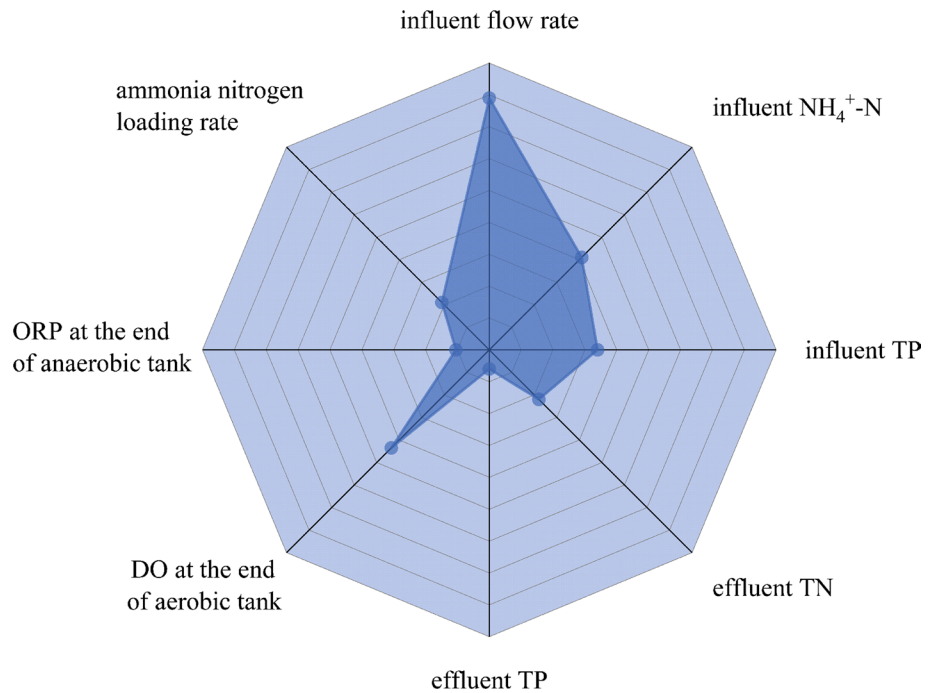


Fig. 4 Comparison between the values predicted by XGB and the actual values. **a** Training set; **b** testing set

Fig. 5 F score of XGB



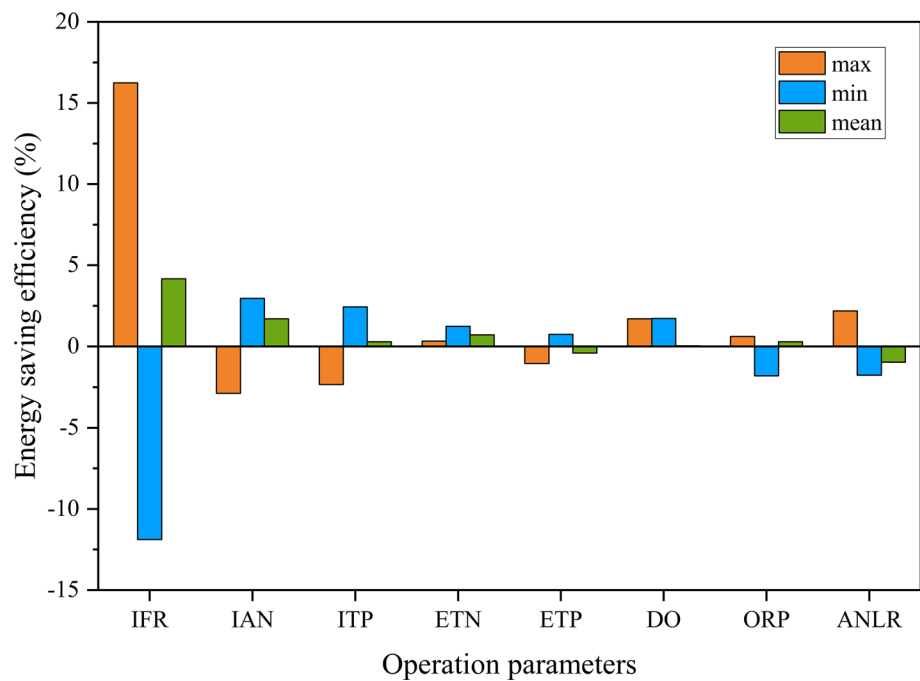
Among the controllable parameters, ETN, DO, and ANLR can influence energy consumption.

Energy saving performance

The energy savings achieved under different conditions were evaluated. For each parameter, the average, maximum and minimum values were taken in the established model, while the other seven variables remained unchanged from their average values (Fig. 6). The energy saving efficiency was

most sensitive to the influent flow rate, which was consistent with its maximal correlation coefficient. Furthermore, the energy saving efficiency variations were also in accord with the Kendall correlation coefficient. In addition, the amount of energy saved under the maximum (or minimum) values was always similar to the amount of energy wasted under the minimum (or maximum) values.

In practical applications, these parameters often change simultaneously. Considering the synergy among the different parameters, the SEGA was used to optimize the energy

Fig. 6 Energy saving efficiency

consumption level, and XGB served as the mapping function between the eight parameters and EC. In the SEGA, eight parameters are the genes of the population, and the EC calculated by XGB is the fitness function. The upper and lower bounds (UBs and LBs, respectively) of the eight input parameters were determined by the discharge standard and the extreme values contained in the historical data. Due to the randomness of the SEGA, several optimization steps were performed (Table 5).

In scenarios 1–3, the UBs and LBs of the eight parameters were set to investigate the resulting changes in the optimal EC. In scenario 1, the UBs of ETN and ETP were the maximum allowed values, which were 15 mg L^{-1} and 0.5 mg L^{-1} , respectively. The other UBs were the historical maximum, and the LBs were zeros (except that of ORP). The UBs of scenario 2 were the same as those of scenario 1, but its LBs were the historical minima. The UB of ORP in scenario 3 was set to 0, and the other boundary conditions of scenario 3 remained the same as those in scenario 2. The restriction imposed on the search area of ORP did not affect the final energy savings. However, the practical water quality and quantity were unfeasible to regulate, so their UBs and LBs were set as the mean values in scenario 4. The management regulation parameters were optimized by setting their boundary conditions according to the historical extreme values.

The DO and ORP probes are widely used in practical applications. Based on previous research, the DO concentration at the end of the aerobic tank was $1\text{--}5 \text{ mg L}^{-1}$ (Qiu et al. 2017), and to achieve biological phosphorus removal, the ORP in an anaerobic environment should be no larger

than -50 mV (Tae et al. 2005; Tang et al. 2012). The optimal parameters obtained from the GA were basically within reasonable ranges. According to the above results, 13–27% of the total energy consumption could be saved (with an average of 22%) by optimizing the management process, and the effluent could meet the discharge standard all the time. When the IFR was maximized, the IAN was close to the average value and the ANLR was high, the minimum energy consumption level could be obtained. Specifically, energy savings could be achieved by setting the management regulation parameters near the optimization results. For ORP, its optimal value could be achieved by adjusting the internal reflow rate. Flexibly opening and closing the air pumps and adjusting the air supply could make DO reach the value given by the GA. For the ANLR, due to the uncontrollable IAN, the difference between IAN and EAN was the determining factor. This showed that energy savings in WWTPs can be achieved by adjusting the operation parameters through the GA, which provides a simple and feasible energy saving strategy.

Conclusions

The energy consumption levels of WWTPs can be predicted and optimized by XGB and the SEGA. In terms of prediction, the XGB model achieved good performance, which was verified by a series of indicators, such as the R^2 , MAE, MAPE, and RMSE metrics. The most important parameter influencing energy consumption is the influent flow rate.

Table 5 Optimized parameters

| Parameter | Water quantity | | | Water quality | | | | Management regulation | | | | | Energy saving efficiency (%) |
|-----------------|---------------------------------------|----------------|--|---------------------------|---------------------------|---------------------------|---------------------------|--------------------------|----------|--|---------------------------|-----|------------------------------|
| | IFR (m ³ d ⁻¹) | Water quantity | | IAN (mg L ⁻¹) | ITP (mg L ⁻¹) | ETN (mg L ⁻¹) | ETP (mg L ⁻¹) | DO (mg L ⁻¹) | ORP (mv) | ANLR (kg m ⁻³ d ⁻¹) | EC (kWh m ⁻³) | | |
| Historical data | Max | 45,672 | | 63.85 | 17.66 | 17.98 | 0.52 | 3.649 | 46.52 | 0.022 | 0.896 | | |
| | Min | 21,238 | | 16.62 | 1.21 | 3.53 | 0.04 | 0.421 | -178.062 | 0.004 | 0.285 | | |
| Scenario 1 | Mean | 29,269.01 | | 36.683 | 5.659 | 8.238 | 0.182 | 0.951 | -40.191 | 0.01 | 0.544 | | |
| | UB | 45,678 | | 70 | 20 | 15 | 0.5 | 5 | 50 | 0.03 | | | |
| Scenario 2 | LB | 0 | | 0 | 0 | 0 | 0 | 0 | -50 | 0 | | | |
| | Optimized value | 43,492.767 | | 28.609 | 2.515 | 2.589 | 0.324 | 1.639 | 9.248 | 0.022 | 0.396 | 27% | |
| Scenario 3 | UB | 45,678 | | 70 | 20 | 15 | 0.5 | 5 | 50 | 0.03 | | | |
| | LB | 21,238 | | 16.62 | 1.21 | 3.53 | 0.04 | 0.421 | -178.062 | 0.004 | | | |
| Scenario 4 | Optimized value | 43,731.547 | | 29.107 | 1.789 | 4.806 | 0.328 | 1.651 | -0.917 | 0.023 | 0.395 | 27% | |
| | UB | 45,678 | | 70 | 20 | 15 | 0.5 | 5 | 0 | 0.03 | | | |
| Scenario 4 | LB | 21,238 | | 16.62 | 1.21 | 3.53 | 0.04 | 0.421 | -178.062 | 0.004 | | | |
| | Optimized value | 43,929.915 | | 28.934 | 2.482 | 4.067 | 0.361 | 1.572 | -3.395 | 0.021 | 0.395 | 27% | |
| Scenario 4 | UB | 29,269.01 | | 36.683 | 5.659 | 8.238 | 0.182 | 5 | 50 | 0.03 | | | |
| | LB | 29,269.01 | | 36.683 | 5.659 | 8.238 | 0.182 | 0.421 | -178.062 | 0.004 | | | |
| Scenario 4 | Optimized value | 29,269.01 | | 36.683 | 5.659 | 8.238 | 0.182 | 2.108 | -9.305 | 0.018 | 0.476 | 13% | |
| | Mean | | | | | | | | | | 0.422 | 22% | |

Therefore, compared with small-scale WWTPs, large-scale WWTPs with high IFR values need less EC and lower operating costs. In terms of optimization, 13–27% of the total energy consumption (with an average of 22%) could be saved by the optimized management regulation parameters obtained from the SEGA model. This research provides a convenient and reliable strategy for saving energy in WWTPs, which can be used in other treatment processes in practical applications.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability The authors do not have permission to share data.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bagherzadeh F, Nouri AS, Mehrani M-J, Thennadil S (2021) Prediction of energy consumption and evaluation of affecting factors in a full-scale WWTP using a machine learning approach. *Process Saf Environ Prot* 154:458–466. <https://doi.org/10.1016/j.psep.2021.08.040>
- Chang J, Jing Y, Geng Y, Song X (2021) Promote the low-carbon transformation of municipal sewage treatment industry and facilitate the realization of emission peak and carbon neutrality. *China Environ Prot Ind*. <https://doi.org/10.1016/j.scitotenv.2023.165201>
- Chen T, Guestrin C (2016) XGBoost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp 785–794
- Chu Y, Huang Z, Hahn J (2009) Improving prediction capabilities of complex dynamic models via parameter selection and estimation. *Chem Eng Sci* 64:4178–4185. <https://doi.org/10.1016/j.ces.2009.06.057>
- Das A, Kumawat PK, Chaturvedi ND (2021) A study to target energy consumption in wastewater treatment plant using machine learning algorithms. In: *Computer aided chemical engineering*. Elsevier, pp 1511–1516
- Ding C, Peng H (2003) Minimum redundancy feature selection from microarray gene expression data. In: *Computational systems bioinformatics*. CSB2003. *Proceedings of the 2003 IEEE bioinformatics conference*. CSB2003, pp 523–528
- El-Rawy M, Abd-Ellah MK, Fathi H, Ahmed AKA (2021) Forecasting effluent and performance of wastewater treatment plant using different machine learning techniques. *J Water Process Eng* 44:102380. <https://doi.org/10.1016/j.jwpe.2021.102380>
- Faegh M, Behnam P, Shafii MB, Khiadani M (2021) Development of artificial neural networks for performance prediction of a heat pump assisted humidification-dehumidification desalination system. *Desalination* 508:115052. <https://doi.org/10.1016/j.desal.2021.115052>
- Fan J, Li R (2001) Variable selection via nonconcave penalized likelihood and its oracle properties. *J Am Stat Assoc* 96:1348–1360. <https://doi.org/10.1198/016214501753382273>
- Hanna SM, Thompson MJ, Dahab MF et al (2018) Benchmarking the energy intensity of small water resource recovery facilities. *Water Environ Res* 90:738–747. <https://doi.org/10.2175/106143017X15131012153176>
- Huang H, Wei X, Zhou Y (2022) An overview on twin support vector regression. *Neurocomputing* 490:80–92. <https://doi.org/10.1016/j.neucom.2021.10.125>
- Ingdal M, Johnsen R, Harrington DA (2019) The Akaike information criterion in weighted regression of immittance data. *Electrochim Acta* 317:648–653. <https://doi.org/10.1016/j.electacta.2019.06.030>
- James G, Witten D, Hastie T, Tibshirani R (2021) *An introduction to statistical learning: with applications in R*. Springer, New York
- Khamis H (2008) Measures of association: how to choose? *J Diagn Med Sonogr* 24:155–162. <https://doi.org/10.1177/8756479308317006>
- Liu J, Dong B, Qian Z et al (2022a) Optimizing aeration pattern to improve nitrogen treatment performance of ditch wetlands in polder areas around Chaohu Lake. *China Ecol Eng* 183:106737. <https://doi.org/10.1016/j.ecoleng.2022.106737>
- Liu X, Yang J, Wang L, Wu J (2022b) Bayesian information criterion based data-driven state of charge estimation for lithium-ion battery. *J Energy Storage* 55:105669. <https://doi.org/10.1016/j.est.2022.105669>
- Ly QV, Truong VH, Ji B et al (2022) Exploring potential machine learning application based on big data for prediction of wastewater quality from different full-scale wastewater treatment plants. *Sci Total Environ* 832:154930. <https://doi.org/10.1016/j.scitotenv.2022.154930>
- Nourani V, Asghari P, Sharghi E (2021) Artificial intelligence based ensemble modeling of wastewater treatment plant using jittered data. *J Clean Prod* 291:125772. <https://doi.org/10.1016/j.jclepro.2020.125772>
- Picos-Benítez AR, Martínez-Vargas BL, Duron-Torres SM et al (2020) The use of artificial intelligence models in the prediction of optimum operational conditions for the treatment of dye wastewaters with similar structural characteristics. *Process Saf Environ Prot* 143:36–44. <https://doi.org/10.1016/j.psep.2020.06.020>
- Qiu W, Wu K, Jiang J et al (2017) Optimization of the A2/O technological parameters based on GA-ANN model. *J Harbin Inst Technol* 49:117–121. <https://doi.org/10.11918/j.issn.0367-6234.201607057>
- Sun J, Li J, Fujita H (2022) Multi-class imbalanced enterprise credit evaluation based on asymmetric bagging combined with light gradient boosting machine. *Appl Soft Comput* 130:109637. <https://doi.org/10.1016/j.asoc.2022.109637>
- Tae KH, Kim G-S, Shin S-W et al (2005) Application of ORP and pH as controlling factors in sequencing batch reactor. *KSCE J Civ Eng* 9:73–79. <https://doi.org/10.1007/BF02829061>
- Tang Y, Wu C, Yang H et al (2012) Regulation and control of oxidation reduction potential (ORP) of the activated sludge of modified Carrousel oxidation ditch and its influence on phosphorus absorption/release. *Ind Water Treat* 32:19–22. <https://doi.org/10.3969/j.issn.1005-829X.2012.03.005>

- Vaccari M, Foladori P, Nembrini S, Vitali F (2018) Benchmarking of energy consumption in municipal wastewater treatment plants—a survey of over 200 plants in Italy. *Water Sci Technol* 77:2242–2252. <https://doi.org/10.2166/wst.2018.035>
- Wan X, Li X, Wang X et al (2022) Water quality prediction model using Gaussian process regression based on deep learning for carbon neutrality in papermaking wastewater treatment system. *Environ Res* 211:112942. <https://doi.org/10.1016/j.envres.2022.112942>
- Wang J, Zhao X, Guo Z et al (2022) A full-view management method based on artificial neural networks for energy and material-savings in wastewater treatment plants. *Environ Res* 211:113054. <https://doi.org/10.1016/j.envres.2022.113054>
- Yang G, Wan L, Wang H, Yu D (2021) Energy consumption analysis of sewage treatment plant, measures and application of energy saving and consumption reduction. *Resour Econ Environ Prot*. <https://doi.org/10.16317/j.cnki.12-1377/x.2021.10.004>
- Yang Q, Wang Y, Cao X et al (2022) Research progress of carbon neutrality operation technology in Sewage treatment. *J Beijing Univ Technol* 48:292–305. <https://doi.org/10.11936/bjtxb2021090022>
- Yoshida H, Mønster J, Scheutz C (2014) Plant-integrated measurement of greenhouse gas emissions from a municipal wastewater treatment plant. *Water Res* 61:108–118. <https://doi.org/10.1016/j.watres.2014.05.014>
- Żyłka R, Dąbrowski W, Malinowski P, Karolinczak B (2020) Modeling of electric energy consumption during dairy wastewater treatment plant operation. *Energies* 13:3769. <https://doi.org/10.3390/en13153769>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.