processes

MDPI

Article

Optimization of an Industrial Circulating Water System Based on Process Simulation and Machine Learning

Yingjie Liu [1,2,*], Runjie Shao [1,2], Qing Ye [1,2], Jinlong Li [1,2], Ruiyu Sun [3] and Yifei Zhai [3]

[1] School of Petrochemical Engineering, Changzhou University, Changzhou 213164, China; shao_runjie@163.com (R.S.); huagonglou508@126.com (Q.Y.); lijinlong@cczu.edu.cn (J.L.)
[2] Jiangsu Key Laboratory of Advanced Catalytic Materials and Technology, Changzhou 213164, China
[3] The Yellow River Delta Chambroad Institute Co., Ltd., Binzhou 256500, China; sunruiyu@chambroad.com (R.S.); yifei.zhai@chambroad.com (Y.Z.)
* Correspondence: liuyj@cczu.edu.cn

Abstract: As an important part of industrial production, the optimization of circulating water systems is of great significance for improving energy efficiency and reducing operating costs. However, traditional optimization methods lack real-time and dynamic adjustment capabilities and often cannot fully cope with the complex and changeable industrial environment and energy demands. Advances in computer technology can enable people to use machine learning models to process information and data and ultimately help simplify simulation and optimization. In this paper, the circulating water system of a Fluid Catalytic Cracking (FCC) unit is optimized and evaluated based on process simulation and machine learning, adopting 284 sets of industrial operating data. The cooler network of the system is modified from a parallel structure to a series mode, and the effect is clarified using the ASPEN HYSYS software V12. Meanwhile, the fan power of the cooling tower is predicted by employing an optimized Gradient Boosting Regression (GBR) model, and the influence of the parallel-to-series transformation on the fan power is discussed. It is shown that the computer modeling results are in coincidence with the industrial data. Converting the parallel design to a series arrangement of the cooler network can significantly decrease the water consumption, with a reduction of 11%. The fan power of the cooling tower is also reduced by 8% after the optimization. Considering the changes in both water consumption and fan power, the saved total economic cost is 8.65%, and the decreased gas emission is 2142.06 kg/h. By building the optimization prediction system, the real-time sequencing and monitoring of equipment parameters are realized, which saves costs and improves process safety.

Keywords: cooler network; cooling tower; process simulation; machine learning

check for updates

Academic Editor: Olympia Roeva

Received: 3 January 2025
Revised: 20 January 2025
Accepted: 23 January 2025
Published: 24 January 2025

Citation: Liu, Y.; Shao, R.; Ye, Q.; Li, J.; Sun, R.; Zhai, Y. Optimization of an Industrial Circulating Water System Based on Process Simulation and Machine Learning. *Processes* **2025**, *13*, 332. https://doi.org/10.3390/pr13020332

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The circulating water system is part of the cooling utility system that uses water as a cooling medium to remove the waste heat and to recycle the water. This system is mainly composed of the cooler network, pump network, pipe network and cooling tower, and those elements' operation quality is directly related to the long-term stable operation and economic benefit of the enterprise [1]. The above networks are not independent but affect each other. For example, changes in the operating parameters or equipment connection mode will influence the water consumption, the pump efficiency, and the cooling tower power. Usually, the initial network is established based on the design conditions, but with the operation of the system, some parameters may deviate from the design values and lead

to a reduction in the system's overall efficiency. Thus, the establishment of simulation and optimization methods is of great importance for the improvement of the entire circulating water system.

Many studies have been conducted to improve the circulating water system and conserve energy. Among those studies, optimizing the network connection mode is an effective and popular method because of its low cost and high realizability. For example, Wang et al. [2] proposed a two-step method to convert a parallel configuration of the cooler network to a series–parallel structure. The proposed modification could reduce the energy consumption of the system without changing the cooler structures. Ma et al. [3] proposed a novel multi-loops pump network and an updated main-auxiliary pump network in a cooling water system. In this methodology, the cooler and pump networks, cooling tower and pipeline layouts were optimized simultaneously. Zhang et al. [4] proposed a mathematical model based on the superstructure, under which the cooler pipe network and pump pipe network costs were reduced by 13% and 32%, respectively. Jose et al. [5] used the pressure drop of each cooler as an optimization variable and established a mathematical model. Three examples were used to verify the feasibility of the method in order to find the optimal configuration of the circulating water system. Müller et al. [6] used the MINLP model to address the complexity of pump system design, with the aim of minimizing life cycle costs. The results showed that a life cycle cost savings of up to 21% can be achieved while increasing the net efficiency from 47.2% to 57.8%.

With the development of artificial intelligence, increasing attention has been paid to the study of circulating water systems using machine learning methods. Barigozzi et al. [7] developed a MATLAB algorithm to optimize the net electric power as a function of the way in which the condensation is operated. The optimization of the thermal cycle performance was achieved by fixing the flow rate, temperature and pressure of the steam entering the high-pressure turbine. Liang et al. [8] proposed a MINLP model that considers the fouling in the pipeline, dynamic concentration cycle, and variable frequency drive to optimize the synergy among the heat transfer, pressure drop and fouling. By optimizing the concentration cycle of the circulating water system, water saving and scaling control can be achieved with significant energy/water saving effects. Song et al. [9] proposed a Back-Propagation (BP) neural network to predict 638 sets of field test data in 36 different natural draft counterflow wet cooling towers (NDWCTs) in a power plant and developed a three-layer BP neural network model with a structure of 8-14-2. The results showed that the model has good prediction accuracy for the heat and mass transfer performance of NDWCTs at different scales. Liang et al. [10] proposed a Genetic Algorithm (GA) considering a variable frequency drive (VFD) to optimize the industrial circulating cooling water system to obtain accurate operating parameters. The interaction between the pump and the cooler networks was examined. The results showed that the model can determine the accurate operating parameters of the pump system and valves. Zhang et al. [11] used a coupling algorithm of an artificial neural network (ANN) optimized by establishing the coupling algorithm of the GA–BP neural network with the heat transfer model of the condenser and the air-cooled heat exchanger to obtain the air mass flow rate into the natural draft dry cooling tower (NDDCT). By adjusting the circulating cooling water mass flow to the optimum value, at least 16,515 kWh of circulating pump power consumption could be saved. Bueso et al. [12] proposed a machine learning method based on the Multilayer Perceptron to estimate the thermal performance of cooling towers used in the desalination process. These studies validated the accuracy and efficiency of machine learning methods in process predictions. Table 1 summarizes the optimization methods used in the study.

Table 1. Optimization methods in research.

Object	Method	Advantages	Limitations
cooler network, pump network, pipe network and cooling tower	Two-step method	Save water resources and reduce energy consumption of the system	High cost of pipe network reconstruction
	Novel multi-loops pump network		
	Mathematical model with the pressure drop		
	MINLP		
	MATLAB	Intelligent optimization	Large and diverse data sets are required
	BP neural network		
	Genetic Algorithm		
	ANN		
	Multilayer Perceptron		

The above investigation shows that machine learning is an effective method for solving practical process problems, including solving the problem of cooling tower fan power prediction in process operations. In this paper, the optimization of the circulating water system in a Fluid Catalytic Cracking (FCC) unit at a refinery based on the industrial operating data is conducted, and a prediction system is designed and built. The cooler network is reformed to decrease the water consumption. Next, the fan power of the cooling tower is predicted using an optimized Gradient Boosting Regression (GBR) model. Based on the above studies, the saved water amount in the cooler network and the fan power in the cooling tower after transformation are calculated, and the total economic cost and gas emissions before and after optimization are discussed. The following expected goals are achieved:

- The optimized cooler network scheme is implemented to reduce water consumption.
- The industrial data are analyzed, cleaned and normalized to achieve the purpose of data visualization.
- The fan power of the cooling tower is predicted, which could be used to improve operational efficiency.
- The optimization system is constructed to realize the monitoring of the operation status.

2. Optimization of the Cooler Network

The main purpose of this section is to retrofit the coolers of the circulating cooling water system in an FCC unit at a refinery. These arrangements are usually based on parallel configuration. This design method aims to maximize the cooling water savings and optimize the cooler network structure.

2.1. Optimization Method

In the conventional cooler network, the circulating water delivered from the supply system to the devices generally enters the coolers in the parallel structure, as plotted in Figure 1a. After heat exchanging, the circulating water is collected and returned to the cooling tower. This parallel design of the cooler network has shown the disadvantage of large water consumption and low efficiency of the cooling tower.

In this study, the cooler network of the circulating water system is reformed from the parallel structure to the series scheme, as shown in Figure 1. In the series network structure, the circulating water flows through more than one cooler to reduce its consumption.

In most cases, the parallel-to-series modification of the cooler network only requires some additional cooler connecting pipes, making the work relatively simple, inexpensive and welcomed. The design steps are as follows [2]:

(1) Determine the coolers to be modified;
(2) Determine the series sequence of the coolers;
(3) Calculate the fresh cooling water demand for each cooler;
(4) Calculate the cooling water demand for the entire system;
(5) Determine the network structure of the transformation.

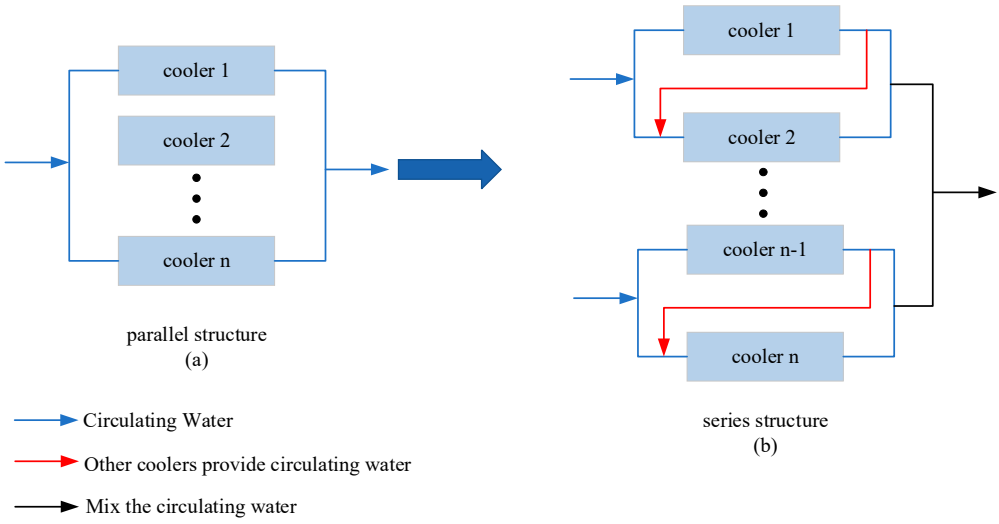


Figure 1. The parallel and series structures of the cooler network.

2.2. Optimization Object

The investigated circulating water system is located in an industrial FCC unit, as shown in Figure 2a. The investigated unit contains a reaction–regeneration system and a fractionation system in which the hot streams, such as naphtha, light diesel and rich gas, enter the coolers in the water using system, as plotted in Figure 2c. Figure 2b illustrates the water supply system, which is composed of a 15,000 m^3 natural draft circulating water tower V-100 and six circulating water pumps, with a total flow rate of 5300 m^3/h. The cooling water in the water supply system enters the coolers in the water using system through different pumps. In the water using system in Figure 2c, the cooling water enters seven coolers and then enters a cooling tower T-101 to take away the obtained heat. Next, the water is returned to the water supply system.

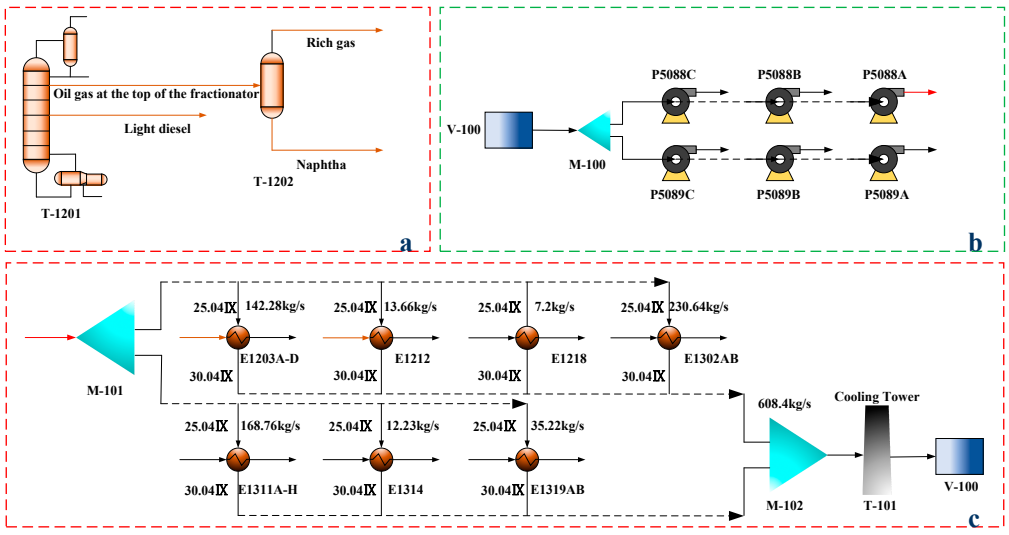


Figure 2. Circulating water system ((**a**)—FCC unit, (**b**)—water supply system, (**c**)—water using system).

The details of the coolers in the water using system are listed in Table 2.

Table 2. Details of the coolers in the parallel network.

Heat Exchanger	Stream		Inlet Temperature	Outlet Temperature	Heat Load/kW	Cooling Water Volume/kg/s	Heat Flow Quantity/kg/s
E1203A-D	Cold Hot	Circulating water Oil gas at top of fractionator	25.04 52	30.04 40	3058.58	142.28	46.67
E1212	Cold Hot	Circulating water Lean absorption oil	25.04 55	30.04 40	294.47	13.66	10.56
E1218	Cold Hot	Circulating water Naphtha	25.04 40.56	30.04 36	155.32	7.20	18.11
E1302AB	Cold Hot	Circulating water Rich gas	25.04 58	30.04 40	4975.4	230.64	83.86
E1311A-D	Cold Hot	Circulating water Liquified petroleum gas	25.04 44	30.04 40	1819.59	84.38	33.58
E1311E-H	Cold Hot	Circulating water Liquified petroleum gas	25.04 44	30.04 40	1819.59	84.38	33.58
E1314	Cold Hot	Circulating water Supplementary absorbent	25.04 40	30.04 36	241.85	12.23	29.86
E1319AB	Cold Hot	Circulating water Light diesel	25.04 123	30.04 40	759.50	35.22	3.87

2.3. Optimization Scheme

The temperature–enthalpy diagram of the heat transfer process for the coolers is shown in Figure 3. The abscissa is the enthalpy value and the ordinate is the stream temperature. The upper solid line with higher temperatures is the process stream line to be cooled, and the lower solid line is a special circulating water line, which is completely parallel to the process stream line and has the same enthalpy. The temperature difference between the two lines is the pinch point temperature difference. This special circulating water line is called the limit cooling water line, which defines the limit value of the cooling water. The actual cooling water line should not be higher than this limit line to avoid the heat transfer temperature difference being smaller than the pinch point temperature difference [13].

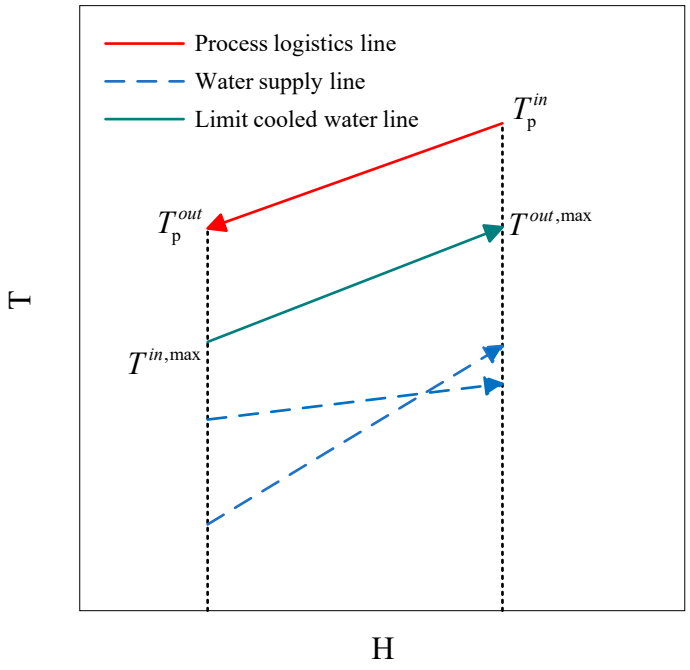


Figure 3. Temperature–enthalpy diagram of cooler.

In order to achieve the global optimization of the cooling water network, the water consumption of the entire system must be considered as a whole. Figure 4 shows the composite temperature–enthalpy diagram of the FCC circulating water system.

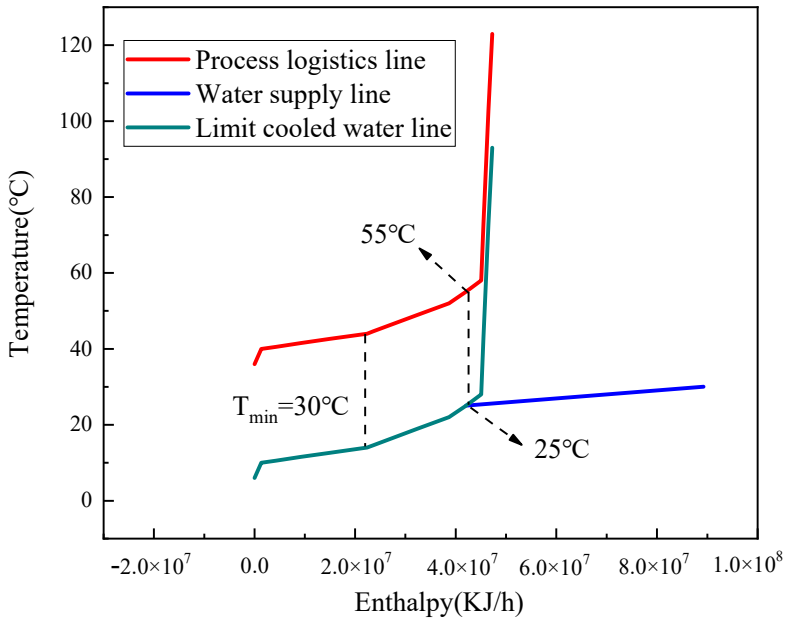


Figure 4. The process heat combination curve.

The red curve in the figure represents the process logistics line, the green curve refers to the limit cooled water line, and the blue curve is the water supply line. The temperature difference at the pinch point is set to 30 °C. In order to minimize the amount of cooling water, the outlet temperature should be increased as much as possible, and the slope of the water supply line should be increased. When the slope of the water supply line increases to a point where it begins to coincide with the limit compound curve, the outlet temperature reaches a maximum, and the water consumption reaches a minimum. The pinch temperature at this point is determined to be 55 °C.

Table 2 shows that the limit inlet temperatures of the coolers E1302AB, E1319AB and E1212 are higher than the pinch point temperature, which means the cooling water here could be from the outlet of other coolers rather than directly from the cooling tower. The limit outlet temperatures of the coolers E1311A-H, E1314, E1203AD and E1218 are lower than the pinch point temperature, indicating that the water here could be utilized before going back to the cooling tower. Therefore, the influence of the pinch point on the modification of the cooler network should be taken into account to obtain the optimized scheme.

Based on the pinch technology and considering the principle of short pipeline transformation and smooth flow, the parallel system in Figure 2c is optimized to the series structure. Three modifications were conducted, as shown in Figure 5. To meet the heat exchanging requirements, the minimum heat transfer temperature difference is set to 5 °C.

There are three modifications for the existing network: (1) The circulating water steams from cooler E1212 and E1218 are combined with the fresh circulating water to supply water for E1203A-D. (2) The circulating water from cooler E1314 is mixed with the fresh water and enters E1311E-H. (3) The circulating water from cooler E1319AB and the fresh water flow into E1311 A-D together. The optimization scheme can decrease the circulating water volume of coolers E1203A-D, E1311E-H and E1311A-D and reduce the water load of the entire system.

In order to investigate the transformation effect of the optimization system, the process simulation of the modified scheme is carried out using ASPEN HYSYS, as shown in Figure 6.

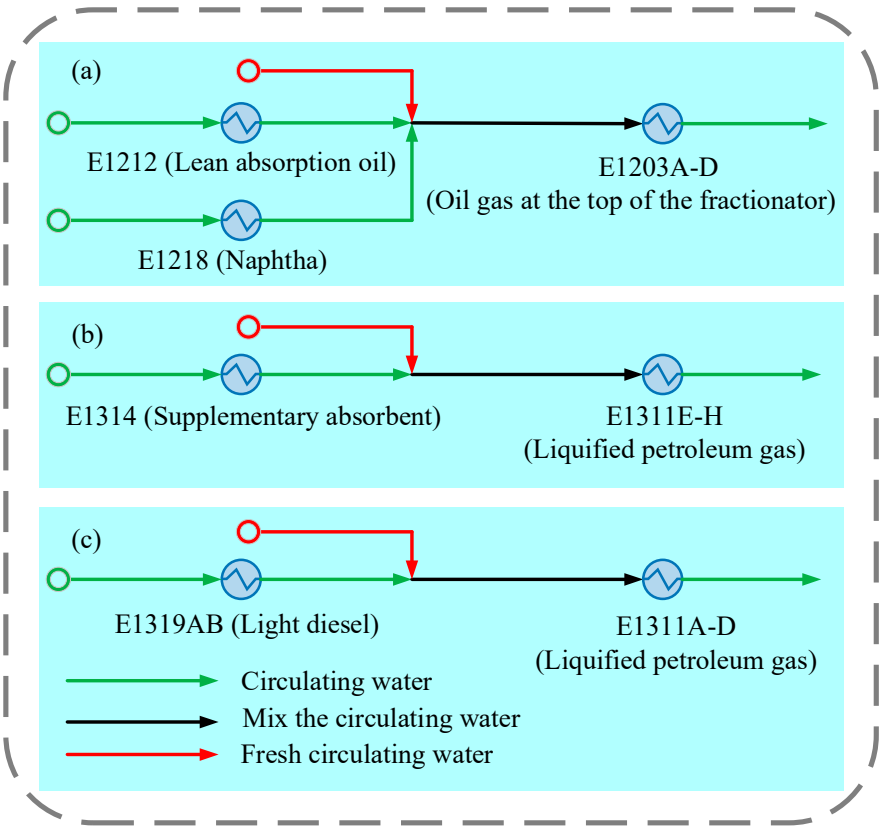


Figure 5. Schematics of the series structure modifications ((**a**)—E1212, E1218 and E1203A-D, (**b**)—E1314 and E1311E-H, (**c**)—E1319AB and E1311 A-D).

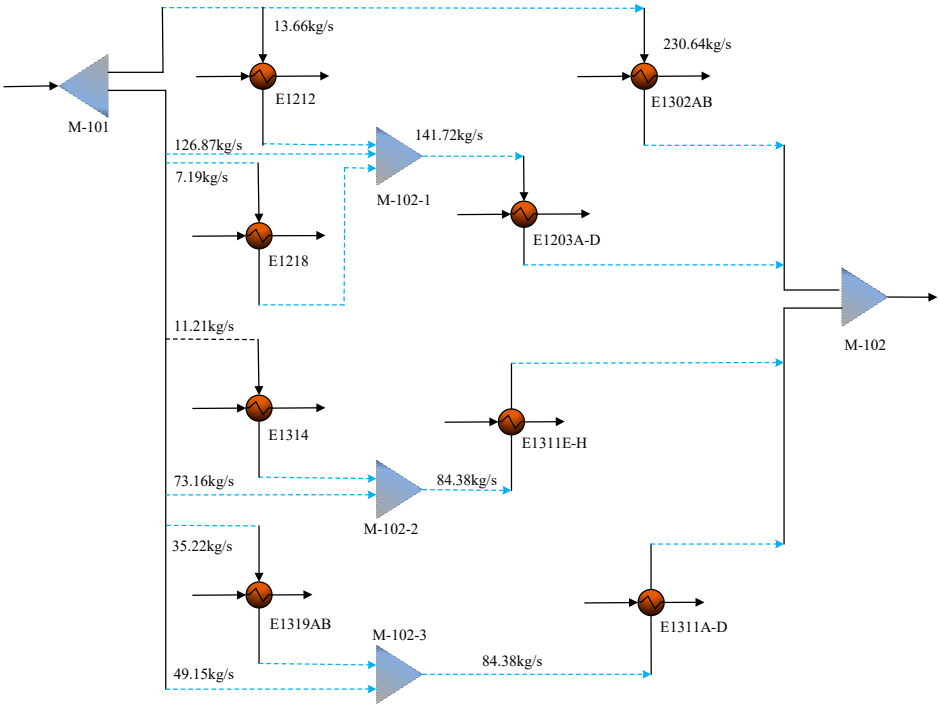


Figure 6. Series scheme of cooler network after optimization.

The water flow rates in the series network obtained by simulation are shown in Table 3 compared with the calculated values using Formula (1).

$$m = \frac{Q}{C_p \times \Delta t} \quad (1)$$

where Q is the heat load, C_p stands for the specific heat capacity of the water, and Δt represents the temperature difference between the inlet and outlet of the cooling water.

Table 3. Comparison of simulated and calculated values.

Cooler Number	Name	Simulated Values (kg/s)	Calculated Values (kg/s)
E1203A-D	Fractionating Tower Top Oil and Gas Cooler	141.72	146.11
E1311A-D	Tower Top Oil and Gas Cooler	84.36	86.86
E1311E-H	Tower Top Oil and Gas Cooler	84.36	86.86

The results show that the error between the simulated and the calculated values is approximately 3%, indicating the reasonability of the simulation. The difference may be due to the fact that the theoretical calculation usually assumes the steady state and unchanged properties of the fluid, while the process simulation can dynamically consider the effect of the operating conditions on the fluid properties.

The simulated total water consumption of the original and optimized systems is 608.4 kg/s and 541.11 kg/s, respectively. The water reduction is 11%, which demonstrates the feasibility of the optimization.

3. Prediction of the Cooling Tower Fan Power

In Figure 2c, the return water after heat exchanging in the cooler network is cooled in the cooling tower. The real time prediction and controlling of the fan power of the cooling tower is important, as the operation of the tower has a great impact on the energy conservation of the entire system. In this section, the fan power of the cooling tower is predicted by adopting the machine learning approach.

3.1. Algorithm Introduction

Six machine learning algorithms are employed to obtain the optimal model, as shown in the following:

(1) Bayesian Regression (BR) Model [14]: This model is a probabilistic framework based on Bayes' theorem and is intended for prediction and classification tasks. It integrates existing prior knowledge and data sets, and it constantly adjusts the probability model by applying Bayes' theorem to predict future events or unknown variables.

(2) Linear Regression (LR) Model [15]: This is a widely used technique in machine learning that aims to predict the linear relationship between a continuous target variable (also known as the response variable) and one or more independent variables.

(3) Elastic Network Regression (EN) Model [16]: The EN model performs well when dealing with highly correlated problems, especially when the number of independent variables is large and the dependent variables are strongly correlated.

(4) Support Vector Regression (SVR) Model [17]: Unlike traditional regression methods, this model's core idea is to find a function that can be as close as possible to the data while ensuring that the deviation between the predicted value and the true value remains within acceptable limits.

(5) Gradient Boosting Regression (GBR) Model [18]: The basic idea of the algorithm is to construct the prediction model through an iterative process, trying to fit the negative

gradient direction of the target variable in each step, to reduce the loss function value of the current model.

(6) Random Forest Regression (RF) Model [19]: The algorithm performs the regression task by integrating multiple decision trees. Its core idea is to aggregate the prediction results of multiple decision trees to determine the final output rather than relying solely on the prediction of a single decision tree.

Four evaluation indicators are used to evaluate the accuracy of the models, i.e., the Explained Variance (EV), Mean Absolute Error (MAE), Mean Square Error (MSE) and R-squared (R^2). The evaluation index is calculated by Formulas (2)–(5), where y_{true} is the true value, y_{pre} is the predicted value, and n stands for the amount of data.

$$\mathrm{EV}\left(y_{true}, y_{pre}\right) = 1 - \frac{\mathrm{Var}(y_{true} - y_{pre})}{\mathrm{Var}(y_{true})} \tag{2}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_{true}^{(i)} - y_{pre}^{(i)}\right| \tag{3}$$

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(y_{true}^{(i)} - y_{pre}^{(i)}\right)^2 \tag{4}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_{true}^{(i)} - y_{pre}^{(i)}\right)^2}{\sum_{i=1}^{n}\left(y_{true}^{(i)} - y_{ave}^{(i)}\right)^2} \tag{5}$$

3.2. Data Processing

Data processing is a crucial pre-processing procedure used to obtain high-quality data sets. It includes two processes, i.e., data cleaning and data normalization.

(1) data cleaning

The main purpose of data cleaning is to eliminate the outliers in the data and avoid the influence of wrong data on the accuracy of the prediction. The operating data of six variables are collected from the factory, including the flowrate of the circulating water, the temperature of the water going in and out of the cooling tower, the daily average ambient temperature, the daily relative humidity of the air and the fan power of the cooling tower. Each variable has 284 pieces of data, as shown in Table 4.

Table 4. Raw data set of cooling tower.

Data Name	Range	Data Size
The flowrate of the circulating water/m3/h	[5500–6900]	284 sets
The temperature of the water going into cooling tower/°C	[10.5–31]	
The temperature of the water going out of cooling tower/°C	[13–35]	
The daily average ambient temperature/°C	[−14–33]	
The daily relative humidity of the air/%	[21–100]	
The fan power of the cooling tower/kW	[0–405]	

A box plot analysis is used to visualize the data [20], providing the median, quartile and outlier information of the data. The definition formula of the outlier is shown in Formula (6). The box plots of the selected parameters are plotted in Figure 7. It is clear that

there is an outlier in the dataset of the temperature of the water out of the cooling tower. Thus, this data set is deleted.

$$x_{out} > Q_3 + 1.5 \times IQR \text{ or } x_{out} < Q_1 - 1.5 \times IQR \quad (6)$$

where x_{out} represents the outlier, Q_3 is the value at the 75% position after the data are sorted from small to large, Q_1 is the value at the 25% position after the data are sorted from small to large, and IQR stands for the difference between Q_3 and Q_1.

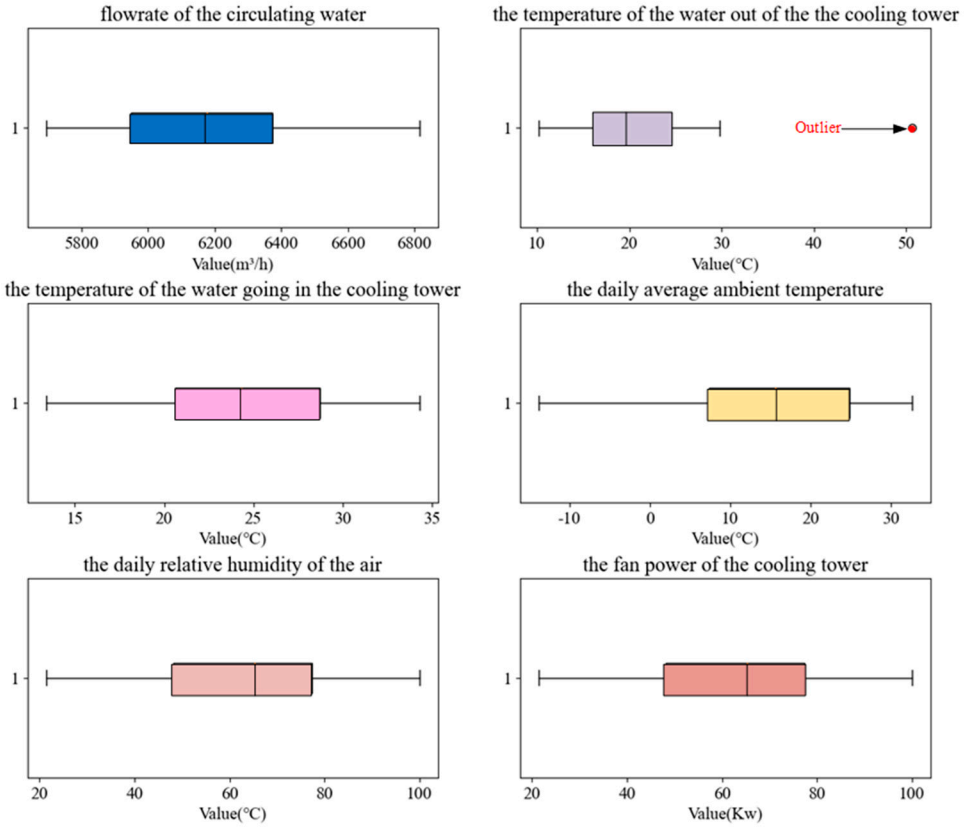


Figure 7. Box plots of the selected parameters.

(2) Normalization

As the collected data have different sizes and data distribution ranges, direct regression training may exaggerate some variables and affect the accuracy of the output target. In order to reduce the subsequent training error, Equation (7) is usually used to normalize the data.

$$x_i' = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (7)$$

where x_i and x_i' represent the values before and after data normalization. x_{max} and x_{min} are the maximum and minimum values in the sample data, respectively.

After the data normalization, 70% of the retained 283 sets of data are used as the training data, and the other 30% are reserved as the test data. The fan power of the cooling tower is set as the target function. The other five parameters are determined as the input variables to carry out the subsequent prediction.

3.3. Algorithm Selection

The six machine learning models introduced in 3.1 are employed to predict the cooling tower fan power. Figure 8 plots the comparisons between the predicted values and the industrial values. It is shown that the fitting effect of the EN model is the worst. The

coincidence degree between the industrial values and the predicted values of the GBR model and the RF model is higher than that of the other algorithms.

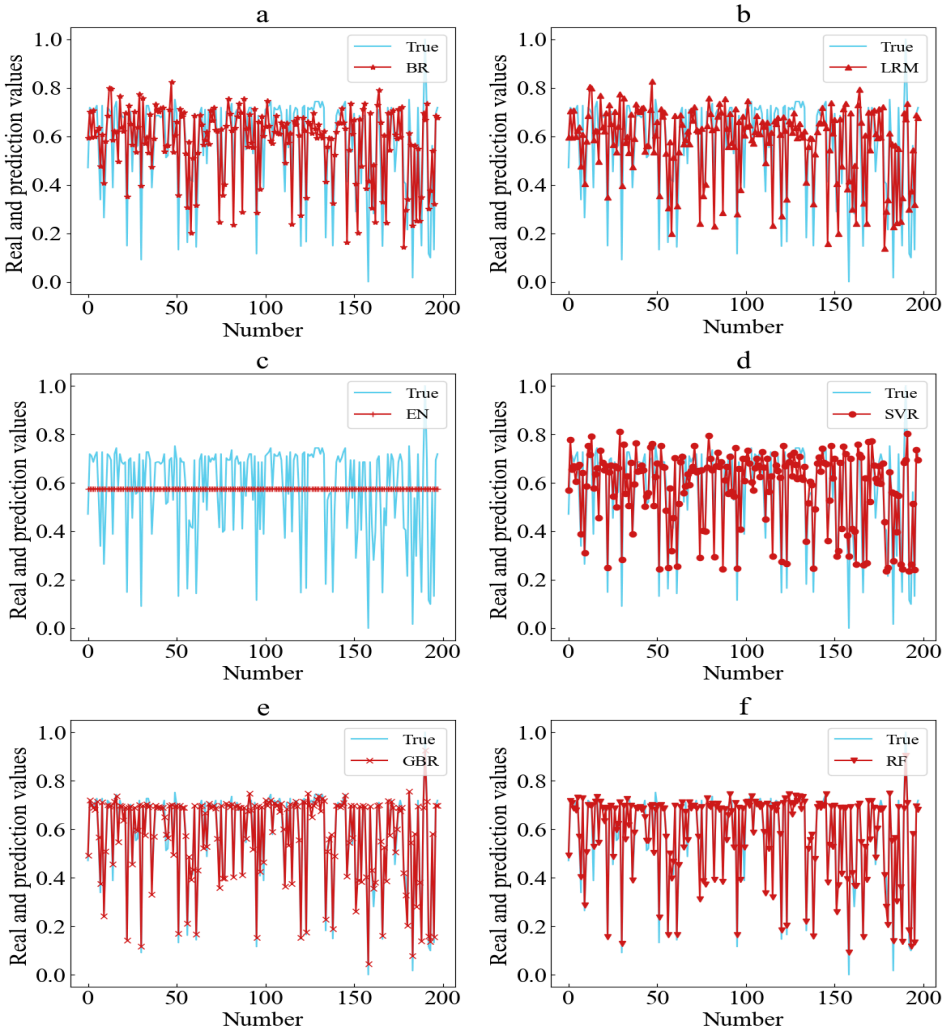


Figure 8. The predicted values versus the industrial values (the blue lines represent the true values, and the red lines represent the predicted values of the (**a**)—BR, (**b**)—LRM, (**c**)—EN, (**d**)—SVR, (**e**)—GBR and (**f**)—RF).

The evaluation indexes of the six prediction algorithms are shown in Table 5. Compared to the other models, the GBR model demonstrates the highest EV and R^2, the lowest MSE and a relatively low MAE. Thus, it can be concluded that the GBR model is superior to the other models, and it is selected as the optimal model for the subsequent prediction.

Table 5. Evaluation indexes of the six prediction models.

Model	EV	MAE	MSE	R^2
BR	0.638	0.088	0.013	0.639
LR	0.640	0.640	0.013	0.639
EN	0.000	0.154	0.036	0.000
SVR	0.768	0.069	0.008	0.767
GBR	0.980	0.019	0.001	0.981
RF	0.965	0.021	0.001	0.964

3.4. Model Prediction

3.4.1. Prediction Method

The GBR model is an ensemble learning method that continuously improves the performance of the model by iteratively training a series of weak learners [21]. Its basic idea is to construct the next model by fitting the difference between the actual value and the predicted value to gradually reduce the prediction error of the model on the training data. The prediction steps are shown in Figure 9.

(1) Select the data set, the appropriate target variables and the characteristics;

(2) Initialize the model. The model function is shown in Formula (8), where y_i is the actual value of the target variable corresponding to each sample in the training set, γ is the predicted value of the model, and $L(y_i, \gamma)$ represents the loss function, which is a standard for measuring the difference between the predicted value of the model and the true value;

$$F_0(x) = \text{argmin}\sum_{i=1}^{n} L(y_i, \gamma) \tag{8}$$

(3) Calculate the residual error, as shown in Formula (9), where $F_{m-1}(x)$ is the current predicted value;

$$b_m = \frac{\partial L[y, F_{m-1}(x)]}{\partial F_{m-1}(x)} \tag{9}$$

(4) Take the residual error calculated in the previous step as a target variable, and fit a new base learning machine;

(5) Multiply the weight of the new base learner by a learning rate and then add it to the current model;

(6) Stop the training if the preset maximum number of iterations is reached or the model error is converged. Otherwise, return to (3) to continue the iteration;

(7) Output the final trained model when the stop condition is met.

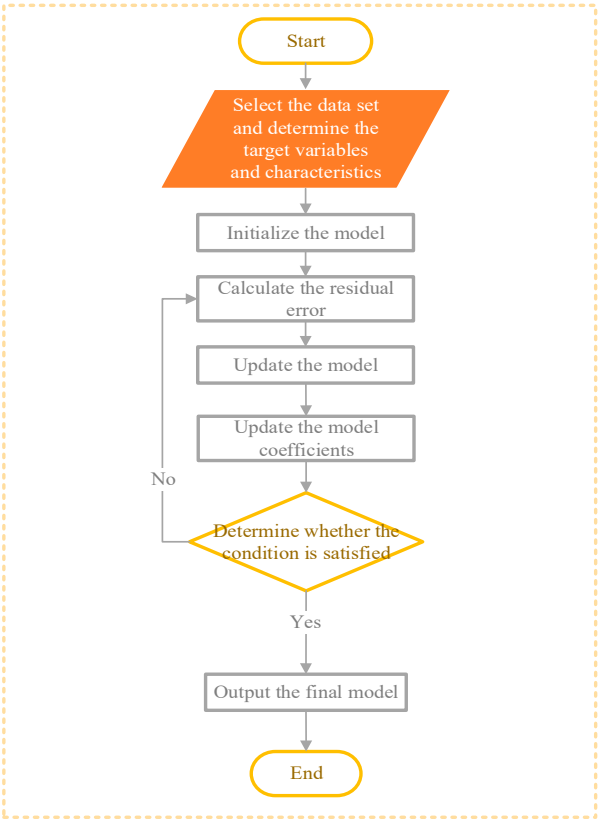


Figure 9. Flow chart of GBR model.

3.4.2. Hyperparameter Optimization

A grid search is a systematic hyperparameter tuning method that searches the predefined hyperparameter space exhaustively to find the best hyperparameter combination. The main parameters affecting the GBR model are the estimators, max depth, min samples split, min samples leaf and learning rate. Since the values of the model parameters are unknown, we set the approximate range of each parameter, as shown in Table 6.

Table 6. Initial parameters of GBR model.

Parameter	Range
Estimators	[20–400]
Min samples split	[1–5]
Max depth	[1–3]
Min samples leaf	[1–5]
Learning rate	[0.1–0.5]

Based on the setting of these parameters, the grid search is added to the GBR model, and the number of cross-validations is set to 10 to perform different parameter combinations to obtain the test scores of multiple groups. Table 7 shows the results of the partial cross-validation. At the same time, the MSE loss function curve is visualized to observe the changes in the scores of the hyperparameter combinations, as shown in Figure 10.

Table 7. Results of partial cross-validation.

Mean Fit Time	Std Fit Time	Std Test Score	Rank Test Score
0.005901	0.000539	0.004549	74
0.011967	0.000695	0.004286	71
0.021208	0.000403	0.003527	52
0.005646	0.000436	0.004549	73
0.011701	0.000779	0.004286	70
0.016838	0.000392	0.001732	1
0.008428	0.005401	0.002613	14
0.017239	0.000933	0.003556	43
0.031380	0.000436	0.002732	8
0.031380	0.000522	0.002613	13

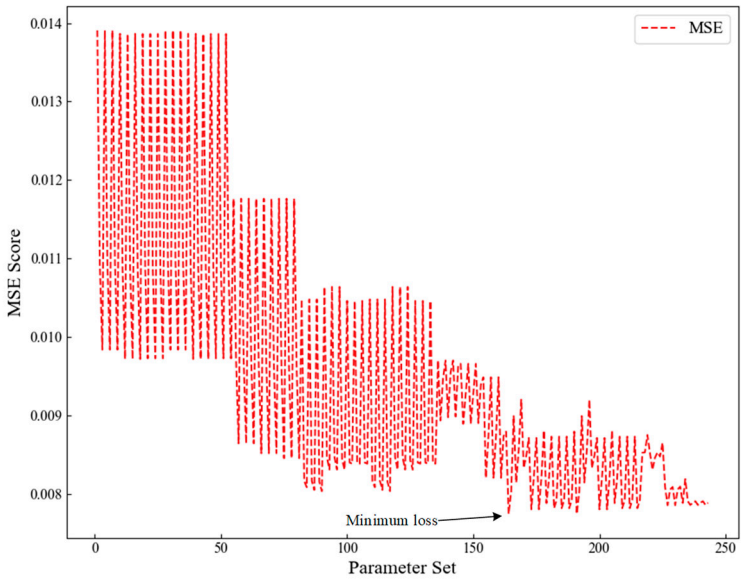


Figure 10. MSE scores of GBR model with different parameter sets.

In the process of the model training, the optimal values of the relevant parameters of the model are found using a single variable method, as shown in Table 8. The results show that the MSE score of the GBR model is reduced from 0.005814 to 0.004900, and R^2 is increased from 0.835860 to 0.897068, indicating that the accuracy of the model is further improved.

Table 8. Optimal parameters of GBR model [22].

Parameter	The Optimal Value
Estimators	100
Min samples split	2
Max depth	3
Min samples leaf	1
Learning rate	0.1

3.4.3. Prediction Results

The fan powers at different conditions are predicted by the selected model after the hyperparameter optimization and then compared with the test data, as shown in Figure 11. It can be seen that the predicted values coincide with the actual data, indicating the accuracy of the prediction.

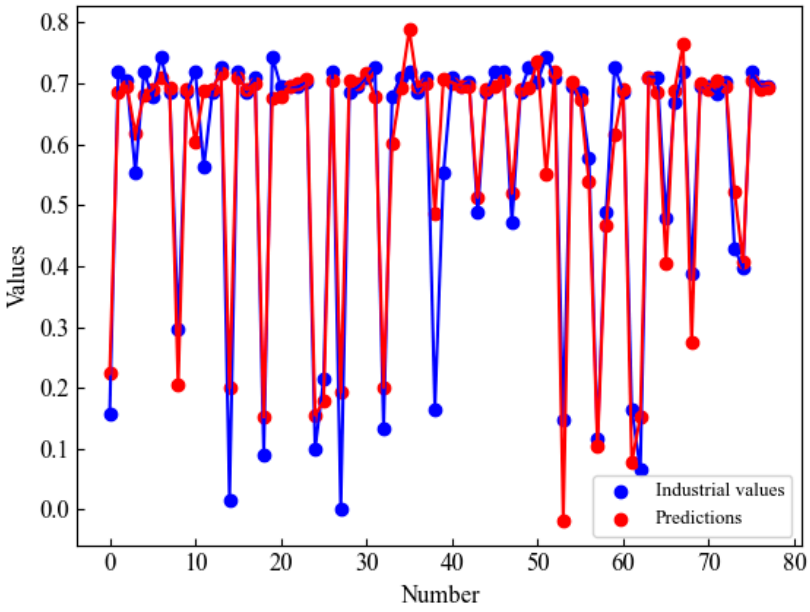


Figure 11. A comparison of the predicted and the industrial values of the fan power.

It is predicted that the total power of the fan power before the optimization is 65,138.54 kW, which is reduced to 59,760.50 kW after the optimization. The fan power reduction is approximately 8%.

The system maintenance interface is developed by using this model. In this interface, to evaluate the influence of the parameters on the stable operation of the cooling tower, the input variables affecting the fan power are sorted, as shown in Figure 12. It can be seen that the daily average ambient temperature has the greatest impact on the fan power, while the circulating water temperature entering the cooling tower has the smallest impact. By ranking the input variables, we can identify the factors with a greater impact on the fan power and prioritize these factors to maximize the power output, and it is possible to develop a more effective preventive maintenance program.

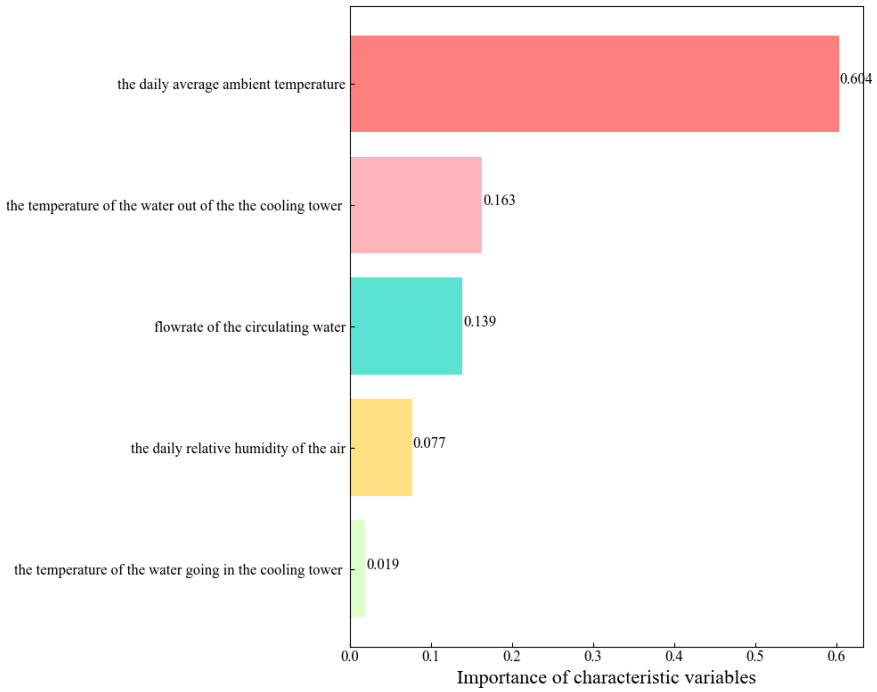


Figure 12. Feature variable ordering.

3.5. Comparison with Literature

In this section, we evaluate the performance of our model by comparing the performance of the GBR model with the findings in the literature on predictive cooling towers. Our goal is to provide valuable insights into the ongoing body of knowledge.

Many researchers use the Root Mean Squared Error (RMSE) [23,24] to evaluate the prediction accuracy, and the prediction results using various methods in different studies are shown in Table 9.

Table 9. Comparison of the prediction results of cooling towers in the literature.

Model	RMSE
ANN [25]	0.178
Poppe [26]	0.17
SVM [27]	0.201
ANN-MLP [28]	0.56
Optimized GBR	0.07

It can be seen from the table that the GBR model has good prediction accuracy in the prediction of different performances of cooling towers, which is of great significance in solving practical problems and promoting technological progress.

4. Economic Analysis and Environmental Evaluation

4.1. Economic Analysis

To evaluate the economic benefit, the water and electricity costs saved by the parallel-to-series modification are calculated. The water cost is estimated based on the local water price of 3.35 yuan/m^3. The electricity cost is calculated by Formula (10) [29]. The calculated results are shown in Figure 13.

$$Electricity\ cost = 16.8 \times (hp/0.6) \times 8000 \times 3.6/1000 \quad (10)$$

where hp represents the horsepower of the compressor(kW).

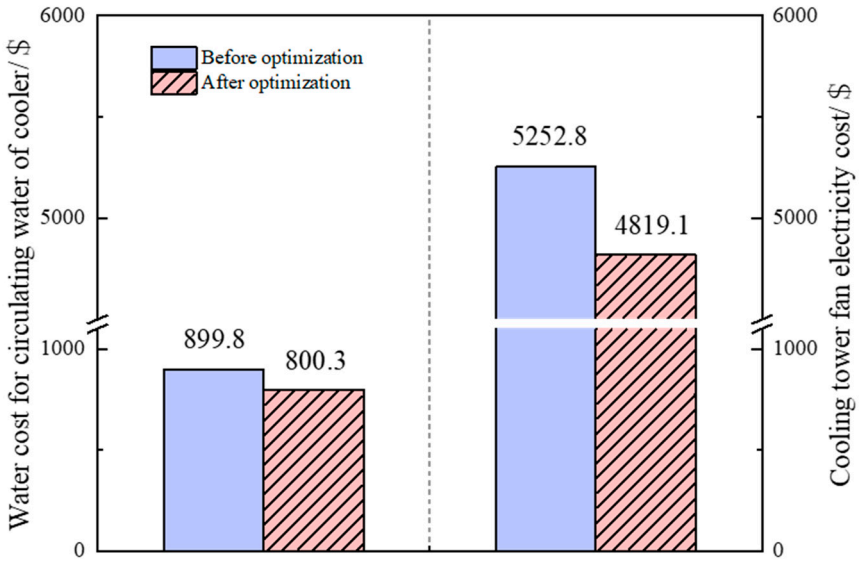


Figure 13. Economic comparison before and after optimization.

It can be seen that the water cost is reduced from $8.99 million/year to $8 million/year, and the electricity cost is reduced from $52.52 million/year to $48.19 million/year. The total saved economic cost is $5.32 million/year, approximately 8.65% of that of the former system.

4.2. Environmental Evaluation

Gas emissions are the result of steam and electricity consumption, which leads to pollution. Therefore, the environmental performance of the optimization prediction can be evaluated by the gas emission as a strong indicator [30]. The FCC process contributes to a large share of the total gas emissions in refineries. Evaluating the environmental footprint of the FCC process via its gas output is crucial for process optimization strategies and ensuring the long-term sustainability of the facility.

Gas emissions mainly include the CO_2 emission and other gas (SO_2 and NO_x) emissions. The methodology for calculating the CO_2 emission is detailed in Equations (11) and (12) [29,31].

$$CO_2 = \left(\frac{Q_{energy}}{NHV}\right) \times \left(\frac{C\%}{100}\right) \times \alpha \tag{11}$$

$$Q_{energy} = \left(\frac{Q_{fan}}{\lambda_{heat}}\right) \times (H_{heat}\text{-}419) \times \left(\frac{T_F\text{-}T_0}{T_F\text{-}T_S}\right) \tag{12}$$

where Q_{energy} denotes the amount of energy liberated during the combustion process. NHV stands for the net heat value generated from the combustion. $C\%$ is the carbon content. α is the molar mass ratio of C in CO_2. Q_{heat} refers to the thermal energy required for the process. λ_{heat} denotes the latent heat associated with the steam formation. H_{heat} is the enthalpy value of the steam. Q_{fan} is the fan power of the cooling tower. T_F, T_0 and T_S denote the temperatures of the flame, stack exhaust, and surrounding environment, respectively.

The calculation formula of the SO_2 and NO_x emissions is shown in (13).

$$M_{gas} = A_{gas}M_A + B_{gas}W_B \tag{13}$$

where A_{gas} is the emission conversion factor of standard coal. B_{gas} stands for the electric emission conversion coefficient. M_A and W_B denote Formulas (14) and (15); the value of M_A is 0 because the cooler's energy consumption does not change in this study.

$$M_A = \frac{Q_{heat} \times N}{Q_{STA} \times \delta_{C,T}} \tag{14}$$

$$M_A = \frac{hp}{\delta_{G,T}} \tag{15}$$

N refers to the operation time, Q_{STA} is the thermal data of standard coal, $\delta_{C,T}$ denotes the entropy efficiency and $\delta_{G,T}$ is the transmission efficiency of the power grid. The detailed values of these variables are illustrated in Table 10.

Table 10. Variable values for calculating gas emission.

Variable		Value
SO_2	A_{gas}	0.075
	B_{gas}	0.030
NO_x	A_{gas}	0.0375
	B_{gas}	0.015
	N	8000 h
	Q_{STA}	29.26 KJ/kg
	$\delta_{C,T}$	0.9
	$\delta_{G,T}$	0.92
	Q_{heat}	the heat duty of process (kW)

After the optimization, the CO_2 emission decreases from 22,754 kg/h to 20,875 kg/h, with a reduction of 8.26%. The SO_2 emission and NO_x emission save 175.37 kg/h and 87.69 kg/h, respectively. The reduction of gas emissions is not only conducive to environmental protection but also brings economic and social benefits to enterprises.

5. Conclusions

In this paper, the optimization and analysis of the circulating water system of an industrial FCC unit are carried out. The parallel-to-series modification of the cooler network is conducted and evaluated by process simulation. The fan power of the cooling tower is predicted based on machine learning. Next, the economic and environmental benefits are discussed. It is revealed that the process of machine learning can effectively predict and optimize the industrial process, and the parallel-to-series modification can not only decrease the water consumption but also reduce the electricity usage. The details are as follows:

(1) Three series modifications of the cooler network are made: E1212 and E1218 are connected to E1203A-D in series, E1314 is directly connected to E1311E-H, and E1319AB is connected to E1311 A-D. The new network consumes 1948 t/h of water, with a reduction of 11% compared to the original structure.

(2) The fan power of the cooling tower is chosen as the target function, and the water flowrate, the temperature of the water going in and out of the cooling tower, the daily average ambient temperature and the daily relative humidity of the air are used as the input variables. A total of 284 industrial data sets with the above parameters are sampled to train six machine learning algorithms. The GBR model, which has the best fitting effect, is determined and optimized as the prediction model.

(3) The optimized GBR model can accurately predict the fan power of the cooling tower, and the calculated power reduction by the series-to-parallel retrofit is approximately

8%. Meanwhile, the machine learning method indicates that the daily average ambient temperature is the input variable that has the greatest influence on the fan power.

(4) Considering the economic and environmental benefits, the economic cost is minimized by 8.65% due to the decrease in water and electric consumption, and the gas emissions decreased by 2142.06 kg/h.

In most cases, the cooling tower does not need to operate at full load, so through the precise control of the fan power, that is, through variable frequency regulation, unnecessary energy consumption can be greatly reduced. The prediction system also enhances the system's reliability and stability. By knowing the possible load of the fan power in advance, it can avoid the impact of sudden load changes on the system and reduce the risk of failure and unexpected shutdown, which is particularly important for industrial cooling systems that need continuous and stable operation.

Author Contributions: Conceptualization, Y.L.; methodology, Y.L. and R.S. (Runjie Shao); software, Y.Z. and R.S. (Ruiyu Sun); validation, Q.Y. and J.L.; formal analysis, R.S. (Runjie Shao), Y.Z. and R.S. (Ruiyu Sun); investigation, Y.L. and R.S. (Runjie Shao); resources, Q.Y. and J.L.; data curation, Y.L.; writing—original draft preparation, Y.L. and R.S. (Runjie Shao); writing—review and editing, Y.L. and R.S. (Runjie Shao); visualization, Q.Y., Y.Z. and R.S. (Ruiyu Sun); supervision, J.L., Y.Z. and R.S. (Ruiyu Sun); project administration, J.L.; funding acquisition, Y.L., Q.Y. and J.L. All authors have read and agreed to the published version of the manuscript.

Funding: The research was funded by the Jiangsu Key Laboratory of Advanced Catalytic Materials and Technology (BM2012110); the National Nature Science Foundation of China (Nos: 22178030, 22078026); and the Ministry of Education's Industry-university Cooperative Education Project (221004650123425).

Data Availability Statement: All data generated or analyzed during this study are included in this published article.

Conflicts of Interest: Authors Ruiyu Sun and Yifei Zhai were employed by The Yellow River Delta Chambroad Institute Co. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The Yellow River Delta Chambroad Institute Co. had no role in the design of the study; in the collection, analyses or interpretation of the data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Wang, F.-H.; Hao, H.-T.; Sun, R.-F.; Li, S.-Y.; Han, R.-M.; Papelis, C.; Zhang, Y. Bench-scale and pilot-scale evaluation of coagulation pre-treatment for wastewater reused by reverse osmosis in a petrochemical circulating cooling water system. *Desalination* **2014**, *335*, 64–69. [CrossRef]
2. Wang, Y.; Chu, K.H.; Wang, Z. Two-Step Methodology for Retrofit Design of Cooling Water Networks. *Ind. Eng. Chem. Res.* **2014**, *53*, 274–286. [CrossRef]
3. Ma, J.; Wang, Y.; Feng, X. Optimization of multi-plants cooling water system. *Energy* **2018**, *150*, 797–815. [CrossRef]
4. Zhang, H.; Feng, X.; Wang, Y.; Zhang, Z. Sequential optimization of cooler and pump networks with different types of cooling. *Energy* **2019**, *179*, 815–822. [CrossRef]
5. Ponce-Ortega, J.M.; Serna-González, M.; Jiménez-Gutiérrez, A. Optimization model for re-circulating cooling water systems. *Comput. Chem. Eng.* **2010**, *34*, 177–195. [CrossRef]
6. Müller, T.M.; Neumann, J.; Meck, M.M.; Pelz, P.F. Sustainable cooling cycles by algorithmically supported design of decentral pump systems. *Appl. Therm. Eng.* **2022**, *217*, 119084. [CrossRef]
7. Barigozzi, G.; Perdichizzi, A.; Ravelli, S. Wet and dry cooling systems optimization applied to a modern waste-to-energy cogeneration heat and power plant. *Appl. Energy* **2011**, *88*, 1366–1376. [CrossRef]
8. Liang, J.; Tian, Y.; Yang, S.; Wang, Y.; Yin, R.; Wang, Y. Long-term operation optimization of circulating cooling water systems under fouling conditions. *Chin. J. Chem. Eng.* **2024**, *65*, 255–267. [CrossRef]
9. Song, J.; Chen, Y.; Wu, X.; Ruan, S.; Zhang, Z. A Novel Approach for Energy Efficiency Prediction of Various Natural Draft Wet Cooling Towers Using ANN. *J. Therm. Sci.* **2020**, *30*, 859–868. [CrossRef]

10. Liang, J.; Li, L.; Li, Y.; Wang, Y.; Feng, X. Operation optimization of existing industrial circulating water system considering variable frequency drive. *Chem. Eng. Res. Des.* **2022**, *186*, 387–397. [CrossRef]
11. Zhang, W.; Ma, L.; Jia, B.; Zhang, Z.; Liu, Y.; Duan, L. Optimization of the circulating cooling water mass flow in indirect dry cooling system of thermal power unit using artificial neural network based on genetic algorithm. *Appl. Therm. Eng.* **2023**, *223*, 120040. [CrossRef]
12. Bueso, M.C.; de Nicolás, A.P.; Vera-García, F.; Molina-García, Á. Cooling tower modeling based on machine learning approaches: Application to Zero Liquid Discharge in desalination processes. *Appl. Therm. Eng.* **2024**, *242*, 122522. [CrossRef]
13. Liu, J.; Xu, Y.; Zhang, Y.; Shuai, Y.; Li, B. Multi-objective optimization of low temperature cooling water organic Rankine cycle using dual pinch point temperature difference technologies. *Energy* **2022**, *240*, 122740. [CrossRef]
14. Moghadam, R.E.; Shafieefar, M.; Akbari, H. A probabilistic approach to predict wave force on a caisson breakwater based on Bayesian regression and experimental data. *Ocean. Eng.* **2022**, *249*, 110945. [CrossRef]
15. H'ng, C.W.; Loh, W.P. A prediction of leaf mechanical properties with data mining. *Comput. Electron. Agric.* **2019**, *162*, 669–676. [CrossRef]
16. Vakharia, V.; Castelli, I.E.; Bhavsar, K.; Solanki, A. Bandgap prediction of metal halide perovskites using regression machine learning models. *Phys. Lett. A* **2022**, *422*, 127800. [CrossRef]
17. Aderyani, F.R.; Mousavi, S.J.; Jafari, F. Short-term rainfall forecasting using machine learning-based approaches of PSO-SVR, LSTM and CNN. *J. Hydrol.* **2022**, *614*, 128463. [CrossRef]
18. Sinha, S.; Rao, C.S.; Kumar, A.; Surya, D.V.; Basak, T. Exploring and understanding the microwave-assisted pyrolysis of waste lignocellulose biomass using gradient boosting regression machine learning model. *Renew. Energy* **2024**, *231*, 120968. [CrossRef]
19. Lillington, J.N.P.; Goût, T.L.; Harrison, M.T.; Farnan, I. Assessing static glass leaching predictions from large datasets using machine learning. *J. Non-Cryst. Solids* **2020**, *546*, 120276. [CrossRef]
20. Moeini, B.; Haack, H.; Fairley, N.; Fernandez, V.; Gengenbach, T.R.; Easton, C.D.; Linford, M.R. Box plots: A simple graphical tool for visualizing overfitting in peak fitting as demonstrated with X-ray photoelectron spectroscopy data. *J. Electron Spectrosc. Relat. Phenom.* **2021**, *250*, 147094. [CrossRef]
21. Guo, G.; Zhu, W.; Sun, Z.; Fu, S.; Shen, W.; Cao, J. An aero-structure-acoustics evaluation framework of wind turbine blade cross-section based on Gradient Boosting regression tree. *Compos. Struct.* **2024**, *337*, 118055. [CrossRef]
22. Fan, J.; Wang, D.; Liu, P.; Xu, J. Research on the Prediction of Sustainable Safety Production in Building Construction Based on Text Data. *Sustainability* **2024**, *16*, 5081. [CrossRef]
23. Jain, S.; Jain, R.; Kumar, V.; Samal, S. Data-driven design of high bulk modulus high entropy alloys using machine learning. *J. Alloys Metall. Syst.* **2024**, *8*, 100128. [CrossRef]
24. Biswas, A.A.; Dhondale, M.R.; Singh, M.; Agrawal, A.K.; Muthudoss, P.; Mishra, B.; Kumar, D. Development and comparison of machine learning models for in-vitro drug permeation prediction from microneedle patch. *Eur. J. Pharm. Biopharm.* **2024**, *199*, 114311. [CrossRef]
25. Jayaweera, C.; Groot, N.; Meul, S.; Verliefde, A.; Nopens, I.; Hitsov, I. Development of a hybrid model for reliably predicting the thermal performance of direct contact countercurrent cooling towers. *Int. J. Heat Mass Transf.* **2022**, *197*, 123336. [CrossRef]
26. Navarro, P.; Serrano, J.M.; Roca, L.; Palenzuela, P.; Lucas, M.; Ruiz, J. A comparative study on predicting wet cooling tower performance in combined cooling systems for heat rejection in CSP plants. *Appl. Therm. Eng.* **2024**, *253*, 123718. [CrossRef]
27. Li, Z.; Li, Z.; Zhang, L.; Chen, C.; Hu, M.; Li, X.; Xu, K. Prediction of calcium concentration in circulating seawater in a closed-cycle seawater cooling system using machine learning models. *Desalination Water Treat.* **2023**, *316*, 744–754. [CrossRef]
28. Khaledi, M.; Mehrabadi, A.R.; Mirabi, M. Developing an innovative corrosion and scaling index for industrial cooling water using artificial intelligence. *J. Water Process Eng.* **2024**, *65*, 105838. [CrossRef]
29. Pan, J.; Ding, Y.; Li, J.; Xie, L.; Xu, Z.; Wu, H.; Ye, Q. Economic, entropy generation and environmental analysis of separation of high-concentration azeotropic mixtures by an innovative extractive distillation configuration based on multi-objective optimization. *Sep. Purif. Technol.* **2024**, *340*, 126729. [CrossRef]
30. Yu, A.; Ye, Q.; Li, J.; Li, X.; Wang, Y.; Rui, Q. Economic; environmental; energy, exergy (4E) analysis and simulated annealing algorithm optimization of dividing-wall column-intensified heterogeneous azeotropic pressure-swing distillation process. *Energy* **2024**, *296*, 131099. [CrossRef]
31. Liu, Y.; Chu, M.; Ye, Q.; Li, J.; Han, D. Multi-objective optimization of FCC separation system based on NSGA-II. *Chem. Eng. Sci.* **2025**, *302*, 120829. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.