Multi-objective Pigeon-inspired optimized feature enhancement soft-sensing model of Wastewater Treatment Process

Chang Peng   Bao Xun   Meng FanChao   Lu RuiWei

Faculty of Information and Technology, Beijing University of Technology, Beijing 100124, China

Chang Peng: (Corresponding author): Investigation, Project administration, Supervision, Writing - original draft, Formal analysis, Validation. Faculty of Information and Technology, Beijing University of Technology, Beijing 100124, China, E-mail:changpeng@bjut.edu.cn. ORCID=0000-0002-7766-5583

Bao Xun: Writing - review & editing. Faculty of Information and Technology, Beijing University of Technology, Beijing 100124, China, E-mail:sherry@emails.bjut.edu.cn

Meng FanChao: Writing - review & editing. Faculty of Information and Technology, Beijing University of Technology, Beijing 100124, China, E-mail:mengfc@emails.bjut.edu.cn

Lu RuiWei: Formal analysis. Faculty of Information and Technology, Beijing University of Technology, Beijing 100124, China, E-mail:Andrewlee@emails.bjut.edu.cn

ARTICLE INFO

Keywords:
Wastewater Treatment Process
Soft-sensing model
Overcomplete Broad Learning System
Pigeon-Inspired Optimization

ABSTRACT

Under the increasingly severe fresh water supply pressure, wastewater treatment is considered to be the optimal strategy to satisfy the current and future water demand, thus being highly valued by most countries. As a complicated process, there are some hard-to-measure effluent indicators in wastewater treatment such as 5-day Biological Oxygen Demand (BOD_5), which brings significant difficulties to the monitoring of key indicators in sewage disposal process, thus imposing massive constraints on evaluation of effluent quality. In order to realize the real time supervision of the water quality, data-driven artificial intelligence soft-sensing models have been widely considered as an active research field. However, the over-parameterization of artificial intelligence methods and the complication of sewage treatment environment lead to obvious non-Gaussian characteristics in the wastewater data, which makes it difficult to artificially set parameters to maintain the optimal values to satisfy the accuracy requirement in wastewater treatment process. In view of the aforementioned problems, a soft-sensing model for super-parameter intelligent setting of Broad Learning System based on Overcomplete Independent Component Analysis (OICA) is proposed in this paper. A two-stage multi-objective optimization algorithm is adopted in the model to set superparameters intelligently, which reduces human intervention and improves the accuracy of the model. Additionally, the adaptability of the proposed model to data is significantly enhanced through improvement of the feature extraction ability of the Broad Learning System (BLS) and the capture of peculiar non-Gaussianity in wastewater data with statistical methods. Comparative experiments are conducted on the sewage simulation platform BSM1 with state-of-the-art artificial intelligence soft measurement models, and the results show that the advantage of the presented model lies both in accuracy and in modeling speed, demonstrating the effectiveness of the proposed method.

1. Introduction

As is well acknowledged, water is an indispensable resource that human beings depend on for existence and development. However, with the soared development of industrialization and the continuous improvement of the production speed, there is a deteriorating pollution of water resources, which has exerted significant influence on the survival and development of mankind and the ecological balance of society (Aarnio and Minkkinen (1986)). Therefore, countries and regions around the world have gradually realized the importance of water resource protection and attached more and more attention to the recycling and reuse of water resources, among which wastewater purification is an effective measure to alleviate this problem. However, the abominable working environment and the interference of multiple uncertainties in Wastewater Treatment Process (WWTP) have imposed massive limitations on the collections of certain sewage indicators. For example, offline measurements are required for Biochemical Oxygen Demand (BOD) and Chemical Oxygen Demand (COD), which lags behind the production process, thus seriously affecting the quality control of WWTP. Accordingly, soft-sensing measurement methods have been applied to the collection of effluent indicators and achieved certain effects. In the study of soft-sensing measurement, Partial Least Squares (PLS) (Hulland (1999)) is unanimously considered as a representative statistical method in the monitoring models, which in the early stage was utilized to measure Total Phosphorus (TP), COD and turbidity in effluent (Blom (1996); Teppola et al. (1999)). However, it should never be ignored that due to the violent biochemical reactions in the WWTP, there are obvious nonlinear characteristics in the data, which means that the commonly-used PLS, a linear method, is incapable of addressing such nonlinear data. Hence, Li et al. (Lee et al. (2006)) proposed Kernel Partial Least Squares (KPLS) to fill in this gap, where linearly inseparable process variables from the low dimensional space are primarily mapped to spaces with higher dimensions to enable them to be linearly separable by kernel tricks, and then regression models are constructed with PLS. The effectiveness of KPLS was proved by the application to the calculation of quality variables in a full-size bioanaerobic filter, which witnessed a huge improvement of calculation performance. Although certain advantages have been possessed by the KPLS in addressing the nonlinearities, the effectiveness of the kernel methods are dependent on the selection of kernel parameters, where no certain theoretically based principle is available, which means that the improper selection of kernel parameters will negatively affect the model accuracy.

In recent years, with the exponential development of Artificial Intelligence (AI) (Yuan et al. (2018, 2019)), various intelligent algorithms have been proposed, where the soft sensor models based on neural network have been favored by researchers with their higher modelling precision and stronger nonlinear processing capability compared

ORCID(s):

to the multivariate statistical methods. A predicting model based on generalized regression neural networks has been established by Heddam et al. to predict the effluent BOD concentration of the urban sewage disposal plants, whose satisfactory accuracy has been proved by experiments to predict the BOD concentration (Heddam et al. (2016)). Zhong et al. (Zhong et al. (2010)) and Huang et al. (Huang et al. (2015)) designed a data-driven compensation fuzzy neural network model for BOD soft sensing measurement, which combined compensatory fuzzy logic and neural network to ensure a higher fault tolerance and more stable system. The experimental results showed that the model had higher prediction accuracy compared with the conventional neural networks, and an appropriate selection of compensation degree can enormously improve the efficiency of the learning algorithm. Qiao et al. (Qiao et al. (2014)) conducted the nonlinear error compensation for the sewage mechanism model through the fuzzy neural network, training the network parameters with Error Back Propagation (EBP) Algorithm, and achieved desirable prediction accuracy. However, an explicit approach to adjusting the network structure remained unavailable with the above method based on gradient descent, which exerted great influence on the network parameters. In view of the aforementioned problem, Han et al. (Han et al. (2017)) designed a Radial Basis Neural Network which can add and delete neurons according to the specific situation, whose stronger generalization ability with more compact structure has been verified by experiments.

Admittedly, the data-driven models for operation indexes mentioned above are capable of accurately describing the relationship between a single operation index and the related process variables. Nevertheless, considering that various operation indicators are contained in the wastewater treatment process simultaneously, it is worth noting that only by considering the characteristics of multiple operation indicators at the same time can the optimal operation effect of sewage treatment be ensured. For instance, wastewater is commonly treated in open-air environment, and thus is susceptible to change, depending on all the conditions such as weather and inflow, which exerts massive constraints on the performance of the shallow networks. Comparatively, there are much more hidden layers in deep neural networks, which enables more nonlinear convergence in the hidden layers, displaying stronger advantages in processing complicated datasets and thus have been widely applied to wastewater treatment process. For a more comprehensive description of the establishment of neural networks, hyper-parameters should be introduced, which are a set of particular parameters required in the establishment of neural networks, preset artificially before the training of neural networks. Designed to define concepts of the model with higher levels, including complexity and learning capabilities, hyper-parameters (Bergstra and Bengio (2012)) are determined not by the learning process of the network but by manual presetting, whose optimization contributes largely to obtaining a neural network with higher accuracy and more compact structure. Benefiting from advantages of global optimization as well as good convergence, the hyper-parameter optimization strategies based on swarm intelligence algorithms have become a research hotspot of neuro-optimization. Qiu et al. (Qiu et al. (2016)) used a deep network to complete the soft measurement of indicators in the process of biochemical wastewater treatment. However, multiple targets are required to be considered when adjusting the hyper-parameters such as accuracy and complexity of networks, which are generally contradictory to each other. Therefore, the issue of how to determine a set of solutions that contains the best trade-offs between objectives is named as Multiple-Objective Problem (MOP). Correspondingly, researchers (Esfe et al. (2017); Chen et al. (2010)) tend to regard adjustment of network parameters such as hyper-parameter adjustment (Bergstra and Bengio (2012)) of the water quality soft sensing measurement model as MOP (Mao et al. (2017)), where multi-objective algorithms, especially Multi-Objective Particle Swarm Optimization (MOPSO) are increasingly utilized for the fine-tuning of hyper-parameters. For example, Xiao et al. (Xiao et al. (2017)) proposed a self-organized neural network for fault detection in wastewater disposal, and experimental result demonstrated that both sensor failures and process failures could be detected at a satisfactory accuracy in different scenarios (sewage treatment plants with various degrees of instrumentation). Qiao et al. (Qiao et al. (2018)) put forward an ammonia nitrogen prediction algorithm based on modified RBF neural network, whose strong approximation ability was verified through the ammonia nitrogen test of the real-world wastewater treatment effluent. A sensor technology based on artificial neural network was presented by Pisa et al. (Pisa et al. (2019)), whose superior prediction performance of the nitrogen concentration was proved by predicting the nitrogen content in sewage with the Long and Short-Term Memory network. Han et al. (Han et al. (2019)) proposed a dissolved oxygen concentration control method based on interval type-2 fuzzy neural network, where adaptive algorithm was developed to adjust the controller parameters online to achieve an accurate concentration control. Multi-objective optimization for hyper-parameters in the model is fully studied in this paper. To be noted, optimization approaches have been widely studied by multiple fields. For example, Tan et al. (Tan et al. (2014)) researched a hybrid Structural Equation ModelingArtificial Neural Networks (SEMANN) approach to empirically investigate on the elements that affect the users intention to adopt mobile learning (m-learning) ,and the Root Mean Square of Errors (RMSE) indicated that the ANN achieved high prediction accuracy. Chiang et al. (Chiang et al. (2014))

constructed a rule-based classification model and reasoning algorithm of Associative Petri Net (APN), achieving automatic detection for arrhythmias. Chen et al. (Chen et al. (2020)) have proposed a novel artificial intelligence-based Evolutionary Bat Algorithm (EBA) controller with machine learning matched membership functions in complex nonlinear systems, achieving stability in fuzzy systems under the influence of parameter uncertainities. Zapta Henry et al. (Zapata et al. (2020)) proposed a hybrid swarm algorithm combining strengths of self-assembly and the particle swarm optimization and successfully decreased the number of iterations in the construction of self-assembly algorithm based on wasp nests. Precup et al. (Precup et al. (2020)) developed a new version of the metaheuristic algorithm SMAF1 to achieve the optimal tuning of interval type-2 fuzzy controllers by inserting the information feedback model F1 in SMA. The same team-Precup et al. (Precup et al. (2021)) presents a novel application of the metaheuristic Slime Mould Algorithm (SMA) to the optimal tuning of interval type-2 fuzzy controllers. The optimisation requires the minimisation of a discrete-time objective function expressed as the sum of time multiplied by squared control errors. All the above researches demonstrate that there is an extensive study history about the thought of optimization. In this paper, high attention is paid to the application of optimization to the soft-sensing model in WWTP.

In recent years, Broad Learning System (BLS) has been established by Chen et al. (Chen and Liu (2017)), where on the premise of ensuring the system accuracy, ridge regression methods are used to adjust the network parameters, thus significantly reducing the complexity of the training process and compacting the training time. More importantly, incremental learning ideas are introduced to expand the network structure by horizontally adding neurons when the accuracy does not satisfy the desirable expectation, thus enhancing the performance of the network. For those trained networks, new environmental data can be utilized to adjust internal parameters when the network accuracy is affected by the newly coming data, without the requirement of an overall retraining, so as to cope with the dynamic changes in the data more efficiently. Chang et al. (Chang et al. (2020b)) used OICA for feature extraction and effectively completed the processing of non-Gaussian data. After that, the team applied the method to multiple models to build the ODRNN monitoring model (Chang et al. (2021) and prediction model (Chang and Li (2021), and achieved good results. The Fuzzy Broad Learning system (FBLS) is conducted into the penicillin fermentation platform and real industrial process by Chang et al. (Chang and Ding (2022)), which can better capture the fuzzified feature and quickly update monitoring model to accomplish the self-increase of fault database without retraining the entire process.Considering that different stages of the penicillin fermentation process have different production characteristics, Chang et al. (Chang et al. (2020a)) used the Affinity Propagation (AP) algorithm to separate the different stages of the production process which conducted research on a multi-stage process monitoring framework that integrates AP and the BLS and achieved a good experimental effect.

Inspired by Chen and Chang et al., in this paper, OICA is used to complete the feature extraction of the data, and the generated data is input into the BLS model, and then the soft measurement of the indicator based on OBLS is completed. For soft measurement of effluent index in sewage treatment process, Chang et al. (Chang et al. (2022)) embedded OICA into BLS, obtaining OBLS modelling method to mitigate the insufficient extraction of nonlinear and non-Gaussian features, and thus completed soft-sensing measurement in WWTP with a high speed.The inspiration of this paper is partly from the above work and further improvement has been done in this paper, where the appropriate settings of the parameters in OBLS network are transformed into a multi-objective optimization problem. As a matter of fact, network parameters is before the OBLS network training and learning process need to set parameters, rather than through training, needing to be defined in advance. Therefore, obtaining a higher accuracy and stability in OBLS soft sensing model, there is definitely an necessity to optimize the parameters that influence the properties of OBLS soft-sensing model, so as to select the optimal network parameters and to acquire remarkable performance. . The Multi-objective Pigeon Inspired Optimization (MOPIO), thanks to its very few optimization parameters, fast convergence speed and easy implementation, has been used by many scholars to address the problem of multi-objective optimization of the industrial process. Therefore, our team chose the MOPIO to optimize the parameters of the OBLS model. By mitigating this problem, the result of the soft sensing in WWTP can be significantly enhanced. One of the primary motivations in this paper is to apply BLS to soft sensing in wastewater treatment, but considering the its inadequacy in feature extraction (Chang and Lu (2021); Chang and Li (2021); Chang et al. (2021) and its randomness in selection of the network weights (Chang et al. (2020a), is proposed in this paper a soft-sensing model for the super-parameter optimization of OBLS, which is able to follow the dynamic changes of operation conditions of WWTP with a satisfied accuracy. Based on BLS, a two-stage multi-objective optimization algorithm for network hyper-parameter tuning is adopted in this model, and the capability of feature extraction of BLS is strongly improved according to specific data characteristics, where the unique non-Gaussian nature in the sewage data is successfully captured with statistical strategies, thus leading to the enhancement of the adaptability to data and the accuracy of the model. Finally,

Table 1
Common biochemical variables in WWTP.

Notation	Definition
S_{NH}	Ammonia nitrogen
S_{ND}	Soluble biodegradable organic nitrogen
S_{ALK}	Alkalinity
T_{SS}	Total suspended solid

experiments were conducted on the wastewater simulation platform, the results of which can sufficiently proved the superiority of the proposed method, and the rationality of hyper-parameter tuning is validated by further comparison between the soft-sensing measurement systems with and without hyper-parameter tuning.

The contributions and innovations of this paper are summarized as follows:

(1) A feature-enhanced BLS method combining Over-complete Independent Analysis (OICA)and Broad Learning System (BLS) is presented to handle the characteristics of nonlinearity, non-Gaussianity and time correlation of the data in Wastewater Treatment Process, thus attaining a satisfactory performance of data feature extraction with high modeling speed.

(2) The Multi-objective Pigeon-inspired Optimization algorithm is adopted to ensure the rationality of the settings of hyper-parameters in OBLS model, thus achieving a desirable accuracy in the soft-sensing measurement of Wastewater Treatment Process.

The remainder of this paper is organized as follows: Section 2 briefly introduces the problem statements and the modeling motivations. Section 3 specifically describes the establishment of MOPIO-OBLS model and the soft sensing strategies. The effectiveness of this algorithm is verified through comparative experiments between the method proposed in this paper and the other three conventional approaches in section 4. Finally, conclusions are given and further discussion is conducted in section 5.

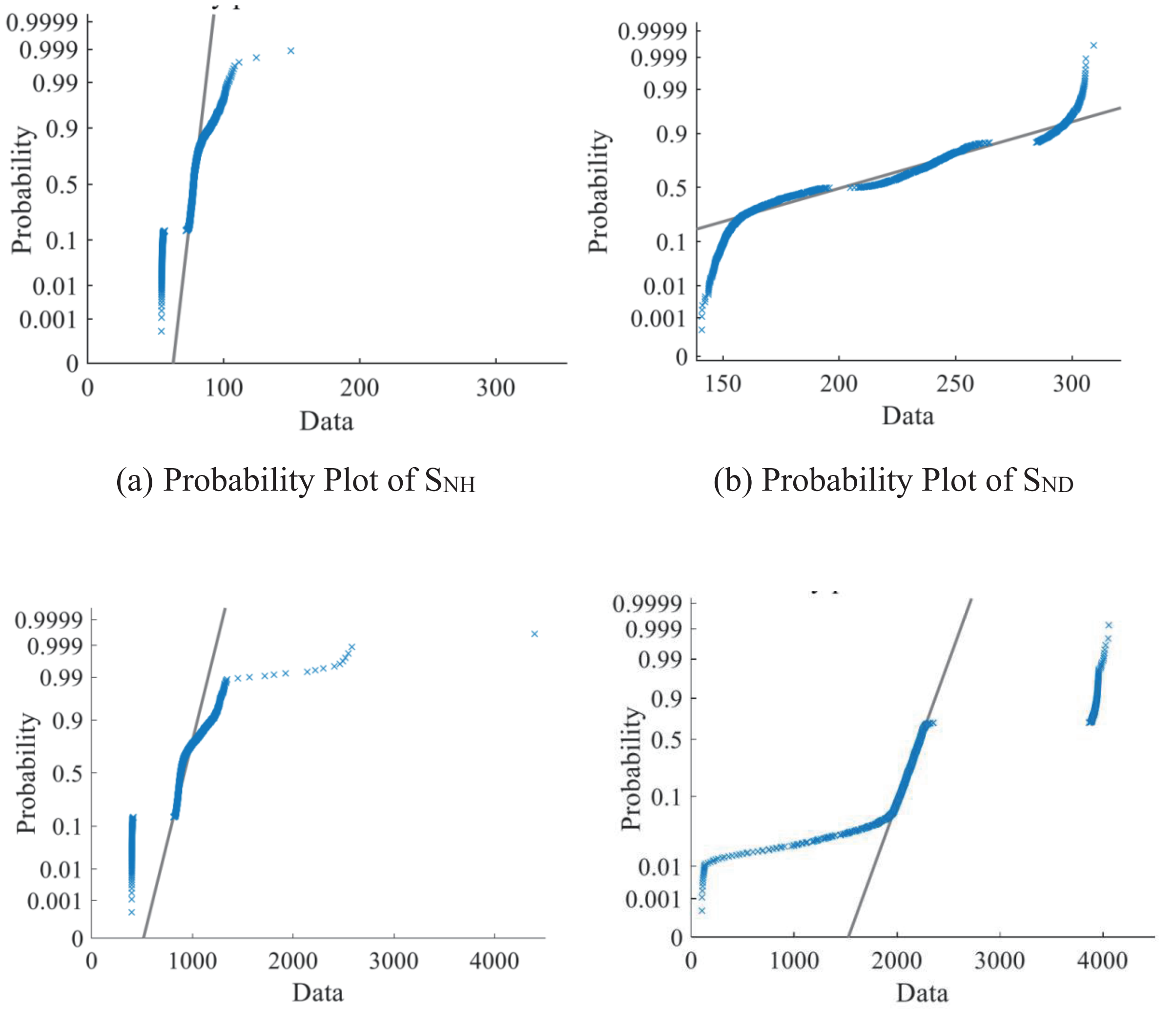
2. Problem Statement and Motivation

As is well acknowledged, there are complex physical and biochemical reactions contained in WWTP, which leads to obvious characteristics of nonlinearity, non-Gaussian and time correlation. The previous classical soft sensing measurements have been successfully applied to various industrial processes in taking nonlinearity into full consideration. However, there is still an insufficient capacity in addressing non-Gaussianity and time correlation, which are verified with four common variables in WWTP as follows:

Statement 1 : Maping the original data to a probabilistic axis by constructing a distribution function is an effective method to verify the non-Gaussianity of the data, and the normal probability plot of the four typical WWTP variables listed in Table 1 are depicted in Figure 1. It can be clearly seen from Figure 1 that when the data represented by the blue dots fit the gray line, the data can be regarded to have Gaussian property, otherwise the data have strong non-Gaussian characteristic.

Statement 2 : The time correlation of the four WWTP variables listed in Table 1 is depicted in Figure 2, where the blue region represents the correlation between adjacent data, and the black line means the correlation control limit. The data are irrelevant when the blue region is within the black line. And the blue region exceeding the black line indicates the correlation between data, which means that there is a strong temporal correlation in the process of WWTP.

It is important to emphasize that the motivation of the optimization algorithm in this paper is the fact that the performance soft-sensing modeling based on Overcomplete Broad Learning System (OBLS) in Wastewater Treatment Process (WWTP) can be improved by finding the optimal number enhancement nodes and feature nodes as well as the bandwidth of the feature nodes. Therefore, the variables of the optimization problem are the optimal number enhancement nodes and feature nodes as well as the bandwidth of the feature nodes in the modelling process based on OBLS.



(a) Probability Plot of S_{NH} (b) Probability Plot of S_{ND}

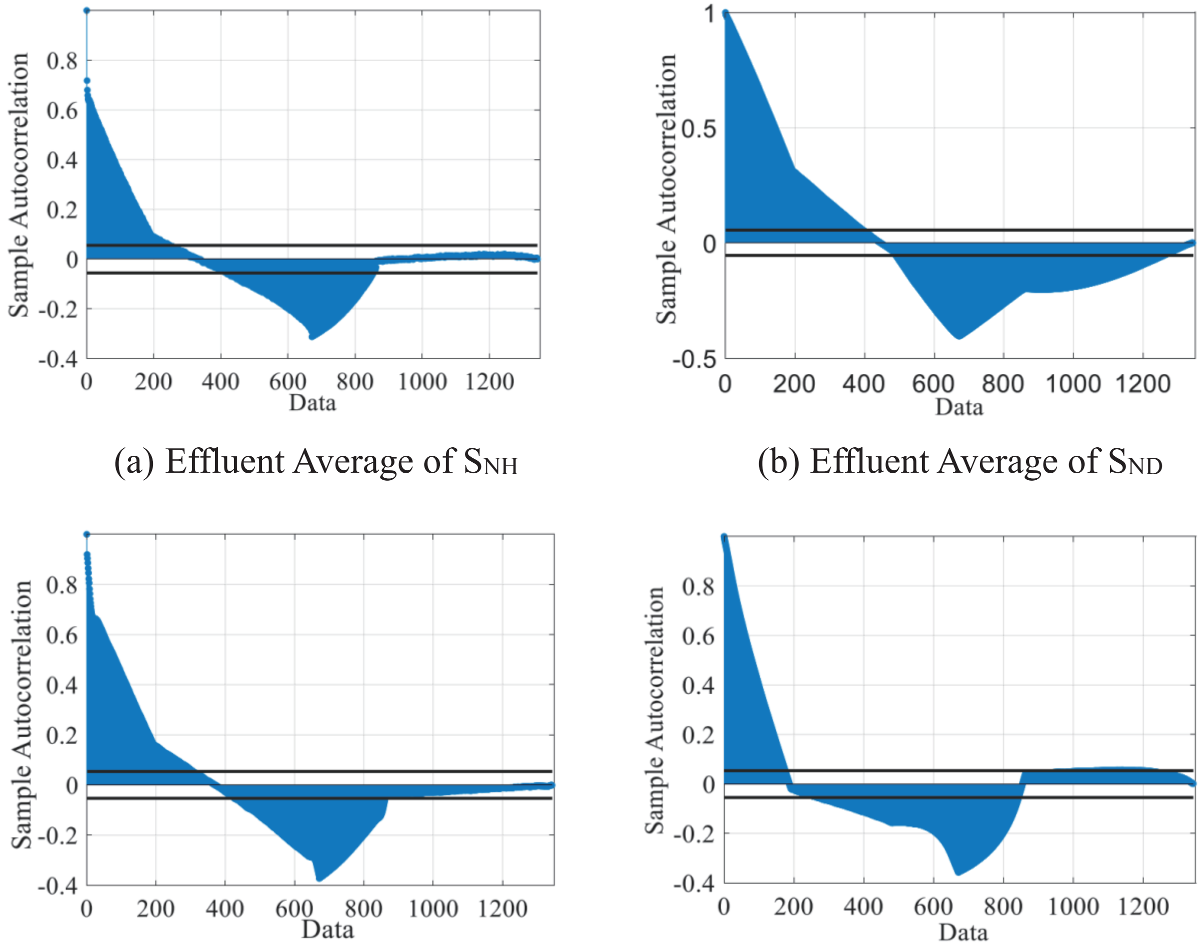
(c) Probability Plot of S_{ALk} (d) Probability Plot of T_{SS}

Figure 1: The normal probability plot of the four WWTP variables.

3. MOPIO-OBLS Model and Soft-sensing Strategies

3.1. BLS Method

It is noted that WWTP is a complicated process with strong nonlinear characteristics, which can be handled by Broad Learning System (BLS) proposed by Chen et al. (Chen and Liu (2017). BLS, a double-layer neural network constructed based on the conventional Random Vector Functional-Link Neural Network (RVFLNN), contains an input layer and an output layer (Han et al. (2019); Chen and Liu (2017). Established in the form of a planar network, BLS takes the original inputs as features to map them to the feature nodes, and the structure is extended in the enhanced nodes. Different from the general neural network, the connection weight of BLS exists between each two adjacent layer units, which is randomly generated by ridge regression, and no changing will take place once introduced into the next hidden layer. The input layer of BLS contains feature nodes and enhancement nodes, among which feature nodes are used to capture the distribution characteristics of the raw data while enhancement nodes are responsible for improving the nonlinear processing capacity of the network. Assuming that the input data set X contains N samples, each with M



(a) Effluent Average of S_{NH} (b) Effluent Average of S_{ND}

(c) Effluent Average of S_{ALk} (d) Effluent Average of T_{SS}

Figure 2: Time correlation test of the four WWTP variables.

dimensions, then $X \in R^{T \times D}$ where T represents all the sampling moments. In the course of feature mapping, l nodes are generated with each map, and the first i feature nodes are constructed as follows.

$$Z_i = \phi\left(XW_{ei} + \beta_{ei}\right) \quad i = 1, 2, \ldots I \tag{1}$$

where W_{ei} and β_{ei} are respectively weights and biases generated randomly while ϕ is the mapping function. The random features are fine-tuned with sparse autoencoders to reduce the randomness.

$$\arg\min_{\widehat{W}} : \|Z\widehat{W} - X\|_2^2 + \lambda\|\widehat{W}\|_1 \tag{2}$$

Where $\widehat{W}$ is the solution of sparse autoencoder and Z represents the desired output of a given equation. Let $f(w) = \|Zw - x\|_2^2, g(w) = \lambda\|w\|_1$, and the above problem can be described as:

$$\arg\min_{w} : f(w) + g(w) \quad w \in R^n \tag{3}$$

Where w is the column vector in $\widehat{W}$.

The above equations can be solved by the following iterative steps.

$$
\begin{aligned}
w_{k+1} &= \left(Z^T Z + \rho I\right)^{-1} \left(Z^T x + \rho \left(o^k - u^k\right)\right) \\
o_{k+1} &= S_{\lambda/\rho} \left(w_{k+1} + u_k\right) \\
u_{k+1} &= u_k + \left(w_{k+1} - o_{k+1}\right)
\end{aligned} \tag{4}
$$

where $\rho > 0$, I is the identity matrix and S is a soft threshold function, which can be defined as follows:

$$
S_{\lambda/\rho}(a) = \begin{cases} a - \lambda/\rho & a > \lambda/\rho \\ 0 & |a| \leq \lambda/\rho \\ a + k & a < \lambda/\rho \end{cases} \tag{5}
$$

Repeat the steps described in Formula (1) for the generation of the required feature nodes and link them as $Z^N = \left[Z_1, \ldots, Z_N\right]$. When the j^{th} enhancement node is constructed as follows.

$$
H_j = \xi\left(Z^n W_{hj} + \beta_{hj}\right) \quad c = 1, 2, \ldots, C \tag{6}
$$

Where W_{hj} and β_{hj} are randomly generated weights and biases, while $\xi(\cdot)$ represents the mapping function. Repeat the steps described in Formula (7) to generate the required feature nodes, linking them as follows.

$$
H^J = \left[H_1, \ldots, H_J\right] \tag{7}
$$

Now that feature nodes and enhancement nodes are readily available, the Broad Learning System can be established by the following formula.

$$
Y = \left[Z_1, \ldots, Z_N \mid H_1, \ldots, H_J\right] W^J = \left[Z^N \mid H^J\right] W^J \tag{8}
$$

Where Y is the output matrix, and $W^J = \left[Z^N \mid H^J\right]^+ Y$ is the connection weight of the network. $[\cdot]^+$ represents the pseudo-inverse of a matrix, which can be obtained with Formula (9)

$$
[A]^+ = \lim_{\lambda \to 0} \left(\lambda I + AA^T\right)^{-1} A^T \tag{9}
$$

Let $[A] = \left[Z^I | H^J\right]$, and the network weight W^J can be obtained correspondingly. The basic structure of Broad Learning System is shown in Figure 3.

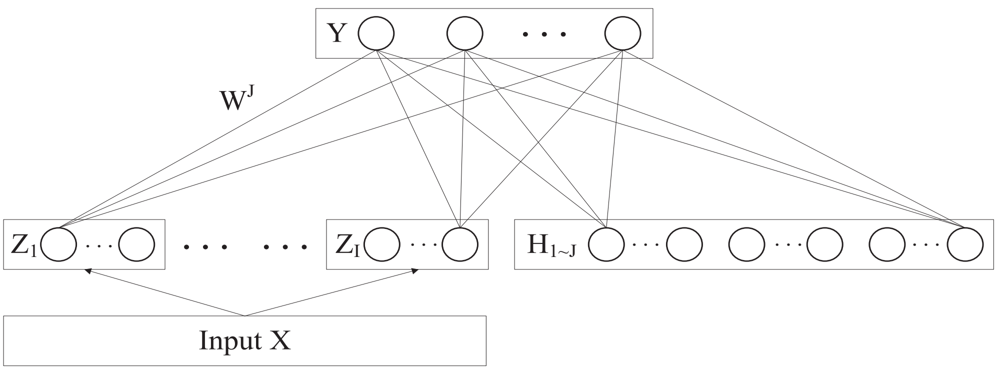


Figure 3: The basic structure of BLS.

3.2. OBLS Method

This section is divided into two parts, which are the introduction of complete independent component analysis and the process of OBLS modeling.

Table 2
Main steps of the OICA-SDP algorithm

Algorithm of OICA-SDP
Input: Matrix X; number of independent components K.
Step1: Construct a subspace formed by atoms $W = \text{span}\left\{d_1 d_1^T, d_2 d_2^T, \dots, d_k d_k^T\right\}$ by X_P.
Step2: Take approximate computation of the atoms $d_i d_i^T, i = 1, 2, \dots, K$, where d_i represents the approximate solution of vectors in hybrid matrix D.
Output: Hybrid matrix D.

3.2.1. Overcomplete Independent Component Analysis (OICA)

As mentioned in Statement 2, there is an obvious non-Gaussian property in WWTP, which is a common data feature in various industrial procedures, and the most classical approach to address non-Gaussian characteristics is Independent Component Analysis (ICA), where non-Gaussian information is contained in higher-order statistics constructed by ICA for further processing.

Specifically speaking, the sample data matrix X_P with P variables can be decomposed into K independent elements:

$$X_P = DS \tag{10}$$

among which the independent component matrix $S = \left(s_1, s_2, \dots, s_K\right)^T$ is composed of independent component vectors $s_k(k = 1, 2, \dots, K)$, and the matrix $D \in R^{P\times K}$ is known as a mixing matrix. According to the relationship between the number of independent component vectors K and the number of variables P in the sample matrix, there are three different cases to classify the ICA algorithm: (1) $K < P$: Undercomplete; (2) $K =$ P : Complete; (3) $K > P$: Overcomplete. Independent component analysis in complete and undercomplete cases has been already widely studied and put into applications, but these commonly-used ICA methods are seldom eligible to be non-Gaussian processing modules for BLS algorithms. Since the principal modelling idea based on neural network in this paper is to achieve a full excavation of the characteristics in data, loss of information should be minimized. For Complete ICA, however, the sampling matrix X_P must be whitened by means of Principal Component Analysis (PCA) before solving the independent components, where certain information loss is unavoidable. When the independent elements are obtained by Complete ICA, further screenings are conducted according to certain indexes such as negative entropy threshold and cumulative contribution, namely Undercomplete ICA, during which human factors can influence the performance of models to a great extent and no recognized reasonable screening principle is readily available so far, which means that regardless of which screening approach is adopted, further loss of information is unavoidable. Therefore, Undercomplete ICA is also excluded from the potential algorithms for the model to be developed in this research. Fortunately, researchers have never stopped marching forward to address this gap. Recently, a novel approach based on Semi-Definite Programming (SDP) to solve the independent elements has been designed by Anastasia Podosinnikova et al. from Massachusetts Institute of Technology (Podosinnikova et al. (2019)), with which independent components can be directly solved from the sampling matrix X_P without whitening, thus avoiding the information loss. Moreover, the complexity of the algorithm is efficiently decreased by replacing all the statistics greater or equal to the fourth order in the original solution scheme with the second order statistics. More importantly, the number of independent elements k can be adjusted arbitrarily to enable the algorithm to work in any case. Specifically speaking, in this research, considering the integrity of the information embedded in raw data, the parameter K is set to be working in an overcomplete circumstance, though there are certain possibility of redundancy caused by excessive independent components (Chang et al. (2020b); Podosinnikova et al. (2019)). Additionally, an innovative design for solving the mixed matrix D based on semi-definite methods, namely Overcomplete Independent Component Analysis via SDP (OICA-SDP), has also been presented in the research by Anastasia Podosinnikova et al., where there is no need for whitening the input matrix X_P in advance. Instead, the training data matrix obtained through sensor sampling is to be calculated by Semi-Definite Independent Component Analysis to obtain the hybrid matrix D for subsequent network operations. The main steps of the OICA-SDP algorithm are shown in Table 2.

(1) Estimation of the Subspace

In order to estimate the subspace, S matrices $H_1, H_2, \dots, H_K \in R^{p\times p}$ are required to be constructed by computing the Heisen matrices of the cumulant generating function (CGF). For the input matrix X with P variables, the Cumulant

Generating Function of any $t \in R^p$ is defined as:

$$\phi_X(t) = logE\left(e^{t^T X}\right) \tag{11}$$

where t is a vector consistent with the Gaussian distribution. In particular, the covariance matrix $cov(X)$ of X can be obtained by substituting t into the second- order Cumulant Generating Function and carrying out the Hessian evaluation. The corresponding formula is expressed as follows.

$$cov(X) = \nabla^2 \emptyset_X(0) \tag{12}$$

Expanding t into a non-zero vector and conducting a Heison evaluation of the Cumulant Generating Function, a generalized covariance matrix $C_X(t)$ can be achieved:

$$\begin{aligned} C_X(t) = \nabla^2 \emptyset_X(t) = \frac{E\left(xx^T e^{t^T x}\right)}{E\left(e^{t^T x}\right)} - \varepsilon_X(t)\varepsilon_X(t)^T \\ \varepsilon_X(t) = \nabla \emptyset_X(t) = \frac{E\left(xe^{t^T x}\right)}{E\left(e^{t^T x}\right)} \end{aligned} \tag{13}$$

The generalized covariance matrix is calculated in the independent component analysis environment, and Formula (10) is substituted into Formula (13) to obtain:

$$\begin{aligned} C_X(t) = DC_\alpha(Y)D^T \\ \varepsilon_X(t) = \frac{\mathrm{DE}\left(\alpha e^{\alpha^T Y}\right)}{E\left(e^{\alpha^T Y}\right)} = D\varepsilon_\alpha(Y) \end{aligned} \tag{14}$$

where $Y = D^T t$. A generalized covariance matrix for the independent elements can then be gained:

$$\begin{aligned} C_\alpha(y) = \nabla^2 \emptyset_\alpha(y) = \frac{E\left(\alpha\alpha^T e^{\alpha^T} x\right)}{E\left(e^{\alpha^T x}\right)} - \varepsilon_\alpha(y)\varepsilon_\alpha(y)^T \\ \varepsilon_\alpha(y) = \nabla \emptyset_\alpha(y) = \frac{E\left(\alpha e^{y^T \alpha}\right)}{E\left(e^{y^T \alpha}\right)} \end{aligned} \tag{15}$$

Since the independent generalized covariance matrix $C_\alpha(Y)$ is a diagonal matrix because of its independence, the generalized covariance matrix $C_X(t)$ in the independent component analysis circumstance can be written as follows.

$$C_X(t) = \sum_{i=1}^{K} \omega_i(t) d_i d_i^T \tag{16}$$

where $\omega_i(t) = \left[C_\alpha\left(D^T t\right)\right]_{ii}$ represents the generalized variance of the i^{th} independent component α_i It can be concluded from Formula (16) that under the circumstances of independent component analysis, the generalized covariance matrix is attributed to the subspace W, so that the space spanned by any number of covariance matrices can be regarded as the subspace of W or the equivalent to W. Hence, choosing a sufficiently large number of vectors $t_1, t_2, \ldots, t_S$ can guarantee that the above inference holds, thus obtaining the equivalent matrix $H_s = C_X\left(t_j\right), s \in [S]$, where H_s is the approximation of the subspace W. Practically speaking, the number S is generally set as a multiple of K.

(2) Estimation of the Atom

After approximating the subspace W through $H_s = C_X\left(t_s\right), s \in [S]$, the approximation of atom $d_i d_i^T, i \in [K]$ should be solved by deflation procedure. The detailed processes are described as follows. First and foremost, the following SDP is to be solved to get an atom.

$$B^*_{sdp} = \arg\max_{B \in S_p} \langle G, B \rangle \tag{17}$$

where $B \in \operatorname{Span}\left\{H_1, H_2, \ldots, H_S\right\} \operatorname{Tr}(B)=1 B \geq 0$, and the matrix G determines the direction of contraction. However, the existence of an optimal solution is not a necessity in practical, so it is highly recommended to let $K \leq P^2/4$, by which there is a considerable possibility that an outcome that is slightly weaker than the optimal solution can be found.
Since the subspace is spanned by h, the following constraint set of Formula (17) should be given full consideration.

$$K=\{B \in W: \operatorname{Tr}(B)=1, B \geq 0\} \tag{18}$$

Construct matrix $\left\{F_1\ F_2, \ldots, F_{M-K}\right\}$, where $M=P(P+1)/2$ represents the base of zero space $N(W)$, and the constraint conditions of Formula (17) can be written as:

$$\left\langle B, F_j\right\rangle=0, j \in[M-K], \operatorname{Tr}(B)=1 B \geqslant 0 \tag{19}$$

In order to alleviate the rigid constraints in the solution process, the relaxation factor is added, and then we can get:

$$B^*=\arg \max_{B \in S_p}\langle G, B\rangle-\frac{\mu}{2} \sum_{j \in[M-K]}\left\langle B, F_j\right\rangle^2 \tag{20}$$

where the regularization parameter $\mu>0$ is conducive to the adjustment the noise expectation level. The semi-definite problem with relaxation factor can be solved by fast iterative shrinkage threshold algorithm (Beck and Teboulle (2009)) or majority maximization principle, and the estimation of a single atom can be obtained approximately.

3.2.2. *OBLS modeling*

In order to obtain multiple rather than single atoms, a contraction process is demanded for Formula (20). However, there is currently no simple and direct shrinkage method that is generally acknowledged, though commonly adopted schemes such as clustering and adaptive shrinkage are available. In view of this barrier in contraction progress, a novel type of semi-adaptive contraction process is proposed in this paper, which combines the advantages of clustering and adaptive contraction to repeat the semi-adaptive shrinkage process continuously(Beck and Teboulle (2009)), each cycle of which can provide one atom $d_i d_i^T$, and all atoms can be obtained after K repetitions. Subsequently, the hybrid matrix can be acquired by the composition of the K vectors according to the number order.

The hybrid matrix D that was obtained in the last step is utilized to solve the independent component matrix of the training data:

$$s=D^T X=\left(s_1,\ s_2, \ldots, s_K\right) \tag{21}$$

using which as an input to the broad learning system, the following feature node is then readily available:

$$Z_i=\phi\left(S W_{ei}+\beta_{ei}\right) i=1,2, \ldots, M \tag{22}$$

Connect the feature nodes to get $Z^M=\left[Z_1, Z_2, \ldots, Z_M\right]$, by which the enhancement node $H_j=\xi\left(Z^M W_{hj}+\beta_{hj}\right)$ can be generated. Henceforth, linking all the enhancement nodes provides us with, and the corresponding feature enhanced broad learning system, namely OBLS, can be expressed as:

$$\begin{aligned} Y & =\left[\phi\left(S W_{e1}+\beta_{e1}\right), \ldots, \phi\left(S W_{eM}+\beta_{eM}\right) \mid \xi\left(Z^M W_{h1}+\beta_{h1}\right), \ldots, \xi\left(Z^M W_{hJ} \quad +\beta_{hJ}\right)\right] W^m \\ & =\left[Z_1, \ldots, Z_M \mid H_1, \ldots, H_J\right] W^J \\ & =\left[Z^M \mid H^J\right] W^J \end{aligned} \tag{23}$$

It is highly worth mentioning that the solution of independent elements by independent component analysis faces obstacles of instability as well as insolubility, and even if the independent components can be successfully solved, the high time expense inevitably puts significant barricades in practical applications. On the other hand, matrix operations are the only required processes in the training process of broad learning system, which makes its training time even shorter than the independent component analysis. Therefore, in terms of time complexity, the conventional independent component analysis is also inappropriate as a feature enhancement approach for the broad learning system. Comparatively, a brand novel algorithm is developed for OICA method based on semi-definite, which essentially reduces the rigid constraints of the formula to be solved, so as to ensure the availability of a solution. Additionally, all the variables used in the operation process are not higher than the second order, which decreases the time complexity significantly. From the experiments to be discussed in the later part of this paper, it can be concluded that little effect is found on the training time of the broad learning system, which enables it to be a massively suitable broad learning system enhancement module. Based on what has been discussed above, the OBLS is established by introducing OICA method, where ICA is utilized to help the network to extract the features in an overcomplete environment, enabling the feature-enhanced-based BLS to achieve a much higher accuracy. The structure of the OBLS model is shown in Figure 4.

Remark 1 : A feature-enhanced BLS method is proposed to update the network at a faster pace and extract the data features more efficiently, coping with the non-Gaussian characteristics in the data. The flexibility nature of OBLS ensures a stable and efficient modeling speed, thus making it possible for the introduction of further optimization algorithms without an undesirable effect of the performance of real-time measurement.

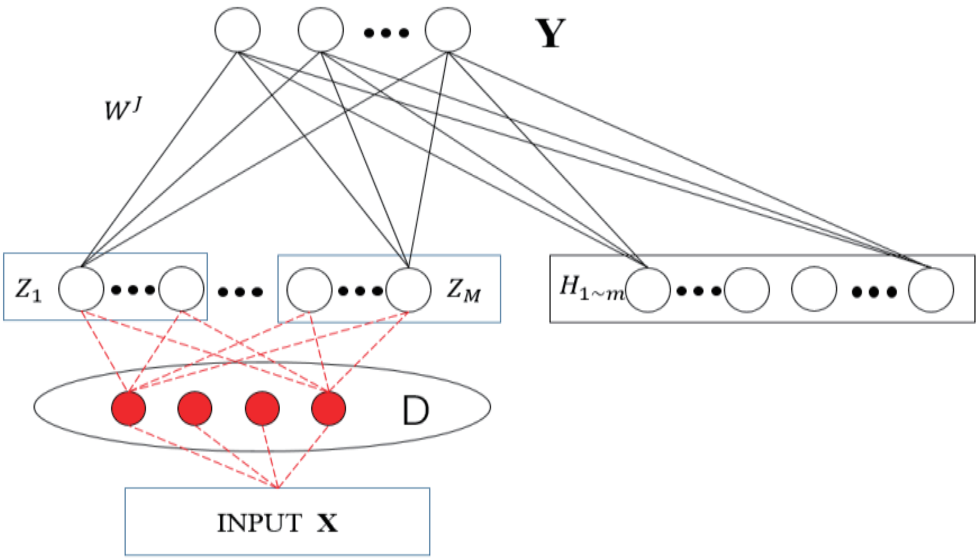


Figure 4: The structure of Overcomplete Broad Learning System(OBLS).

3.3. MOPIO-OBLS Method

Considering that hyper-parameter settings of soft-sensing measurement in WWTP can be modeled as a multi-objective problem, Multi-objective Particle Swarm Optimization (MOPSO) can be employed(Hu and Yen (2013)).

In order to better control the global and local searching capabilities of MOPSO and to improve its optimization efficiency, this paper searches for improvements in the flight equation, presenting a Multi-objective pigeon inspired optimization (MOPIO) with two flight stages, which is applied to optimize the selection of hyper-parameters in the above-mentioned OBLS model, thus further improving the models accuracy.

3.3.1. Parallel unit coordinate system

Firstly, we give an introduction of a method to judge the particles state during the flight. Hu et al. (Hu and Yen (2013)) proposed an optimal solution set state evaluation strategy based on Parallel Cell Coordinate System (PCCS), which defined a series of parameters and evaluates the state of the solution sets from different angles, with the following specific calculation steps.

Primarily, particles are projected into the parallel cells:

$$L_{im} = \left[A \frac{f_m\left(x_i\right) - f_m^{min}}{f_m^{\max} - f_m^{m}} \right] \tag{24}$$

among which $\lceil \cdot \rceil$ is an upward integral function, and L_{im} and $f_m\left(x_i\right)$ respectively represent the projection coordinate value and the adaptation degree of the particle x_i on the m^{th} target, while $f_m^{\min}$ and $f_m^{\max}$ are the minimum and maximum adaptation degrees of the current solution set on the m^{th} target respectively. Combining the projection values of all the targets, the coordinate of the particle x_i in the parallel unit coordinate system $P_i = \left(L_{i1}, L_{i2}, \ldots, L_{iM}\right)$ can be obtained.

The following three parameters are provided by the parallel unit cells for evaluating the particles:

(1) Potential Energy

Based on the coordinate values after particle projection, the state of particles in the population can be analyzed. Define the particle potential energy *Potential* $(\cdot)$to compare the convergence of the particle x_i with other particles in the solution set, and the calculation formula is as follows.

$$Potential\ \left(x_i\right) = \sum_{m=1}^{A} L_{im} \tag{25}$$

Where In the above equation, the convergence situation is well reflected by the potential energy $Potential\ \left(x_i\right)$, where a smaller potential energy represents a more desirable particle convergence.

(2) Density

The particle density *Density* $(\cdot)$ is defined as follows to reflect the diversity of particles:

$$Density\ \left(x_i\right) = \sum_{j=1\neq i}^{A} \frac{1}{PCD\left(x_i, x_j\right)} \tag{26}$$

$$PCD\left(x_i, x_j\right) = \begin{cases} 0.5 & if \quad \forall m \quad L_{im} = L_{jm} \\ \sum_{m=1}^{M}\left|L_{im} - L_{jm}\right| & else \end{cases} \tag{27}$$

Where $PCD\left(x_i, x_j\right)$ represents the particle distance based on parallel unit coordinate system, and a smaller *Density* $\left(x_i\right)$ indicates a greater sparsity around the particle and thus a higher value of exploration of the solution space nearby.

(3) Distribution Entropy

The distribution entropy of populations *Entropy*$(\cdot)$ and Δ *Entropy* $(\cdot)$ are defined to reflect the change of the current solution set, namely tending to convergence, diversity or stagnation, the definition formula of which is exhibited as follows.

$$Entropy\ (t) = -\sum_{i=1}^{N}\sum_{m=1}^{M} \frac{Cell_{\ im}(t)}{NM} log \frac{Cell_{\ im}(t)}{NM} \tag{28}$$

where $Cell_{im}(t)$ represents the number of points in the unit cell located at the i^{th} line, j^{th} column of the parallel cell coordinate system at the tth iteration, and when the cell is empty, $Cell_{im}(t) = 0$ correspondingly.

$$\Delta\ Entropy\ (t) = \ Entropy\ (t) - \ Entropy\ (t-1) \tag{29}$$

By comparing the intensity of the entropy variation $|\Delta Entropy(t)|$ and the threshold δ, the tendency of the current solution set can be diagnosed, where δ can be computed with the following equation:

$$\delta = -2M \frac{1}{NM} \log \frac{1}{NM} \tag{30}$$

Therefore, the states of the solutions or the solution sets can be evaluated from various aspects through the above-mentioned three parameters (potential energy, density and distribution entropy) so as to guide the adjustment of flight parameters.

3.3.2. Multi-objective pigeon-inspired optimization modeling process

Thorough research of the homing behaviors of pigeons conducted by Duan et al. (Duan and Qiao (2014)) have achieved an idealization of this process, which can be summarized to the following two stages:

(1) Map & Compass Stage:

At this stage, pigeons can perceive the earth's magnetic field through the magnet inside their bodies and then shape the map in the brains. Especially when far away from the destination, pigeons rely mostly on magnets and the sun, and the height of the sun is exploited as a compass for the homing pigeons to adjust the directions, thus flying back to their nests.

(2) Landmark Stage: When close to their destination, the nearby landmarks rather than magnet or the sun are what pigeons depend on to locate their nests. Pigeons that are familiar with the landmarks can fly directly to the destination, while for pigeons who are unfamiliar with the landmarks, those experienced pigeons are what they need to follow. Fully considering the homing behavior of pigeons, it can be summarized that when far from the destination, the essential goal of pigeons flight is to approach the destination, that is, to search for convergence through global search. In comparison, w hen close to the target, the pigeons begin to explore around the destination, which means an expansion of the diversity of solutions. Inspired by the habit of pigeon swarms, a two-stage multi-objective pigeon-inspired algorithm is designed, where the flight updating formula is improved, so that the first stage focuses on global searching to ensure convergence, while the second phase concentrates on local searching to enrich the diversity of solutions. The way the OBLS model works is train before test. To optimize the accuracy of OBLS models, reasonable parameter Settings are required. Therefore, a MOPIO is proposed in this paper. Inspired by the homing behavior of pigeons, the particle optimization process is divided into two stages to ensure convergence and diversity of solution sets. Hence, the two stages of the multi-target pigeon-inspired algorithm are designed as follows:

A. Global Searching Stage

Step 1: Population initialization.

Step 2: Project the optimal solution set, calculate the potential energy, and select the particle with the smallest potential energy as the global optimal solution g_b.

Step 3 : Calculate the corresponding distribution entropy Entropy $\left(t_1\right)$, Δ Entropy $\left(t_1\right)$ and the threshold δ, based on which to update the flight parameters.Where $\omega\left(t_1-1\right)$ and $c_2\left(t_1-1\right)$ represents the flight parameters at the previous iteration, which takes the initial value when t_1-1 equals to zero, and ω and c_2 express the updating step length of the parameters, which are the ratios of the upper and lower bound difference to the total number of iterations.

Step 4 : Conduct the updating of the particle speed and position, whose formulas are shown as follows.

$$
\begin{aligned}
& v_i\left(t_1\right)=\omega\left(t_1\right) \cdot v_i\left(t_1-1\right)+c_2\left(t_1\right) r_2 \cdot\left(g_b-x_i\left(t_1-1\right)\right) \\
& x_i\left(t_1\right)=x_i\left(t_1-1\right)+v_i\left(t_1\right)
\end{aligned} \tag{31}
$$

Importantly, Formula (31) takes the particles with the smallest potential energy, i.e. the best convergence in the current solution set, as the global optimum to accelerate the speed of particles approaching the Pareto front. At this time, g_b can be analogous to the sun or the magnetic field in the first stage of pigeons' homing, which is used to guide the population convergence.

Step 5 : The updated population is saved to the optimal solution set and non- dominant ranking mechanism is adopted if its size exceeds the upper limit. If the iteration times in the first stage is equal to the maximum number of iterations, enter the second stage. Otherwise, if the iteration times t_1 in the first stage is inferior to the maximum value, let $t_1=t_1+1$ and jump to *Step 2*.

B. Local Searching Stage

Step 1 : The state of the population and the optimal solution set is determined to the same with that in the first stage.

Step 2 : The density is calculated from the projection of the optimal solution set, and the particle with the smallest density is selected as the global optimal solution.

Step 3: The distribution entropy Entropy (t) and Δ Entropy $(\cdot)$ as well as the threshold δ are computed to update the flight parameters.

$$\omega\left(t_2\right)=\begin{cases}\omega\left(t_2-1\right) & if \quad \Delta Entropy=0 \\ \omega\left(t_2-1\right)+2Step_{\omega}(1+\mid \Delta Entropy \mid) & if \quad 0<\mid \Delta Entropy \mid<\delta \\ \omega\left(t_2-1\right)-Step_{\omega}\mid \Delta Entropy \mid & if \quad \mid \Delta Entropy \mid \geq \delta\end{cases}$$
$$c_1\left(t_2\right)=\begin{cases}c_1\left(t_2-1\right) & if \quad \Delta Entropy=0 \\ c_1\left(t_2-1\right)+2Step_{c_1}(1+\mid \Delta Entropy \mid) & if \quad 0<\mid \Delta Entropy \mid<\delta \\ c_1\left(t_2-1\right)-Step_{c_1}\mid \Delta Entropy \mid & if \quad \mid \Delta Entropy \mid \geq \delta\end{cases} \tag{32}$$

where $\omega\left(t_2-1\right)$ and $c_2\left(t_2-1\right)$ are the flight parameters of the previous iteration, and when $t_2-1=0, \omega$ takes the value at the end of the first stage and and c_2 takes the initial values, in which $Step_{\omega}$ and $Step_{c_1}$ are the ratios of upper and lower bound difference of parameter values to total iteration times.

Step 4 : Update the speeds and positions of the particles with the following formulas respectively:

$$\begin{aligned}&v_i\left(t_2\right)=\omega\left(t_2\right)\cdot v_i\left(t_2-1\right)+c_1\left(t_2\right)r_2\cdot\left(g_b-x_i\left(t_2-1\right)\right)\\&x_i\left(t_2\right)=x_i\left(t_2-1\right)+v_i\left(t_2\right)\end{aligned} \tag{33}$$

In formula (33), particles with the smallest density are considered to be the global optimum, because the degrees of exploration around these particles are proved to be the lowest, which means the highest exploration values. Therefore, the population should be guided to fully explore these regions and thus enrich the diversity of solutions in the solution set.

Step 5 : The updated population is saved into the optimal solution set, and non-dominant ranking mechanism is necessary when the upper limit is exceeded. Moreover, if the iteration times in the second stage t_2 is equal to the maximum number of iterations, the second stage is finished and the solution set is to be output. Otherwise, if the iteration numbers of the second stage t_2 is less than the upper limit, then let $t_2=t_2+1$ and jump to *Step 2*.

Remark 2 : A higher reliability of the selection of hyper-parameters is obtained by optimizing the OBLS with MOPIO algorithm, which tackles a practical problem in the establishment of the soft sensing measurement models in WWTP.

3.4. MOPIO-OBLS Soft Sensor Strategy

The soft sensor of the MOPIO−OBLS algorithm mainly includes an offline part and an online part, of which the flowchart is depicted in Figure 5. This chapter introduces the modeling process, divided into offline modeling and online modeling. The detailed process is as follows:

Offline modeling

The purpose of offline modeling is to determine the parameters and weights of the model, so that the constructed model can attain better performance. The detailed steps are as below:

Step 1 : Obtain the offline data matrix X_P with p-dimensional variables. Then standardize the offline process data X_P.

Step 2 : Construct a spatial subspace W formed by atoms $d_i d_i^T, i \in [K]$.

Step 3 : Calculate the $C_X(t)$ and the $\varepsilon_x(t)$ according to the cumulant generating functions in Formula (13) to estimate the subspace W.

Step 4 : Extend the $C_X(t)$ and $E_X(t)$ as the generalized covariance matrix of ICA by Formula (14).

Step 5 : Through $Y=D^T t$, Obtain the y's generalized covariance matrix $C_\alpha(y)$ that formulated as Formulas (15).

Step 6 : Based on the $\omega_i(t)=\left[C_\alpha\left(D^T t\right)\right]_{ii}$, the generalized covariance matrix of C_X is obtained by Formula (16), then the subspace W is determined.

Step 7 : Estimate atoms $d_i d_i^T$ of the W through the semi-definite method with relaxation element that expressed as Formula (20).

Step 8 : Through the Formulas (21) − (22), the model of OBLS is as constructed as Formula (23).

Step 9 : Take the number of feature nodes and enhancement nodes in the OBLS system as the hyper-parameters, and take the accuracy and complexity of the system as the two objectives of optimization. The two stages of MOPIO

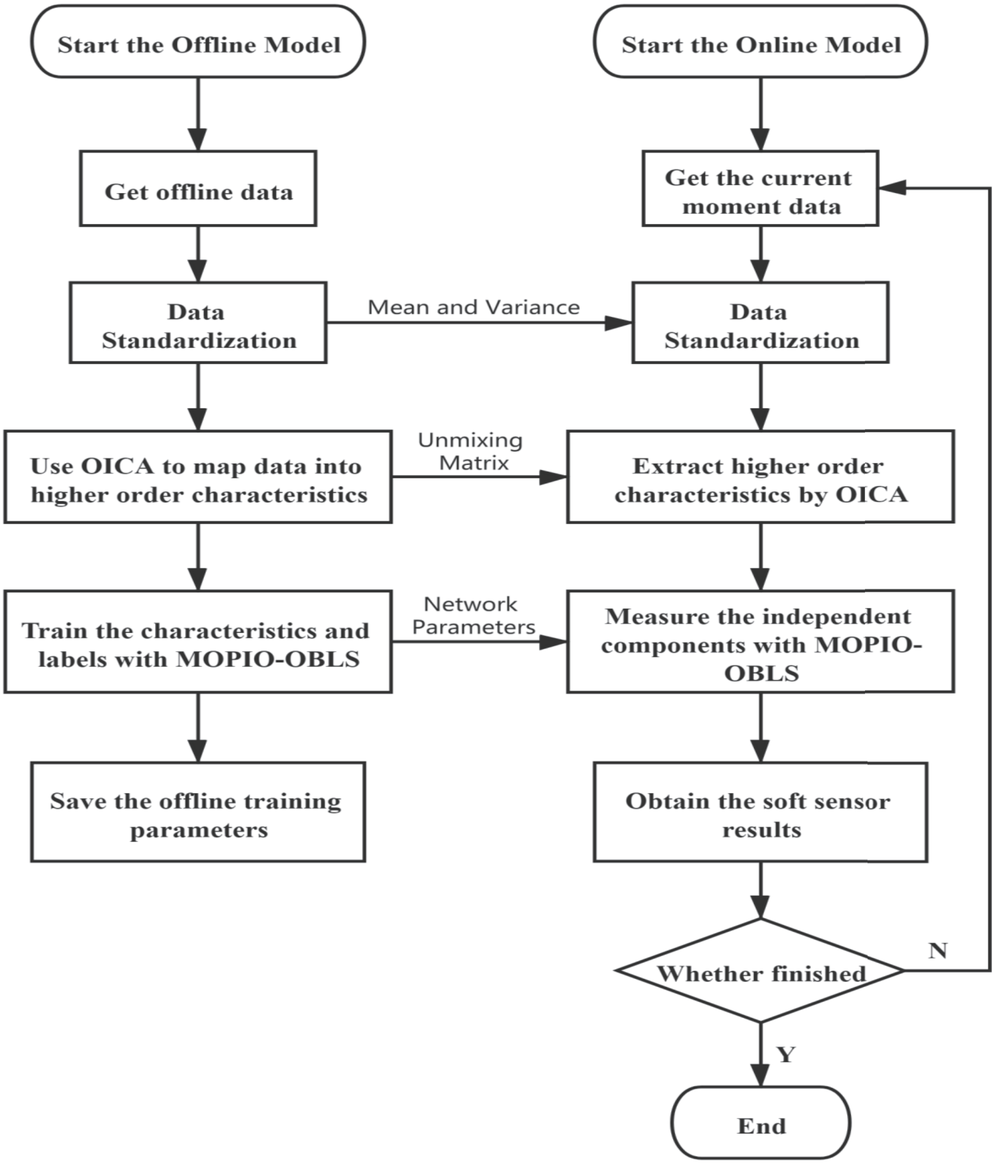


Figure 5: The soft sensor flowchart of MOPIO-OBLS.

described in Section 3.3 are performed to select the optimal hyper-parameters. Then the parameters are saved for the online model. So far, offline modeling is over.

Online modeling

The purpose of online modeling is to use the model built by offline modeling to apply the model to new data to determine whether the model can well complete the task of soft sensing.

Step 1: Standardize the online data of the WWTP by the Mean and variance.

Step 2 : Map the online data into independent components by the W in *step* 7 in offline modeling.

Step 3 : The independent components of online data is computed by the MOPIO-OBLS structure of *Step 9* in offline modeling to get the soft-sensing target. So far, Online modeling is over. To be more specific, the number of feature nodes and enhancement nodes as well as the bandwidth of the feature nodes in OBLS network are difficult to be set in optimal values by manual adjustment, opposing constrains on the accuracy in soft measurement in WWTP.

And considering the fact that WWTP is complex and changeable, there are certain obstacles to guaranteeing that the parameter values set artificially are reasonable and reliable, or that the effluent water quality reaches the standards for discharge of sewage. The extent to which the accuracy of OBLS soft-sensing measurement is affected by the above three parameters is variable. Therefore, how to determine the parameters to the optimal value is one of the main focuses in our manuscript. The Multi-Objective Pigeon Inspired Optimization (MOPIO) is fully adopted in this paper smartly setting up the above mentioned three parameters and reducing the subjective influence of human intervention, so as to improve the precision of the OBLS network model. A much clearer explanation about the relation of optimization-learning architecture and algorithm-optimization algorithms-implementation, can be accessible in diagram in Figure 6.

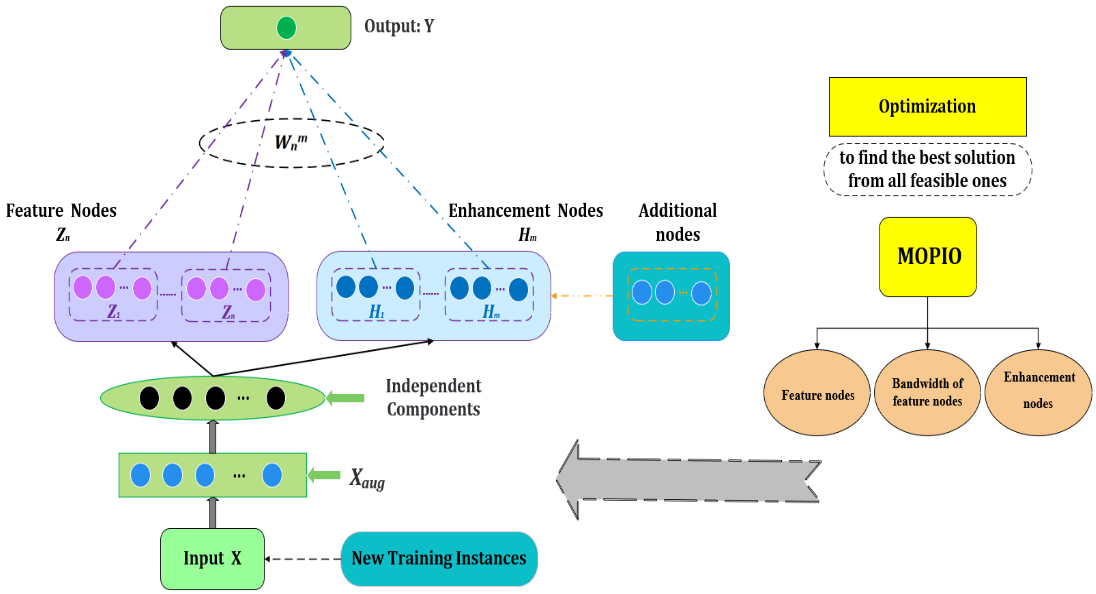


Figure 6: The relation of optimization-learning architecture and algorithm-optimization algorithms-implementation

4. Experiment and Discussion

This paper is targeted on optimize the parameters in OBLS soft-measuring strategy, i.e. the bandwidth of the feature nodes of the network and the number of feature nodes and enhancement nodes, with improved MOPIO. Therefore, the experiment has been designed to compare the performance in conducting soft measurement on effluent water quality BOD_5 of the constructed model in this manuscript with the OBLS soft-measuring strategy without parameter optimization as well as other state-of-the-art networks. In our experiment Section, we have provided necessary parameter settings and data source descriptions. The MOPIO is used to adjust the nodes of the OBLS network to optimize the accuracy of the OBLS model. In addition, the systems were modeled five times in each three environments (sunny day, rainy day and rainstorm) provided by BSM1 platform to reduce the impact of randomness. In order to verify the effectiveness of the proposed soft-sensing measurement model, the benchmark simulation platform presented by the International Water Association (IWA), namely the Benchmark Simulation Model 1 (BSM1) (Alex et al. (2008)), is exploited by this section to conduct comparisons with two state-of-the-art measurement models in wastewater treatment process, Support Vector Machine (SVM) and Recurrent Neural Network (RNN), to enhance the persuasiveness of the experiment. In addition, the optimized version of the algorithm is tested in sunny days, whose results show that there is a more desirable performance and a more satisfactory accuracy compared to the unoptimized algorithms, thus evaluating the feasibility of this method.

4.1. Benchmark Simulation Model 1 (BSM1)

Presented by the International Water Association, BSM1 provides a benchmark simulation environment, where a sewage treatment plant layout and the simulation model are defined, and indexes including effluent loads, test procedures and evaluating criteria are designed to provide a non-deviation benchmark system to facilitate the comparison of different control strategies without reference to specific facilities.

4.2. Settings of data and parameters

In order to verify the the model proposed in this paper, BOD_5 of the effluent is selected as the soft sensor measuring object, which is acknowledged to be a significant indicator to reflect the water quality, expressed as mg/L. Since in reality, the sewage disposal plants operate continuously in different weather conditions, three test scenarios including sunny, rainy and rainstorm environment were simulated in BSM1, where closed-loop simulations were carried out for 14 days, sampling data every 15 minutes, obtaining a total of 1344 samples. As a prerequisite before the experiment, the parameters of the model are set as follows. The data in the three environments are grouped into training set and test set, which are respectively derived from the first 1244 sample points and the last 100 groups. All the state variables supplied by BSM1 are collected as the auxiliary variables, and the number of independent components is determined to be $\lfloor P^2/4 \rfloor$. Additionally, the performances of the soft sensing measurement of the BOD_5 in effluent water among the four strategies are compared according to Root Mean Squared Error (RMSE), which can be computed by the following equation:

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(y_t - \hat{y}_t\right)^2} \tag{34}$$

Where y_t is the real value the moment t, and $\hat{y}_t$ is the estimated value provided by the soft measurement system at the moment t, while T represents all the estimation time. Due to the uncertainty of the neural network training results, the accuracy may be different in different training processes. In view of minimizing the impact of randomness on the experimental results, all the algorithms are repeated for five times and the average of RMSE in the five operations is utilized for comparison.

4.3. Experimental results

This paper has completed the test of the model under three working conditions, namely, sunny, rainy and heavy rain. These three working conditions can cover 99% of the whole year, so they are very representative. The details are as follows:

(1) Discussion of sunny results

In order to test the optimization ability of MOPIO for the soft measurement system, the performance of MOPIO-optimized and non-optimized soft measurement systems are compared in three weather conditions provided by BSM1 platform, i.e. sunny, rainy and rainstorm. Nodes in the broad learning system are adjusted by MOPIO to optimize the accuracy and to reduce the complexity of the system, and the effects of randomness are mitigated by repeatedly modelling the systems for 5 times in each condition. It can be witnessed by Figure 7 that there is no unique optimal solution in multi-objective optimization problems. Therefore, the MOPIO is designed to get the optimal solution set , and can be selected one final solution as the parameters of the soft measurement modeling system, and the solutions in the Pareto frontier are evaluated based on slope. The final solution selection scheme based on slope is shown by the experimental results in sunny environments, where the red block describes the final solution set obtained by the multi-objective pigeon-inspired algorithm in this experiment, and the blue curve represents the approximate Pareto front obtained by polynomial fitting the final solution set. The number of neurons are contained in x-axis, representing the complexity of the system, while the RMSE between the system estimation value and the real effluent BOD_5 value is expressed in the y-axis, reflecting the accuracy of the system. In the real-world working circumstances, accuracy is considered as the primarily significant performance, and thus the optimal solutions with the slope close to zero should be chosen, where in this experiment the optimal solution set with the slope that is less than 1.0E-5 is ultimately opted, with which the network can be established, whose accuracy is difficult to be improved even if a considerable amount of additional nodes are invested. Finally, among all the solutions whose slopes are within the requirement, one with the minimum number of nodes, which means the lowest complexity, is elected as the ultimate optimal solution.

It can be seen from Figure 8 that the effluent BOD_5 shows a dramatic fluctuation intensity, but in terms of magnitude, its fluctuation range is acceptable, which is between 2.676 and 2.681. More importantly, OBLS performs best among the four comparative models and better than the non-optimized BLS model, achieving a satisfactory fitting effect of the real value and an accurate prediction of the fluctuation of BOD_5, while in contrast, there is a general positive

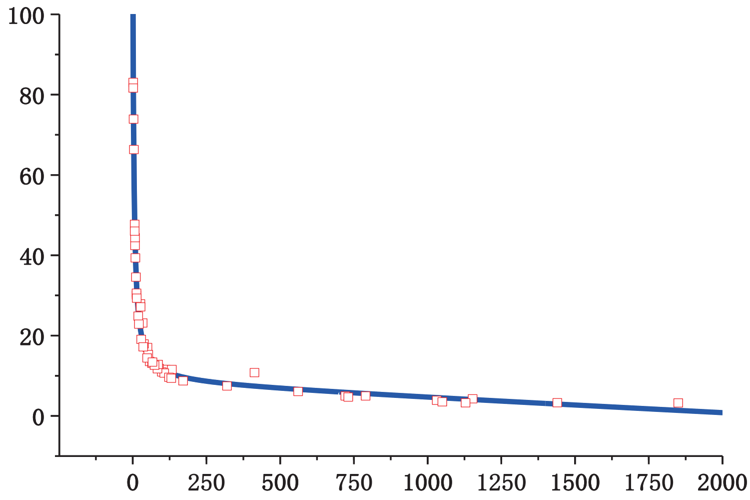


Figure 7: The optimal solution set and Pareto front in sunny weather.

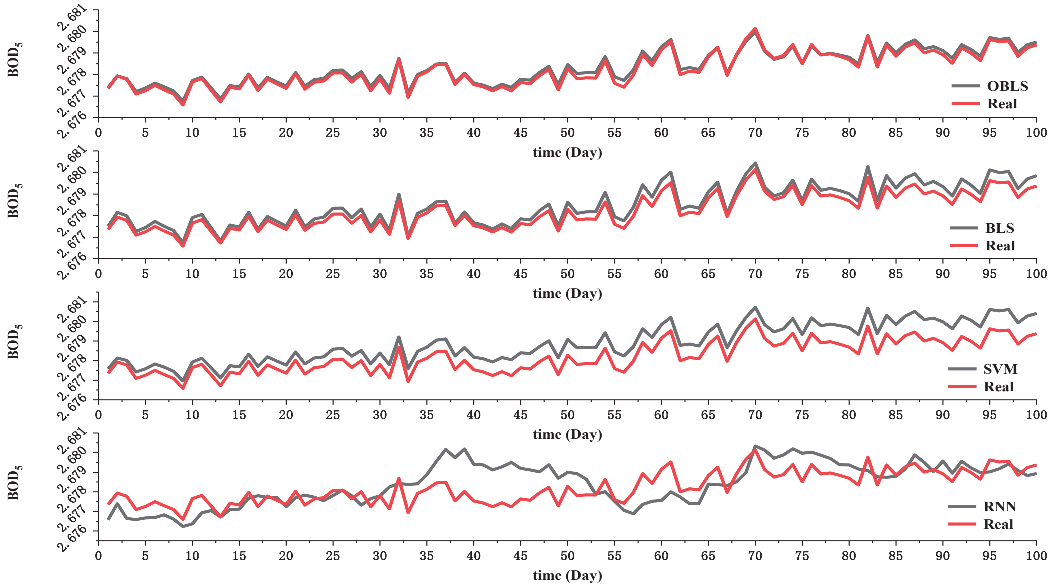


Figure 8: Measurement results of BOD_5 in sunny environment.

deviation in estimation values to the real values in the SVM-based method, whereas the fitting results of the RNN-based strategy is far from satisfactory from the 35th day to the 80th day. Moreover, in order to verify the effectiveness of the optimization algorithm, prudent comparisons between the optimized OBLS and the unoptimized one and other three algorithms are conducted, of which the specific results are described in Table 3. It can be found that the RMSE value of the optimized OBLS is significantly less than the optimized version, illustrating the desirable accuracy and robustness of the optimized OBLS. Whats more, the Average and Standard Deviation of BLS are slightly less than that of OBLS, and these two indicators of SVM and RNN are both weaker than those of OBLS as well.

(2) Discussion of rain results

Table 3
values of RMSE of the algorithms under sunny conditions

Algorithm	RMSE			
	Average	Standard Deviation	Min	Max
MOPIO-OBLS	4.1E-5	7.2E-5	2.5E-5	5.4E-5
OBLS	12.5E-4	7.4E-5	1.5E-4	3.4E-4
BLS	3.3E-4	9.1E-5	2.0E-4	4.3E-4
SVM	2.0E-3	8.1E-5	1.6E-3	2.1E-4
RNN	6.0E-4	3.0E-4	2.4E-4	9.2E-4

Table 4
values of RMSE of the algorithms under rainy conditions

Algorithm	RMSE			
	Average	Standard Deviation	Min	Max
MOPIO-OBLS	7.08E-3	1.9E-3	5.8E-3	8.3E-3
OBLS	1.1E-2	2.1E-3	8.8E-3	1.4E-2
BLS	2.1E-2	1.0E-2	1.0E-2	3.2E-2
SVM	1.8E-2	7.3E-3	1.5E-2	1.9E-2
RNN	1.5E-2	4.8E-3	1.2E-2	2.3E-2

Table 5
values of RMSE of the algorithms under rainstorm conditions

Algorithm	RMSE			
	Average	Standard Deviation	Min	Max
MOPIO-OBLS	8.26E-3	3.5E-3	6.7E-3	9.4E-3
OBLS	2.3E-2	3.7E-3	1.8E-2	2.7E-2
BLS	2.4E-2	9.0E-3	1.0E-2	3.3E-2
SVM	1.9E-2	6.9E-3	1.7E-2	2.2E-2
RNN	2.6E-2	8.4E-3	1.2E-2	3.4E-2

The performance of the proposed soft sensing measurement model is further tested in rainy conditions. Figure 9 exhibits the soft measurement results of the effluent BOD_5 of the method designed in this paper and the comparative methods, from which we can find that the fluctuation range of BOD_5 in rainy days is more extensive compared to that in sunny conditions (2.0-3.0), imposing a significant challenge to the performance of the soft sensing measurement model. It can be seen from the curve of the measurement results that under rainy conditions, the difference among the four models is too obscure to directly run the comparison. Therefore, the RMSE is introduced to further compare the performances of the four models, and the optimized MOPIO-OBLS model is also added to the comparison. From Table 4, it can be found that the RMSE of OBLS still maintains the lowest average and standard deviation, showing the best performance and robustness compared to the other three conventional methods, while the performance of the optimized MOPIO-OBLS is further enhanced based on OBLS. As the working environment becomes more complex, there is a certain decline of the robustness of BLS, but the performance of SVM and RNN is still weaker than that of OBLS.

(3) Discussion of storm results

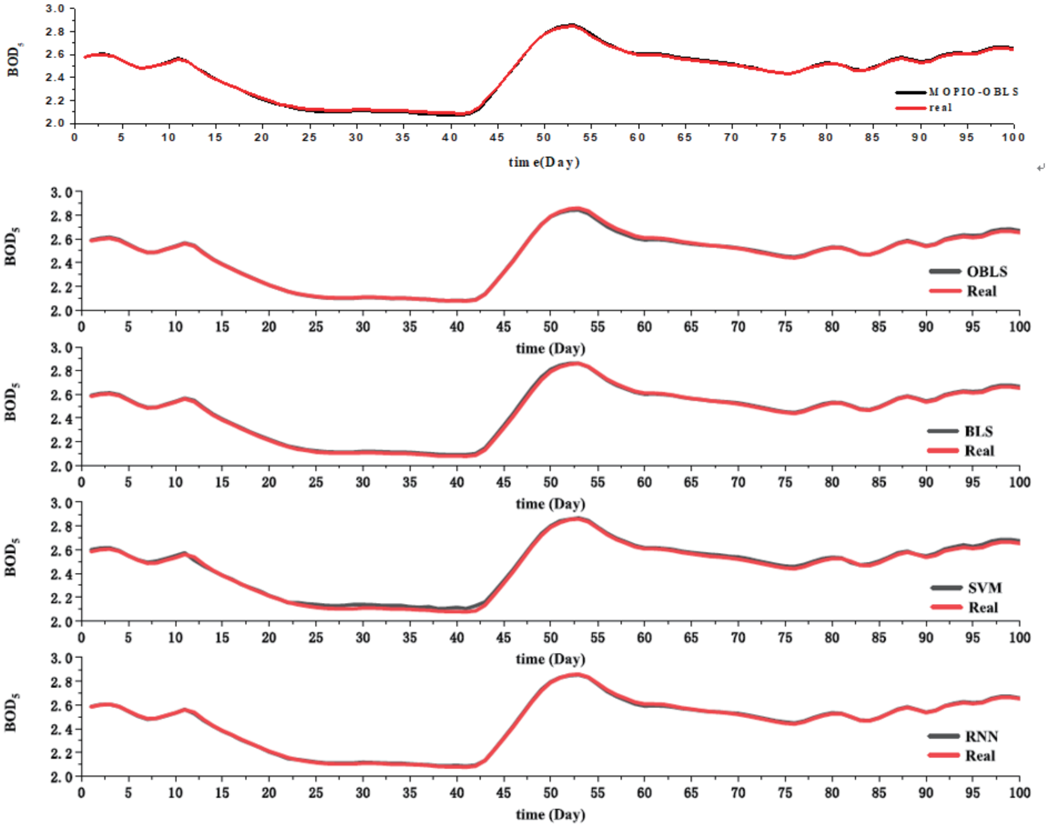


Figure 9: Measurement results of BOD_5 in rainy environment.

Test the performance of the soft sensor system in a more extreme working environment (rainstorm). Figure 10 depicts the soft measurement results of BOD_5 in effluent by the proposed strategy and the comparison methods under the condition of rainstorm. The change of the collected data in the rainstorm condition is very dramatic, which further imposes challenges on the data processing abilities of the models. It can be seen that OBLS still performs well in this weather condition, but it is difficult to directly compare the advantages and disadvantages of the four methods only through the result curve. Therefore, the performance of the optimized MOPIO-OBLS is supplemented to comparison with that of the unoptimized OBLS and the other three models, where the RMSE values of the five methods are observed and compared, as shown in Table 5. It can be clearly concluded that in extreme weather conditions, the performance of optimized MOPIO-OBLS is better than that of the unoptimized version, and the accuracy the robustness of OBLS are better than those of the other three models.

4.4. Discussion

By observing the number of nodes in sunny environment, it can be judged that the effluent BOD_5 value changes slightly on sunny days, so more nodes are set by the optimization algorithm for the system to better capture the small changes of parameters. Comparatively, since the fluctuation of the effluent BOD_5 is more obvious in rainy days, there is a minor demand for a large number of nodes to capture small changes. By prudently weighing the accuracy and complexity, the total number of nodes set by MOPIO for rainy days is half of that in sunny days. As for the extreme rainstorm environment, the total number of optimized nodes is basically the same as the default value, but more feature nodes are allocated to the system by MOPIO, emphasizing the feature extraction ability of the system. In view of the experimental results, the values of RMSE have been significantly improved under all the three conditions, which verifies that the MOPIO algorithm does make an essential improvement of the OBLS soft sensing measurement system.

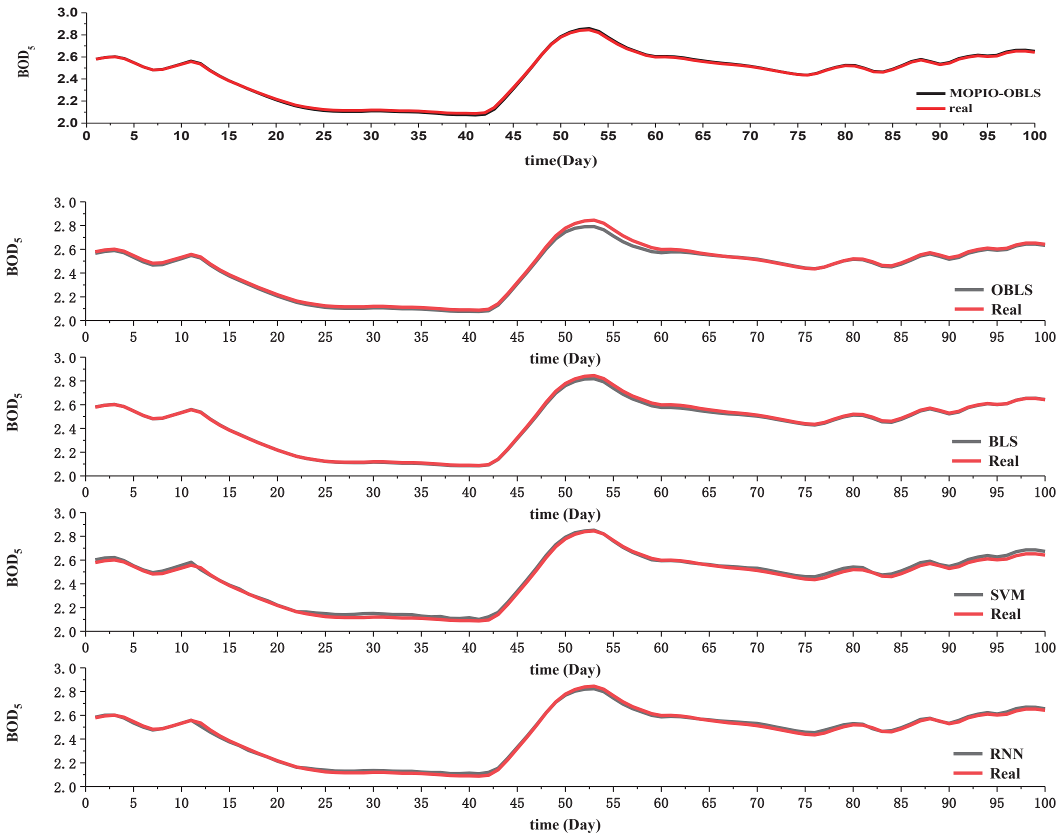


Figure 10: Measurement results of BOD_5 in rainstorm environment.

5. Conclusion

In this paper, a soft sensing measurement strategy based on feature-enhanced Broad Learning System with intelligent setting of hyper-parameters, namely MOPIO-OBLS, is proposed to conducting soft sensing on the effluent qualities. To be specific, the feature extraction ability of BLS is vastly improved with the introduction of OICA, capturing the non-Gaussian characteristics more efficiently in the sewage data. In addition, on the basis of the relatively high speed of modeling in OICA, MOPIO is carried out to optimize the determinations of hyper-parameters, thus essentially enhancing the adaptability of the model. Finally, the MOPIO-OBLS proposed in this paper and the state-of-the-art methods, OBLS, BLS and RNN, as well as the classically-used models, SVM, are simulated and compared in the simulation platform BMS1. The experimental results show that the Root Mean Square Error of the proposed OBLS network model is smaller than that of methods based on BLS, SVM and RNN, and the introduction of MOPIO can lead to a preferable optimization of the systems performance.

However, it is undeniable that comparative experiments were only carried out on simulation platform instead of the real-world industrial sites. In search of continuous improvement of the industrial wastewater treatment process, it is considered for future work to apply the proposed soft sensing strategy to the actual wastewater treatment scenarios. Moreover, scaling the prediction of the innovative model to other industrial domains such as microbial pharmaceutical field as well as utilizing our proposed model into fault warnings are also the possible directions of our future work.

References

Aarnio, P., Minkkinen, P., 1986. Application of partial least-squares modelling in the optimization of a waste-water treatment plant. Analytica Chimica Acta 191, 457–460.

Alex, J., Benedetti, L., Copp, J., Gernaey, K., Jeppsson, U., Nopens, I., Pons, M., Rieger, L., Rosen, C., Steyer, J., et al., 2008. Benchmark simulation model no. 1 (bsm1). Report by the IWA Taskgroup on benchmarking of control strategies for WWTPs , 19–20.

Beck, A., Teboulle, M., 2009. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM journal on imaging sciences 2, 183–202.

Bergstra, J., Bengio, Y., 2012. Random search for hyper-parameter optimization. Journal of machine learning research 13.

Blom, H.A., 1996. Indirect measurement of key water quality parameters in sewage treatment plants. Journal of Chemometrics 10, 697–706.

Chang, P., Ding, C., 2022. Monitoring multi-domain batch process state based on fuzzy broad learning system. Expert Systems with Applications 187, 115851.

Chang, P., Li, Z.Y., 2021. Over-complete deep recurrent neutral network based on wastewater treatment process soft sensor application. Applied Soft Computing 105, 107227.

Chang, P., Li, Z.Y., Wang, G.M., Wang, P., 2021. An effective deep recurrent network with high-order statistic information for fault monitoring in wastewater treatment process. Expert Systems with Applications 167, 114141.

Chang, P., Lu, R.W., 2021. Process monitoring of batch process based on overcomplete broad learning network. Engineering Applications of Artificial Intelligence 99, 104139.

Chang, P., Lu, R.W., Olivia, K., Wang, K., 2020a. Batch process fault detection for multi-stage broad learning system. Neural Networks 129, 298–312.

Chang, P., Wang, K., Wang, P., 2020b. Quality relevant over-complete independent component analysis based monitoring for non-linear and non-gaussian batch process. Chemometrics and Intelligent Laboratory Systems 205, 104140.

Chang, P., Zhao, L.L., Meng, F.C., Xu, Y., 2022. Soft measurement of effluent index in sewage treatment process based on overcomplete broad learning system. Applied Soft Computing 115, 108235.

Chen, C.P., Liu, Z.L., 2017. Broad learning system: An effective and efficient incremental learning system without the need for deep architecture. IEEE transactions on neural networks and learning systems 29, 10–24.

Chen, S., Hong, X., Harris, C.J., 2010. Particle swarm optimization aided orthogonal forward regression for unified data modeling. IEEE Transactions on Evolutionary Computation 14, 477–499.

Chen, T., Babanin, A., Muhammad, A., Chapron, B., Chen, C., 2020. Modified evolved bat algorithm of fuzzy optimal control for complex nonlinear systems. Romanian Journal of Information Science and Technology 23, 28–40.

Chiang, H.S., Shih, D.H., Lin, B., Shih, M.H., 2014. An apn model for arrhythmic beat classification. Bioinformatics 30, 1739–1746.

Duan, H.B., Qiao, P.X., 2014. Pigeon-inspired optimization: a new swarm intelligence optimizer for air robot path planning. International journal of intelligent computing and cybernetics .

Esfe, M.H., Razi, P., Hajmohammad, M.H., Rostamian, S.H., Sarsam, W.S., Arani, A.A.A., Dahari, M., 2017. Optimization, modeling and accurate prediction of thermal conductivity and dynamic viscosity of stabilized ethylene glycol and water mixture al2o3 nanofluids by nsga-ii using ann. International Communications in Heat and Mass Transfer 82, 154–160.

Han, H.G., Liu, Z., Qiao, J.F., 2019. Fuzzy neural network-based model predictive control for dissolved oxygen concentration of wwtps. International Journal of Fuzzy Systems 21, 1497–1510.

Han, H.G., Wu, X.L., Zhang, L., Tian, Y., Qiao, J.F., 2017. Self-organizing rbf neural network using an adaptive gradient multiobjective particle swarm optimization. IEEE transactions on cybernetics 49, 69–82.

Heddam, S., Lamda, H., Filali, S., 2016. Predicting effluent biochemical oxygen demand in a wastewater treatment plant using generalized regression neural network based approach: a comparative study. Environmental Processes 3, 153–165.

Hu, W., Yen, G.G., 2013. Adaptive multiobjective particle swarm optimization based on parallel cell coordinate system. IEEE Transactions on Evolutionary Computation 19, 1–18.

Huang, M.Z., Ma, Y.W., Wan, J.Q., Chen, X.H., 2015. A sensor-software based on a genetic algorithm-based neural fuzzy system for modeling and simulating a wastewater treatment process. Applied Soft Computing 27, 1–10.

Hulland, J., 1999. Use of partial least squares (pls) in strategic management research: A review of four recent studies. Strategic management journal 20, 195–204.

Lee, D.S., Lee, M.W., Woo, S.H., Kim, Y.J., Park, J.M., 2006. Multivariate online monitoring of a full-scale biological anaerobic filter process using kernel-based algorithms. Industrial & engineering chemistry research 45, 4335–4344.

Mao, C.Y., Lin, R.R., Xu, C.F., He, Q., 2017. Towards a trust prediction framework for cloud services based on pso-driven neural network. IEEE Access 5, 2187–2199.

Pisa, I., Santín, I., Vicario, J.L., Morell, A., Vilanova, R., 2019. Ann-based soft sensor to predict effluent violations in wastewater treatment plants. Sensors 19, 1280.

Podosinnikova, A., Perry, A., Wein, A.S., Bach, F., dAspremont, A., Sontag, D., 2019. Overcomplete independent component analysis via sdp, in: The 22nd International Conference on Artificial Intelligence and Statistics, PMLR. pp. 2583–2592.

Precup, R.E., David, R.C., Roman, R.C., Szedlak-Stinean, A.I., Petriu, E.M., 2021. Optimal tuning of interval type-2 fuzzy controllers for nonlinear servo systems using slime mould algorithm. International Journal of Systems Science , 1–16.

Precup, R.E., Hedrea, E.L., Roman, R.C., Petriu, E.M., Szedlak-Stinean, A.I., Bojan-Dragos, C.A., 2020. Experiment-based approach to teach optimization techniques. IEEE Transactions on Education 64, 88–94.

Qiao, J.F., Li, W., Han, H.G., 2014. Soft computing of biochemical oxygen demand using an improved t–s fuzzy neural network. Chinese Journal of Chemical Engineering 22, 1254–1259.

Qiao, J.F., Sun, Y.Q., Han, H.G., 2018. Prediction of effluent ammonia nitrogen based on improved k-means algorithm optimizing rbf neural network [j]. Control Engineering of China 25, 375–279.

Qiu, Y., Liu, Y.Q., Huang, D.P., 2016. Date-driven soft-sensor design for biological wastewater treatment using deep neural networks and genetic algorithms. Journal of chemical engineering of Japan 49, 925–936.

Tan, G.W.H., Ooi, K.B., Leong, L.Y., Lin, B., 2014. Predicting the drivers of behavioral intention to use mobile learning: A hybrid sem-neural networks approach. Computers in Human Behavior 36, 198–213.

Teppola, P., Mujunen, S.P., Minkkinen, P., 1999. Kalman filter for updating the coefficients of regression models. a case study from an activated sludge waste-water treatment plant. Chemometrics and intelligent laboratory systems 45, 371–384.

Xiao, H.J., Huang, D.P., Pan, Y.P., Liu, Y.Q., Song, K., 2017. Fault diagnosis and prognosis of wastewater processes with incomplete data by the auto-associative neural networks and arma model. Chemometrics and Intelligent Laboratory Systems 161, 96–107.

Yuan, X.F., Huang, B., Wang, Y.L., Yang, C.H., Gui, W.H., 2018. Deep learning-based feature representation and its application for soft sensor modeling with variable-wise weighted sae. IEEE Transactions on Industrial Informatics 14, 3235–3243.

Yuan, X.F., Li, L., Wang, Y.L., 2019. Nonlinear dynamic soft sensor modeling with supervised long short-term memory network. IEEE transactions on industrial informatics 16, 3168–3176.

Zapata, H., Perozo, N., Angulo, W., Contreras, J., 2020. A hybrid swarm algorithm for collective construction of 3d structures. International Journal of Artificial Intelligence. 18, 1–18.

Zhong, W.H., Guan, H.W., Ma, X.S., Peng, X.H., 2010. Compensatory fuzzy neural network modeling in a wastewater treatment process, in: 2010 IEEE International Conference on Intelligent Systems and Knowledge Engineering, IEEE. pp. 39–42.