Complex & Intelligent Systems (2021) 7:2199–2210
https://doi.org/10.1007/s40747-021-00395-w

ORIGINAL ARTICLE

Reinforcement learning-based particle swarm optimization for sewage treatment control

Lu Lu[1] · Hui Zheng[1] · Jing Jie[1] · Miao Zhang[1] · Rui Dai[1]

Received: 21 January 2021 / Accepted: 6 May 2021 / Published online: 28 May 2021

Abstract

To solve the problem of high-energy consumption in activated sludge wastewater treatment, a reinforcement learning-based particle swarm optimization (RLPSO) was proposed to optimize the control setting in the sewage process. This algorithm tries to take advantage of the valid history information to guide the behavior of particles through a reinforcement learning strategy. First, an elite network is constructed by selecting elite particles and recording their successful search behavior. Then the network is trained and evaluated to effectively predict the particle velocity. In the periodic wastewater treatment process, the RLPSO runs repeatedly according to the optimized cycle. Finally, RLPSO was tested based on Benchmark Simulation Model 1 (BSM1) of sewage treatment, and the simulation results showed that it could effectively reduce the energy consumption on the premise of ensuring qualified water quality. Furthermore, the performance of RLPSO was analyzed using the benchmarks with higher dimension, which verifies the effectiveness of the algorithm and provides the possibility for RLPSO to be applied to a wider range of problems.

Keywords Wastewater treatment · Reinforcement learning · Particle swarm optimization (PSO) · Cycle optimization

Introduction

The activated sludge method is a biological sewage treatment method commonly used in the wastewater treatment processes (WWTP) [1, 2]. Through biochemical reaction, the pollutants in the sewage are adsorbed, decomposed and oxidized, so the pollutants are degraded and separated from the sewage to achieve the purification of the sewage [3–6]. To ensure that the effluent water quality reaches the standard, it is necessary to fill the aeration tank with appropriate oxygen through the blower to maintain the concentration of dissolved oxygen (S_O) in the aerobic area, and use the reflux pump to maintain the concentration of nitrate nitrogen (S_{NO}) in the anoxic zone [7]. However, the operation of blower and reflux pump requires a large amount of energy loss, which inevitably increases the operation cost. At the same time, from the perspective of biochemical reaction mechanism, suitable S_O and S_{NO} are helpful to ensure the successful progress of nitrification and denitrifying reactions [8, 9]. Therefore, it is necessary to dynamically optimize S_O and S_{NO} and construct the control strategy aiming at reducing the energy consumption (*EC*) in the sewage treatment process on the premise of ensuring qualified effluent quality (*EQ*).

With the characteristics of nonlinearity, time variation and strong coupling, the control issues in the WWTP have been extensively investigated. The main challenge of WWTP is to construct an optimal control strategy with the aim of reducing *EC* while ensuring qualified *EQ*. For example, Vrečko presented a PI-based control strategy including feedforward control and a step-feed procedure, which was applied to WWTP [10]. Furthermore, Vrečko et al. presented a model predictive controller (MPC) for ammonia nitrogen, which gives better results in terms of ammonia removal and aeration energy consumption than PI controller [11]. Mulas proposed a dynamic matrix-based predictive control algorithm, which is able to decrease the energy consumption costs and, at the same time, reduce the ammonia peaks and nitrate concentration [12]. Han et al. proposed an efficient self-organizing sliding-mode controller (SOSMC) to suppress the disturbances and uncertainties of WWTP [13]. However, in the above algorithm, the concentration setting

✉ Hui Zheng
111003@zust.edu.cn

✉ Jing Jie
jingjie@zust.edu.cn

[1] Zhejiang University of Science and Technology, Hangzhou 310023, China

values of the key variables in sewage process are fixed or changed according to the preset trajectories, without considering the real-time influence of sewage quality and flow rate.

Sewage treatment is a complex dynamic reaction process. To reduce *EC* under the statue of meeting *EQ* standards, more and more intelligent algorithms are presented to dynamically optimize the setting values of key variables in WWTP. For examples, Hakanen et al. designed a multi-objective interactive wastewater treatment software based on differential evolution (DE), using variables such as the *So* setpoint in last aerobic zone and the methanol dose as decision variables [14]. Han et al. proposed a Hopfield neural network method (HNN) based on Lagrange multiplier for the optimal control of pre-denitrification WWTP [15]. Yang used an artificial immune network-based combinatorial optimization algorithm (Copt-ai Net) to determine the optimal set values of *So* and S_{No} [16]. In [17], an adaptive multi-objective evolutionary algorithm based on decomposition (AMOEA/D) is developed with the usage of *EC* and *EQ* as objectives to be optimized.

However, sewage treatment is a cyclical process, that is, optimization calculations should be performed in intervals, which can result in high fitness evaluations (*FEs*) cost for optimization. In the above intelligent control algorithm, sewage treatment information is not fully utilized. The subsequent optimization does not extract useful information from the previous optimization process, and the previous optimization does not play a guiding role for the subsequent optimization.

In the cycle optimization process, information storage and reuse can improve computing efficiency and sewage treatment effect. Inspired by reinforcement learning mentioned in [18, 19], and considering the simple operation and fast convergence of particle swarm optimization algorithm (PSO) [20–23], we propose a wastewater treatment control method based on reinforcement learning particle swarm optimization (RLPSO). This method introduces a reinforcement learning strategy in the particle update. First, select the elite particles, record their concentration setting values and adjustment trends, and construct an elite particle set. Then an elite network was trained and used as the strategy function to predict the particle velocity. Finally, a simplified evaluation method is utilized to calculate the state value function which is used to update the elite network model.

The remainder of the paper is organized as follows. The next section introduces the international Benchmark Simulation Model 1 (BSM1) of WWTP and optimization objective function. The subsequent section describes RLPSO in detail. Then the experiment results and analysis are shown. The final section provides the conclusion and outlook.

Wastewater treatment processes optimization

In WWTP, the main reaction is carried out by biological reactor and secondary sedimentation tank. The biological reactor consists of five units. The first two units are anaerobic zones, which mainly complete denitrification reaction, while the last three are aerobic zones, which mainly complete nitrification reaction. To evaluate and compare different optimal control strategies, the Benchmark Simulation Model 1 (BSM1) [24–26] was developed by the IWA (International Water Association) and COST (European Cooperation in the Field of Science and Technology), shown in Fig. 1. In BSM1, there are two control loops, S_O and S_{NO}. The first control loop tunes the dissolved oxygen concentration in the fifth unit S_O by changing the oxygen transfer coefficient K_{La5}. The second control loop tunes the nitrate nitrogen level in the second unit S_{NO} by changing the internal recirculation flow rate Q_a. The two control loops adopt proportional integral controller (PI). However, due to the influence of weather or users, sewage quality keeps changing. If S_O or S_{NO} is set at a constant value, it is difficult to maintain the optimal balance between *EQ* and *EC*. Therefore, it is necessary to dynamically optimize the set values of S_O and

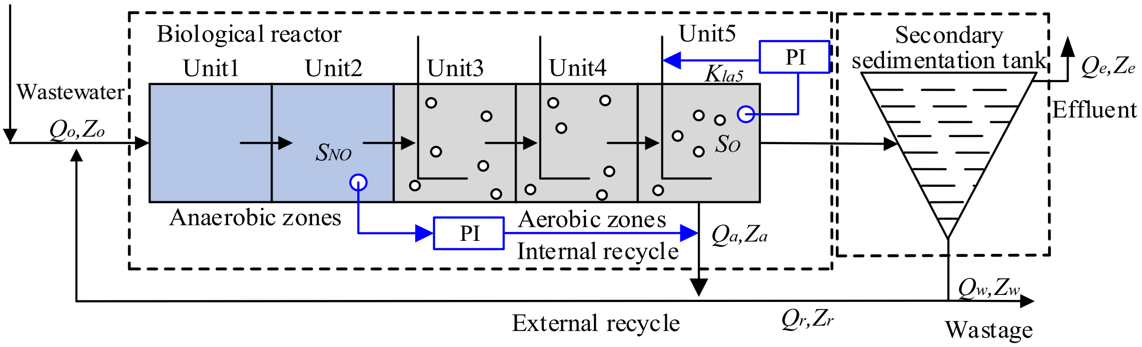


Fig. 1 The architecture of the BSM1

S_{NO} and construct the optimized control strategy aiming at reducing *EC* on the premise of ensuring qualified *EQ*.

Aeration energy (*AE*) and pumping energy (*PE*) consumption accounts for more than 70% of total energy consumption, so the *EC* of the optimization problem is defined as the sum of *AE* and *PE*:

$$EC = AE + PE. \tag{1}$$

According to the BSM1 mechanism model, *AE* and *PE* are defined as follows, respectively [27]:

$$AE = \frac{S_{O.\mathrm{sat}}}{T \times 1.8 \times 1000} \int_t^{t+T} \sum_{i=1}^{5} V_i \cdot K_{\mathrm{Lai}}(t)\mathrm{d}t, \tag{2}$$

$$PE = \frac{1}{T} \int_t^{t+T} (0.004Q_a(t) + 0.05Q_w(t) + 0.008Q_r(t))\mathrm{d}t, \tag{3}$$

where K_{Lai} is the oxygen conversion coefficient and V_i is the volume of the *i*th biological reactor, respectively. $\mathrm{S}_{o,\mathrm{sat}}$ is the saturation concentration for oxygen. *T* is the evaluation cycle. Q_a, Q_r, and Q_w denote the internal recycle flow rate, return sludge recycle flow rate and waste sludge flow rate, respectively. Z_a, Z_r, and Z_w are the corresponding components concentrations.

EQ represents the fine to be paid for the discharge of water pollutants to the receiving water body. According to the definition of BSM1, the equation of *EQ* is [27]

$$EQ = \frac{1}{T \times 1000} \int_t^{t+T} (2SS(t) + COD(t) + 30S_{NO}(t) + 10S_{Nkj}(t) + 2BOD_5(t))\mathrm{d}t, \tag{4}$$

where *SS*, *COD*, S_{NO}, S_{NKj} and BOD_5 are suspended solid concentration, chemical oxygen demand, nitrate concentration, Kjeldahl concentration, and biochemical oxygen demand, respectively. *EQ* value will impact the operation cost of WWTP if the effluent discharge fee is executed strictly.

In addition to *EQ*, the five effluent parameters should meet the following standards specified in BSM1 [28]:

$$\begin{aligned} &N_{\mathrm{tot}} \le 18\mathrm{mg/L}, COD \le 100\mathrm{mg/L}, \\ &S_{NH} \le 4\mathrm{mg/L}, SS \le 30\mathrm{mg/L}, \\ &BOD_5 \le 10\mathrm{mg/L}, \end{aligned} \tag{5}$$

where $N_{\mathrm{tot}} = S_{NO} + S_{NKj}$. S_{NH} denotes influent ammonium.

In summary, the constrained objective optimization function of the WWTP is

$$\min\ \mathrm{f} = c \cdot EC + EQ\,, \tag{6}$$

where *c* is the weight coefficient and the set values of S_O and S_{NO} are the decision variables. Since sewage treatment is a dynamic and periodic optimization process, we proposed a RLPSO control strategy to minimize the objective optimization function (6) by dynamically adjusting the set values of S_O and S_{NO}, to improve the sewage treatment efficacy and reduce the operating cost.

Reinforcement learning-based particle swarm optimization

Particle swarm optimization

PSO originated from the study of the behavior of preying on birds, its basic idea is that whole swarm of birds will tend to follow the bird which found the best path to food [29]. To search an optimum, PSO defines a swarm of particles to represent the potential solutions to an optimization problem. Each particle begins with an initial position randomly and flies through the *D*-dimensional solution space. The flying behavior of each particle can be described by its velocity and position as the following.

$$v_{id}(k+1) = \omega v_{id}(k) + c_1 r_1 (p_{id}(k) - x_{id}(k)) + c_2 r_2 (p_{gd}(k) - x_{id}(k)), \tag{7}$$

$$x_{id}(k+1) = x_{id}(k) + v_{id}(k+1), \tag{8}$$

where $V_i = (v_{i1}, v_{i2}, \ldots, v_{id}, \ldots, v_{iD})$ is the velocity vector of the *i*th particle; $X_i = (x_{i1}, x_{i2}, \ldots, x_{id}, \ldots, x_{iD})$ is the position vector of the *i*th particle; $P_i = (p_{i1}, p_{i2}, \ldots, p_{id}, \ldots, p_{iD})$ is the best position found by the *i*th particle; $P_g = (p_{g1}, p_{g2}, \ldots, p_{gd}, \ldots, p_{gD})$ is the global best position found by the whole swarm. c_1, c_2 are two learning factors, usually $c_1 = c_2 = 2$ [29]; r_1, r_2 are random numbers between (0, 1) [30]; ω is the inertia weight to control the velocity, which may decrease linearly starting at 0.9 and ending at 0.4 or $\omega \in (0, 1)$ [30].

WWTP is a process of periodic optimization. This is because the WWTP is a complex system with large lag, which is difficult to operate in real time. Therefore, it is necessary to set the cycle time and carry out the optimization calculation in each cycle. However, PSO has the characteristics of random initialization to improve the diversity, and only consider the individual optimum and global optimum during the optimization, ignoring the inherent properties of the system. If PSO is directly applied to WWTP, information from previous cycles does not provide any guidance for subsequent optimization processes, which will lead to low efficiency. To improve the treatment effect, it is necessary to record the influence of the set values of S_O and S_{NO} on the sewage parameters, and reuse the information to the optimization process, which will provide reference data for the next optimization calculation. So we consider adding a prediction item to Eq. (7), as shown in the following:

$$v_{id}(k+1) = \omega v_{id}(k) + c_1 r_1(p_{id}(k) - x_{id}(k)) + c_2 r_2(p_{gd}(k) - x_{id}(k)) + r_\mu v_{id\mu}(k+1), \tag{9}$$

where $v_{id\mu}$ is the dth dimensional predicted velocity of particle i by the strategy function μ, and r_μ is the prediction coefficient. According to Eq. (9), the velocity direction of particles is determined by four parts: inertial velocity, individual historical optimum, global optimum, and prediction item. On the one hand, it draws the advantages of PSO, which is both self-cognition and group sociality, on the other hand, the prediction item infuses PSO with historical information, which is more suitable for repeated cycle optimization problems. To determine the prediction item $v_{id\mu}$, we introduce reinforcement learning (RL) [31] strategy to PSO.

Reinforcement learning strategy

Reinforcement learning interacts with the environment through a trial-and-error mechanism and learns optimal strategies by maximizing cumulative rewards. Reinforcement learning agent mainly includes four basic elements: environment, state (s), action (a) and reward (R) [31]. During operation, the agent determines an action a according to the current state s through the strategy function μ, executes the action, and enters the next state. At the same time, the system returns the value R to reward or punish the action. The process runs repeatedly to maximize the expected benefits of the agent.

In the similar way, reinforcement learning-based PSO (RLPSO) includes four basic elements shown in Fig. 2. The agent is a particle in population and the environment is the WWTP in the paper. The state s is the position X of each particle in the population; the action a is the velocity V prediction strategy, which is determined by the strategy function μ. The reward value R is related to the fitness value f of the optimization problem. Therefore, to obtain the particle velocity prediction $v_{id\mu}$, we need to establish the strategy function μ according to reward value R.

In the RLPSO, the particle agent predicts the speed according to the strategy function μ:

$$v_{id\mu}(k+1) = \mu(X_i(k)). \tag{10}$$

In this paper, the strategy function μ is described as an elite network model. By learning the information of elite particles, the elite network model was trained. The process is mainly divided into three steps: elite particle set construction, strategy function training and elite network model evaluation. The details are described as follows.

Elite particle set construction

Elite network model is trained with elite particle information to guide the search of offspring population. The first step in the kth iteration is to select elite particles based on the reward value $R(k)$. In the iteration process, the reward value $R(k)$ is determined according to the fitness variation value, as shown in the following equation:

$$\begin{cases} if\ f(k+1) - f(k) > 0,\ R(k) = 1; \\ if\ f(k+1) - f(k) \le 0,\ R(k) = -1; \end{cases} \tag{11}$$

where $f(k)$ is the fitness value of the kth iteration, $k = 0, ..., K-1$. K is the maximum number of iterations of each run. Only the particles with reward value $R(k) = 1$ are selected as the elite particles, and then the position $X_i(k)$ before the update of the elite particles and the speed $V_i(k)$ after the update are saved to construct the elite particle set Ω_e.

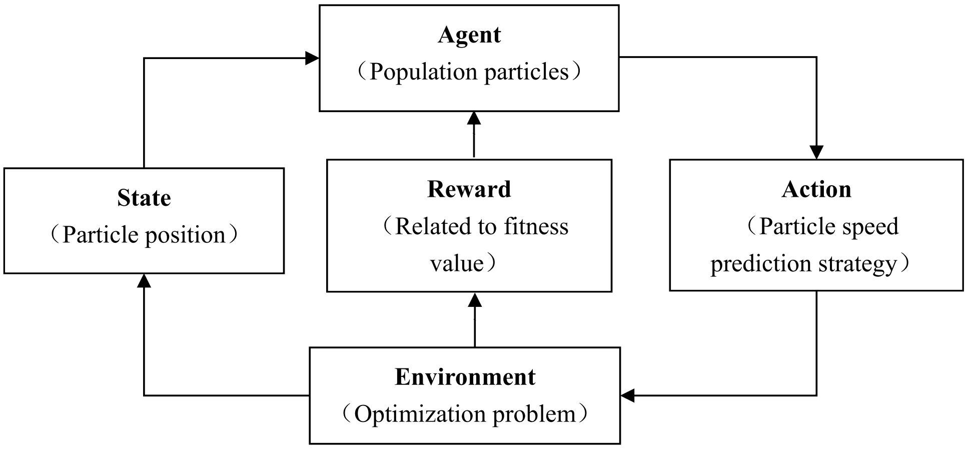


Fig. 2 Schematic diagram of interaction between particle agent and environment

Strategy function training

The elite particle set Ω_e is used to save the position x of the elite particle before the update and the speed v after the update. RLPSO uses a limited capacity elite particle set to store elite particles. Suppose the number of elite particle set Ω_e is N_e. $\Omega_e^{'}$ is the newly generated elite particle set, and its number is $N_e^{'}$. If $N_e + N_e^{'}$ exceeds the finite capacity value N_{em}, all the elite particles $\Omega_e + \Omega_e^{'}$ are sorted according to the fitness value, only the first N_e items are stored into Ω_e again, and the original data is overwritten. The elite particle set Ω_e is used as a data set. In the data set, the particle position is input and the speed is output, and then a neural network model is trained to obtain the elite network model Φ. The trained elite network model Φ is used as the strategy function μ to guide the particle operation. With the elite network model Φ, the particle velocity can be predicted according to the particle position $X_{i:}$

$$v_{id\mu}(k+1) = \Phi(X_i(k)). \tag{12}$$

Elite network model evaluation

With the continuous update and change of the elite particle set Ω_e, the elite network model will be evaluated after the training. During the evaluation process, the new model and the original model are used to guide the particle optimization process. To better reflect the influence of the strategy function μ on the particle velocity update, RLPSO velocity update equation is simplified as

$$v_{id}(k+1) = \omega v_{id}(k) + r_\mu v_{id\mu}(k+1). \tag{13}$$

When the termination conditions $k \geq K$ is satisfied, the optimal fitness value obtained by the guidance of the new elite network model is set to f_1^*, and the optimal fitness value obtained by the original network model is set to f_2^*. If $f_1^* > f_2^*$, it means that the prediction effect of the new network model is better. Set the new network reward value $R(K) = 1$ after the iteration, otherwise $R(K) = -1$.

Considering the randomness of particles, the above evaluation process is repeated M times to estimate the state value function $\widehat{V}_\mu(X)$:

$$\widehat{V}_\mu(X) = \sum_{m=1}^{M} R^m(K), \tag{14}$$

where $\widehat{V}_\mu(X)$ represents the average reward that can be obtained after the particle X moves through the strategy function μ. If $\widehat{V}_\mu(X) > 0$, the new model is considered better than the original model, and the new network is used to replace the original network. If $\widehat{V}_\mu(X) \leq 0$, keep the original network. By comparing the two models, we determine the prediction model required by the subsequent algorithm.

Algorithm procedure

The algorithm procedure is described below.

```
RLPSO
1. Initialize particle position $X_i$ and velocity $V_i$, $i = 1, 2, ..., N$
2. Let Run = 1. Update the particle position and velocity according to Eqs. (7) and (8). In the iterative process ($k < K$), select the particle with reward value $R(k) = 1$ as the elite particle, establish the elite particle set $\Omega_e$ and train the elite network model $\Phi$
3. Randomly generate $N$ particles. Let $r_\mu \neq 0$, use the elite network model $\Phi$ to predict the particle velocity $v_{id\mu}$, and update the particle position and velocity according to Eqs. (8) and (9). At the same time, continue to select particles with reward value $R(k) = 1$ as elite particles. If the number of elite particles exceeds the limited capacity, establish a new elite particle set $\Omega_e^{'}$ and train a new elite network model $\Phi'$
4. Evaluation of the elite network model. According to Eqs. (8) and (13), the original model $\Phi$ and the new model $\Phi'$, respectively, instruct the particle swarm to run $M$ times to calculate the estimated value $\widehat{V}_\mu(X)$. If $\widehat{V}_\mu(X) > 0$, the new model $\Phi'$ replaces the original model $\Phi'$
5. Run = Run + 1. If terminal condition is not satisfied, return 3
```

Experiments

Simulation experiment of RLPSO based on BSM1

The proposed RLPSO is simulated on BSM1 platform and compared with PI controller, CPSO [32], SLPSO [33], PSO [34], APSO [35], DE [36], HNN [15], Copt-ai Net [16] and AMOEA/D [17]. The simulation conditions are based on the sunny and good weather in the BSM1. The parameters of the HNN, Copt-ai Net and AMOEA/D algorithms are determined by the original papers. Besides, the other algorithms parameters are set as follows.

The selection time of simulation data is 14 days. The sampling interval is 15 min, and the optimization period is 2 h. So a total of 168 runs are conducted for each algorithm. In the PI strategy, the set values $S_O = 2$ and $S_{NO} = 1$. The ranges of S_O and S_{NO} are 0.5–2 mg/L and 0.8–2 mg/L, respectively, in RLPSO, CPSO, SLPSO, PSO, APSO and DE. In the objective optimization function of Eq. (6), $c = 0.1$.

During the optimization, one difficulty in employing RLPSO in BSM1 is the huge time consumption for fitness evaluations (*FEs*), the algorithm is required not only to satisfy the optimization accuracy, but also to accelerate the convergence speed. Therefore, for RLPSO, the population size N is set to 10, $r_\mu = 0.3$, $\omega = 0.4$, $D = 2$, and $K_{\max} = 40$.

c_1 and c_2 are 2. We can figure out that *FEs* is 67,200. Early experiments show that RLPSO with the setting parameters is nearly convergent in the 40th iteration, and meets the requirements of *EQ* and *EC*. For the convenience of comparison, the inertia weights of the PSO-based algorithms are all selected as 0.7. In DE algorithm, the mutation rate is 0.5 and the crossover probability is 0.9. The population size and iterations number of these algorithms are the same as RLPSO.

Table 1 shows the comparison of *EQ* and *EC* under several strategies. As can be seen from Table 1, compared with PI strategy, all of these intelligent algorithms can reduce *EC* by optimizing the set values of S_O and S_{NO}. Among them, the *EC* obtained by PSO algorithm is lower than RLPSO, which is 3652.40 kWh/d. However, S_{NH} concentration via PSO is 4.19 mg/L, which exceeds the limit of 4 mg*/L*. Similarly, S_{NH} concentration obtained by DE and APSO also exceeds the standard. Besides, *EC* obtained by CPSO, SLPSO, HNN, and Copt-ai Net is obviously higher than that of RLPSO, which proves that RLPSO is superior to these algorithms.

We can also see from Table 1 that the *EC* obtained by AMOEA/D is slightly lower than RLPSO, but its *EQ* is higher than RLPSO. The performance of the two algorithms is comparable. But it should be noted that, in the AMOEA/D strategy, the population size *N* is 100, and $K_{\max} = 300$. We can calculate that in each optimization cycle, the *FEs* of RLPSO is just 1/75 of AMOEA/D. RLPSO can obtain *EC* similar to AMOEA/D with significantly lower *FEs*, which prove that RLPSO is more suitable for sewage treatment process.

RLPSO simulation experiment based on benchmark functions

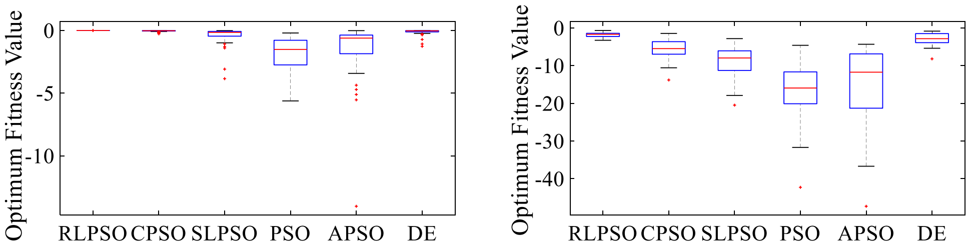
To further study the performance of RLPSO, the RLPSO algorithm is analyzed on the high dimensional general benchmarks. Six different types of benchmark functions (Rastrigin, Griewank, Ellipsoid, Rosenbrock, Sphere, and Ackley functions) are used to study the performance of RLPSO compared with CPSO, SLPSO, PSO, APSO and DE. In the algorithm, the population size $N = 10$, and the dimension D is set to 10 and 20 dimensions, respectively. Each algorithm runs 50 times. The maximum number of iterations per run is 200. Other parameters of different algorithms are the same as the BSM1 experiment. In the experiment process, f_j^* represents the optimal fitness value obtained in the *j*th run, $j = 1,2,\ldots50$.

Figures 3, 4, 5, 6, 7 and 8 show the boxplots comparison of f^* obtained by various algorithms. As can be seen from

Table 1 Comparison of *EQ* and *EC* of different control strategies in fine weather

Method	S_{NH} mg/L	N_{tot} mg/L	*BOD* mg/L	*COD* mg/L	*SS* mg/L	*AE* kWh/d	*PE* kWh/d	*EC* kWh/d	*EQ* kg poll./d
RLPSO	3.48	16.09	2.69	47.76	12.62	3414.30	263.50	3677.80	6312.0
CPSO	2.56	17.49	2.69	47.73	12.62	3600.80	221.53	3822.33	6236.5
SLPSO	3.99	15.42	2.70	47.78	12.62	3364.48	320.54	3685.02	6379.4
PSO	***4.19***	15.68	2.70	47.79	12.62	3347.98	304.42	3652.40	6493.5
APSO	***4.15***	15.52	2.70	47.79	12.62	3348.04	317.50	3665.54	6446.1
DE	***4.09***	15.61	2.70	47.78	12.62	3353.85	304.44	3658.29	6441.3
PI	2.39	16.84	2.68	47.71	12.62	3695.41	241.46	3936.87	6067.5
HNN[15]	3.24*	14.92*	2.69*	47.55*	12.62*	3435.1*	267.2*	3702.3*	–
Copt-ai Net[16]	3.48*	14.83*	2.63*	47.55*	12.51*	3540.1*	245.8*	3785.9*	–
AMOEA/D[17]	2.75*	15.36*	2.68*	47.54*	12.58*	3384.27*	254.10*	3638.35*	6345.4*

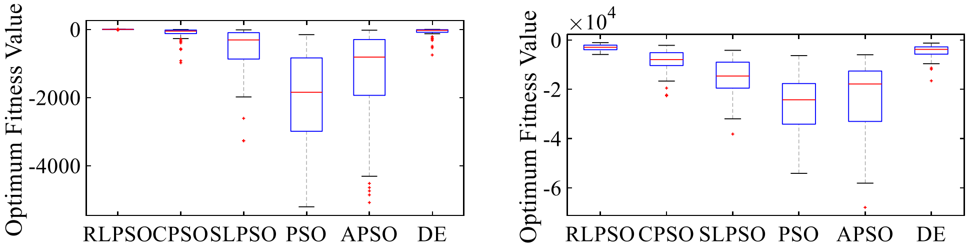
*Results are listed in the original papers, – denotes none, and the bold italic symbol indicates that the effluent parameters exceed the standard specified in BSM1



a. 10- dimension

b. 20- dimension

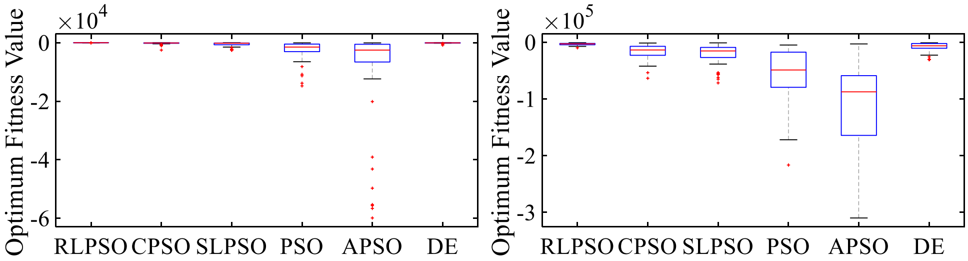
Fig. 3 Boxplot comparison of various algorithms on the Sphere benchmark



a. 10- dimension

b. 20- dimension

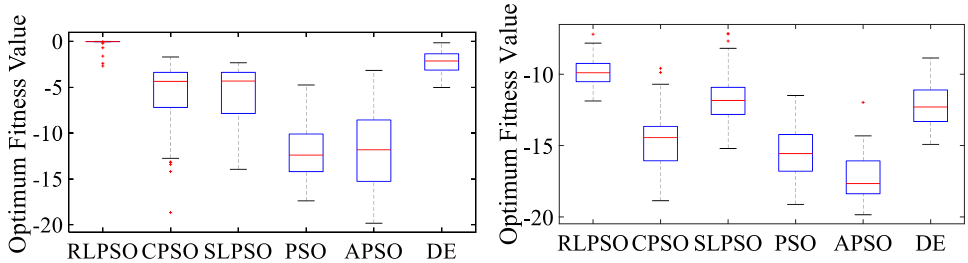
Fig. 4 Boxplot comparison of various algorithms on the Ellipsoid benchmark



a. 10- dimension

b. 20- dimension

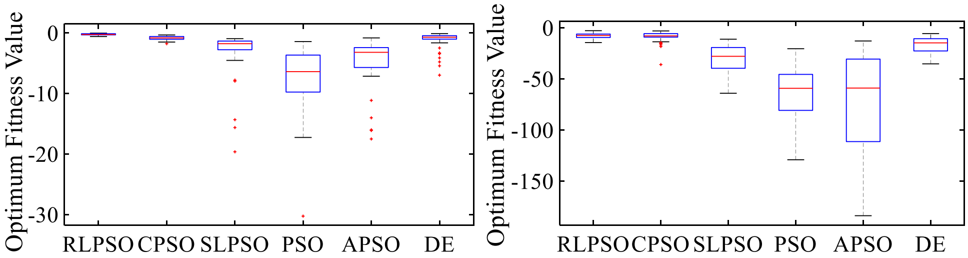
Fig. 5 Boxplot comparison of various algorithms on the Rosenbrock benchmark



a. 10- dimension

b. 20- dimension

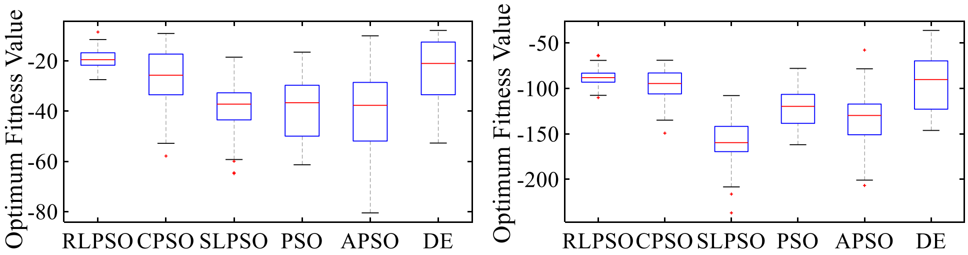
Fig. 6 Boxplot comparison of various algorithms on the Ackley benchmark



a. 10- dimension

b. 20- dimension

Fig. 7 Boxplot comparison of various algorithms on the Griewank benchmark

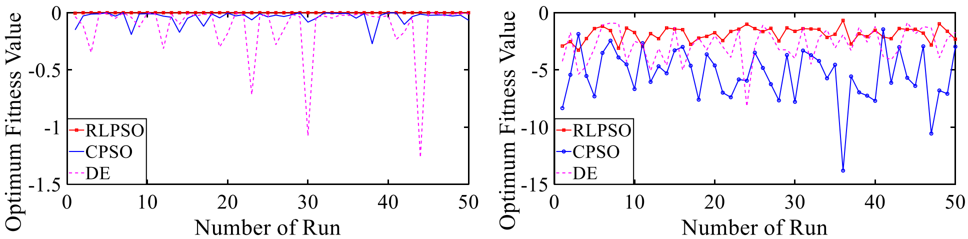


a. 10- dimension

b. 20- dimension

Fig. 8 Boxplot comparison of various algorithms on the Rastrigin benchmark

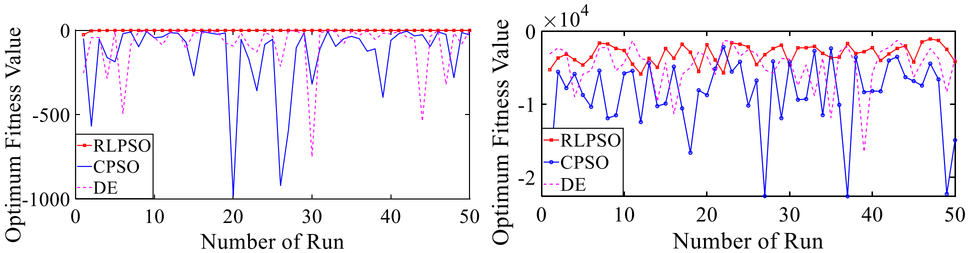
Fig. 9 Curve plot comparison of various algorithms on the Sphere benchmark



a. 10- dimension

b. 20- dimension

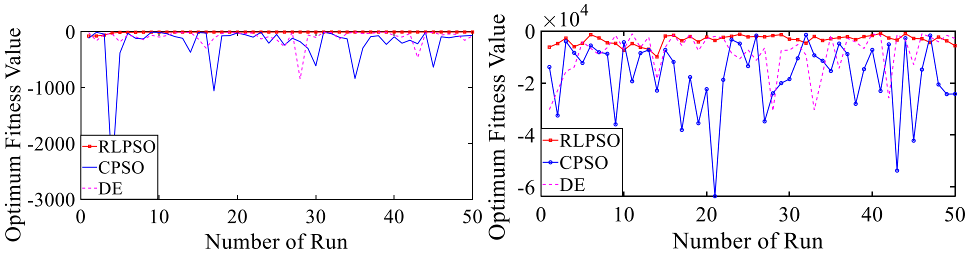
Fig. 10 Curve plot comparison of various algorithms on the Ellipsoid benchmark



a. 10- dimension

b. 20- dimension

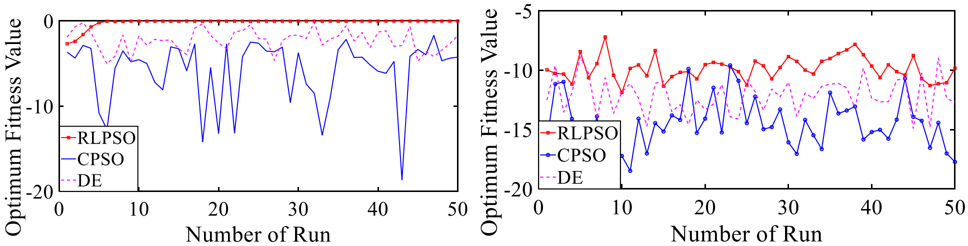
Fig. 11 Curve plot comparison of various algorithms on the Rosenbrock benchmark



a. 10- dimension

b. 20- dimension

Fig. 12 Curve plot comparison of various algorithms on the Ackley benchmark



a. 10- dimension

b. 20- dimension

the figures, median and interquartile range of RLPSO are obviously better than SLPSO, PSO and APSO, which proves that the performance of RLPSO is superior to these algorithms. Besides, RLPSO has almost no outliers, which also proves the stability of the RLPSO.

To further observe the performance of RLPSO compared with CPSO and DE, Figs. 9, 10, 11, 12, 13 and 14 show the f^* trend of these three algorithms during 50 runs. For CPSO and DE, there is no data connection between different runs, and each run is randomly initialized, so f^* trend fluctuates significantly. However, it can be seen from Figs. 9a–12a that f^* of RLPSO tends to converge. This is because RLPSO relies on elite neural network to transmit information between different runs, and the previous optimization can play a guiding role in the subsequent optimization. It should be noted that, as can be seen from Figs. 13a–14a and 9b–14b, RLPSO still has fluctuations. This is because elite network with fixed structure was selected in the training process, which resulted in a decline in RLPSO performance for

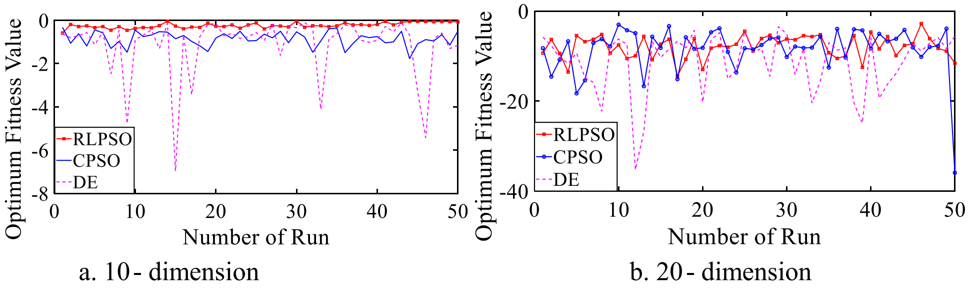


Fig. 13 Curve plot comparison of various algorithms on the Griewank benchmark

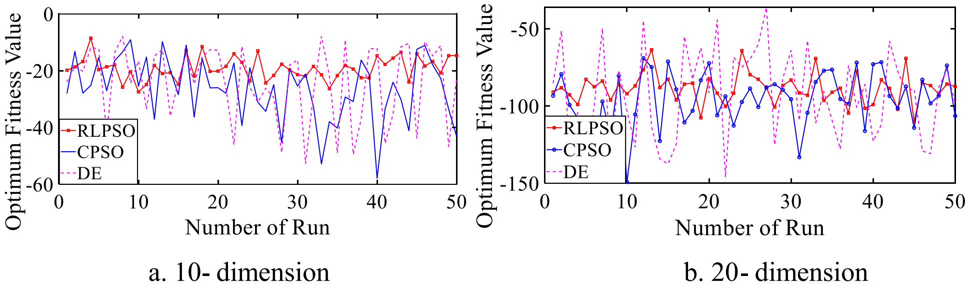


Fig. 14 Curve plots comparison of various algorithms on the Rastrigin benchmark

more complex or high-dimensional benchmarks. Nevertheless, the fluctuation range of RLPSO is significantly weaker than that of CPSO or DE.

Tables 2 and 3 list the best, worst, mean and standard deviation value of f^* for RLPSO, CPSO, SLPSO, PSO, APSO and DE. It can also be seen from the tables that the best value of RLPSO is weaker than that of DE in Rastrigin function, but its mean or standard deviation is better. For the other benchmarks, all performance statistics of RLPSO are optimal, which proves the accuracy, robustness and effectiveness of the RLPSO algorithm.

Conclusions

In this paper, we proposed an RLPSO algorithm to solve the WWTP problem. On the one hand, this method is based on the theory of reinforcement learning. Through continuous interactive attempts of environment and action, the method adjusts the strategy according to the feedback information, and finally judges the optimal concentration setting value under various conditions. On the other hand, this method is based on swarm intelligence algorithm PSO. In the application of WWTP, it is helpful to improve the diversity distribution of solutions to find the global optimal concentration setting value. Besides, the method has an elite network with memory function. In order to improve the treatment effect, it is necessary to record the influence of the set values of S_O and S_{NO} on the sewage parameters, and reuse the information, which provides reference data for the next optimization calculation.

In summary, the RLPSO algorithm proposed in this paper can not only meet the effluent standard, but also reduce the operating cost and provide a feasible solution for the actual sewage treatment plant. In the further, we will continue to study the sewage treatment system, carry out data mining [37, 38], and seek for a better optimal control method. In addition, we will use RLPSO to solve more practical problems, such as robot control [39, 40] and sEMG-based human–machine interaction [41].

Table 2 Optimal fitness values comparison of various algorithms on the ten-dimensional benchmarks

Benchmarks	Algorithm	Best	Worst	Mean	Std
Sphere	RLPSO	**−0.0000**	**−0.0040**	**−0.0001**	**0.0006**
	CPSO	−0.0003	−0.2708	−0.0398	0.0543
	SLSPO	−0.0140	−3.8460	−0.4396	0.7158
	PSO	−0.1999	−5.6188	−1.8527	1.2757
	APSO	−0.0135	−14.0262	−1.5015	2.2759
	DE	−0.0002	−1.2657	−0.1219	0.2522
Ellipsoid	RLPSO	**−0.0045**	**−0.0252e+3**	**−0.0006e+3**	**0.0036e+3**
	CPSO	−4.3230	−0.9799e+3	−0.1375e+3	0.2159e+3
	SLSPO	−13.2826	−3.2691e+3	−0.6086e+3	0.7743e+3
	PSO	−154.4323	−5.2060e+3	−1.9674e+3	1.2540e+3
	APSO	−23.9730	−5.0818e+3	−1.4994e+3	1.5780e+3
	DE	−0.3091	−0.7504e+3	−0.0857e+3	0.1514e+3
Rosenbrock	RLPSO	**−0.0121e+3**	**−0.0079e+3**	**−0.0121e+3**	**0.0016e+4**
	CPSO	−0.2132e+3	−0.2517e+3	−0.2132e+3	0.0394e+4
	SLSPO	−0.5299e+3	−0.2570e+3	−0.5299e+3	0.0693e+4
	PSO	−2.6526e+3	−1.4766e+3	−2.6526e+3	0.3491e+4
	APSO	−9.8681e+3	−5.9917e+3	−9.8681e+3	1.7573e+4
	DE	−0.0943e+3	−0.0847e+3	−0.0943e+3	0.0138e+4
Ackley	RLPSO	**−0.0096**	**−2.6772**	**−0.1786**	**0.5434**
	CPSO	−1.6990	−18.6465	−5.8465	3.7386
	SLSPO	−2.3367	−13.9372	−5.7514	3.0068
	PSO	−4.7522	−17.3965	−11.8340	2.8871
	APSO	−3.1722	−19.8252	−2.2681	4.0848
	DE	−0.1568	−5.0463	−2.2681	1.2706
Griewank	RLPSO	**−0.0260**	**−0.5855**	**−0.2333**	**0.1266**
	CPSO	−0.3212	−1.7691	−0.8502	0.3325
	SLSPO	−0.9384	−19.6293	−3.0762	3.7830
	PSO	−1.4165	−30.2562	−7.4790	5.2791
	APSO	−0.8104	−17.5080	−4.6235	3.9492
	DE	−0.1051	−6.9637	−1.1641	1.4424
Rastrigin	RLPSO	−8.5248	**−27.4713**	**−19.1717**	**4.0504**
	CPSO	−9.1003	−57.8890	−26.4194	11.1090
	SLSPO	−18.5467	−64.8677	−39.2199	10.6655
	PSO	−16.5423	−61.3610	−39.1907	12.0584
	APSO	−10.0329	−80.5182	−39.4284	16.2354
	DE	**−7.8872**	−52.7196	−24.0427	13.3836

The bold symbol denotes the optimal value

Table 3 Optimal fitness values comparison of various algorithms on the twenty-dimensional benchmarks

Benchmarks	Algorithm	Best	Worst	Mean	Std
Sphere	RLPSO	**−0.6717**	**−3.2759**	**−1.8618**	**0.5992**
	CPSO	−1.4531	−13.8093	−5.4161	2.2718
	SLSPO	−2.8312	−20.4800	−8.9089	4.3282
	PSO	−4.6102	−42.2766	−16.8668	7.5637
	APSO	−4.3319	−47.3077	−15.7421	11.3461
	DE	−0.8476	−8.1967	−2.8236	1.4919
Ellipsoid	RLPSO	**−1.0399*e*+3**	**−0.5840*e*+4**	**−0.3061*e*+4**	**0.1251 *e*+4**
	CPSO	−2.1369*e*+3	−2.2601*e*+4	−0.8642*e*+4	0.5060 *e*+4
	SLSPO	−4.1573*e*+3	−3.8145*e*+4	−1.5508*e*+4	0.7894 *e*+4
	PSO	−6.2966*e*+3	−5.4099*e*+4	−2.5631*e*+4	1.1086 *e*+4
	APSO	−5.9527*e*+3	−6.7928*e*+4	−2.3705*e*+4	1.4887 *e*+4
	DE	−1.2265*e*+3	−1.6536*e*+4	−0.4695*e*+4	0.3030 *e*+4
Rosenbrock	RLPSO	**−0.8770*e*+3**	**−0.0990*e*+5**	**−0.0336*e*+5**	**0.1843 *e*+4**
	CPSO	−1.4090*e*+3	−0.6359*e*+5	−0.1701*e*+5	1.3656 *e*+4
	SLSPO	−1.1084*e*+3	−0.7176*e*+5	−0.2156*e*+5	1.8348 *e*+4
	PSO	−4.9263*e*+3	−2.1662*e*+5	−0.5804*e*+5	4.8334 *e*+4
	APSO	−2.9769*e*+3	−3.1000*e*+5	−1.1315*e*+5	7.7027 *e*+4
	DE	−1.0148*e*+3	−0.3077*e*+5	−0.0825*e*+5	0.8021 *e*+4
Ackley	RLPSO	**−7.2032**	**−11.8781**	**−9.8578**	**0.9775**
	CPSO	−9.5876	−18.8654	−14.5778	2.2082
	SLPSO	−7.1612	−15.1980	−11.5509	1.9809
	PSO	−11.5042	−19.1147	−15.5985	1.6954
	APSO	−11.9756	−19.8470	−17.2302	1.5447
	DE	−8.8675	−14.9062	−12.1245	1.5043
Griewank	RLPSO	**−2.7916**	**−14.4050**	**−7.9022**	**2.5382**
	CPSO	−3.0191	−35.9245	−8.5155	5.3316
	SLPSO	−11.1398	−64.0096	−30.3635	14.3900
	PSO	−20.4285	−129.0643	−63.0142	23.9921
	APSO	−12.8475	−183.7946	−73.6182	47.7158
	DE	−3.4753	−35.2718	−11.3498	6.6519
Rastrigin	RLPSO	−63.5822	**−110.1032**	**−88.4879**	**9.9199**
	CPSO	−68.9832	−149.1848	−96.1483	18.6308
	SLPSO	−107.8212	−236.9322	−158.7753	25.0560
	PSO	−77.9249	−161.8871	−121.4302	19.9995
	APSO	−57.7596	−206.7736	−132.1759	29.8269
	DE	**−36.1493**	−146.2063	−92.9517	29.4824

The bold symbol denotes the optimal value

Author contributions Conceptualization, HZ; methodology, HZ and JJ; validation, LL and RD; writing—original draft preparation, LL and HZ; writing, review and editing, JJ and HZ; supervision, HZ and MZ.

Funding This work was supported by National Natural Science Foundation of China (Grant No. 62003306) and Natural Science Foundation of Zhejiang Province (LQ21F030009, LY19F030003).

Availability of data and material All data generated or analyzed during this study are included in this published article.

Code availability If the code is necessary, please email 111003@zust.edu.cn for it.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Wan J, Gu J, Zhao Q et al (2016) COD capture: a feasible option towards energy self-sufficient domestic wastewater treatment[J]. Sci Rep 6(1):1–9
2. Oturan MA, Aaron JJ (2014) Advanced oxidation processes in water/wastewater treatment: principles and applications. A review[J]. Crit Rev Environ Sci Technol 44(23):2577–2641
3. Iratni A, Chang NB (2019) Advances in control technologies for wastewater treatment processes: status, challenges, and perspectives[J]. IEEE/CAA J Autom Sinica 6(2):337–363
4. Skouteris G, Saroj D, Melidis P et al (2015) The effect of activated carbon addition on membrane bioreactor processes for wastewater treatment and reclamation–a critical review[J]. Biores Technol 185:399–410
5. Sadeghassadi M, Macnab CJB, Gopaluni B et al (2018) Application of neural networks for optimal-setpoint design and MPC control in biological wastewater treatment[J]. Comput Chem Eng 115:150–160
6. Hai R, He Y, Wang X et al (2015) Simultaneous removal of nitrogen and phosphorus from swine wastewater in a sequencing batch biofilm reactor[J]. Chin J Chem Eng 23(1):303–308
7. Santín I, Pedret C, Vilanova R et al (2015) Removing violations of the effluent pollution in a wastewater treatment process[J]. Chem Eng J 279:207–219
8. Newhart KB, Holloway RW, Hering AS et al (2019) Data-driven performance analyses of wastewater treatment plants: a review[J]. Water Res 157:498–513
9. Åmand L, Carlsson B (2012) Optimal aeration control in a nitrifying activated sludge process[J]. Water Res 46(7):2101–2110
10. Vrečko D, Hvala N, Kocijan J (2002) Wastewater treatment benchmark: what can be achieved with simple control? [J]. Water Sci Technol 45(4–5):127–134
11. Vrečko D, Hvala N, Stražar M (2011) The application of model predictive control of ammonia nitrogen in an activated sludge process[J]. Water Sci Technol 64(5):1115–1121
12. Mulas M, Tronci S, Corona F et al (2015) Predictive control of an activated sludge process: an application to the Viikinmäki wastewater treatment plant[J]. J Process Control 35:89–100
13. Han H, Wu X, Qiao J (2018) A self-organizing sliding-mode controller for wastewater treatment processes[J]. IEEE Trans Control Syst Technol 27(4):1480–1491
14. Hakanen J, Sahlstedt K, Miettinen K (2013) Wastewater treatment plant design and operation under multiple conflicting objective functions[J]. Environ Model Softw 46:240–249
15. Han G, Qiao JF, Han HG, Chai W (2014) Optimal control for wastewater treatment process based on Hopfield neural network. Control Decis 29(11):2085–2088
16. Yang WJ (2018) Research on optimization of wastewater treatment process based on improved artificial immune algorithm[D]. Lanzhou University of Technology
17. Zhou H, Qiao J (2019) Multi-objective optimal control for wastewater treatment process using adaptive MOEA/D[J]. Appl Intell 49(3):1098–1126
18. Syafiie S, Tadeo F, Martinez E et al (2011) Model-free control based on reinforcement learning for a wastewater treatment problem[J]. Appl Soft Comput 11(1):73–82
19. Silver D, Schrittwieser J, Simonyan K et al (2017) Mastering the game of go without human knowledge[J]. Nature 550(7676):354–359
20. Sun C, Jin Y, Cheng R et al (2017) Surrogate-assisted cooperative swarm optimization of high-dimensional expensive problems[J]. IEEE Trans Evol Comput 21(4):644–660
21. Guo Y, Zhang X, Gong D et al (2020) Novel interactive preference-based multi-objective evolutionary optimization for bolt supporting networks[J]. IEEE Trans Evol Comput 24(4):750–764
22. Jie J, Zhang J, Zheng H et al (2016) Formalized model and analysis of mixed swarm based cooperative particle swarm optimization[J]. Neurocomputing 174:542–552
23. Cheng S, Lu H, Lei X et al (2018) A quarter century of particle swarm optimization[J]. Complex Intell Syst 4(3):227–239
24. Alex J, Benedetti L, Copp J, et al (2018) Benchmark simulation model no. 1 (BSM1)[J]. Report by the IWA Taskgroup on benchmarking of control strategies for WWTPs, 19–20
25. Jeppsson U, Pons MN (2004) The COST benchmark simulation model—current state and future perspective[J]. Control Eng Pract 12(3):299–304
26. Qiao JF, Hou Y, Han HG (2019) Optimal control for wastewater treatment process based on an adaptive multi-objective differential evolution algorithm[J]. Neural Comput Appl 31(7):2537–2550
27. Qiao J, Zhang W (2018) Dynamic multi-objective optimization control for wastewater treatment process[J]. Neural Comput Appl 29(11):1261–1271
28. Han HG, Qian HH, Qiao JF (2014) Nonlinear multiobjective model-predictive control scheme for wastewater treatment process[J]. J Process Control 24(3):47–59
29. Kennedy J, Eberhart R (1995) Particle swarm optimization[C]. In: Proceedings of ICNN'95-International Conference on Neural Networks. IEEE 4: 1942–1948.
30. Jie J, Xu LX (2016) Intelligent particle swarm optimization computing: control methods, collaborative strategies, and optimization applications [M]. Science press
31. Sutton RS, Barto AG (2018) Reinforcement learning: an introduction[M]. MIT press
32. Zheng H, Jie J, Wu XL (2014) Chaotic particle swarm optimisation for fitting magnetic spin parameters of transition metal complexes[J]. Int J Comput Sci Math 5(2):165–173
33. Cheng R, Jin Y (2015) A social learning particle swarm optimization algorithm for scalable optimization[J]. Inf Sci 291:43–60
34. Shi Y, Eberhart RC (1999) Empirical study of particle swarm optimization[C]. In: Proceedings of the 1999 congress on evolutionary computation-CEC99 (Cat. No. 99TH8406). IEEE 3: 1945–1950
35. Qiao JF, Pang ZF, Han HG (2012) Neural network optimal control of wastewater treatment process based on Improved Particle Swarm Optimization Algorithm [J]. CAAI Trans Intell Syst 07(5):429–436 (**in Chinese**)
36. Price KV (2013) Differential evolution[M]//Handbook of optimization. Springer, Berlin, pp 187–214
37. Zhou L, Zheng J, Ge Z et al (2018) Multimode process monitoring based on switching autoregressive dynamic latent variable model[J]. IEEE Trans Industr Electron 65(10):8184–8194
38. Zhou L, Wang Y, Ge Z et al (2018) Multirate factor analysis models for fault detection in multirate processes[J]. IEEE Trans Industr Inf 15(7):4076–4085
39. Cao J, Liang W, Wang Y et al (2019) Control of a soft inchworm robot with environment adaptation[J]. IEEE Trans Industr Electron 67(5):3809–3818
40. Liang W, Cao J, Ren Q et al (2019) Control of dielectric elastomer soft actuators using antagonistic pairs[J]. IEEE Trans Mechatron 24(6):2862–2872
41. Jie J, Liu K, Zheng H et al (2021) High dimensional feature data reduction of multichannel sEMG for gesture recognition based on double phases PSO[J]. Complex & Intelligent Systems

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