IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 15, 2021, accepted June 30, 2021, date of publication July 5, 2021, date of current version July 13, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3094466

Deep Reinforcement Learning-Based Smart Joint Control Scheme for On/Off Pumping Systems in Wastewater Treatment Plants

GIUP SEO[1], (Graduate Student Member, IEEE), SEUNGWOOK YOON[2], (Graduate Student Member, IEEE), MYUNGSUN KIM[1], (Graduate Student Member, IEEE), CHANGHO MUN[3], AND EUISEOK HWANG[1], (Member, IEEE)

[1]Department of Electrical Engineering and Computer Science, Gwangju Institute of Science and Technology (GIST), Gwangju 61005, Republic of Korea
[2]Department of Mechanical Engineering, Gwangju Institute of Science and Technology (GIST), Gwangju 61005, Republic of Korea
[3]Sewage Business Department, Busan Environmental Corporation, Busan 46754, Republic of Korea

Corresponding author: Euiseok Hwang (euiseokh@gist.ac.kr)

This research was supported in part by Energy AI Convergence Research & Development Program through the National IT Industry Promotion Agency of Korea (NIPA) funded by the Ministry of Science and ICT (No. 1711120811) and in part by GIST Research Institute (GRI) Grant funded by the GIST in 2021.

ABSTRACT In this paper, we propose a deep reinforcement learning (DRL) based predictive control scheme for reducing the energy consumption and energy cost of pumping systems in wastewater treatment plants (WWTP), in which the pumps are operated in a binary mode, using on/off signals. As global energy consumption increases, the efficient operation of energy-intensive facilities has also become important. A WWTP in Busan, Republic of Korea is used as the target of this study. This WWTP is a large energy-consuming facility, and the pumping station accounts for a significant portion of the energy consumption of the WWTP. The framework of the proposed scheme consists of a deep neural network (DNN) model for forecasting wastewater inflow and a DRL agent for controlling the on/off signals of the pumping system, where proximal policy optimization (PPO) and deep Q-neural network (DQN) are employed as the DRL agents. To implement smart control with DRL, a reward function is designed to consider the energy consumption amount and electricity price information. In particular, new features and penalty factors for pump switching, which are essential for preventing pump wear, are also considered. The performance of our designed DRL agents is compared with those of WWTP experts and conventional approaches such as scheduling method and model predictive control (MPC), in which integer linear programming (ILP) optimization is employed. Results show that the designed agents outperform the other approaches in terms of compliance with operating rules and reducing energy costs.

INDEX TERMS Predictive control, deep neural network, reinforcement learning, pumping system, cost-effective energy efficiency.

I. INTRODUCTION

As energy demand increases around the world, there have been many efforts to reduce energy consumption and costs, along with efforts to mitigate carbon dioxide emissions from energy production and the consequent impacts of climate change. In particular, the industrial use of energy accounts for about half of global energy consumption, according to the International Energy Agency [1], and many energy-intensive industrial facilities are being researched to increase energy efficiency through smart control. In the case of the water industry, water demand is expected to double by 2035 [2]; therefore, large amounts of additional energy are expected to be consumed for water supply and wastewater treatment, unless the energy efficiency of plants is increased. Wastewater treatment plants (WWTP) have been found to have considerable potential for reducing energy consumption and costs [3]–[5], and several strategies for the energy-efficient operation of WWTPs are being introduced [6]. The main energy-intensive tasks in a WWTP include pumping and

The associate editor coordinating the review of this manuscript and approving it for publication was Shadi Alawneh.

aeration processes. In this paper, the pumping process is targeted, and we investigate the control scheme of the energy-intensive pumping station of a WWTP, which consumes a huge amount of energy through the process as it delivers and purifies the wastewater generated by common households and industries.

Many researchers have proposed methods for reducing energy consumption, cost, or both in the pumping system, including WWTPs and water supply system. These were entirely focused on scheduling the operation of pumps considering long-term flexibility [7]. To efficiently schedule the operation of pumps, the proper combination of pumps should be chosen for each time interval. This requires not only reducing energy consumption and but also observing the operating rules. Baran *et al.* proposed a pump schedule optimization method based on multi-objective evolutionary algorithms for water supply systems with four objectives to be minimized [8]. Energy cost, maintenance cost, maximum peak power and variation in a reservoir level were considered. However, as the number of pumps being used and objectives increase, scheduling the operation of pumps requires a tremendous amount of computing time [9], [10]. To find faster feasible sets of solutions for scheduling pumps, approximation methods have been proposed. Puleo *et al.* and Kim *et al.* simplified the pump scheduling problem as a linear programming problem [10], [11]. Ghaddar *et al.* proposed an approximation approach using Lagrangian decomposition and showed better performance, compared with an approach that used a mixed integer linear programming problem by piecewise-linearization [12]. Fooladivanda and Taylor proposed another approximation method, which transformed a mixed integer non-linear programming problem into a mixed integer second-order cone programming problem, which also takes into account the hydraulic characteristics of variable-speed pumps [13].

Generally, scheduling methods based on solving optimization problems generate plans for efficient operation of targets. This necessarily requires the forecasting of relevant features such as wastewater inflow amount to solve the problem with respect to future operation. For this reason, Cheng *et al.* proposed deep learning-based models to forecast WWTP key features such as influent flow and influent biochemical oxygen demand [14]. However, predictive models can bring uncertainties caused by forecasting errors when scheduling future operations. For variable targets, there may be a huge difference between forecasted and real values, which can lead to an unexpected situation due to improper plans. Therefore, to apply more stable operation to actual plants, online forecasting and scheduling is important to compensate for discrepancies between forecasted and real values. Van Staden *et al.* proposed an online optimization method based on model predictive control (MPC) for binary mode pumping systems [15]. This method means repeatedly solving optimization problems and using the first index of plans, which was more robust to model uncertainty than a scheduling method that solves an optimization problem once. However, this considered only one pump and assumed that the inflow of wastewater was constant, which excluded several conditions for the operation of pumps and a situation in which the inflow and water demand were variable over time.

Recently, as sensors and networked systems increase in plants, it becomes possible to collect large amount of data from the plants. And, this provides opportunities that data-driven scheduling or control (i.e. real-time scheduling) framework can deal with decision-making problems without designing complex models considering high dimensional states. Shiue *et al.* proposed a Q-learning based real-time scheduling approach for a smart factory [16]. The reinforcement learning (RL) module is used to select a proper multiple dispatching rules strategy for manufacturing system, which outperformed heuristic individual dispatching rules. Xia *et al.* proposed a digital twin approach for smart manufacturing, in which a deep Q-neural network (DQN) agent is trained in virtual systems to establish an optimal policy and can drive decision makings for operation in a real-world system [17]. Huang *et al.* proposed an RL-based demand response (DR) scheme for steel power manufacturing, where actor-critic-based deep reinforcement learning (DRL) is utilized for efficient scheduling [18]. The agent reduced energy costs of the manufacturing process with efficient manufacturing schedule through DR. RL-based scheduling frameworks have been applied to not only manufacturing processes but also various plants, such as vinyl acetate monomer plant, circulating fluidized bed plant, coal-fired power plant, nuclear power plant and WWTP [19]–[24]. In particular, Filipe *et al.* proposed a RL-based control framework for variable-frequency pumps in a WWTP [7]. The framework consists of a predictive model and a DRL method. It requires only data for training, without any mathematical model of the pumping system. The model is used with gradient boosting trees (GBT) for forecasting the wastewater inflow, and then the inflow forecast is contained in the state of the DRL. Proximal policy optimization (PPO) is utilized as the DRL agent, which is one of the policy gradient methods of DRLs [25].

State-of-the-art pumping systems can be composed of on/off pumps (i.e. fixed-speed pumps) or variable-frequency pumps (i.e. variable-speed pumps), and the existing DRL-based data-driven control approach was proposed only for variable-frequency pumps [7]. This cannot be directly applied to on/off pumping systems because there are several different constraints (e.g. turning on/off pumps properly without being damaged, selecting a efficient pump combination), which requires different state information and reward design for the DRL-based framework to properly operate. In particular, limiting the number of turning on/off pumps is necessary to prevent the pumps from being damaged [8], [26], [27]. To that end, new state features and a reward function should be designed. Even though the DRL-based control approach for variable-frequency pumps showed better performance than that of WWTP experts, this has not been compared to the previously proposed approaches solving optimization problems, such as scheduling and model predictive control (MPC)

approaches based on linear programming (LP). In addition, in [7], the reward function was designed to reduce only energy consumption. However, such reduction of energy consumption does not always lead to energy cost savings owing to variations of electricity prices [11]. Thus, Time of Use (ToU) tariff has to be considered to ensure that energy cost can be reduced by decreasing the energy consumption at peak times of ToU tariff. Scheduling based on ToU can also contribute to alleviating energy peak loads [28]. Thus, designing a smart DRL-based control framework for on/off pumps is required. And then, performance differences between the DRL-based control approach and LP-based approaches should be compared.

In this paper, we propose a DRL-based control scheme for binary mode pumping systems in WWTPs. The energy consumption amount, energy cost, and number of turning on/off pumps are jointly considered in a reward function. To this end, new features for limiting the number of switching pumps are designed and electricity price information of a ToU tariff is exploited as an element of state. DQN and PPO are utilized as DRL agents for controlling on/off pumps. To identify the performance differences between the proposed scheme and some existing methods, we compare the proposed DRL-based control method with a scheduling method that solves an optimization problem with integer linear programming (ILP) and a control method such as MPC that repeatedly solves the optimization problems for each time interval. The same predictive model is used to generate wastewater inflow forecasts for operation of pumps. As a result, we show that the proposed method for on/off pumps properly works and outperforms the ILP methods and WWTP experts. The contributions of this paper are summarized as follows:

- A DRL-based pump control scheme is designed for pumping systems that are operated in a binary mode.
- New features and a reward function are designed to take into account the constraints of on/off pumping system, such as the number of turning on/off pumps and selection of a proper pump combination.
- The performance of the proposed control scheme is contrasted against WWTP experts and the approaches based on ILP.

The remainder of this paper is structured as follows: Section II describes the pumping station to be targeted and the general framework of reinforcement learning. Section III introduces the proposed scheme and benchmark schemes. Section IV discusses the performance comparison. Section V concludes the paper.

II. BACKGROUND

A. WASTEWATER TREATMENT PLANT

A WWTP encompasses various treatment phases, including pretreatment and primary treatment which are physical and biological treatments, respectively [29]. In this study, a WWTP in Busan, South Korea is selected as the target to be efficiently managed. In the WWTP, the pumping system that we cover is between pretreatment and primary treatment

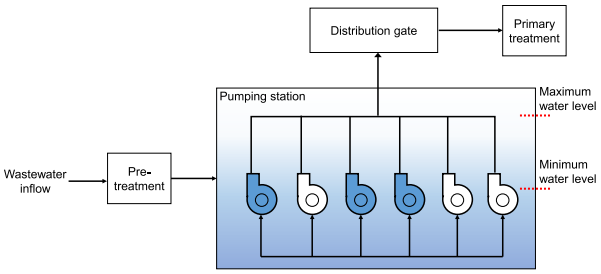


FIGURE 1. Schematic diagram of WWTP processes.

TABLE 1. Specifications of the pumping system.

Pump index	Pump speed (m^3/h)	Pump power (kW)	Season in which the corresponding pump is used
1	3767.6	200	Spring, Autumn, Winter
2	3638.1	200	Summer
3	3920.5	160	Spring, Summer, Autumn, Winter
4	4506.4	200	Spring, Summer, Autumn, Winter
5	5364.5	300	Autumn

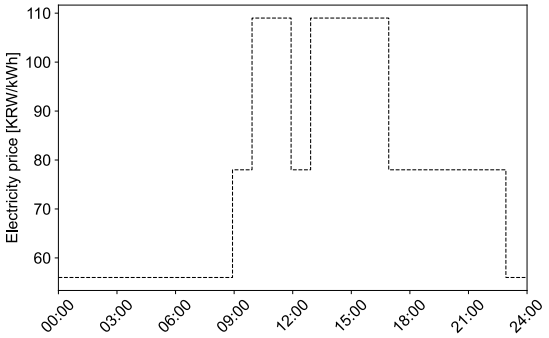


FIGURE 2. The ToU tariff profile.

phases; it moves wastewater from the pumping station to the distribution gate. Fig. 1 shows that it consists of six pumps (one for backup), which are binary mode fixed-speed pumps. Among these pumps, the available ones are changed according to the season. This is described in Table 1, which contains the detailed specifications of the pumps. The pump operation is usually controlled by WWTP experts under the condition that the water level of the pumping station should be kept between the specified minimum and maximum of water levels. In the case of management by WWTP experts, it is stated in [3] that the energy use in WWTPs is generally not being optimally managed, which implies some potential to improve efficiency by reducing redundant energy consumption. In addition, a control strategy with ToU tariffs can be useful for reducing electricity costs. Fig. 2 shows the ToU tariff profile that applies to the target WWTP. The Korea Electric Power Corporation (KEPCO), a South Korean power provider, supplies power to the WWTP based on this profile [30].

B. DATASET

We deal with the data observed in the period from Nov. 2018 to Nov. 2019, which contains the actual operational history of the WWTP experts. The time interval of the data is

five minutes. The inflow rate is not measured in the WWTP of the target due to structural limitations, but it can be estimated through the mass-balance equation [31], [32].

$$I_t = \{O_t \cdot \Delta t + (L_{t+1} - L_t) \cdot A\}/\Delta t \tag{1}$$

The values of $O_t, L_{t+1}, L_t, A, \Delta t$ are all known from the data, where O_t is the outflow rate of the pumps, L_t is the water level of the pumping station, A is the area of the pumping station, and Δt is used to convert the units of wastewater inflow rate to the volume of wastewater inflow.

During the given period, abnormal measurements such as missing values were interpolated by averaging the two nearest points or deleted if the length of consecutive missing values was larger than two. The dataset was divided into training sets (270 days) and test sets (57 days), of which the test set contains the on/off pattern of pump control by the WWTP experts. The training set was utilized to train a predictive model for forecasting future inflow rates and DRL agents for controlling the pumps of the target. The test set was used for a performance comparison between the WWTP experts, DRL agents, and ILP-based approaches.

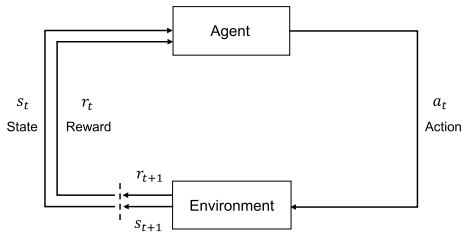


FIGURE 3. General framework of reinforcement learning.

C. REINFORCEMENT LEARNING

A decision making process based on reinforcement learning is generally formalized in the Markov decision process (MDP) framework, as shown in Fig. 3. The framework comprises an agent and its environment, and interactions occur between them through three signals (action, state, reward) [33]. In a nutshell, the MDP is described as a 4-tuple (S, A, P, R). The agent chooses an action $a_t \in A$ from an observed state $s_t \in S$ and then the environment determines the reward $r_{t+1} \sim R(s_t, a_t)$ and the next state $s_{t+1} \sim P(s_t, a_t)$ [34]. The agent continuously learns how to make an optimal decision (action) at each state through the interactions with the environment to achieve the maximum return, maximum cumulative reward, while taking into account immediate and future rewards. A simple expected return, g_t, can be defined as follows:

$$g_t = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \gamma^3 r_{t+4}, \ldots, +\gamma^T r_{t+T} \tag{2}$$

where T denotes a final time step and γ is a discount factor, which is a value between 0 and 1, to reflect the present value of future rewards.

III. METHODOLOGIES

A. PROPOSED SCHEME

Fig. 4 shows the framework of the proposed control scheme. This framework was inspired by [7], [20], [35]. In [7], the authors added a predictive model to the general MDP framework to take into account wastewater inflow forecasts as features of state, which used GBT as the predictive model and PPO as the agent for controlling the pumping system. Similarly, in [20], [35], the frameworks are composed of an artificial neural network (predictive model) and Q-learning (agent). In this study, our framework consists of DNN (predictive model), and PPO or DQN (agent), which cover the continuous state space with more features (inflow forecast, electricity price, and pump usage time). In particular, we modified the structure of the PPO used in [7] to apply to discrete setting (selecting a pump combination), therefore softmax function was utilized for the policy of PPO instead of Gaussian or Beta distribution [36]. We denote the modified PPO as discrete setting PPO (DPPO). And, to distinguish the our designed DRL agents, we denote the designed DPPO as A-DPPO and the designed DQN as A-DQN. In this framework, the predictive model generates the wastewater inflow forecasts, and electricity price is generated from the ToU information from a power provider in South Korea. In particular, the WWTP counts the pump usage time and uses it to limit the frequency of switching pumps. The features serve as important elements of the state when making a decision for efficient pump control.

1) PREDICTIVE MODEL

With online updating of the policy for controlling the pumping system, online updating of the predictive model should be also considered. This is important because the inflow pattern of wastewater is variable over time and by season. By exploiting DNNs, we can apply online updates to the predictive model. We use the features (inflow rate, date) that were used by [7] to forecast the future inflow rate.

$$\hat{I}_{t+1}, \hat{I}_{t+2}, \ldots, \hat{I}_{t+N} = f(I_t, I_{t-1}, \ldots, I_{t-n}, month, day, hour) \tag{3}$$

In Eq. (3), n is the number of lags and N is the number of forecasted inflows from the current time. The metrics for evaluating the predictive model are the mean absolute percentage error (MAPE) and the mean absolute error (MAE).

$$MAPE = \frac{100}{N} \sum_{t=1}^{N} \frac{|\hat{I}_t - I_t|}{|I_t|} \tag{4}$$

$$MAE = \frac{1}{N} \sum_{t=1}^{N} |\hat{I}_t - I_t| \tag{5}$$

2) DEEP REINFORCEMENT LEARNING AGENTS

The existing DRL methods include DQN, PPO, advantage actor critic (A2C), and deep deterministic policy gradient (DDPG) which are commonly used in many research fields

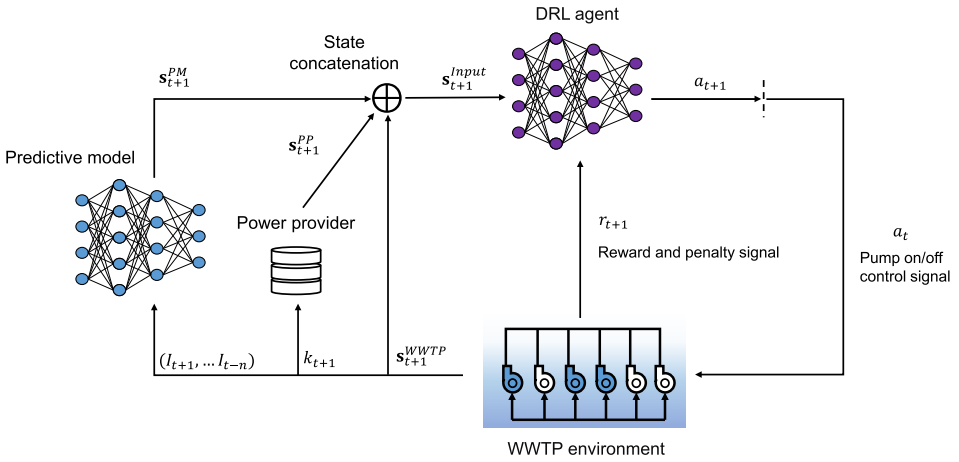


FIGURE 4. The framework of the proposed scheme for the pumping system.

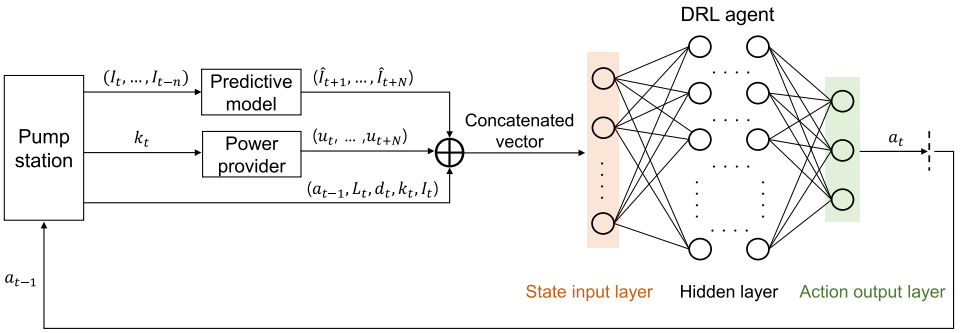


FIGURE 5. Decision making process for controlling the pumping system.

as model-free methods that do not require any mathematical modeling but do require thousands of interactions for training [25], [37]–[39]. PPO outperformed almost all other methods for continuous control and was competitive with value-based methods in discrete settings [40]. Among the DRL methods, we used DPPO (i.e. discrete setting PPO) and DQN as the agent to interact with the environment (we denoted the designed DPPO as A-DPPO and the designed DQN as A-DQN). Fig. 5 illustrates the detailed process of making a decision and reflecting the changed state information in the data flow. At a given time, the DRL agent performs an action for controlling the pumping system and the WWTP checks the changed state compared with the previous one. Some elements of the state vector from the WWTP are used to make electricity price and wastewater inflow forecasts. A concatenated state vector that contains information about the WWTP, electricity price, and wastewater inflow forecasts is used as an input vector to the state input layer of the DRL. As interactions between the environment and the agent increase, the neural networks of the DRL are updated, which provides an optimal control policy for the pumps.

Action: During the test days in spring, three pumps were operated. Therefore, we constructed an action space as in Eq. (6) for controlling the pumping system. All of the pump combinations can also be considered as in [41] without considering season but we used the action space to compare the performance of agents with that of WWTP experts under the condition in which the same pumps are operated. To simplify the process of switching the pumps, it is assumed that each pump was turned on or off in order of efficiency. The system efficiency is important to set optimal pump combinations, which can be found by identifying the outflows of the pumps [42]. The variable a_t represents the number of pumps being used. We excluded the uncommon case in which all of the pumps are turned off.

$$a_t \in \{1, 2, 3\} \tag{6}$$

State: In response to an action from the agent, the environment provides an observation vector, s_t^{WWTP}, which includes the previous action a_{t-1}, water level L_t, pump use duration d_t, current time k_t, and current inflow rate I_t. Then, vectors, s_t^{PM},s_t^{PP}, which contain some external features such as inflow forecasts ,$\hat{I}_{t+1}, \hat{I}_{t+2}, \hat{I}_{t+3}, \ldots, \hat{I}_{t+N}$ and electricity price $u_t, u_{t+1}, u_{t+2}, \ldots, u_{t+N}$ are integrated with the observation vector. Finally, the concatenated vector, s_t^{input}, is given as the state vector to the agent.

$$\mathbf{s}_t^{WWTP} = (a_{t-1}, L_t, d_t, k_t, I_t) \tag{7}$$

$$\mathbf{s}_t^{PM} = (\hat{I}_{t+1}, \hat{I}_{t+2}, \hat{I}_{t+3}, \ldots, \hat{I}_{t+N}) \tag{8}$$

$$\mathbf{s}_t^{PP} = (u_t, u_{t+1}, u_{t+2}, u_{t+3}, \ldots, u_{t+N}) \tag{9}$$

$$\mathbf{s}_t^{input} = (a_{t-1}, L_t, d_t, k_t, I_t, \hat{I}_{t+1}, \hat{I}_{t+2}, \hat{I}_{t+3}, \ldots, \hat{I}_{t+N}, u_t, u_{t+1}, u_{t+2}, u_{t+3}, \ldots u_{t+N}) \quad (10)$$

Reward: Generally, reward functions are designed by trial and error [33]. To construct reward signals that are closer to the real costs of pump operation, we use Korean Won (KRW) units when each reward occurs. First, we consider r_{t+1}^{rule} which occurs according to the water level change by an action of the agent.

$$r_{t+1}^{rule} = \begin{cases} 0, & L_{min} \le L_t \le L_{max} \\ -P\delta, & \text{otherwise} \end{cases} \quad (11)$$

If the water level is within the given operating range, this value is zero. However, if it deviates from the given range, the penalty cost is added, which takes into account damage to the pumping system by pump purchase price, P and penalty coefficient, δ. In [27], when the operating rules are violated, the penalty factor included pump and water cost.

The pump switching is closely related to the age of the pumps. Therefore, if switching is not limited during pump operation, controlling the pumps could be abnormal or even fatal to the system. In addition, the number of switches is used as a factor to calculate the maintenance cost [8], [26], [27]. In [15], when scheduling the operation of a pump, the method includes a constraint in which the frequency of switches is limited to four times per hour. In the reinforcement learning framework, there is no specific method for establishing the constraints for satisfying the given conditions [43]. To solve this problem, we generate new features that indicate the duration of pump use and design the corresponding penalty for frequent pump switching. d_t represents the duration for which the current pumps are used. $d_{len.}$ is the preferred duration.

$$1 \le d_t \le d_{len.} \quad (12)$$

In a nutshell, if a pump switch occurs before the required duration length is reached, the penalty is applied to avoid wearing out the pumps. r_{t+1}^{switch} depends on the duration of pump operation and the amount of energy changed by switching the pumps, where e_t represents the energy consumption at the current time, taking into account pump's power times the time interval. If d_t is equal to $d_{len.}$, r_{t+1}^{switch} is zero because the switching frequency causes no damage to the pumping system. The value of d_t can be between 1 and $d_{len.}$.

$$r_{t+1}^{switch} = -\left(1 - \frac{d_t}{d_{len.}}\right)|e_t - e_{t-1}|u_t \quad (13)$$

Lastly, we design r_{t+1}^{cost}, which contains the information of how to reduce energy consumption and cost. In Eq. (15), c is the electricity price that corresponds to the ToU tariff provided by KEPCO. The tariff requires a different price depending on the profile, as shown in Fig. 2. Moreover, we add new parameters λ_1 and λ_2 to the ToU tariff to regulate the impact of the ToU information on the control of the pumps; this can be used to reduce the power usage during times when the load is heavy.

$$\lambda_2 = \frac{1}{\lambda_1} \quad (0 < \lambda_1 \le 1) \quad (14)$$

$$u_t \in \{c_{lowest}\lambda_1, c_{middle}, c_{highest}\lambda_2\} \quad (15)$$

When continuing the training for constructing a policy of controlling the pump system, it is necessary to give positive reward to the agent to keep controlling the pump system for the maximum return. Therefore, the value of r_{t+1}^{cost} is set to the amount of reduced the energy consumption and cost, compared with the maximum consumption, e_{max}, which was captured.

$$r_{t+1}^{cost} = u_t(e_{max} - e_t) \quad (16)$$

We consider all the conditions by summing all of the reward factors, which can be coordinated by properly setting w_1, w_2, and w_3 for each purpose. As a result, the agent learns the optimal policy for maximizing the return from the reward signals.

$$r_{t+1} = w_1 r_{t+1}^{rule} + w_2 r_{t+1}^{switch} + w_3 r_{t+1}^{cost} \quad (17)$$

B. SCHEDULING APPROACH FOR BENCHMARK

Scheduling approaches are designed to compare the performance of the proposed method with that of an optimization model, such as a ILP model. The objective function and the optimization problem are defined as follows:

$$\phi(x_t, y_t, z_t) = \sum_{t=1}^{T} c_t x_t E^X + c_t y_t E^Y + c_t z_t E^Z \quad (18)$$

$$\min_{x_t, y_t, z_t} \phi(x_t, y_t, z_t) \quad (19)$$

The decision variables are defined as x_t, y_t, and z_t which are set according to the number of pumps being activated. These variables reflect the action space of the proposed method for a fair comparison. x_t represents a decision of activating a pump, y_t represents a decision of activating two pumps, and z_t represents a decision of activating three pumps, which are all binary, x_t, y_t, and $z_t \in \{0, 1\}$. E denotes the energy consumption of each pump combination and c_t is the electricity price according to the ToU tariff. The constraints of the optimization problem are as follows:

$$x_t + y_t + z_t = 1 \quad (20)$$

$$L_{min} \le L_0 + \sum_{k=1}^{t} I_k - (x_k O_k^X + y_k O_k^Y + z_k O_k^Z) \le L_{max} \quad (21)$$

In Eq. (21), L_0 denotes the initial water level at the pumping station. The given operating range is between L_{max} and L_{min}. Also, I_k denotes the inflow of wastewater. The outflow of each pump combination is denoted as O_k^X, O_k^Y, and O_k^Z.

To take into account the pump switching, we designed the features and penalty factors to the proposed control method. In the ILP approach, the resolution of the data is downsampled to limit the switching count while scheduling the optimal plan for activating the pumps.

The scheduling method generates a pattern of pump use for a day. There is no feedback from the target, even though there are discrepancies between predictive and actual values. The pump use pattern is generated in two ways. First, to find the potential maximum gain from optimizing the target, a pattern of pump use is scheduled under the assumption that the scheduler already has information about future wastewater inflows. Then, another pump use pattern is scheduled with no information about those inflows, where predictive model is only used with historical information.

C. CONTROL APPROACH FOR BENCHMARK

In this section, an MPC model is designed to compensate for the discrepancies between the predictive values and actual values. Control approaches that utilize MPC have already been exploited in industrial applications [15], [44]–[47]. The MPC model repeatedly solves optimization problems each time interval to generate plans for the operation of the target, in which the first index elements of the plans are used. Through the process, an online optimization, it can reflect a changed state, such as the water level per time interval, which compensates for the above-mentioned discrepancies. However, when the discrepancy occurs, it can drive the target to a state in which an optimization problem is infeasible because of constraint violations. In this case, simply removing the constraints or re-solving the previous problem could cause an unexpected control behavior [48]. To deal with this challenge, slack variables to soften constraints are added to the optimization problems as a more systematic method [48]–[50]. These variables serve as penalty factors in the cost functions of the optimization problems, which cause the optimizer to find a solution that minimizes the original cost function and simultaneously keeps the number of violations as small as possible [48]. The modified objective function and the optimization problem are defined as

$$\begin{aligned}\phi(x_t, y_t, z_t, \epsilon_t^{upper}, \epsilon_t^{lower}) &= \sum_{t=1}^{T} c_t x_t E^X + c_t y_t E^Y \\ &\quad + c_t z_t E^Z + P\delta(\epsilon_t^{upper} + \epsilon_t^{lower}) \qquad (22)\end{aligned}$$

$$\min_{x_t, y_t, z_t, \epsilon_t^{upper}, \epsilon_t^{lower}} \phi(x_t, y_t, z_t, \epsilon_t^{upper}, \epsilon_t^{lower}) \qquad (23)$$

where P and δ used in the reward function of the proposed scheme are utilized as penalty coefficients. The slack variables ϵ_t^{upper} and ϵ_t^{lower} in the objective function are added to the constraints for the operating rules as follows:

$$0 \le \epsilon_t^{upper} \qquad (24)$$

$$0 \le \epsilon_t^{lower} \le L_{min} \qquad (25)$$

$$\begin{aligned}L_{min} - \epsilon_t^{lower} &\le L_0 + \sum_{k=1}^{t} I_k - (x_k O_k^X + y_k O_k^Y + z_k O_k^Z) \\ &\le L_{max} + \epsilon_t^{upper} \qquad (26)\end{aligned}$$

As a result, at each time interval, the optimizer repeatedly solves the optimization problems for a fixed time window T without any infeasible areas, thereby achieving an online optimization.

IV. RESULTS AND DISCUSSION

A. MODEL TRAINING

1) PREDICTIVE MODEL

Some of the training datasets (202 days) were used for predictive model training, whereas the others (68 days) were used for validation, which both were used as training datasets to establish the policy of the DRL agents. Finally, to test the performance of the predictive model, the test dataset (57 days) is used. Fig. 6 shows the changes in training and validation loss by epoch, and Table 2 summarizes the performance of the model.

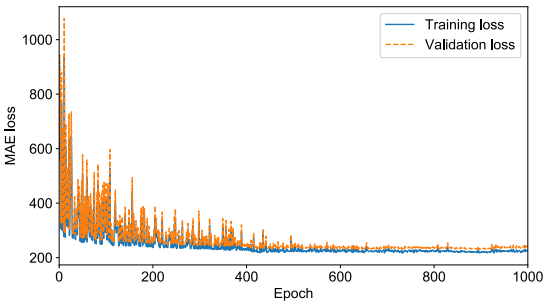


FIGURE 6. The change in training and validation loss.

TABLE 2. Performance of the predictive model based on DNNs.

	Training error	Validation error	Test error
MAE (m^3/h)	215	242	277
MAPE (%)	2.88	3.72	4.03

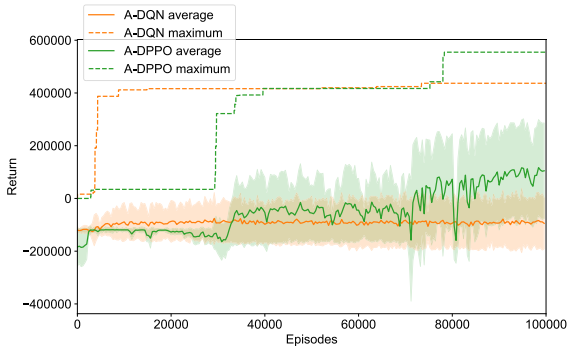


FIGURE 7. Changing return over episodes during training.

2) DRL AGENT MODELS

The agent-environment interactions can be divided into several episodes, which are also called trials. The episodes end in the terminal state where a violation occurs or the final time step of a day comes. Each return denotes the cumulative reward during each episode. Fig. 7 shows the process of learning the policy for controlling the pumping system. The average return is the mean value per 400 episodes. Reinforcement learning has the characteristic of high variance because of

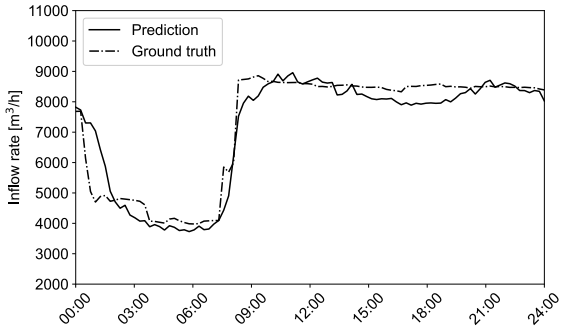


FIGURE 8. The comparison of the predictive inflow rates and actual inflow rates.

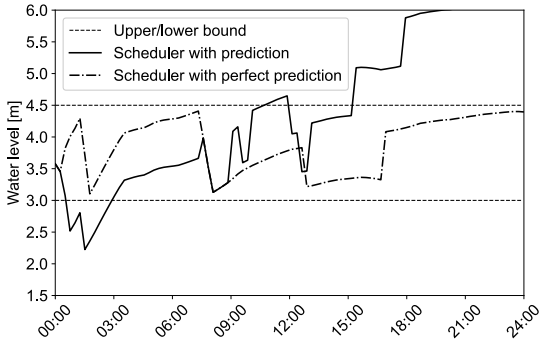


FIGURE 9. The changing water level of the pump station by the ILP schedulers.

stochasticity caused when exploring and making the policy. The wastewater inflow is also highly variable on some days because of different weather conditions, such as rain, dryness, or different seasons. By identifying the increase in the maximum and average returns, we can confirm that the policy is improved. After the agents had learned the policy for controlling the pumping system through the training dataset (270 days), the learned policy was applied to the test dataset (57 days).

B. PERFORMANCE COMPARISON DURING THE TEST DAYS

1) OPERATING RULE ANALYSIS

Fig. 8 illustrates the predictive and actual inflow rates for a test day. The two schedulers generate a pattern of using the pumps based on the predicted and real values, respectively. In Fig. 9, the results show that the scheduler with prediction seriously violates the given operation range. These violations are caused by errors between the predicted and actual inflow rates when generating the plan for pump operation. As the errors accumulate, the severity of the violations also increases. Therefore, to apply this scheduler to a WWTP, it is necessary to compensate for discrepancies between forecasts and ground truths in real time while generating a decision per time interval. Unless the accuracy of prediction is 100%, it would be difficult to use the scheduling method for the efficient operation of targets without compensating for errors from the states of targets in real time.

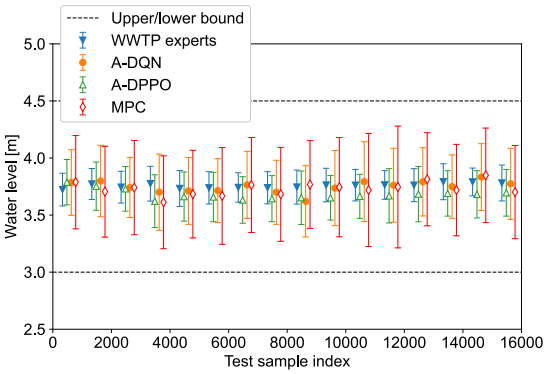


FIGURE 10. Error bars of water level change during the test days.

TABLE 3. Water level comparison between WWTP experts, DRL agents (A-DQN, A-DPPO), MPC, and scheduler).

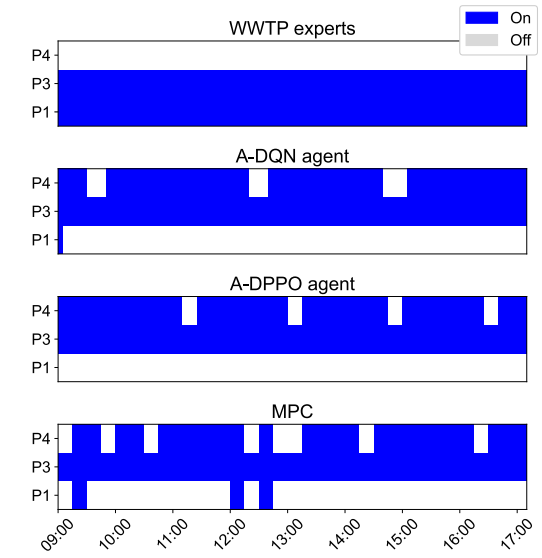
	WWTP experts	A-DQN	A-DPPO	MPC	Scheduler
Water level mean (m)	3.76	3.75	3.68	3.73	2.45
Water level standard deviation (m)	0.14	0.31	0.22	0.43	7.34
Water level violation number	1	1	0	**276**	**4405**

Fig. 10 shows the average and variation of the changing water level during the test days. In Table 3, the details of the figures are identified. They indicate that the DRL agents showed equal or better performance compared with the WWTP experts in terms of the operating rules. On the other hand, the scheduler severely violated the operational rules, which was caused by the accumulated errors of the forecasted and actual inflows. The MPC significantly alleviated violations of the operating rules by taking into account the errors in real time. However, there were still many violations compared with the DRL agents and WWTP experts.

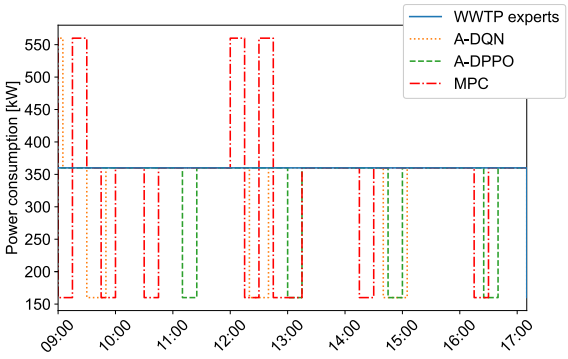
The MPC usually tended to respond to violations only when they had already occurred, which means that it could not prevent the violations beforehand. The MPC made a decision for the operation of pumps without taking into account that the changing water level was approaching the boundaries, unless it deviated from the boundaries. It considered only constraints in forecasted inflow rates without taking into account uncertainties caused by forecasting errors. In contrast with the MPC, the DRL agents could cope with the challenge. Whenever the DRL agents made a decision, they evaluated the cumulative reward at each state, which takes into account rewards from future states. At this time, a discounting factor is considered to apply different weights to future rewards as a function of the time from the current state. If there is a risk of violations in the near future, the agents choose the most stable decision to prevent the violations even though there are chances to reduce energy and cost. Therefore, the MPC caused a higher standard deviation and violation numbers than the DRL agents, as shown in Table 3.

2) ENERGY CONSUMPTION AND COST ANALYSIS

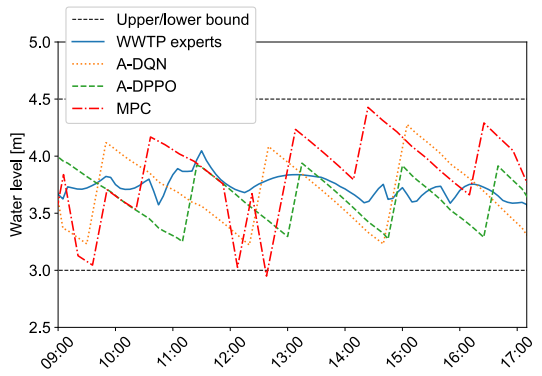
The patterns of pump use, the power consumptions, and water levels are illustrated in Fig. 11. It can be seen that the WWTP expert mainly activated pumps 1 and 3. On the other hand,



(a) Patterns of pump use

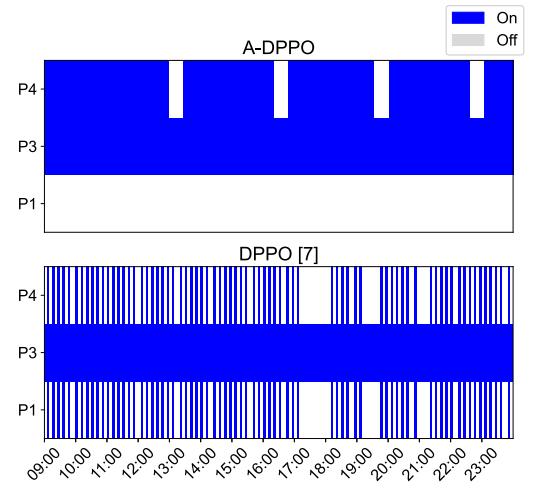


(b) Power consumptions as a function of the patterns

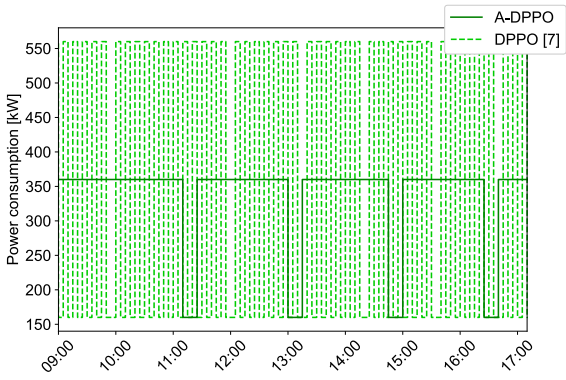


(c) Water levels as a function of the patterns

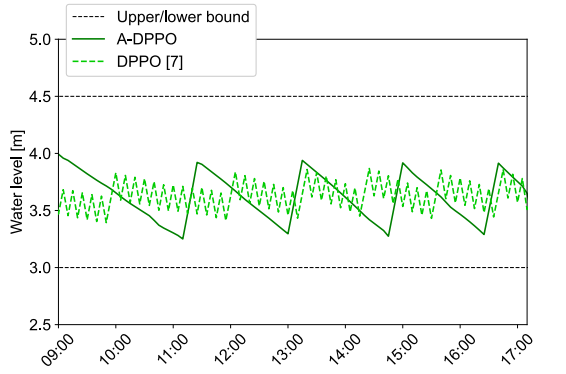
FIGURE 11. Operation results of each control strategy during a test day.



(a) Patterns of pump use



(b) Power consumptions as a function of the patterns



(c) Water levels as a function of the patterns

FIGURE 12. Operation comparison between A-DPPO and DPPO [7].

the A-DQN agent, A-DPPO agent, and MPC typically used pumps 3 and 4, which created an opportunity to better utilize the capacity of the pumping station. In addition, the DRL agents and MPC turned off a pump during the highest price periods to reduce energy consumption and costs. The highest price periods were from 10:00 am to 12:00 pm and from 1:00 pm to 5:00 pm during the test days.

In terms of switching number, all of the approaches showed an increase to better utilize the capacity of the pumping station. However, if switching pumps occurred very frequently it could cause tremendous degradation of pumps. In [15], the possible switching interval was set as 15 minutes (4 times per one hour). Here, we assume the same switching interval. Thus, the maximum allowable switching number is 5472 during the test days (57 days). In the case of the scheduler and MPC, to satisfy this condition, a pattern of pump use over time was generated with 15 minute intervals. Through r_{t+1}^{switch}, d_t, $d_{len.}$, which were proposed in this

TABLE 4. Energy consumption and cost comparison between WWTP experts, DRL agents (A-DQN, A-DPPO), MPC and scheduler.

	WWTP experts	A-DQN	**A-DPPO**	MPC	Scheduler	DPPO [7]
Energy consumption (kWh)	427,586 (-)	412,497 (-3.53%)	411,513 **(-3.76%)**	413,330 (-3.33%)	421,830 (-1.35%)	426,580 (-0.23%)
Energy cost (KRW)	34,190,686 (-)	32,910,493 (-3.74%)	32,844,810 **(-3.94%)**	32,869,537 (-3.86%)	33,534,187 (-1.92%)	34,103,210 (-0.25%)
Switching number	488	931	1040	1873	1450	**10960**
Water level violation number	1	1	**0**	276	4405	0

paper, the DRL agents (A-DPPO and A-DQN) prevented any deviation in maintaining the appropriate switching frequency range. In Fig.12, during a test day, it is identified that a DPPO agent [7] without the proposed features and reward function abnormally changed pump combination, causing highly frequent on/off transition, while the A-DPPO agent with the features and reward function showed a proper pattern. During the test days, the switching number of the DPPO agent was 10960, which deviated seriously from the given switching constraint. The details are described in Table.4.

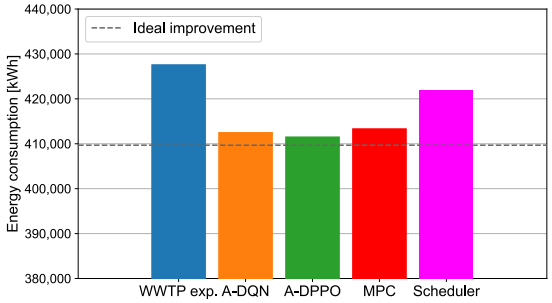


FIGURE 13. Energy consumption comparison.

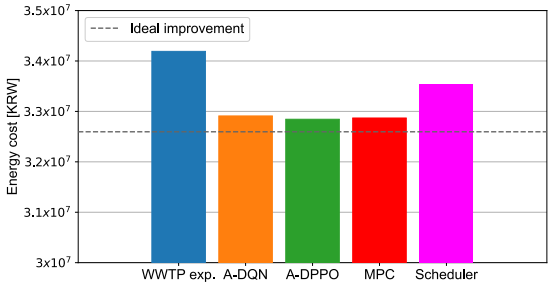


FIGURE 14. Energy cost comparison.

Energy consumption and cost comparison are shown in Fig. 13 and Fig. 14, respectively. The details are presented in Table 4. It can be seen that the scheduler with prediction shows a reduction in energy consumption and costs of up to 1.35% and 1.92% respectively, which indicates severe violations in the operating rule analysis. The MPC shows a reduction in energy consumption and costs of up to 3.33% and 3.86%, respectively. This compensated significantly for the weakness of the scheduler and simultaneously improved its performance. The A-DQN could reduce the energy consumption and costs by up to 3.53% and 3.74%, respectively. The A-DPPO could reduce the energy consumption and costs by up to 3.76% and 3.94%, respectively. In the case of [7], without the features and reward function, the DPPO was stuck in sub-optimal policy, showing insignificant energy and cost reduction up to 0.23% and 0.25%, respectively. The performance of the scheduler with perfect prediction was used as an ideal improvement on reducing the energy consumption and costs. It shows a reduction in energy consumption and costs of up to 4.18% and 4.66%, respectively. The highest gain in optimization of the target could potentially be under the assumption that all future inflow rates are known. The A-DPPO showed the most similar performance to the ideal improvement.

The proposed DRL agents (i.e. A-DPPO and A-DQN) could achieve the increase of operating efficiency without seriously violating the given operating rules or damaging the pumping system, compared with the WWTP experts, scheduler, MPC, and DPPO [7]. The MPC showed almost the same performance in reducing the energy consumption and costs, but it caused severe violations compared with the WWTP experts and DRL agents. In addition, the switching numbers are lower in the proposed DRL agents than in the scheduler and MPC, even though the DRL agents showed the better performance in reducing energy consumption and costs. As a result, it was confirmed that the proposed scheme outperformed the ILP-based approaches in the efficient operation of the target.

V. CONCLUSION AND DISCUSSION

The existing researches on pumping systems focused mainly on scheduling the operation of the pumps. Online optimization approaches such as MPC repeatedly solving optimization problems and data-driven predictive control based on reinforcement learning can compensate for the weakness of scheduling the operation of the pumps. In this study, we designed a deep reinforcement learning (DRL) based predictive control scheme and integer linear programming (ILP) based MPC for binary mode fixed-speed pumps. To this end, a reward function and new features were proposed to limit the frequency of switching pumps. The pumping station of a WWTP in the Republic of Korea was set as the target to be efficiently controlled. During the test days, the result showed that the ILP-based scheduling method severely violated the operating rules and the ILP-based MPC could alleviate significantly the number of violations by compensating for forecasting errors. However, there were still many violations because the MPC could respond to the violations almost after those occurred. On the other hand, the DRL-based control schemes could prevent violations beforehand, which showed

equal or better performance compared with the WWTP experts in terms of the operating rules. In terms of energy consumption and cost, the MPC and DRL based scheme showed similar performance, which outperformed significantly the WWTP experts and scheduling method. As a result, we confirmed that the DRL-based scheme was most suitable for the operation of pumps in uncertainties caused by forecasting errors.

We utilized DRL agents such as PPO and DQN based on model-free algorithms to efficiently control on/off pumps. Model-free algorithms gradually search for an optimal policy through exploration and require a lot of training samples to find a proper policy, compared to model-based algorithms such as ILP-based MPC. To improve the scheme, Model-based DRL agent can be considered regarding the future direction. For some applications, it was identified that model-based DRL agents could learn a control policy with much less data and quickly adapt to unseen situations and sudden changes [51], [52]. We will try to build on a model-based DRL scheme for on/off pumping system and compare it with other schemes such as the model-free-based DRL scheme and ILP-based MPC.

ACKNOWLEDGMENT

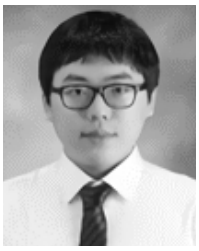
This research was supported in part by Energy AI Convergence Research & Development Program through the National IT Industry Promotion Agency of Korea (NIPA) funded by the Ministry of Science and ICT (No. 1711120811) and in part by GIST Research Institute (GRI) Grant Funded by the GIST in 2021.

REFERENCES

[1] J. Henriques and J. Catarino, "Sustainable value–an energy efficiency indicator in wastewater treatment plants," *J. Cleaner Prod.*, vol. 142, pp. 323–330, 2017.
[2] M. Lavelle and T. K. Grose, "Water demand for energy to double by 2035," Nat. Geographic News, 2013.
[3] L. Castellet and M. Molinos-Senante, "Efficiency assessment of wastewater treatment plants: A data envelopment analysis approach integrating technical, economic, and environmental issues," *J. Environ. Manage.*, vol. 167, pp. 160–166, Feb. 2016.
[4] D. Kirchem, M. Á. Lynch, V. Bertsch, and E. Casey, "Modelling demand response with process models and energy systems models: Potential applications for wastewater treatment within the energy-water nexus," *Appl. Energy*, vol. 260, Feb. 2020, Art. no. 114321.
[5] D. Torregrossa, J. Hansen, F. Hernández-Sancho, A. Cornelissen, G. Schutz, and U. Leopold, "A data-driven methodology to support pump performance analysis and energy efficiency optimization in waste water treatment plants," *Appl. Energy*, vol. 208, pp. 1430–1440, Dec. 2017.
[6] V. Zejda, V. Máša, Š. Václavková, and P. Skryja, "A novel check-list strategy to evaluate the potential of operational improvements in wastewater treatment plants," *Energies*, vol. 13, no. 19, p. 5005, Sep. 2020.
[7] J. Filipe, R. J. Bessa, M. Reis, R. Alves, and P. Póvoa, "Data-driven predictive energy optimization in a wastewater pumping station," *Appl. Energy*, vol. 252, Oct. 2019, Art. no. 113423.
[8] B. Barán, C. von Lücken, and A. Sotelo, "Multi-objective pump scheduling optimisation using evolutionary strategies," *Adv. Eng. Softw.*, vol. 36, no. 1, pp. 39–47, Jan. 2005.
[9] P. W. Jowitt and G. Germanopoulos, "Optimal pump scheduling in water-supply networks," *J. Water Resour. Planning Manage.*, vol. 118, no. 4, pp. 406–422, 1992.
[10] V. Puleo, M. Morley, G. Freni, and D. Savić, "Multi-stage linear programming optimization for pump scheduling," *Procedia Eng.*, vol. 70, pp. 1378–1385, Apr. 2014.
[11] Y. Kim, S. Yoon, C. Mun, T. Kim, D. Kang, M. Sim, D. Choi, and E. Hwang, "Smart day-ahead pump scheduling scheme for electricity cost optimization in a sewage treatment plant," in *Proc. Int. Conf. Inf. Commun. Technol. Converg. (ICTC)*, Oct. 2019, pp. 565–567.
[12] B. Ghaddar, J. Naoum-Sawaya, A. Kishimoto, N. Taheri, and B. Eck, "A Lagrangian decomposition approach for the pump scheduling problem in water networks," *Eur. J. Oper. Res.*, vol. 241, no. 2, pp. 490–501, Mar. 2015.
[13] D. Fooladivanda and J. A. Taylor, "Energy-optimal pump scheduling and water flow," *IEEE Trans. Control Netw. Syst.*, vol. 5, no. 3, pp. 1016–1026, Sep. 2018.
[14] T. Cheng, F. Harrou, F. Kadri, Y. Sun, and T. Leiknes, "Forecasting of wastewater treatment plant key features using deep learning-based models: A case study," *IEEE Access*, vol. 8, pp. 184475–184485, 2020.
[15] A. J. van Staden, J. Zhang, and X. Xia, "A model predictive control strategy for load shifting in a water pumping scheme with maximum demand charges," *Appl. Energy*, vol. 88, no. 12, pp. 4785–4794, Dec. 2011.
[16] Y.-R. Shiue, K.-C. Lee, and C.-T. Su, "Real-time scheduling for a smart factory using a reinforcement learning approach," *Comput. Ind. Eng.*, vol. 125, pp. 604–614, Nov. 2018.
[17] K. Xia, C. Sacco, M. Kirkpatrick, C. Saidy, L. Nguyen, A. Kircaliali, and R. Harik, "A digital twin to train deep reinforcement learning agent for smart manufacturing plants: Environment, interfaces and intelligence," *J. Manuf. Syst.*, vol. 58, pp. 210–230, Jan. 2021.
[18] X. Huang, S. H. Hong, M. Yu, Y. Ding, and J. Jiang, "Demand response management for industrial facilities: A deep reinforcement learning approach," *IEEE Access*, vol. 7, pp. 82194–82205, 2019.
[19] L. Zhu, Y. Cui, G. Takami, H. Kanokogi, and T. Matsubara, "Scalable reinforcement learning for plant-wide control of vinyl acetate monomer process," *Control Eng. Pract.*, vol. 97, Apr. 2020, Art. no. 104331.
[20] X. Xu, Y. Jia, Y. Xu, Z. Xu, S. Chai, and C. S. Lai, "A multi-agent reinforcement learning-based data-driven method for home energy management," *IEEE Trans. Smart Grid*, vol. 11, no. 4, pp. 3201–3211, Jul. 2020.
[21] J. Fu, H. Xiao, H. Wang, and J. Zhou, "Control strategy for denitrification efficiency of coal-fired power plant based on deep reinforcement learning," *IEEE Access*, vol. 8, pp. 65127–65136, 2020.
[22] D. Lee, A. M. Arigi, and J. Kim, "Algorithm for autonomous power-increase operation using deep reinforcement learning and a rule-based system," *IEEE Access*, vol. 8, pp. 196727–196746, 2020.
[23] F. Hernández-del-Olmo, E. Gaudioso, R. Dormido, and N. Duro, "Tackling the start-up of a reinforcement learning agent for the control of wastewater treatment plants," *Knowl.-Based Syst.*, vol. 144, pp. 9–15, Mar. 2018.
[24] F. Hernandez-del-Olmo, E. Gaudioso, and A. Nevado, "Autonomous adaptive and active tuning up of the dissolved oxygen setpoint in a wastewater treatment plant using reinforcement learning," *IEEE Trans. Syst., Man, Cybern., C (Appl. Rev.)*, vol. 42, no. 5, pp. 768–774, Sep. 2012.
[25] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," 2017, *arXiv:1707.06347*. [Online]. Available: http://arxiv.org/abs/1707.06347
[26] K. E. Lansey and K. Awumah, "Optimal pump operations considering pump switches," *J. Water Resour. Planning Manage.*, vol. 120, no. 1, pp. 17–35, Jan. 1994.
[27] D. Torregrossa and F. Capitanescu, "Optimization models to save energy and enlarge the operational life of water pumping systems," *J. Cleaner Prod.*, vol. 213, pp. 89–98, Mar. 2019.
[28] Y. Cao, S. Tang, C. Li, P. Zhang, Y. Tan, Z. Zhang, and J. Li, "An optimized EV charging model considering TOU price and SOC curve," *IEEE Trans. Smart Grid*, vol. 3, no. 1, pp. 388–393, Mar. 2012.
[29] C. K. Yoo, D. S. Kim, J.-H. Cho, S. W. Choi, and I.-B. Lee, "Process system engineering in wastewater treatment process," *Korean J. Chem. Eng.*, vol. 18, no. 4, pp. 408–421, 2001.
[30] *Time Use Tariff Provided by Korea Electric Power Corp. (KEPCO)*. Accessed: Jul. 2019. [Online]. Available: https://cyber.kepco.co.kr/ckepco/front/jsp/CY/E/E/CYEEHP00301.jsp
[31] Z. Zhang, Y. Zeng, and A. Kusiak, "Minimizing pump energy in a wastewater processing plant," *Energy*, vol. 47, no. 1, pp. 505–514, Nov. 2012.
[32] Z. Zhang, A. Kusiak, Y. Zeng, and X. Wei, "Modeling and optimization of a wastewater pumping system with data-mining methods," *Appl. Energy*, vol. 164, pp. 303–311, Feb. 2016.
[33] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 2018.
[34] M. Hausknecht and P. Stone, "Deep recurrent Q-Learning for partially observable MDPs," 2015, *arXiv:1507.06527*. [Online]. Available: http://arxiv.org/abs/1507.06527

[35] R. Lu, S. H. Hong, and M. Yu, "Demand response for home energy management using reinforcement learning and artificial neural network," *IEEE Trans. Smart Grid*, vol. 10, no. 6, pp. 6629–6639, Nov. 2019.
[36] C. Ching-Yun Hsu, C. Mendler-Dünner, and M. Hardt, "Revisiting design choices in proximal policy optimization," 2020, *arXiv:2009.10897*. [Online]. Available: http://arxiv.org/abs/2009.10897
[37] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing atari with deep reinforcement learning," 2013, *arXiv:1312.5602*. [Online]. Available: http://arxiv.org/abs/1312.5602
[38] V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu, "Asynchronous methods for deep reinforcement learning," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 1928–1937.
[39] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," 2015, *arXiv:1509.02971*. [Online]. Available: http://arxiv.org/abs/1509.02971
[40] P. Henderson, R. Islam, P. Bachman, J. Pineau, D. Precup, and D. Meger, "Deep reinforcement learning that matters," 2017, *arXiv:1709.06560*. [Online]. Available: http://arxiv.org/abs/1709.06560
[41] A. Kusiak, Y. Zeng, and Z. Zhang, "Modeling and analysis of pumps in a wastewater treatment plant: A data-mining approach," *Eng. Appl. Artif. Intell.*, vol. 26, no. 7, pp. 1643–1651, Aug. 2013.
[42] V. K. Arun Shankar, S. Umashankar, S. Paramasivam, and N. Hanigovszki, "A comprehensive review on energy efficiency enhancement initiatives in centrifugal pumping system," *Appl. Energy*, vol. 181, pp. 495–513, Nov. 2016.
[43] R. Nian, J. Liu, and B. Huang, "A review on reinforcement learning: Introduction and applications in industrial process control," *Comput. Chem. Eng.*, vol. 139, Aug. 2020, Art. no. 106886.
[44] J. Richalet, A. Rault, J. L. Testud, and J. Papon, "Model predictive heuristic control," *Automatica*, vol. 14, no. 5, pp. 413–428, Sep. 1978.
[45] C. R. Cutler and R. B. Hawkins, "Application of a large predictive multivariable controller to a hydrocracker second stage reactor," in *Proc. Amer. Control Conf.*, Jun. 1988, pp. 284–291.
[46] S. Qin and T. Badgwell, "An overview of industrial model predictive control technology," in *Proc. AIche Symp. Ser.*, vol. 93, no. 316, 1997.
[47] G. M. Zeng, X. S. Qin, L. He, G. H. Huang, H. L. Liu, and Y. P. Lin, "A neural network predictive control system for paper mill wastewater treatment," *Eng. Appl. Artif. Intell.*, vol. 16, no. 2, pp. 121–129, Mar. 2003.
[48] E. C. Kerrigan and J. M. Maciejowski, "Soft constraints and exact penalty functions in model predictive control," in *Proc. United Kingdom Autom. Control Council (UKACC) Int. Conf.*, Sep. 2000.
[49] N. M. de Oliveira and L. T. Biegler, "Constraint handing and stability properties of model-predictive control," *AIChE J.*, vol. 40, no. 7, pp. 1138–1155, 1994.
[50] P. O. M. Scokaert and J. B. Rawlings, "Feasibility issues in linear model predictive control," *AIChE J.*, vol. 45, no. 8, pp. 1649–1659, Aug. 1999.
[51] A. Nagabandi, I. Clavera, S. Liu, R. S. Fearing, P. Abbeel, S. Levine, and C. Finn, "Learning to adapt in dynamic, real-world environments through meta-reinforcement learning," 2018, *arXiv:1803.11347*. [Online]. Available: http://arxiv.org/abs/1803.11347
[52] A. Nagabandi, G. Kahn, R. S. Fearing, and S. Levine, "Neural network dynamics for model-based deep reinforcement learning with model-free fine-tuning," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2018, pp. 7559–7566.

GIUP SEO (Graduate Student Member, IEEE) received the B.S. degree from the School of Mechanical and Control Engineering, Handong Global University, Pohang, Gyeongsangbuk, South Korea, in 2019, and the M.S. degree from the School of Electrical Engineering and Computer Science, Gwangju Institute of Science and Technology, Gwangju, South Korea, in 2021, where he is currently pursuing the Ph.D. degree with the School of Electrical Engineering and Computer Science. His research interests include signal processing, intelligent systems, schedule optimization, and energy informatics.



SEUNGWOOK YOON (Graduate Student Member, IEEE) received the B.S. degree from the Department of Electric Engineering, Kwangwoon University, Seoul, South Korea, in 2014. He is currently pursuing the integrated M.S. and Ph.D. degree with the School of Mechatronics, Gwangju Institute of Science and Technology, Gwangju, South Korea. His research interests include energy informatics, vehicle grid integration, and data channel array signal processing.

MYUNGSUN KIM (Graduate Student Member, IEEE) received the B.S. degree from the School of Mechanical Engineering, Gwangju Institute of Science and Technology, Gwangju, South Korea, in 2019, where she is currently pursuing the M.S. degree with the School of Electrical Engineering and Computer Science. Her research interests include energy informatics, signal processing, and data mining.

CHANGHO MUN received the B.S. degree from the Department of History, Korea University, Seoul, South Korea, in 2008, and the M.S. degree from the Department of Electrical Engineering, Pukyong National University, Busan, South Korea, in 2020.

He joined Busan Environmental Corporation, in 2012. He is currently managing electrical equipment for energy-intensive processes in the wastewater treatment plant.

EUISEOK HWANG (Member, IEEE) received the B.S. and M.S. degrees from the School of Engineering, Seoul National University, Seoul, South Korea, in 1998 and 2000, respectively, and the M.S. and Ph.D. degrees in electrical and computer engineering from Carnegie Mellon University, Pittsburgh, PA, USA, in 2010 and 2011, respectively.

He was with the Digital Media Research Center, Daewoo Electronics Company Ltd., South Korea, from 2000 to 2006, and the Channel Architecture Group, LSI Corporation (currently, Broadcom), San Jose, CA, USA, from 2011 to 2014. Since 2015, he has been an Assistant/Associate Professor with the School of Mechatronics/Electrical Engineering and Computer Science/Artificial Intelligence, Gwangju Institute of Science and Technology (GIST), South Korea. His research interests include data channel signal processing and coding, energy informatics and intelligence implementations for smart grid, and information processing for system intelligence in emerging ICT/IoT applications.