# XAI-ASSISTED MULTI-AGENT DEEP REINFORCEMENT LEARNING FOR A GUARANTEED AUTONOMOUS CONTROL SYSTEM OF SEQUENCING BATCH REACTOR FOR SUSTAINABLE WASTEWATER MANAGEMENT

SungKu Heo, TaeYong Woo, SangYoun Kim, and ChangKyoo Yoo*
Integrated Engineering Major, Dept. of Environmental Science and Engineering,
College of Engineering, , Kyung Hee University, Seocheon-dong 1, Giheung-gu,
Yongin-Si, Gyeonggi-Do 446-701, Republic of Korea

*Abstract*

Wastewater treatment plants (WWTP) were invented to treat wastewater pollutants from various industrial sector; in nowadays, it is also required to deal with sustainable efficiency considering energy consumption and environmental benefits under climate change. In this context, autonomous control system for sequencing batch reactor (SBR), which is one of the advanced WWTP, was proposed using multi-agent reinforcement learning (MARL). To train the MARL agents, the SBR was modeled based on the activated sludge model No. 1 (ASM1); then various dataset of influent characteristic was generated. A game abstraction method based on a two-stage attention network (G2ANET) algorithm was employed to search the setpoints of aeration and extra carbon (EC) injection for SBR operation. To explain control performance of G2ANET, layer-wise relevance propagation (LRP) which is one of the explainable AI (XAI) methods was adopted. The result indicated that the G2ANET agents control the SBR to reduce 14.6% of energy consumptions, while maintaining effluent quality criteria. Furthermore, it was guaranteed that the G2ANET agents can recognize the mechanism in SBR operation without human intervention by LRP. Hence the XAI-assisted MARL approach to control SBR system can be applicated to real WWTP with guaranteed control performance considering sustainable efficiency.

*Keywords*

Explainable AI, Multi-agent deep reinforcement learning, Sequencing batch reactor, wastewater, Autonomous control

## Introduction

Environmental process systems are devised to prevent and remove the pollution from industrial, commercial, domestic, and agricultural sectors by human behaviors. Among the environmental process systems, a wastewater treatment system is a remarkable invention from the 19th century to solve an anthropological crisis that alleviates sanitation and protects human health from waterborne epidemic disease. In the 21st century, global water challenges have shifted and intensified from water quality to water affordability, scarcity, and resilience to climate change and the increasing demand for a growing human population and urbanization. Under climate change, wastewater treatment plant (WWTP) has been highly required to be operated with sustainable efficiency in energy consumption and environmental resilience.

* To whom all correspondence should be addressed

In this context, wastewater utilities require innovative solutions to tackle these issues by adopting novel digital technologies including big data, artificial intelligence (AI), and machine learning (ML). Especially, reinforcement learning related to optimal control is highlighted machine learning algorithm which is feasible to solve the task without human intervention.

Numerous investigations have already been studied to control the WWTP using the RL. Nam et al. (2019) applied deep Q-network to the membrane bio-reactor process and it can reduce 34% of aeration energy consumption. Chen et al. (2022) developed dynamic control strategy of conventional WWTP by utilizing a multi-agent deep deterministic policy gradient. This novel strategy can guide the operation of WWTP considering energy consumption, eutrophication potential, and greenhouse gas emission (GHG) simultaneously. The previous researches showed the RL can be applicated to WWTP for sustainable operation; furthermore, the feasibility of autonomous operation is accomplished. However, the previous researches cannot represent how to solve the limitation of RL which do not guarantee the reliable operation of RL agent due to the black-box problem in training a policy of RL agent. Hence, to accomplish the sustainable and autonomous operation in wastewater sector, guaranteed RL algorithm should be developed.

To tackle those issues, we suggested multi-agent RL (MARL) assisted by explainable AI (XAI) and applicated to the sequencing batch reactor (SBR) process which is one of the novel WWTP with complex operational tasks. This research represented that XAI-assisted MARL can autonomously operate the SBR system by considering sustainable objectives including energy efficiency and effluent quality.

## Materials and methods

### *Proposed method*

Figure 1 depicted a proposed research framework to develop the guaranteed autonomous control for SBR process based on MARL and XAI algorithms.

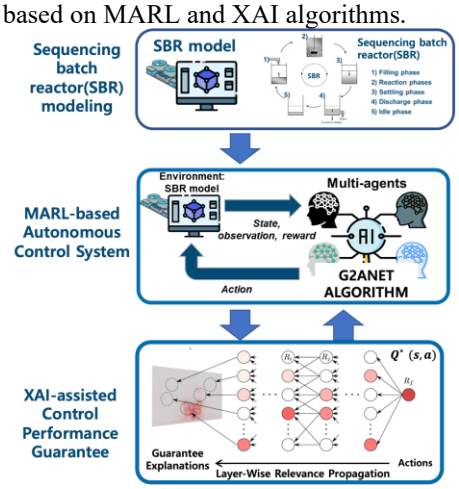


*Figure 1. Research framework for developing the guaranteed autonomous control of SBR process based on XAI-assisted MARL*

### *Sequencing batch reactor (SBR) modeling*

A mathematical model for SBR system was developed to evaluate the autonomous control strategy. This SBR model modified the rules established in the COST 624 benchmark to transform a continuous WWT process into a batch operation system with a buffer tank (Pons et al., (2004)). Activated sludge model No. 1 was adopted to model the general biological removal of wastewater pollutants in terms of bacteria (i.e., hetero- and autotrophic bacteria). The core mechanisms of the process kinetic rate and stoichiometric matrix related to the group behavior of the microbial population are *represented* in *Figure 2*.

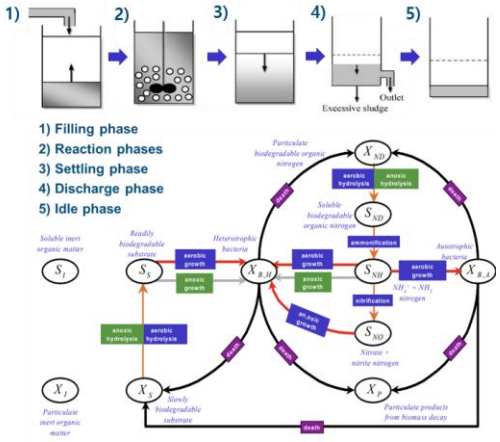


*Figure 2. Operational phases and mechanism of activated sludge model No.1 (ASM1) in SBR system (modified from Henze et al.(2000))*

The operation of SBR process consists of sequential phases as filling, reaction, settling, discharge, and idle. The mass balance of filling, non-filling, and discharge phases follows the equation (1) to (3).

$$Q_{in} + r(Z)V = Q_{out}Z + \frac{d(VZ)}{dt} \quad (1)$$

$$\frac{d(Z)}{dt} = 0, \; dV/dt = -Q_{out} \quad (2)$$

$$K_L a \cdot V \cdot (S_O^* - S_O) + r(S_O)V = \frac{d(V \cdot S_O)}{dt} \quad (3)$$

where, $Q$ is volume flowrate in SBR, $V$ is volume, $r$ is reaction processes, $K_L a$ is oxygen transportation coefficient, and Z is state variables. Eq.(4) indicate the mass balance of dissolved oxygen (DO) named as $S_o$ in filling and non-filling phases.

By those equation, this SBR model can deal with the operation phases. The Table 1 indicated the assigned phases of the proposed SBR model with time length of total operation. It is assumed that complete mixing effect was

considered in filling and reaction phases; and both phases are under aerobic and anaerobic conditions.

*Table 1. Information of operation phases in SBR*

| length(%) | Phase | Feeding | Aeration | Mixing | Discharge |
|---|---|---|---|---|---|
| 1 (4.2) | Filling | Yes | No | Yes | No |
| 2 (8.3) | Reaction | No | No | Yes | No |
| 3 (37.5) | Reaction | No | Yes | Yes | No |
| 4 (31.2) | Reaction | No | No | Yes | No |
| 5 (2.1) | Reaction | No | Yes | Yes | No |
| 6 (8.3) | Settling | No | No | No | No |
| 7 (2.1) | Discharge | No | No | No | Yes |
| 8 (6.3) | Idle | No | Yes | No | No |

Carbon and nitrogen are the two major characteristics of wastewater influent. Their variation in WWTP plant operations give rise to a critical impact on the removal performance of the WWTP system (Henze et al., 2000). The variability of carbon and nitrogen are usually measured as ammonia ($NH_4$) and chemical oxygen demand (COD). Accordingly, the autonomous control of the MARL should suggest the optimized operation performance under various $NH_4$ and COD circumstances. To consider the varying influent condition in developing MARL-based autonomous operation in SBR system, the influent data was generated by referring to the information of influent characteristics from South Korea (Heo et al., 2021). For realistic consideration on actual WWTP, the composition ratio of carbon and nitrogen in generated influent loads was changed within the 10% variations, considering diurnal patterns.

*Multi-agent reinforcement learning (MARL): G2ANET*

The MARL was adopted to develop the autonomous control of SBR process, while considering cost-effective and environment-friendly performance. For autonomous control, two RL agents were assigned to search for the optimal setpoints of DO concentration in the aerobic phase, and injection of external carbon (EC) dosage in the anaerobic phase. For this, A game abstraction method based on a two-stage attention network (G2ANet) algorithm was employed to simultaneously search for two operational setpoints: DO, and EC. Two operating variables were manipulated by the G2ANET and were determined to maintain the reliable performance of nitrification and denitrification processes.

G2ANET is comprised of graph and attention mechanisms (Liu et al.(2020)). The graph structure represents the connectivity between nodes and edges as shown in Eq.(4). Each node mapped the state information of all agents; and then, it is contributed to each agent as represented in Eq.(5). Herein, the adjacency matrix represents the set of edges as shown in Eq.(6). This series of mechanism is encoded as hard and soft abstraction to express G2ANET policy.

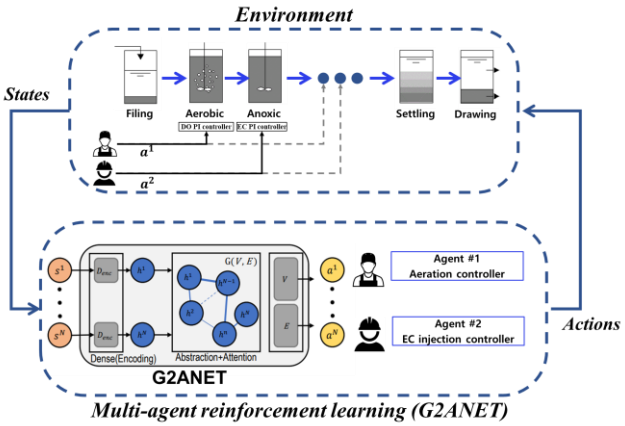


*Figure 3. Interfaces between SBR environment and multi-agent from G2ANET algorithm for aeration and EC injection controllers*

$$Q_i(O_i, a_i) = f_i(g_i(O_i, a_i), x_i) \quad (4)$$

$$x_i = \sum_{j \neq i} w_j v_j \quad (5)$$

$$w_j = W_h^{i,j} W_s^{i,j} \propto \exp\left(\mathrm{h}(\mathrm{BiLSTM}_j(e_i, e_j)) e_j^T W_k^T W_q e_i\right) \quad (6)$$

The attention mechanism is a core of G2ANETand it includes hard and soft attention. Hard attention is to set the connectivity between nodes, and soft attention is a mechanism to decide which information to propagate between nodes. For hard attention, encoded nodes summarized information from all nodes and propagated those features in the attention layer. After that, Gumbel-softmax processed propagated information as 1 or 0. A treated values of 0 and 1 indicated the status of connectivity between nodes. Then, an adjacency matrix can be obtained which represented significant connectivity between nodes for the communication of agents.

In soft attention mechanism, the informative features are decoded as query, key, and value from encoded nodes. For each encoded node, the scaled score is calculated by comparing to the query, while keys and values are collected. The scaled score indicated information to be propagated to the other nodes. By veiling the connectivity from hard attention, the node that will deliver informative features is feasible to be simplified for selective communication between agents.

The cooperative agents of G2ANET should decide the controller setpoints to maximize the total rewards by interfacing with the SBR model by the Q function in Eq. (7). For this, the structure of G2ANET algorithm including state, observations, actions, and total reward functions was designed as table.

$$Q_i^{\pi}(s, \vec{a}) = E_{\pi}\left\{\sum_{k=0}^{\infty} \gamma^k r_i^{t+k} | s_t = s, \vec{a}_t = \vec{a}\right\} \quad (7)$$

*Table 2. Structure of G2ANET for SBR control system*

| Structures | Descriptions |
|---|---|
| Agent 1 (Aeration controller) | - Action: [-1, -0.5, 0, +0.5, and +1 $m^3$/d] of Air flowrate<br>- Observations: $[t, X_H, X_A, S_O, S_{NO}, \dot{X}_H, \dot{X}_A, \dot{S}_O, \dot{S}_{NH}]$ |
| Agent 2 (EC injection controller) | - Action: [change of -1, -0.5, 0, +0.5, +1 $m^3$/d] for external carbon flow rates<br>- Observations: $[t, X_H, X_A, S_{NH}, S_{NO}, \dot{X}_H, \dot{X}_A, \dot{S}_O, \dot{S}_{NO}]$ |
| State | $[t, S_S, S_I, S_{NH}, S_{NO}, S_O, X_S, X_I, X_H, X_A, \cdots, X_H, X_A, X_P, X_{ND}]$ |
| Reward | $R_t = 1 - {(EQI^2 + OCI^2)}/{2}$ |
| Hyperparameters | - Encoding: 5-32 (linear), 32-32 (GRU)<br>- Hard attention: 64-32 (Bidirectional GRU (Bi-GRU))<br>- Soft attention: 32-32 (query, linear), 32-32 (key, linear), 32-32 (value, linear)<br>- Graph neural network (GNN): 64-5 (linear)<br>- Central critic: 14-64-64-1 (linear) |

To achieve the sustainable operation of SBR through the MARL training, the total reward function was made up of effluent quality index (EQI) for environmental-friendly benefit and operational cost index (OCI) of the cost-effective in Eqs. (8) and (9), respectively.

$$EQI = \frac{1}{1000 \cdot t_{obs}} \int_{t_0}^{t_f} \left[\beta_{TSS} TSS_e(t) + \beta_{COD} COD_e(t) + \beta_{NKj} S_{NKj,e}(t) + \beta_{NO} S_{No,e}(t) + \beta_{BOD} BOD_e(t)\right] \quad (8)$$

$$OCI = AE + PE + 5 \cdot SP + 3 \cdot EC + ME \quad (9)$$

where, AE is aeration energy (kWh/d), PE is the pumping energy (kWh/d), SP is the sludge production for disposal (kg/d), EC is external carbon addition (kgCOD/d), ME is mixing energy (kWh/d), and is the net heating energy needed to heat the anaerobic digester (kWh/d).

*Control performance guarantee using LRP*

Layer-wise Relevance Propagation (LRP) was utilized as an Explanation AI (XAI) method to assign importance scores (i.e., relevance) to the different input variables of an actor-critic network that represent the contribution of an input variable to the control policies of G2ANET agents. The importance score is backpropagated through actor-critic networks and assigned to neurons in all layers. These importance scores indicate the weighted significance of neurons that more contributed to the feature propagation through the actor-critic network, hence it can formulate a relevance metric for the regularized criterion. Considering local propagation rules, the basic LRP rule is used to propagate relevance scores ($R_k$) at a layer $k$ to neurons of the lower layer $j$ by backward layer-by-layer as shown in Eq. (10) (Montavon et al. (2019)).

$$R_j^l = \sum_k \frac{z_{jk}}{\sum_j z_{jk}} R_k^{l+1} \quad (10)$$

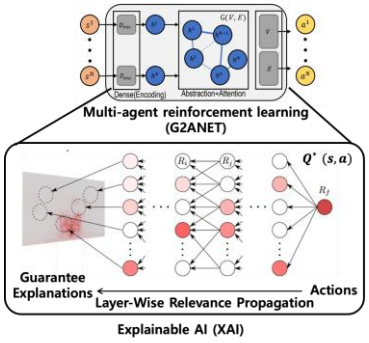


*Figure 4. XAI-assisted explanations on guaranteed control performance of G2ANET in SBR system using layer-wise relevance propagation (LRP)*

## Results and discussion

*Development of G2ANET multi-agent for SBR system*

The multi-agents for the SBR control system were trained from the generated training dataset as shown in Fig. 5. The proposed SBR control system was fully trained after 3,000 episodes. G2ANET agents learning progress is represented by total reward functions as depicted in Fig. 5(a). When the agent identifies the optimal control strategy, the reward increases. Moreover, the reward convergence indicates that the multi-agents of G2ANET model were optimized. For this case, the rewards converge to an approximate value of 0.3 after 3,000 episodes. This means that the agent can find the optimal control method for effluent quality and energy consumption simultaneously. To optimize the neural network model, no regularization technique was used, but it showed acceptable performance because the augmented dataset acted in a regularization role to prevent the network to fall into overfitting issues.

Fig. 5(b) to (d) shows the control performance of the agent through the episodes. In early episodes in Fig.5(b), the agent controlled the SBR system unsteadily based on a random selection of air flowrate and EC without any policy. Therefore, the agent was not able to learn from the causality between inverter frequency and the environment of the SBR system. Progressively, the RL agents learned how to select correct actions to control the environment.

Following, at transition stage in episode 1,000 to 2,000 as shown in Fig. 5(c), the action profiles of multi-agents in each phases represented to be operated to minimize the $S_{no}$ and $S_{nh}$. The results showed that, by sustaining the air flowrate close to around 2 mg/L, the SBR control system could reduce energy consumptions while considering the carbon and nitrogen components such as $S_S$, $S_{NO}$, and $S_{nh}$ during reaction phases. However, the control performance

does not satisfy the effluent quality criteria due to the increasing airflowrate in the last reaction phases by fully untrained agents.

After 5,000 episode, the agent found the policy to properly control the SBR system as shown in Fig. 5(d). The SBR was autonomously controlled by multi-agents of G2ANET; those RL agents can control the EC and airflowrate to minimize the energy consumptions and satisfy the effluent quality simultaneously. Hence, it can be concluded that the trained AI-based SBR control system finds the optimal method to control airflowrate and EC from the complex characteristics of influent and sequencing batch processes under aerobic and anaerobic conditions.

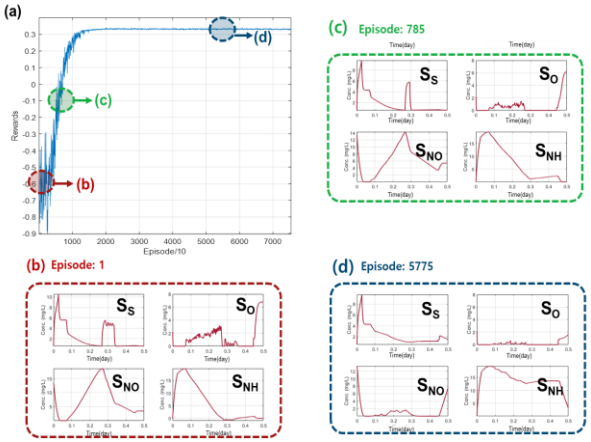


*Figure 5. Training result of the G2ANET-based SBR control system for (a) average rewards per 10 episodes of the training result and control performances in (b) 1, (c) 785, and (d) 5,775 episodes*

Fig. 6 represented reward profile of G2ANET agents into t-SNE space. Fig. 6(a) indicated the original training procedure of RL agents for SBR control system, and Fig. 6(b) to (e) represented training procedure in t-SNE space. Note that the t-SNE space indicated continuous variations while the G2ANET agents were trained. The first and second t-SNE variables was changed from lower to higher value, although the third t-SNE variable was fluctuated. This figure verified that the proposed G2ANAET-based SBR control system was trained by converging into higher rewards function; this conclusion is same as previous studies including Chen et al.(2021).

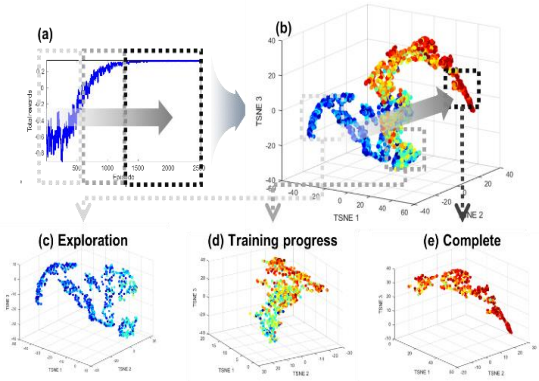


*Figure 6. Visualization of training progress of G2ANET-based SBR control system using t-SNE*

*Autonomous control performances of G2ANET algorithm*

As representative, Figure 7 indicated the control performance of MARL in SBR system. Note that the MARL can reduce the aeration energy as 14.6 % comparing to the base case. In base case, the DO ($S_O$) was increased to 4mg/L without considering the influent characteristics and SBR operations. While consuming the aeration energy, the SBR usually maintain high DO level to minimize the pollutants in effluent including the ammonia; because high concentration of ammonia can result in eutrophication of water sphere. On contrary, the proposed MARL-based SBR control system can reduce the aeration energy while considering the effluent criteria. The trained G2ANET agents operated the SBR as 0.5mg/L in the first reaction phase to minimize the aeration energy. Then, in sequential anoxic reaction phase the trained agent increased EC to treat the ammonia. To satisfy the effluent quality, the G2ANET increased the DO up to 2mg/L; the ammonia and total nitrogen (TN) can be lower than 4 and 20 mg/L, while the $S_{NO}$ was increased due to the last aeration in reaction phase.

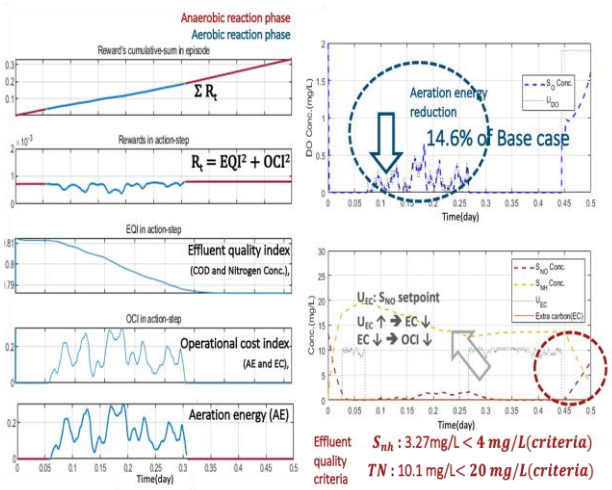


*Figure 7. Control performance of G2ANET-based SBR control system with reward functions*

In respect to reward functions, the instant rewards considering EQI and OCI continuously increased in a cycle of SBR operation. It indicated that the G2ANET agents can operate the SBR considering the trade-off perspective of environmental and economic values, simultaneously. Furthermore, the G2ANET agents can find the optimal control policy from the influent characteristics and process information of SBR system without human intervention. Hence, the result verified that the G2ANET can be utilized to autonomously operate the SBR system.

*XAI-assisted control performance guarantee of MARL*

Figure 8 represented the explanation for autonomous control of SBR system by G2ANET agents using LRP. Using the LRP the relevance and important features for MARL agents to find the autonomous control policy can be identified. Figure 8 indicated the relevant features from states and two observations for each agent. The propagation of state variables was separated by the operational phases; in aerobic phase, the concentrations of DO, heterotrophs, and autotrophs are more significantly propagated. Likewise, for observation of aeration controller agent, *So* was highly relevant features to determine the DO control setpoint by agent. Furthermore, the derivations of autotrophs and DO were contributed to find the DO control policy. For EC injection agent, the concentrations of $X_H, S_{nh}, \dot{X}_H$, and $\dot{S}_{no}$ are highly relevant to the control policy of EC injection. The objective of EC injection is to reduce the $S_{nh}$ through the $X_H$, and then $S_{nh}$ is converted to $S_{no}$. Hence by considering those variables including the increment of $S_{no}$ by $\dot{S}_{no}$, the MARL agents decided the EC injection control policy. Thus, the proposed MARL control system can control the SBR autonomously without human intervention, based on the physical, chemical, and biological characteristics in SBR operation.

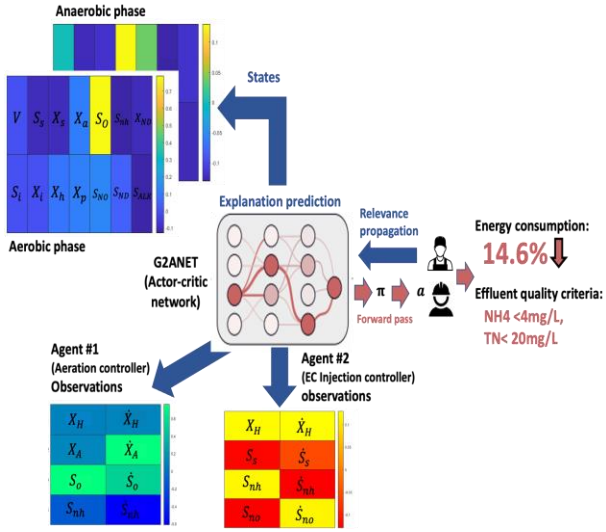


*Figure 8. LRP explanations of relevant features from state and observation variables for control polices of G2ANET agents*

## Conclusion

We developed MARL-based autonomous control system for SBR using G2ANET algorithm. For this, we generated various dataset of influent characteristic to train the G2ANET; and the SBR was modeled based on the ASM1. G2ANET was compromised of two agents to control the aeration and EC injection. Furthermore, LRP which is one of the XAI methods was implemented to explain the control performance guarantee of proposed G2ANET algorithm. The result indicated that the G2ANET agents control the SBR to reduce the energy consumptions as 14.6% comparing to the base case, while maintaining effluent quality criteria. Furthermore, XAI explained that the improved control performance of G2ANET agents comes from the understanding of mechanism in SBR operation without human intervention. Hence the proposed G2ANET-based SBR control system can be implemented into real WWTP with guaranteed control performance to assist the practitioner.


## Acknowledgement

This work was supported by the National Research Foundation (NRF) grant funded by the South Korean government (MSIT) (No. 2021R1A2C2007838), the BK21 FOUR program of NRF of Korea, and project for Collabo R&D between Industry, Academy, and Research Institute funded by Korea Ministry of SMEs and Startups in 2022 (Project No.S3105519).


## References


Nam, K., Heo, S., Loy-Benitez, J., Ifaei, P., & Yoo, C. (2020). An autonomous operational trajectory searching system for an economic and environmental membrane bioreactor plant using deep reinforcement learning. Water Science and Technology, 81(8), 1578-1587.

Chen, K., Wang, H., Valverde-Pérez, B., Zhai, S., Vezzaro, L., & Wang, A. (2021). Optimal control towards sustainable wastewater treatment plants based on multi-agent reinforcement learning. Chemosphere, 279, 130498.

Pons, M. N., Casellas, M., & Dagot, C. (2004). Definition of a benchmark protocol for sequencing batch reactors (B-SBR). IFAC Proceedings Volumes, 37(3), 439-444.

Henze, M., Gujer, W., Mino, T., & van Loosdrecht, M. C. (2000). Activated sludge models ASM1, ASM2, ASM2d and ASM3. IWA publishing.

Heo, S., Nam, K., Tariq, S., Lim, J. Y., Park, J., & Yoo, C. (2021). A hybrid machine learning–based multi-objective supervisory control strategy of a full-scale wastewater treatment for cost-effective and sustainable operation under varying influent conditions. Journal of Cleaner Production, 291, 125853.

Liu, Y., Wang, W., Hu, Y., Hao, J., Chen, X., & Gao, Y. (2020, April). Multi-agent game abstraction via graph attention neural network. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 34, No. 05, pp. 7211-7218).

Montavon, G., Binder, A., Lapuschkin, S., Samek, W., & Müller, K. R. (2019). Layer-wise relevance propagation: an overview. Explainable AI: interpreting, explaining and visualizing deep learning, 193-209.