2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

Reinforcement Learning-based Operational Decision-Making in the Process Industry Using Multi-View Data

Chenliang Liu, Yalin Wang, Chunhua Yang, Weihua Gui

Abstract—**Owing to the frequent fluctuations encountered in raw material characteristics and operational conditions in the process industry, traditional data-driven approaches prove inadequate in adapting the adjustment of operational variables. Furthermore, the potential of multi-view data, including images, audio, and sensor data, remains underexploited in industrial processes. This study proposes an operational decision-making method based on feedstock-guided multi-view actor-critic (FMAC-ODM) using multi-view data to address these issues. This method utilizes the idea of reinforcement learning (RL) for enhanced decision-making. First, the problem of optimizing operational variables is reformulated into a continuous RL problem to acquire an improved decision-making policy aligned with the current operational conditions. Subsequently, the inclusion of feedstock properties in the state space is implemented to provide essential guidance for the decision-making process. Finally, in pursuit of a comprehensive understanding and bolstering the precision of the decision-making strategy, multi-view data sourced from the industrial site is harnessed as a surrogate for human observation. The effectiveness of the proposed decision-making method is substantiated through its practical application in the industrial flotation process.**

I. Introduction

The process industry assumes a paramount significance in the economic advancement of modern society, including a diverse array of fields such as steel, petroleum, and chemicals [1-3]. In the production process of the process industry, the intelligent decision-making of operating variables is crucial to augment product quality and yield to a significant degree. However, the decision-making process is often influenced by the experience levels of on-site workers, which can significantly impact the achievement of overall production goals. Moreover, due to the existence of physical and chemical reactions in the production process, it is difficult to establish complex nonlinear relationship models between operational variables and production metrics via mechanism analysis. Hence, optimizing operational variables remains a complex and daunting problem in process industries [4, 5].

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 92267205, in part by the Science and Technology Innovation Program of Hunan Province in China under Grant 2021RC4054, in part by the Fundamental Research Funds for the Central Universities of Central South University under Grant 2022ZZTS0181, and in part by the China Scholarship Council under Grant 202206370157. *(Corresponding author: Yalin Wang.)*

Chenliang Liu is with the School of Automation, Central South University, Changsha 410083, China, and also with the School of Chemistry, Chemical Engineering and Biotechnology, Nanyang Technological University, Singapore 637459 (e-mail: lcliang@csu.edu.cn).

Yalin Wang, Chunhua Yang and Weihua Gui are with the School of Automation, Central South University, Changsha 410083, China (e-mail: ylwang@csu.edu.cn; ychh@csu.edu.cn; gwh@csu.edu.cn).

Several data-driven decision-making methods have been developed recently to generate decision-making values of operational variables in industrial processes [6-8]. For example, Zhou et al. [9] developed a supervised monitoring strategy to adjust the operational variables of the industrial grinding process based on changes in boundary conditions. Lu et al. [10] developed a data-driven optimization scheme for operational variable control without requiring complete knowledge of industrial process dynamics. Yang et al. [11] developed a multi-objective evolutionary method for adjusting operational variables in the mineral processing process. In addition, model predictive control (MPC) is a commonly used optimization method in the process industry. Amrit et al. [12] developed a method for optimizing process economics to address the nonlinear model predictive control problems. Chai et al. [13] developed a network-based model predictive control method to achieve the prescribed performance goals for setpoint compensation in industrial processes. However, the fitness evaluation of evolutionary computation results in high computational costs. While for the application of MPC, the mismatch between the mathematical description and the actual process may lead to significant performance loss. Given that reinforcement learning (RL) enables us to derive an optimal decision-making policy based on real-time information, it provides a means to mitigate the performance loss arising from model mismatch due to the uncertainty of reality. Hence, a model-free RL algorithm presents a promising solution for industrial processes [14].

RL is an innovative and efficient approach to obtaining optimal decision-making policies in industrial processes by interacting with agents and situations approaching real-world complexity [15, 16]. Ding et al. [17] developed a dynamic RL-based multi-objective optimization to identify the optimal values of operational variables in industrial plants. Ma et al. [18] developed an adaptive reference vector RL approach for industrial copper-burdening optimization. He et al. [19] developed an industrial decision-making system based on deep Q-networks for optimizing vital operational variables.

It is noteworthy that in industrial processes, the optimal strategy of the operational variables is conventionally designed by engineers based on historical data and experience, resembling an expert system grounded on the knowledge of operators. Analogous to expert systems, RL has the potential to continuously enhance operational decision-making policies based on reward data that update the performance metrics function. This attribute renders the application of RL algorithms in industrial processes more reasonable.

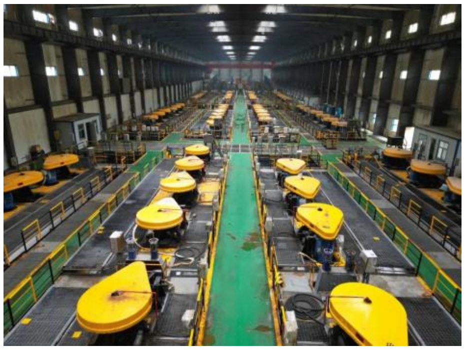

Fig. 1. Industrial flotation process.

Inspired by the above ideas, this paper proposes an operational decision-making method based on feedstock-guided multi-view actor-critic (FMAC-ODM) for the decision-making of operational variables in industrial processes. The main contributions of this paper are summarized as follows:

a) The multi-view data of the industrial process is utilized to enhance the adaptability of operational decision-making strategy by fully simulating the overall perception of the operators at the industrial site.

b) To address the issue of frequent fluctuations in raw materials in the process industry, the feedstock conditions of the production process have been introduced as the state space of the algorithm to enhance its accuracy.

c) The advantages of online learning of the actor-critic framework are used to reduce the calculation basis and accelerate the convergence speed.

d) The proposed operational decision-making method is substantiated by utilizing data collected from flotation plants in real industrial processes.

II. Problem Formulation

In this section, a brief description of the industrial flotation process is provided. Then, the operational decision-making framework for the flotation process is introduced accordingly.

A. Industrial Flotation Process Description

The flotation process plays a significant role in the mineral processing of the process industry, which entails the separation of minerals from raw ores through physicochemical surface properties. A diagram of the flotation process line as applied in an industrial field is presented in Fig. 1. The flotation process has been widely used in the process industry due to its ability to recover various minerals, including metals, non-metals, and coal. The objective of the flotation process is to concentrate the valuable minerals from the raw ores by selectively attaching the desired mineral particles to air bubbles. These air bubbles then ascend to the surface of the flotation cell and create a froth layer that contains the mineral concentrate. Then, the froth is collected and further processed.

To achieve effective flotation in the industrial process, it is necessary to adjust the operating variables in real time based on the working condition fluctuations. These operating variables include the slurry level, aeration, flotation agent, and

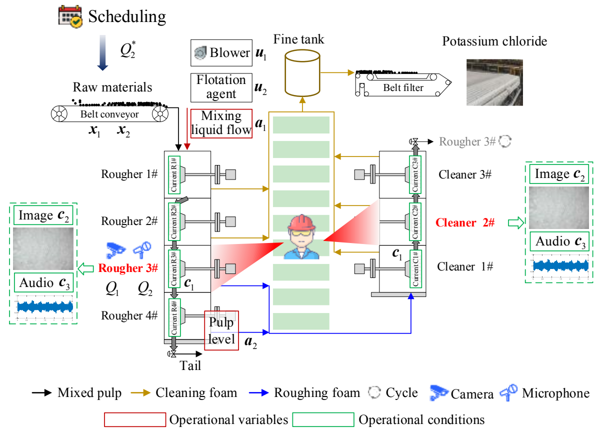

Fig. 2. Schematic of flotation process in a worker view.

TABLE I
DISCRIPTION OF OPERATIONAL VARIABLES

Symbol	Description
a_1	Mixed mother liquid flow
a_2	Roughing flotation pulp level
u_1	Ventilation volume
u_2	Flotation agent

agitation rate. In the current industrial process, the values of these operating variables are determined by operators based on their experience, with the aim of achieving the desired concentrate yield and grade within the target range. However, due to frequent changes in feedstock and operating conditions, manual selection of setpoints by operators is prone to errors resulting in significant fluctuations in both concentrate output and concentrate grade. This phenomenon can also ultimately lead to product degradation. One prospective remedy to this issue involves circumventing the selection of setpoints based on manual experiential knowledge and instead utilizing intelligent decision-making strategies. The proficient integration of such strategies harbors the capability to markedly amplify the utilization value of raw ore and elevate the overall efficiency of the mineral processing process.

B. Operational Decision-Making Framework

In this study, the potassium chloride flotation process at one factory in China was utilized as a representative industrial process, as depicted in Fig. 2. It mainly includes the roughing flotation stage and the cleaning flotation stage. On-site operators mainly observed the flotation performance of the Rougher 3# and indirectly observed Cleaner 2# to assess the flotation efficiency of the entire flotation group. Usually, two crucial performance metrics, the concentration of flotation froth Q_1 and the grade of flotation froth Q_2, are used to evaluate the flotation performance. The concentration of the flotation froth indicates the ratio of the solid in the flotation froth to the slurry (solid-liquid ratio), which can indirectly evaluate the froth yield of the flotation process. The grade of flotation froth indicates the proportion of ions in the flotation froth in the filtered solid, which can indirectly evaluate the froth quality of the flotation process.

Usually, the operational conditions of the flotation devices, such as stirring capacity, may directly impact the flotation performance, which is commonly represented by the stirring current in practice. In addition, to enable real-time monitoring of the industrial process, an industrial camera and microphone were installed to facilitate the acquisition of real-time images and audio data of the flotation froth in the vicinity of the Rougher 3# and Cleaner 2# instead of relying on manual observation by on-site operators near the flotation tank. Hence, the operational conditions of flotation process are expressed as $\boldsymbol{c}=[c_1,c_2,c_3]$, where c_1 denotes the stirring current, c_2 denotes the froth image, and c_3 denotes the froth audio.

Moreover, frequent changes in the feedstock properties can result in the working condition fluctuations of the flotation production process. Therefore, the feedstock ore conditions are also the operational conditions that must be considered. They include the flow and grade of feedstock ore, denoted by $\boldsymbol{x}=[x_1,x_2]$, where x_1 denotes the feedstock ore flow and x_2 denotes the potassium grade of the feedstock ore.

Table I presents the primary operational variables, which are summarized by extensive investigation of industrial sites and active discussions with experts. In the actual industrial application, the ventilation volume and flotation agent are generally fixed at constant values based on the human operators using their experiences and process knowledge. Hence, the operational variables are denoted as $\boldsymbol{a}=[\boldsymbol{a}_1,\boldsymbol{a}_2]$.

Since the setting and adjustment of operational variables are crucial to the production performance of the entire process, this study aims to propose an intelligent decision-making method that fuses the operational conditions based on the multi-view data and feedstock ore conditions obtained from the industrial process to provide optimal decision-making values of the operational variables. Then, the objective of decision-making problem of flotation performance metrics Q is described as

$$Q=Q_1+\rho\left(Q_2 \middle| Q_2^*\right) \tag{1}$$

where Q_1, Q_2 and Q_2^* represent the actual concentration of flotation froth, the actual grade of flotation froth, and the target grade of flotation froth, respectively. The penalty function $\rho(\bullet)$ assumes a negative scalar value if Q_2 is less than Q_2^* or $\rho(\bullet)$ equals zero under all other circumstances. Q_1 and Q_2 are described as

$$Q_1=f_1\left(\boldsymbol{a}\middle|\boldsymbol{c},\boldsymbol{x}\right) \tag{2}$$

$$Q_2=f_2\left(\boldsymbol{a}\middle|\boldsymbol{c},\boldsymbol{x}\right) \tag{3}$$

where $f_1(\bullet)$ and $f_2(\bullet)$ represent the embedding models of the concentration and the grade of the flotation froth, respectively.

Over the past few decades, most research efforts have investigated decision-making methods for operational variables in the flotation process. However, these approaches typically assume that feedstock ore conditions remain constant. This assumption is flawed since variations in raw material properties and equipment can lead to frequent fluctuations in the working conditions of the entire flotation process, negatively impacting production performance. Therefore, this

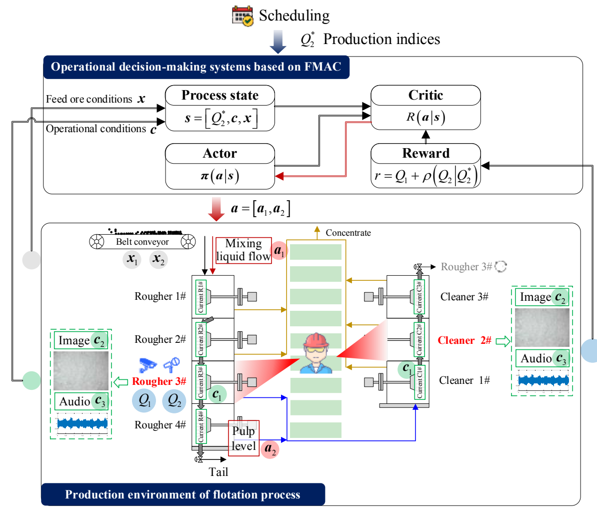

Fig. 3. Operational decision-making framework based on FMAC.

study proposes a dynamic optimization problem to describe the decision-making problem of operational variables in the flotation process and aims to achieve better performance.

$$\mathrm{E}_{(c,Q_2^*)\sim \boldsymbol{e}}[Q]=\mathrm{E}_{(c,Q_2^*)\sim \boldsymbol{e}}\left[f_1\left(\boldsymbol{a}\middle|\boldsymbol{c},\boldsymbol{x}\right)+\rho\left(f_2\left(\boldsymbol{a}\middle|\boldsymbol{c},\boldsymbol{x}\right)\middle|Q_2^*\right)\right] \tag{4}$$

where $\boldsymbol{e}$ represents the probability distribution of operational conditions.

Notably, most of the current research endeavors related to operational decision-making methods focus on flotation froth images. To the best of our knowledge, the use of froth audio to provide intelligent decision-making of operational variables for field workers is the first attempt.

III. Proposed Operational Decision-Making Method

To overcome the impact of feedstock ore fluctuations and effectively utilize multi-view data obtained from industrial processes, this study designs a FMAC algorithm for industrial operational decision-making. Fig. 3 further provides a visual framework for its application in industrial flotation process.

A. Reinforcement Learning

Reinforcement Learning (RL) is an effective learning approach based on behavioral psychology principles, which aims to find a balance between exploration and exploitation by using a trial-and-error mechanism with the environment [15]. Active and model-free RL methods have found extensive application in addressing various optimization and control challenges, particularly in complex environments [20]. Usually, the fundamental components of RL consist of the agent and the environment, where actions and rewards are the critical elements that connect them. Specifically, the agent is responsible for selecting the actions to interact with the environment, while the rewards received in response to the actions taken are used to evaluate their efficacy.

The mathematical formulation of RL is presented as a three-tuple $(s,\boldsymbol{a},r)$, which includes three key components:

(1) State (s) characterizes the environment;

(2) Action ($\boldsymbol{a}$) represents the decision-making output, which is determined based on the current state of the system and the associated reward;

(3) Reward (r) provides a quantitative evaluation.

The fundamental concept of RL is to enhance the optimal strategy by obtaining the maximum reward function in the long term, which is expressed as

$$\max_{\boldsymbol{\pi}} J(\boldsymbol{\pi}) = \max_{\boldsymbol{\pi}} \sum_{t=0}^{T} \mathrm{E}_{(\boldsymbol{s}_t,\boldsymbol{a}_t)\sim\boldsymbol{\pi}}\left[\gamma^t r(\boldsymbol{s}_t,\boldsymbol{a}_t)\right] \tag{5}$$

where T represents the maximum step. $\gamma \in (0,1]$ represents the discount factor, which is helpful to the convergence of the function. $\pi(\boldsymbol{a}|\boldsymbol{s})$ represents the conditional distribution, which is a criterion of selecting actions. Then, the optimal decision-making policy is calculated by optimizing the long-term reward function.

$$\tilde{\boldsymbol{\pi}} = \arg\max_{\boldsymbol{\pi}} \sum_{t=0}^{T} \mathrm{E}_{(\boldsymbol{s}_t,\boldsymbol{a}_t)\sim\boldsymbol{\pi}}\left[\gamma^t r(\boldsymbol{s}_t,\boldsymbol{a}_t)\right] \tag{6}$$

where $\tilde{\boldsymbol{\pi}}$ represents the optimal decision-making policy. The above function is often solved using an algorithm that searches for extreme values.

B. *Operational Decision-Making Method based on Feedstock-Guided Multi-View Actor-Critic*

In the context of the industrial flotation process, the formulation of a rational program and the precise definition of states, rewards, and actions are fundamental to achieving a global optimal decision-making strategy. As discussed in the previous section, different operational conditions and grades of flotation froth requirements require different decision-making values of operational variables. For instance, when the raw materials are high quality, it is necessary to select the decision-making values of operational variables that increase product yield to improve production efficiency. Conversely, for low-quality raw materials, it is necessary to select the decision-making values of the operational variables that improve the grade of flotation froth.

Therefore, based on industrial process mechanism and prior knowledge, the state space of the RL algorithm comprises operational conditions and feedstock ore conditions, as well as the target grade of flotation froth, which is denoted as $s = \left[Q_2^*, \boldsymbol{c}, \boldsymbol{x}\right]$. In particular, we use industrial cameras and microphones to collect flotation froth images and audio from actual industrial sites to assist in operational decision-making. This can ensure the real-time acquisition of flotation process multi-view data, circumventing the need for on-site workers to remain constantly at the industrial site for observation.

Inspired by the above analysis and reinforcement learning algorithm, an operational decision-making method based on feedstock-guided multi-view actor-critic (FMAC-ODM) is proposed for effective operational decision-making in the flotation process. This operational decision-making strategy aims to obtain relatively optimal decision-making values of the operational variables to ensure that the concentration and grade of flotation froth remain within the desired range. To achieve this, the reward function of the proposed FMAC-ODM is defined as

$$r = Q_1 + \rho\left(Q_2 \middle| Q_2^*\right) \tag{7}$$

$$r\left(Q^*, \boldsymbol{a}, \boldsymbol{c}, \boldsymbol{x}\right) = f_1\left(\boldsymbol{a}|\boldsymbol{c},\boldsymbol{x}\right) + \rho\left(f_2\left(\boldsymbol{a}|\boldsymbol{c},\boldsymbol{x}\right) \middle| Q_2^*\right) \tag{8}$$

where r is a nonpositive scalar function, $f_1\left(\boldsymbol{a}|\boldsymbol{c},\boldsymbol{x}\right)$ represents the concentration of flotation froth, $\rho\left(f_2\left(\boldsymbol{a}|\boldsymbol{c},\boldsymbol{x}\right)\middle|Q_2^*\right)$ is a penalty function.

The decision-making framework for operational variables, as presented in Eq. (4), can be transferred to Eq. (5) in the RL algorithm framework, which is given as

$$J(\boldsymbol{\pi}) = \mathrm{E}_{\boldsymbol{s}\sim\boldsymbol{e},\boldsymbol{a}\sim\boldsymbol{\pi}}\left[r(\boldsymbol{s}_t,\boldsymbol{a}_t)\right] \tag{9}$$

It is worth noting that, unlike other sequential decision processes, the decision-making of operational variables in this context is not sequential. Therefore, the step size T is selected to be one in each episode. The iterative approach is frequently employed to refine the optimal decision-making policy, which can be characterized as a continuous process. Thus, the derivation of this policy can be described as follows:

$$\boldsymbol{\pi}_{\text{new}} = \arg\max_{\boldsymbol{\pi}} \mathrm{E}_{\boldsymbol{s}\sim\boldsymbol{e}(s),\boldsymbol{a}\sim\boldsymbol{\pi}(\boldsymbol{a}|\boldsymbol{s})}\left[r(\boldsymbol{s},\boldsymbol{a})\right] \tag{10}$$

where $\boldsymbol{\pi}(\boldsymbol{a}|\boldsymbol{s})$ is assumed as a conditional distribution belongs to Gaussian distribution.

Considering the high-dimensional and continuous nature of the state and action spaces involved in the optimal operational decision-making problem, an actor network is employed using a neural network implementation denoted as $\boldsymbol{\pi}_\theta(\boldsymbol{a}|\boldsymbol{s})$, where parameter θ is used to approximate the Gaussian distribution. The actor network utilizes the state as input and produces the action as output.

In addition, the critic network $R_\varphi(\boldsymbol{a}|\boldsymbol{s})$ with parameter φ is used to estimate the reward generated by $\boldsymbol{\pi}_\theta(\boldsymbol{a}|\boldsymbol{s})$. The input of the critic network consists of the state and action. During the training process, the loss function of the critic network is defined as follows:

$$J(\varphi) = \frac{1}{2}\left[R_\varphi(\boldsymbol{s},\boldsymbol{a}) - r(\boldsymbol{s},\boldsymbol{a})\right]^2 \tag{11}$$

where $r(\boldsymbol{s},\boldsymbol{a})$ represents the actual reward of the production data. Then, by minimizing the loss function, $R_\varphi(\boldsymbol{s},\boldsymbol{a})$ can be replaced by $r(\boldsymbol{s},\boldsymbol{a})$ when the training accuracy is satisfied. Hence, the policy is updated as

$$\boldsymbol{\pi}_{\theta\text{new}} = \arg\max_{\boldsymbol{\pi}_\theta} \mathrm{E}_{\boldsymbol{s}\sim\boldsymbol{e}(s),\boldsymbol{a}\sim\boldsymbol{\pi}_\theta(\boldsymbol{a}|\boldsymbol{s})}\left[R_\varphi(\boldsymbol{s},\boldsymbol{a})\right] \tag{12}$$

Furthermore, integrating experience replays into the FMAC-ODM method allows for repeated learning from experiential data with benefits such as reduced costs, fewer trials and errors, and faster learning speeds. By doing so, the utilization efficiency of industrial process data can be enhanced, thereby addressing the challenge of insufficient data commonly encountered in industrial processes. In the experience replay method, a set of experiences consisting of the state, action, and immediate reward obtained during the interaction between the FMAC-ODM method and the

TABLE II
DISCRIPTION OF VARIABLES IN THE FLOTATION PROCESS

Tag	Description
Feedstock condition	Feedstock ore flow ($\boldsymbol{x}_1$), feedstock ore grade ($\boldsymbol{x}_2$)
Operational condition	Stirring current ($\boldsymbol{c}_1$), froth image ($\boldsymbol{c}_2$), froth audio ($\boldsymbol{c}_3$)
Operational variable	Mixed mother liquid flow ($\boldsymbol{a}_1$), roughing flotation pulp level ($\boldsymbol{a}_2$)
Performance metric	Froth concentration (Q_1), froth grade (Q_2)

flotation production process is stored in the experience pool. These experiences are randomly selected to form a batch and are used as a training set for the actor network. By minimizing the loss function defined based on the criterion, the decision-making policy can be improved as

$$\boldsymbol{\pi}_{\theta\text{new}} = \arg\max_{\boldsymbol{\pi}_\theta} \mathrm{E}_{\boldsymbol{s}\sim P,\boldsymbol{a}\sim\boldsymbol{\pi}_\theta(\boldsymbol{a}|\boldsymbol{s})}\left[R_\varphi(\boldsymbol{s},\boldsymbol{a})\right] \tag{13}$$

where P denotes the experience replay pool. It should be noted that during the training process of the critic network, a batch gradient descent method is employed to train the network using randomly selected experiences. Subsequently, the loss function is reformulated as shown below:

$$J(\varphi) = \frac{1}{2}\mathrm{E}_{(\boldsymbol{s},\boldsymbol{a},r)}\left[R_\varphi(\boldsymbol{s},\boldsymbol{a}) - r(\boldsymbol{s},\boldsymbol{a})\right]^2 \tag{14}$$

The above loss function can effectively alleviate the poor performance of the stochastic gradient descent method due to the presence of noise in a single sample. Subsequently, the FMAC-ODM method is used to obtain the relatively optimal decision-making policy based on the realizations of actor and critic networks based on iteratively updating Eqs. (13) and (14) in an alternating manner. Finally, the optimal decision-making values of the operational variables are obtained from the actor network, denoted as

$$\tilde{\boldsymbol{a}} = \arg\max_{\boldsymbol{a}} R_\varphi(\boldsymbol{s},\boldsymbol{a}) \tag{15}$$

where $\tilde{\boldsymbol{a}}$ represents the optimal decision-making values of the operational variables.

IV. EXPERIMENTS

In this section, the proposed FMAC-ODM method is applied to an actual industrial flotation process to validate its effectiveness.

A. Data Description and Experiment Setup

All experimental data sets are collected from the largest potassium chloride flotation plant of a mineral processing enterprise in China. The description of the potassium chloride flotation process is introduced in Section II. A total of 223 data sets were collected, including the feedstock ore conditions, operational conditions, operational variables, and performance metrics. The detailed description of these variables is given in Table II. The first 180 data sets were used for training, while the remaining 43 were reserved for validation. In this study, the three-layer neural networks are used to approximate the models $f_1(\bullet)$ and $f_2(\bullet)$ of the flotation process. These networks are designed to map the operational and feedstock ore conditions to the performance metrics. Given the focus of our study on the practical implementation of RL for operational decision-making in industrial processes, we assume that the current system has the capability to track and implement the decision-making outcomes for operational variables. To simulate the flotation production process at an industrial site, a dynamic model is designed based on the human experience and industrial mechanism.

TABLE III
COMPARISION RESULTS OF FOUR DECISION-MAKING METHODS

Method	Concentration (%)	Grade (%)
Manual operation	36.36-44.37 (40.64)	27.07-31.81 (29.40)
DQN-ODM	40.71-46.70 (43.63)	27.47-32.33 (29.76)
AC-ODM	39.08-47.77 (44.12)	27.05-32.12 (29.90)
FMAC-ODM	40.53-47.71 (44.90)	27.66-32.48 (30.42)

In the proposed FMAC-ODM method, the state vector is composed of the feedstock ore conditions $\boldsymbol{x}$, operational conditions $\boldsymbol{c}$, and target flotation froth grade Q_2^*. The action vector is obtained from the proposed operational decision-making method based on the FMAC-ODM method. The production goal of the industrial flotation process is to maximize the flotation froth concentration while meeting the flotation froth grade specifications. Notably, the concentration of flotation froth can indirectly reflect the flotation froth yield of the overall production process. Hence, the reward function is designed as

$$r = r_1 + r_2 \tag{16}$$

$$r_1 = Q_1,\ \text{and}\ r_2 = \begin{cases} -0.8, & Q_2 < Q_2^* \\ 0, & Q_2 \geq Q_2^* \end{cases} \tag{17}$$

Comparative experiments are designed to assess the effectiveness of the proposed method. Manual operations collected at industrial sites were used as a baseline for comparison. In addition, operational decision-making methods based on the deep Q-Network (DQN-ODM) [15] and the standard actor critic (AC-ODM) [21] are used as additional comparisons. For unbiased and impartial experimentation, all actor networks use three-layer neural networks comprising 128 hidden-layer neurons and are trained using a learning rate of 0.06.

B. Results and Discussion

The experimental results of the flotation froth performance metrics under four comparison operational decision-making methods are presented in Table III and Fig. 4. Table III reports the minimum, maximum, and average values (in parentheses) of the performance metrics. Fig. 4 intuitively depicts the trajectories of two performance metrics. As can be seen from the results, the proficiency of on-site operators lies primarily in regulating the froth grade, while their control of froth concentration has no significant advantages. Other operational decision-making methods based on the idea of RL framework, including DQN-ODM, AC-ODM, and FMAC-ODM, have significantly improved froth concentration, which indirectly guarantees an increase in flotation yield. Despite the increase in froth concentration observed under the DQN-ODM method, the froth grade tends to decrease, especially there is no high froth grade exists. On the other hand, the froth grade obtained

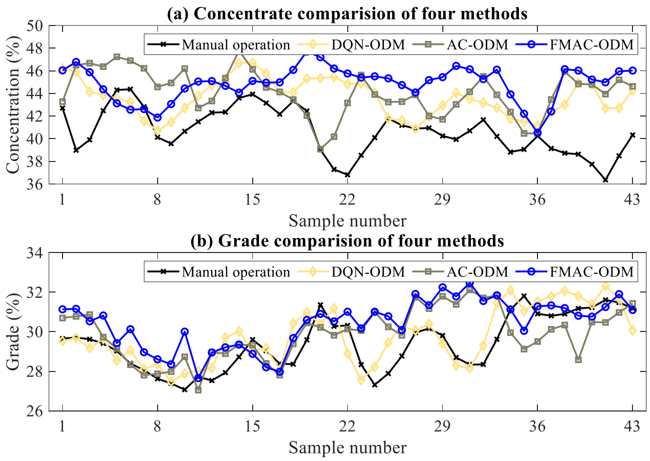

Fig. 4. Comparision results of froth concentration and grade based on different operational decision-making methods.

by the AC-ODM method fluctuates significantly, which may be attributed to the frequent adjustment of operational variables. Although the maximum froth grade achieved with the proposed FMAC-ODM method is lower than manual operation, it demonstrates the highest mean froth grade among the compared methods. In addition, the proposed method improves froth concentration while ensuring the froth grade, which also demonstrates its effectiveness in the operation optimization of industrial processes. Based on the above experimental results and analysis, it is concluded that the RL-based operational decision-making methods have the favorable potential for guiding the production of industrial processes, especially in scenarios where multi-view data and feedstock conditions from industrial sites are factored into the decision-making process.

V. Conclusion

This study introduces a novel operational decision-making method named FMAC-ODM for the optimization of operational variables in industrial processes. Compared to existing methods, the proposed FMAC-ODM method exhibits superior utilization of multi-view data derived from industrial sites, consequently affording it an enhanced capacity for comprehensive perception. Moreover, it effectively attenuates the impact of feedstock conditions on operational fluctuations. The superiority of the proposed method is demonstrated through simulation experiments using actual data sourced from the industrial flotation process.

References

[1] Tao Yang, Xinlei Yi, Shaowen Lu, Karl H. Johansson, and Tianyou Chai, "Intelligent manufacturing for the process industry driven by industrial artificial intelligence," *Engineering*, vol. 7, no. 9, pp. 1224-1230, 2021.

[2] Yishun Liu, Chunhua Yang, Keke Huang, Weihua Gui, and Shiyan Hu, "A systematic procurement supply chain optimization technique based on industrial internet of things and application," *IEEE Internet of Things Journal*, vol. 10, no. 8, pp. 7272-7292, 2023.

[3] Daniel Hodouin, "Methods for automatic control, observation, and optimization in mineral processing plants," *Journal of Process Control*, vol. 21, no. 2, pp. 211-225, 2011.

[4] Chenliang Liu, Kai Wang, Yalin Wang, and Xiaofeng Yuan, "Learning deep multimanifold structure feature representation for quality prediction with an industrial application," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 9, pp. 5849-5858, 2022.

[5] Yalin Wang, Silong Li, Chenliang Liu, Kai Wang, Xiaofeng Yuan, Chunhua Yang, and Weihua Gui, "Multi-scale feature fusion and semi-supervised temporal-spatial learning for performance monitoring in the flotation industrial process," *IEEE Transactions on Cybernetics*, early access, 2023, doi: 10.1109/TCYB.2023.3295852.

[6] Shuai Mao, Bing Wang, Yang Tang, and Feng Qian, "Opportunities and challenges of artificial intelligence for green manufacturing in the process industry," *Engineering*, vol. 5, no. 6, pp. 995-1002, 2019.

[7] Yi Jiang, Jialu Fan, Tianyou Chai, JinNa Li, and Frank L. Lewis, "Data-driven flotation industrial process operational optimal control based on reinforcement learning," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 5, pp. 1974-1989, 2018.

[8] Diju Liu, Yalin Wang, Chenliang Liu, Kai Wang, Xiaofeng Yuan, and Chunhua Yang, "Blackout missing data recovery in industrial time series based on masked-former hierarchical imputation framework," *IEEE Transactions on Automation Science and Engineering*, early access, 2023, doi: 10.1109/TASE.2023.3287895.

[9] Ping Zhou, Tianyou Chai, and Jing Sun, "Intelligence-based supervisory control for optimal operation of a DCS-controlled grinding system," *IEEE Transactions on Control Systems Technology*, vol. 21, no. 1, pp. 162-175, 2013.

[10] Xinglong Lu, Bahare Kiumarsi, Tianyou Chai, and Frank L. Lewis, "Data-driven optimal control of operational indices for a class of industrial processes," *IET Control Theory & Applications*, vol. 10, no. 12, pp. 1348-1356, 2016.

[11] Cuie Yang, JinLiang Ding, Yaochu Jin, Chengzhi Wang, and Tianyou Chai, "Multitasking multiobjective evolutionary operational indices optimization of beneficiation processes," *IEEE Transactions on Automation Science and Engineering*, vol. 16, no. 3, pp. 1046-1057, 2019.

[12] Rishi Amrit, James B. Rawlings, and Lorenz T. Biegler, "Optimizing process economics online using model predictive control," *Computers & Chemical Engineering*, vol. 58, pp. 334-343, 2013.

[13] Tianyou Chai, Lin Zhao, Jianbin Qiu, Fangzhou Liu, and Jialu Fan, "Integrated network-based model predictive control for setpoints compensation in industrial processes," *IEEE Transactions on Industrial Informatics*, vol. 9, no. 1, pp. 417-426, 2013.

[14] Dongqi Wu, Xiangtian Zheng, Dileep Kalathil, and Le Xie, "Nested reinforcement learning based control for protective relays in power distribution systems," in *2019 IEEE 58th Conference on Decision and Control (CDC)*, 2019, pp. 1925-1930.

[15] Richard S Sutton and Andrew G Barto, *Reinforcement learning: An introduction*: MIT press, 2018.

[16] Daniele Gammelli, Kaidi Yang, James Harrison, Filipe Rodrigues, Francisco C. Pereira, and Marco Pavone, "Graph neural network reinforcement learning for autonomous mobility-on-demand systems," in *2021 60th IEEE Conference on Decision and Control (CDC)*, 2021, pp. 2996-3003.

[17] Jinliang Ding, Hamidreza Modares, Tianyou Chai, and Frank L. Lewis, "Data-based multiobjective plant-wide performance optimization of industrial processes under dynamic environments," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 2, pp. 454-465, 2016.

[18] Lianbo Ma, Nan Li, Yinan Guo, Xingwei Wang, Shengxiang Yang, Min Huang, and Hao Zhang, "Learning to optimize: reference vector reinforcement learning adaption to constrained many-objective optimization of industrial copper burdening system," *IEEE Transactions on Cybernetics*, vol. 52, no. 12, pp. 12698-12711, 2022.

[19] Zhenglei He, Kim-Phuc Tran, Sebastien Thomassey, Xianyi Zeng, Jie Xu, and Changhai Yi, "A deep reinforcement learning based multi-criteria decision support system for optimizing textile chemical process," *Computers in Industry*, vol. 125, Art. no. 103373, 2021.

[20] Rui Nian, Jinfeng Liu, and Biao Huang, "A review on reinforcement learning: Introduction and applications in industrial process control," *Computers & Chemical Engineering*, vol. 139, Art. no. 106886, 2020.

[21] Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel, and Sergey Levine, "Soft actor-critic: off-policy maximum entropy deep reinforcement learning with a stochastic actor," in *Proc. 35th Int. Conf. Mach. Lear.*, 2018, pp. 1861-1870.