

# Machine learning-based prediction and optimization of plasma-catalytic dry reforming of methane in a dielectric barrier discharge reactor

Jiayin Li<sup>1,\*</sup>, Jing Xu<sup>2</sup>, Evgeny Rebrov<sup>3,4</sup> and Annemie Bogaerts<sup>1,\*</sup>

<sup>1</sup> Research Group PLASMANT, University of Antwerp, Department of Chemistry, Antwerp, 2610, Belgium

<sup>2</sup> School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan, Hubei, 430074, People's Republic of China

<sup>3</sup> School of Engineering, University of Warwick, Coventry, CV4 7AL, the United Kingdom

<sup>4</sup> Department of Chemical Engineering and Chemistry, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, the Netherlands

\*Corresponding author: [jiayin.li@uantwerpen.be](mailto:jiayin.li@uantwerpen.be) (J. Li), [annemie.bogaerts@uantwerpen.be](mailto:annemie.bogaerts@uantwerpen.be) (A. Bogaerts).

## Abstract

We developed an innovative machine learning (ML) model, including a supervised learning (SL) and reinforcement learning (RL) model, to predict and optimize the plasma-catalytic dry reformation of methane (DRM) over Ni/Al<sub>2</sub>O<sub>3</sub> catalysts in a dielectric barrier discharge (DBD) reactor based upon experimental data. To tackle its intricate and non-linear characteristics, the SL model uses artificial neural networks (ANN) to accurately predict the performance, achieving excellent consistency with the experimental results. The RL model subsequently investigates the optimal optimization policy, namely starting with a coarse tuning of the more influential parameters, followed by fine-tuning of the less important parameters, to obtain the best performance. The optimal results show that a discharge power at lowest bond (i.e., 20 W) but CO<sub>2</sub>/CH<sub>4</sub> ratio at highest bond (i.e., 1.5) result in the minimum energy cost (21 eV/molec), validated by our SL model and experimental data. Furthermore, we also investigated the simultaneous optimization of total conversion and energy cost, resulting in a maximum total conversion of 36%, combined with a minimum energy cost of 34 eV/molec, at a Ni loading of 9.5 wt%, discharge power of 60 W, and total flow rate of 74 mL/min. Our ML model showcases an impressive capacity to derive advantageous insights from existing datasets, thereby advancing and optimizing plasma-catalytic chemical processes.

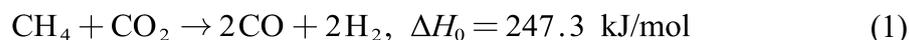
## Keywords

Plasma catalysis, Dry reforming of methane, Machine learning, Process optimization, Syngas production

## 1. Introduction

Carbon dioxide (CO<sub>2</sub>) and methane (CH<sub>4</sub>) are the two major greenhouse gasses that significantly contribute to climate change [1]. Currently, there is an urgent need for their conversion into value-added chemicals [2,3]. Dry reforming of methane (DRM) presents an emerging method for the concurrent conversion of both CO<sub>2</sub> and CH<sub>4</sub>,

yielding valuable syngas (i.e., CO and H<sub>2</sub>), as shown in Eq. (1). The syngas can serve as a hydrogen source or to further process to produce higher-value chemicals, e.g., methanol and formaldehyde [4,5].



Non-thermal plasma (NTP), a cutting-edge technology, enables this reaction operating at mild temperatures and ambient pressure [6–8], providing a valuable approach to traditional methods. Indeed, a recent techno-economic analysis, applied to a pilot plant producing 100 tCO/day, revealed that the energy cost of plasma-based CO<sub>2</sub> conversion is 43% less than for electrolysis and conventional CO<sub>2</sub> conversion methods, and that using a post-plasma carbon bed is crucial for this cost-effectiveness, to facilitate additional CO production from O<sub>2</sub> and enhancing the CO<sub>2</sub> conversion [9].

Plasma is a partially (or fully) ionized gas, containing charged species, radicals and excited species that exhibit collective behavior. The presence of radicals and highly energetic electrons enables reaction pathways that would otherwise be impossible, making plasma reactors valuable for converting thermodynamically stable molecules like CO<sub>2</sub> and CH<sub>4</sub>. The catalyst can further promote the selective syngas production, while combining it with plasma holds significant promise for generating a synergistic effect that boosts overall efficiency, although synergy is not always reached and more insight is needed in the underlying mechanisms [10,11]. Furthermore, plasma is generated with electricity, and can be rapidly switched on and off, thus facilitating its integration with the fluctuating nature of renewable energy sources [12].

Over the past few decades, a wide range of plasma reactors have been thoroughly tested for plasma-based DRM applications, such as glow discharge [13], corona discharge [14], gliding arc [15], microwave discharge (MW) [16] and dielectric barrier discharge (DBD) [17–20]. Because of its simple design and user-friendly nature, especially when combined with packing materials, most plasma catalysis research so far has utilized DBDs, within a coaxial reactor [7]. In general, the reaction performance and cost-effectiveness of plasma reactors are considered for evaluation, both of which are very important. Although the total conversion for DBD reactors can be quite high, the energy cost is often excessively high [21,22]. On the other hand, it is complicated to select for optimal performance as the input parameters also requires balancing chemical performance and cost-effectiveness [7]. Therefore, combining high conversions with relatively low energy costs is really challenging due to the many interacting parameters in complex plasma-catalytic systems [10,11].

To address these challenges, machine learning (ML) methods offer a novel pathway for exploring intricate scientific phenomena, and they are receiving growing attention for plasma processes, such as in plasma medicine [23–25], for the synthesis of chemicals [26,27] and in pollution control [28]. In general, ML methods can be broadly divided into two categories: (a) supervised learning (SL), which uses both input (predictors) and output (predicted variables) data, to effectively learn a mapping between them, but it requires large amounts of data; and (b) unsupervised learning, which relies solely on input data to learn patterns or correlations for finding hidden patterns or structures in input data, but the results are often harder to interpret and

validate. In plasma catalysis, previous studies have primarily utilized SL (referred to as ML in many papers) models, typically by artificial neural network (ANN) algorithm, for chemical performance prediction. For example, Liu et al. [29] successfully predicted the reaction performance in plasma-based CH<sub>4</sub> conversion to hydrocarbons. Similarly, Zhu et al. [30] elucidated the relationships between operation parameters and performance of CH<sub>3</sub>OH oxidation. Wang et al. [31] revealed that both higher gas conversion and energy efficiencies were favor at optimal conditions of 10-20 W and 5-20 °C. Recently, Cai et al. [32] found the optimal condition led to the maximum energy yield in plasma-catalytic DRM by a hybrid SL model. Despite the promise of ML in catalysis, the optimization of plasma-catalytic DRM by ML is however still in its infancy.

The complex plasma–catalyst interactions can be divided into two categories: the effects of plasma on the catalyst, and the effects of the catalyst on the plasma. Hence, changing one of the effects may have unpredictable influence on the other effects. In addition, the optimization process for the operating conditions likely leads to trade-offs among conflicting objectives (e.g., higher conversion and product yields, or higher energy efficiency), and the SL approach is not always accurate and robust enough for solving such multi-target problems. To achieve a cost-effective plasma-catalytic system with good reaction performance, it is crucial to obtain a thorough understanding on how the interconnected operating parameters can be optimized [13]. However, no studies consider the optimization of both total conversion and energy cost in plasma-catalytic DRM. Considering traditional trial-and-error efforts in experiments have reached their limit, and the simultaneous multi-objectives optimization, based on many interacting parameters, is challenging.

Reinforcement learning (RL), another major subset of ML, is a goal-direct approach, which involves how to maximize the outcome by mapping actions, to steer future experimental work [33]. The RL agents can learn to achieve specific goals by actively interacting with the environment, bypassing the necessity to a priori develop an accurate mathematical physico-chemical model [34]. This approach leverages RL's two strengths - obtaining performance via experience and preventing reliance on exact models, are especially beneficial for complex plasma-surface reactions, where it is not yet feasible to construct a precise dynamic model based on first principles, and the dynamic characteristics are highly variable [35]. However, the quality and quantity of training dataset matter most in the performance of RL models. In previous work, Cai et al., [32], carried out a comprehensive investigation, involving 100 distinct reaction conditions, to ensure the robustness and relevance for further RL model development.

In this work, we designed a very first RL model to simultaneously maximize the total conversion and minimize the energy cost in plasma-catalytic DRM, in an attempt to develop an artificial intelligence (AI) method for plasma catalysis. Firstly, we developed the prediction model for reaction performance (CO<sub>2</sub> and CH<sub>4</sub> conversion, H<sub>2</sub> and CO yield, total conversion) and cost-effectiveness (energy cost) by the ANN algorithm. Subsequently, we designed RL controllers (agents) by using the prediction model as simulation environment for learning interactions. Finally, the output from the RL model is compared with the actions chosen from four operating parameters,

including Ni loading, total gas flow rate, CO<sub>2</sub>/CH<sub>4</sub> ratio and discharge power. Therefore, by mapping these operating parameters, the RL model can determine the fine-tuning steps that achieve cost-effectiveness as well as good reaction performance in plasma-catalytic DRM.

## 2. Methodology

In this work, we used two ML methods, i.e., SL and RL, to develop the model describing the plasma-catalytic DRM process, as schematically illustrated in Figure 1, and detailly elaborated in the subsequent sections.

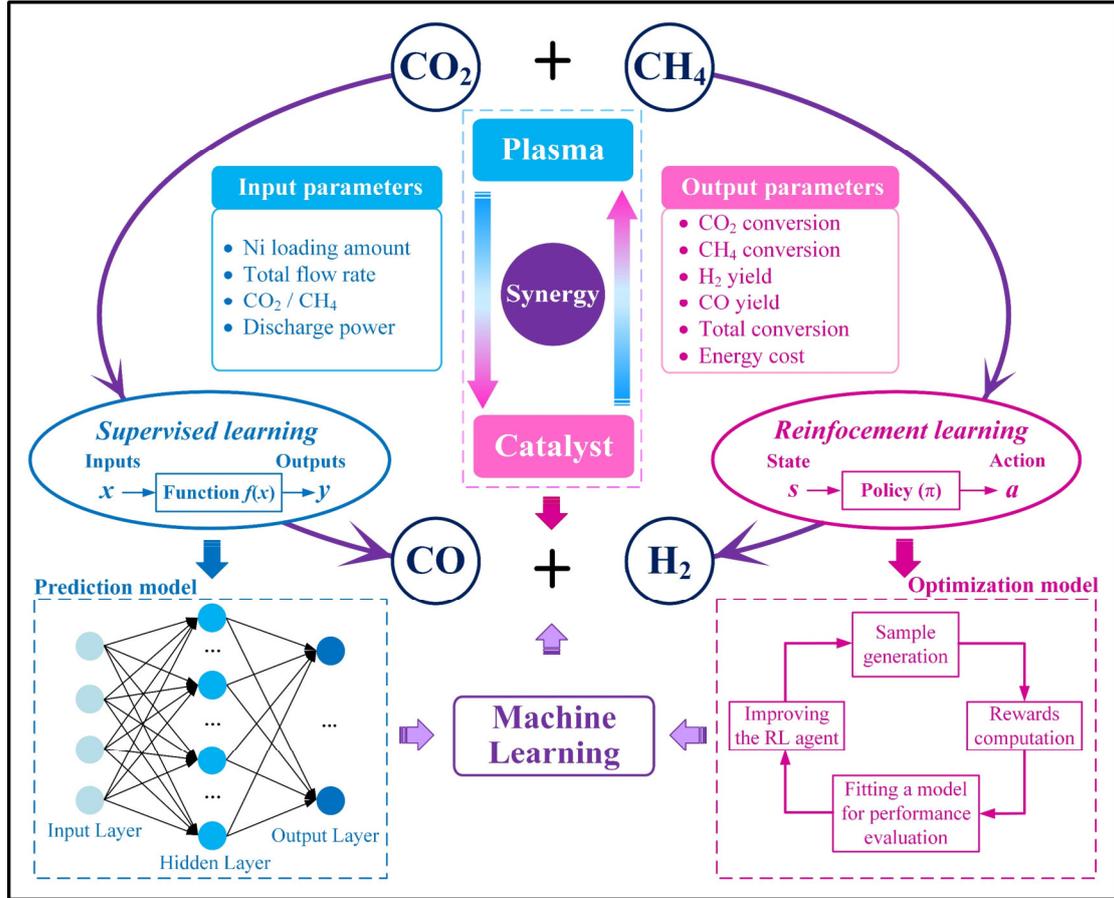


Fig.1 Overview of the ML model for plasma-catalytic DRM. A SL model predicts the output variable  $y$  by mapping input variables  $x$  through a function  $f(x)$ ; A RL optimization model involves how to map states to actions based on policy ( $\pi$ ).

### 2.1 Data collection and processing

Fig. S1 shows the experimental setup for DRM, which is detailly described in Ref. [20]. The total conversion  $\chi_{total}$  is of great importance, and obviously more complicated for optimization, which is defined by the summing of the effective CO<sub>2</sub> and CH<sub>4</sub> conversions [36]:

$$\chi_{total} = \sum_i \chi_i^{eff} = \chi_{CO_2}^{abs} \cdot y_{CO_2}^{in} + \chi_{CH_4}^{abs} \cdot y_{CH_4}^{in} \quad (2)$$

where the absolute CO<sub>2</sub> conversion  $\chi_{CO_2}^{abs}$  times the inlet CO<sub>2</sub> fraction  $y_{CO_2}^{in}$  and the

absolute CH<sub>4</sub> conversion  $\chi_{CH_4}^{abs}$  times the inlet CH<sub>4</sub> fraction  $y_{CH_4}^{in}$ . The energy cost (EC in unit eV/molec) is defined as [7]:

$$EC \text{ (eV/molec)} = \frac{SEI \text{ (eV/molec)}}{\chi_{total}} = \frac{Power(kW) \cdot 60 \text{ s/min} \cdot (24.05 \text{ L/mol} \cdot 6.24 \cdot 10^{21} \text{ eV/kJ})}{Total \text{ gas flow rate (L/min)} \cdot \chi_{total} \cdot 6.02 \cdot 10^{23} \text{ molec/mol}} \quad (3)$$

By varying the above-mentioned process parameters, the actual reaction performance and EC were collected as the dataset, which can be found in the supplementary material (Table S2, S3). Before model development, all datasets were pre-process by a Min-Max normalization method [37]. The normalized data is confined to the same interval, which accelerates the convergence of the network and avoids the saturation of neurons, as shown in Eq. S1 and Eq. S2 in supplementary information (SI). The complete dataset was divided into 70% training subset and 30 % testing subset. If the ratio between training and test set would be reduced, the ML prediction capability and quality will be lower, as less data will be used for training and more data will be used for testing. The ML model may fail to learn the underlying structure of the data, leading to underfitting and poor generalization ability. Also, if the ratio would be increased, the ML prediction capability and quality will be also lower. Although more data is used for training to identify learning patterns, there will be less test data, which may affect the reliability and stability of the evaluation results, as a smaller test set may not fully represent the diversity of the overall dataset.

## 2.2 Description of the artificial neural network

The artificial neural network (ANN), a conventional SL algorithm, was utilized to estimate the performance of plasma-based DRM under various operating conditions, as it exhibits remarkable predictive accuracy to solve the nonlinear problems. The network contains numerous artificial neurons inside, with connection weights between them, serving as learnable parameters. Typically, an ANN model comprises multiple layers, each of which consists of numerous nodes. Each node corresponds to one dimension of input and output data. In this work, each layer is fully connected to the previous and subsequent layers. The four above-mentioned operating parameters are used as inputs, while six above-mentioned performances are selected as the targets for prediction.

Table 1. Detailed parameters of the ANN model

Parameter	ANN model
Number of input layers	4
Number of hidden layer 1	15
Number of hidden layer 2	10
Number of output layers	4
Activation function	tanh
Optimizer	lbfgs
Loss function	MSE
Evaluation indicator	MSE and R <sup>2</sup>

We used the backpropagation (BP) algorithm to optimize the network parameters by gradient descent [38]. To address the nonlinear problem and mitigate the gradient

disappearance issue, we used the tanh function as the model’s activation function. To enhance the gradient descent algorithm’s efficiency, we used the mean square error (MSE) as loss function, which ensures faster convergence and performs well in solving regression tasks. Meanwhile, MSE and coefficient of determination ( $R^2$ ) as evaluation metrics to measure the performance of the ANN model, respectively [39]:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (4)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (5)$$

where  $y_i$  denotes the actual value,  $\hat{y}_i$  denotes the predicted value and  $\bar{y}$  indicates the average of the actual value. Generally, higher  $R^2$  and lower MSE are desirable to enhance the model accuracy [40]. Furthermore, we employed the grid search method to fine-tune the hyperparameters for the ANN algorithm, as it effectively processes intricate models with multiple parameters [41]. After the optimization (see MSE plotted in Fig. S2), we listed the parameters of the ANN model in Table 1.

### 2.3 Methods of significance analysis

The Pearson’s Correlation Coefficient (PCC) was utilized to access the linear dependency between different input variables, which is given by [42,43]:

$$\rho_{xy} = \frac{\Sigma(x_i - x_{mean})(y_i - y_{mean})}{\sqrt{\Sigma(x_i - x_{mean})^2 \Sigma(y_i - y_{mean})^2}} \quad (6)$$

where  $\rho_{xy}$  represents PCC value between the input feature and output target ranging from -1 (negative) to +1 (positive). The  $x_{mean}$  denote the averages of input feature  $x$  and  $y_{mean}$  denotes the averages of output target  $y$ . According to the absolute values of the PCC, we can evaluate the relative importance of four operating parameter for all reactor performance, as shown in Fig. S3.

### 2.4 Description of the RL model

SL models are designed to predict patterns, but they do not involve decision-making process. In contrast, RL models develops the optimal decision policy by interacting with the environment [34]. Since the RL system’s actions will impact its future inputs, it can be treated as a feedback control system based on rewards. The learner, so-called agents developed by RL, is not guided on specific actions, rather it must explore which actions maximize the rewards. A reward signal determines the goal in an RL problem, which is affected by the agent’s current action and the current state of the environment, and in turn, the agent can directly influence on reward, or indirectly influence through changing the environment’s state. Accordingly, a policy is used to guide the agent on the appropriate action to take in those states, to maximize the return (the expected accumulated discounted reward over the course of an episode). It is the job of the

algorithm to automatically figure out how to choose good actions.

As shown in Fig.1, RL typically involves four essential steps to train an agent. More details on the RL agent are described in Fig. S4. Our goal is to design such RL agents that determine the four above-mentioned operating parameters to explore the maximum reaction performance and minimum EC both within and outside the investigated range. In this work, the agent was trained by the Proximal Policy Optimization (PPO) algorithm built on the Actor-Critic (AC) framework (see Fig. S5 and S6 in SI), which has good performance within the continuous data space [44]. By observing the agent's behavior, one can conclude the optimal policy of reaction performance and cost-effectiveness based on its decision process. The detailed parameters of the RL model are listed in Table S4.

### 3. Results

#### 3.1 ANN model evaluation

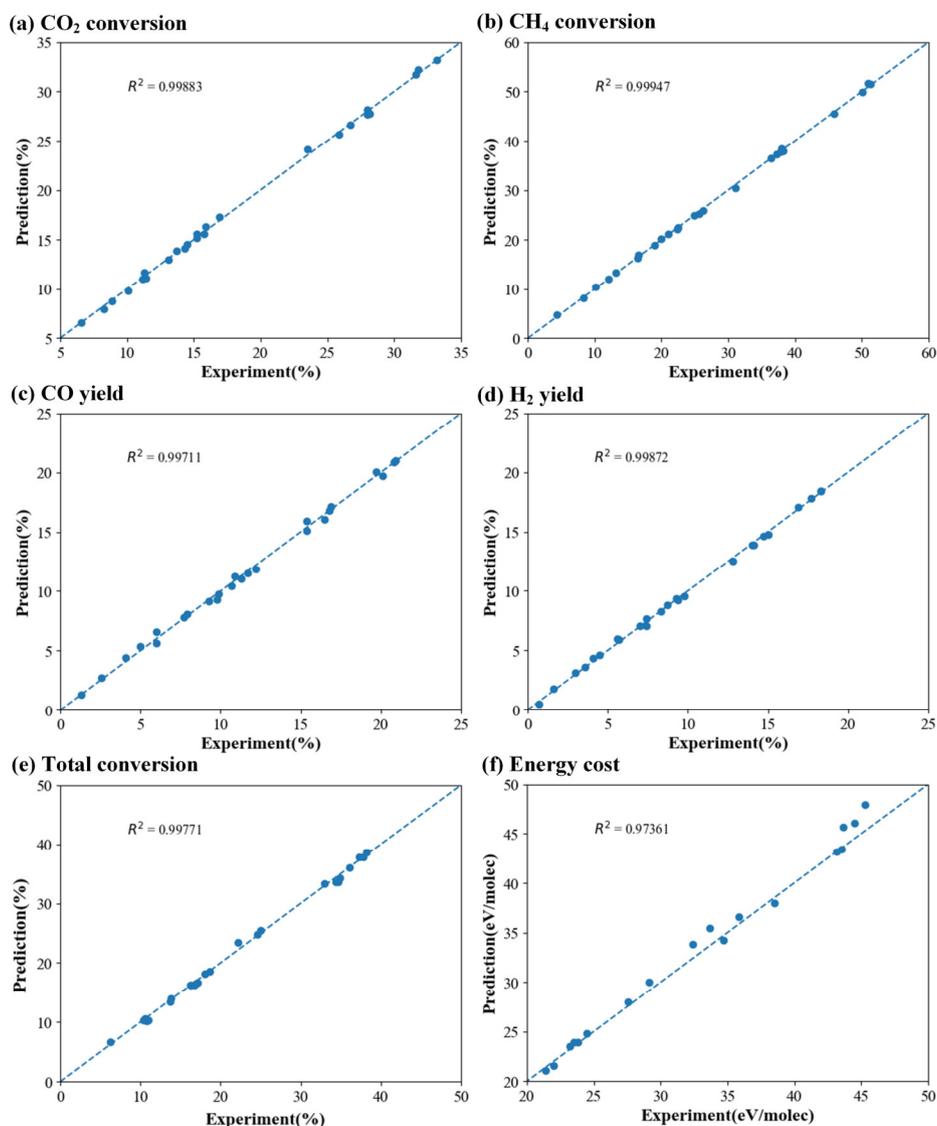


Fig. 2. Predicted data versus experimental results ( $R^2$  plot). (a)  $\text{CO}_2$  conversion; (b)  $\text{CH}_4$  conversion; (c) CO yield; (d)  $\text{H}_2$  yield; (e) total conversion and (f) energy cost.

We assess the ANN model’s performance by comparing its prediction values and actual values for reaction performance and cost-effectiveness, as shown in Fig. 2 and Fig. S7. This ANN prediction model exhibits impressive accuracy through the complete dataset, achieving an  $R^2$  consistently near 1 and an MSE of only 0.000026. This highlights the ANN model has exceptional ability in forecasting the plasma-catalytic DRM process, providing a robust basis for developing the RL model. To further evaluate the model’s generalizability, we conducted additional experiments using new operating parameters within the investigated ranges. As presented in Fig. S8, the model's predictions on unseen data aligned well with the actual data, thereby confirming its reliability.

### 3.2 Significance and trend analysis of parameters

The relative significance of each factor and the trend between input operating parameters and output performance parameters are demonstrated in Table 2. The most important factor is total flow rate for CO<sub>2</sub> conversion (56.6 %), CH<sub>4</sub> conversion (43.5%), CO yield (49.5%) and H<sub>2</sub> yield (51.0%), as well as for total conversion (57.3%). Additionally, the discharge power has a considerable positive influence (>20%) on all these reaction performances. For the EC, the relative significance of total flow rate (43.1%) and discharge power (39.1%) are very close to each other. The CO<sub>2</sub>/CH<sub>4</sub> ratio and discharge power exerts similar influences on gas conversion and product yield. It should be noted that Ni loading has minimal influence on overall process performance indicators (less than 7%), indicating that the plasma characteristic plays a more critical role.

Table 2. Relative significance of different input parameters on various output parameters

Output Parameter	1 <sup>st</sup> IF	2 <sup>nd</sup> IF	3 <sup>rd</sup> IF	4 <sup>th</sup> IF
CO <sub>2</sub> conversion	Flow (56.6%) (-)	Power (22.8%) (+)	Ratio (17.2%) (-)	Loading (3.3%) (↗ ↘)
CH <sub>4</sub> conversion	Flow (43.5%) (-)	Ratio (26.6%) (+)	Power (23.5%) (+)	Loading (6.4%) (↗ ↘)
CO yield	Flow (49.5%) (-)	Ratio (25.6%) (+)	Power (21.0%) (+)	Loading (3.9%) (↗ ↘)
H <sub>2</sub> yield	Flow (51.0%) (-)	Ratio (23.5%) (+)	Power (21.6%) (+)	Loading (3.8%) (↗ ↘)
Total conversion	Flow (57.3%) (-)	Power (27.9%) (+)	Ratio (9.1%) (↗ ↘)	Loading (5.8%) (↗ ↘)
Energy cost	Flow (43.1%) (↘ ↗)	Power (39.1%) (+)	Ratio (14.6%) (-)	Loading (3.2%) (↘ ↗)

List of the abbreviations included in the table: Important factor (IF), Total flow rate (Flow), Discharge power (Power) and CO<sub>2</sub>/CH<sub>4</sub> ratio (Ratio). Positive factor and negative factor are represented by (+) and (-), respectively. Some output parameters first increased and then decreased with rising input parameters, indicated as (↗ ↘), while other output parameters first decreased and then increased with rising parameter, indicated as (↘ ↗).

### 3.3 RL model evaluation

The investigated range of the four input parameters is as follows: Ni loading (5-15 wt%), CO<sub>2</sub>/CH<sub>4</sub> ratio (0.5-1.5), discharge power (20-60 W) and total flow rate (25-125 mL/min). It would be interesting to first discover the theoretically (or potentially) maximum performance RL agents could reach without any physical limitation of input parameters (i.e. outside the investigated range). We will now establish the RL models, based on input parameters first outside and then within the investigated range.

### 3.3.1 Input parameters outside the investigated range

Fig. 3 shows the testing curve of the RL models of total conversion and EC with its corresponding actions, respectively. The total conversion (Fig. 3 (a) and 3 (b)) can reach a maximum of 55 %, while the agent first optimized the total flow rate until it converges to a negative value. Next, the agent optimized the discharge power, followed by the CO<sub>2</sub>/CH<sub>4</sub> ratio, which converge to positive values. The Ni loading parameter also converges at a negative value, but it has low impact (< 7%) on the outputs compared with the other input parameters. Fig. 3 (c) shows that the minimum value of EC can reach almost 12 eV/molec, while the agent first optimized the discharge power to converge at a negative value, followed by the CO<sub>2</sub>/CH<sub>4</sub> ratio. Figure 3 (d) shows the Ni loading and total flow rate already had their optimal value (i.e., not converged to the boundary of the operating range), while the CO<sub>2</sub>/CH<sub>4</sub> ratio reached 1.9, which exceeds the current operating range (0.5-1.5). We notice that the actions on Ni loading and total flow rate are in line with the trend on EC shown in Table 2. Their trends are first decreasing and then increasing, so the actions of the agent fluctuated to find its optimal value. Since the EC decreases with increasing CO<sub>2</sub>/CH<sub>4</sub> ratio, the action on CO<sub>2</sub>/CH<sub>4</sub> ratio increases and even exceeds the current operating range to obtain the minimum EC. It should be noted that the actions of total flow rate for the total conversion and of discharge power for the EC are not physical, as their value should be higher than 0 in reality.

We can also see that the discharge power converges to positive and negative values for total conversion and EC, respectively. Even though the input parameters are not restricted within their boundary, the positively and negatively correlated parameters reach their boundary value, as the dataset in the training process has upper and lower bounds as the time step progresses.

In real application of plasma-catalytic DRM, both total conversion and EC should be considered, and they are often affected in different ways by the input parameters (cf. also Table 2 above), so the best trade-offs should be determined. We simulate three cases, setting different weights between total conversion and EC, i.e., 1 (total conversion and EC are equally important), 2 (total conversion is twice as important), and 0.5 (EC is twice as important). The weighted superposition of total conversion and EC with its corresponding actions is shown in Fig. 4. The CO<sub>2</sub>/CH<sub>4</sub> ratio is still proportional to the weighted superposition of total conversion and EC, since it monotonically converges to a positive value, even changing the weight (Fig. 4 (b), (d) and (f)). A similar pattern can also be seen for the discharge power in Fig. 4 (b) and (d). However, when the weight is 0.5, the power shows inverse effect on the weighted superposition of total conversion and EC. Moreover, the actions of flow rate and Ni

loading present little difference, which means changing the flow rate and Ni loading is not effective for both total conversion and EC optimization.

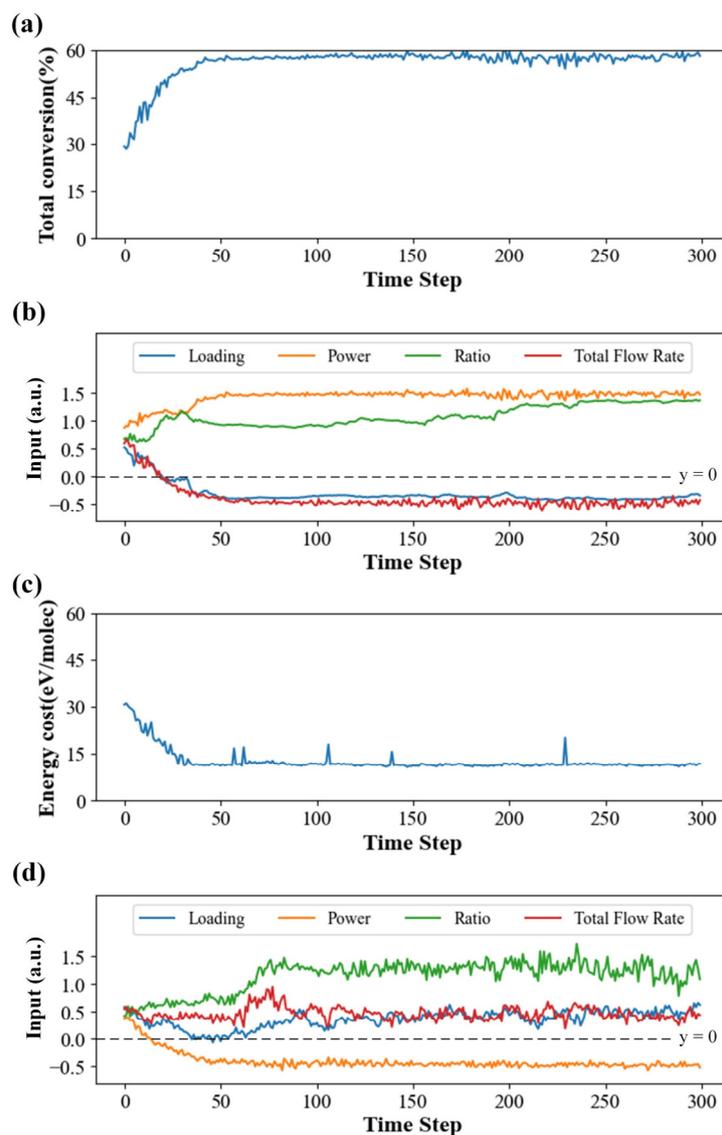


Fig. 3. Testing curve of the RL models of (a) total conversion and (c) energy cost, by plotting them as a function of time step, as well as the corresponding actions (b) and (d), outside the investigated range. The y-axis representing the input parameters (for (b) and (d)) shows the normalized values.

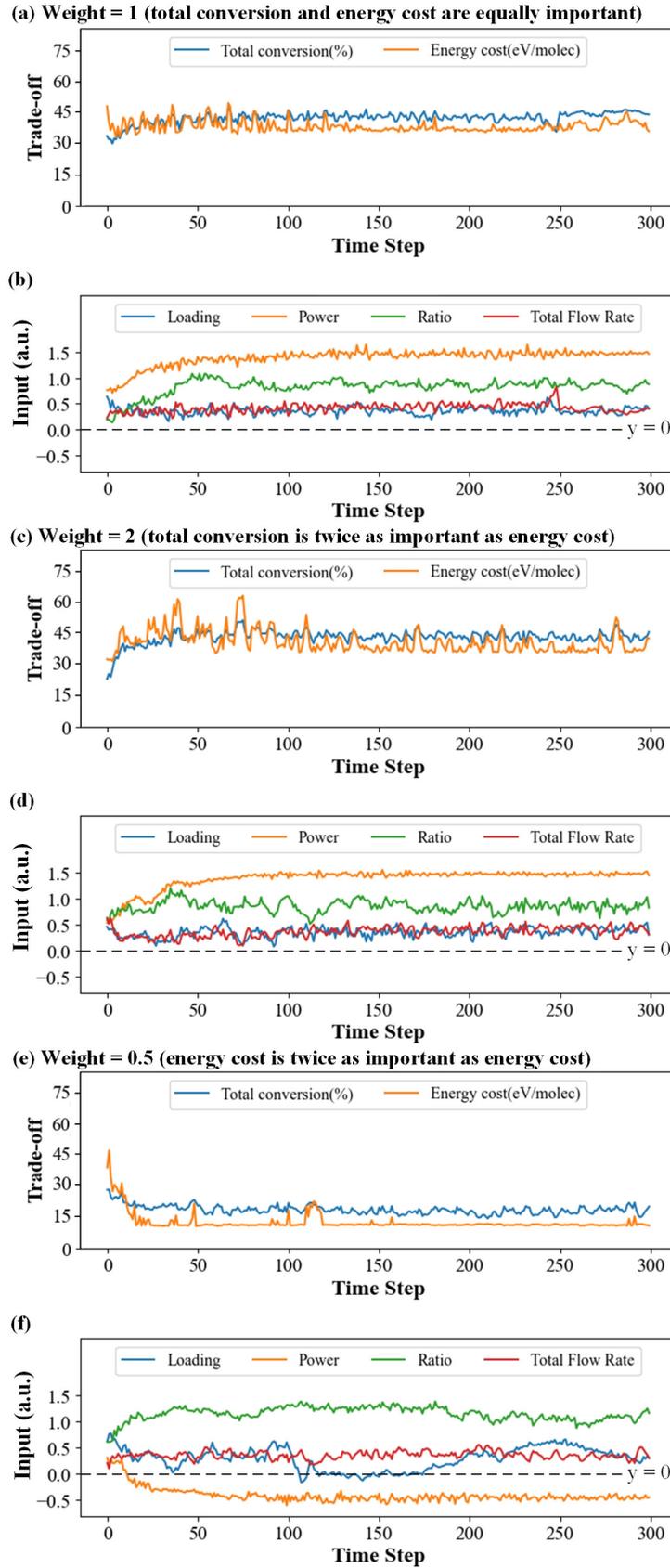


Fig. 4. Superposition of total conversion and energy cost, showing the best trade-offs (a,c,e), as well as the corresponding actions of the four input parameters outside the investigated range (b,d,f), for different weights, i.e., weight = 1: (a) and (b), weight = 2: (c) and (d), weight = 0.5: (e) and (f).

### 3.3.2 Input parameters within the investigated range

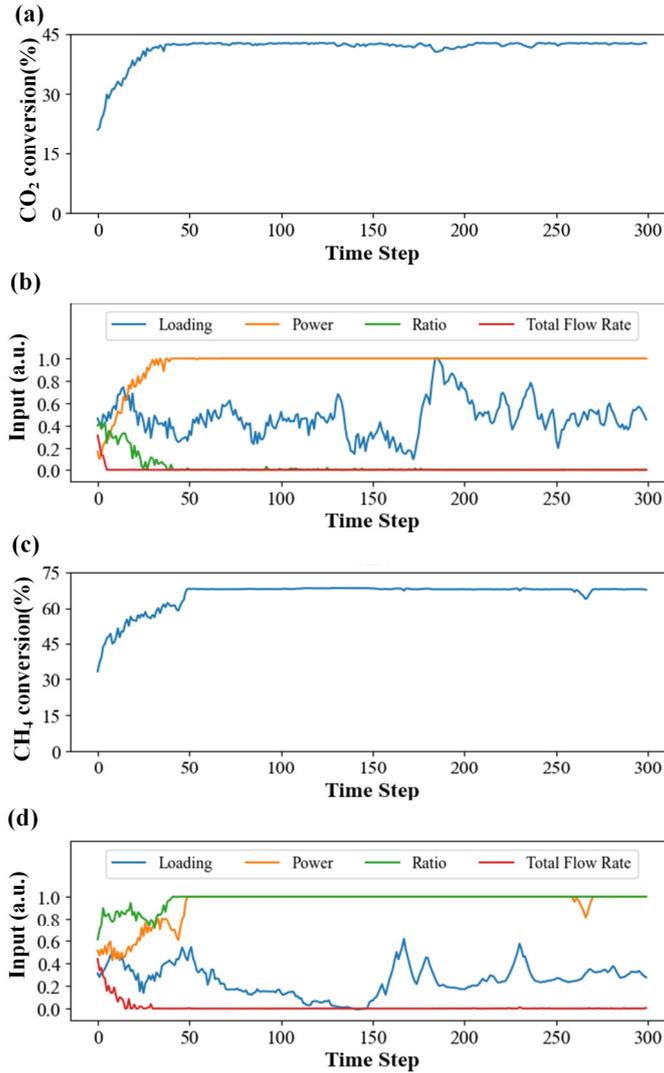


Fig. 5. Testing curve of the RL models of (a) CO<sub>2</sub> conversion and (c) CH<sub>4</sub> conversion, by plotting them against time step, as well as the corresponding actions (b) and (d), within the investigated range. The y-axis representing the input parameters (for (b) and (d)) shows the normalized values.

To avoid the actions of agents reaching non-physical conditions, we now limit the input parameters within the investigated range of the training dataset (the reason will be further discussed in section 4.1). Fig. 5 shows the testing curve of the RL models of CO<sub>2</sub> conversion and CH<sub>4</sub> conversion, including their corresponding actions within the range of input parameters. Fig. 5 (a) and Fig. 5 (b) present that the CO<sub>2</sub> conversion can reach its maximum value of 42% when the total flow rate first reaches its lower boundary (i.e. 25 mL/min), and then the discharge power reaches its upper boundary (i.e. 60 W), followed by the CO<sub>2</sub>/CH<sub>4</sub> ratio reaching its lower boundary (i.e. 0.5). Nevertheless, the CH<sub>4</sub> conversion can reach its maximum value of 68 % when the total flow rate first reaches its lower boundary (i.e. 25 mL/min), followed by the CO<sub>2</sub>/CH<sub>4</sub> ratio and discharge power reaching their upper boundary (i.e. 1.5 and 60 W). Furthermore, the actions of discharge power coincide very closely with the trajectory

of  $\text{CO}_2$  conversion and  $\text{CH}_4$  conversion, indicating a fine-tuning step. As shown from the relative significance in Table 2, the most significant factor is the total flow rate for gas conversion, so the agent first optimized the total flow rate. While the discharge power is the second important factor for the  $\text{CO}_2$  conversion, the agent optimized it as the second parameter. Moreover, the Ni loading keeps fluctuating within its range to seek its optimal value for both the  $\text{CO}_2$  and  $\text{CH}_4$  conversion, in line with the trend in Table 2 above, because of its low impact.

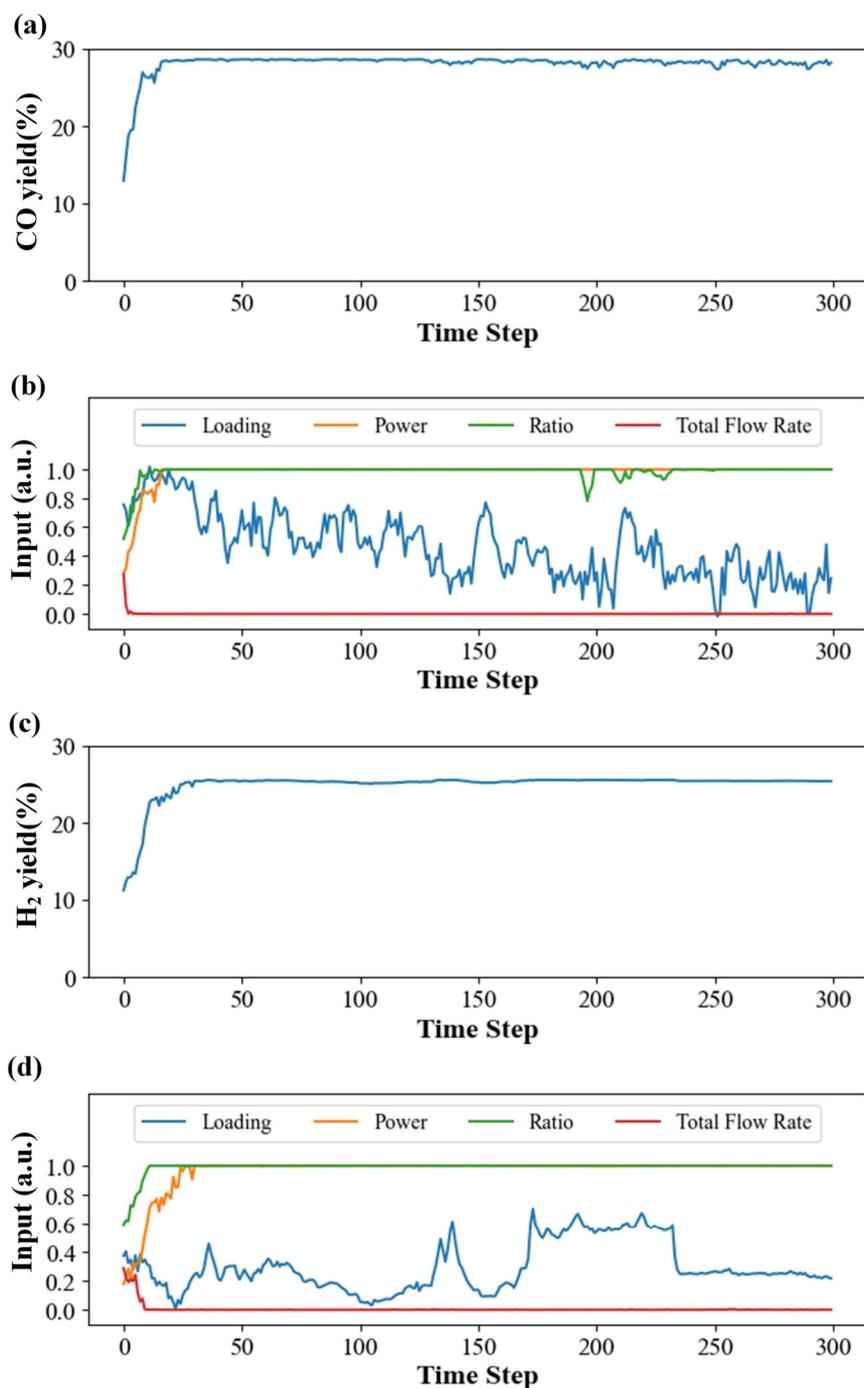


Fig. 6. Testing curve of the RL models of (a)  $\text{CO}$  yield and (c)  $\text{H}_2$  yield, by plotting them as a function of time step, as well as the corresponding actions (b) and (d), within the investigated range. The y-axis representing the input parameters (for (b) and (d)) shows the normalized values.

A similar regulation policy can be found in Fig. 6, which presents the time step-dependence of the CO yield and H<sub>2</sub> yield, with the corresponding actions, within the range of input parameters. Fig. 6 (a) and Fig. 6 (c) show that the CO yield can reach its maximum value of 28 % and the H<sub>2</sub> yield can reach its maximum value of 26 % when the total flow rate first reaches its lower boundary (i.e. 25 mL/min), followed by the CO<sub>2</sub>/CH<sub>4</sub> ratio and the discharge power reaching their upper boundary (i.e. 1.5 and 60 W) (Fig. 6 (b) and Fig. 6 (d)), respectively. On the other hand, the Ni loading keeps fluctuating within its range to seek its optimal value for both the CO and H<sub>2</sub> yield, in line with the trend in Table 2 above, again because of its low impact.

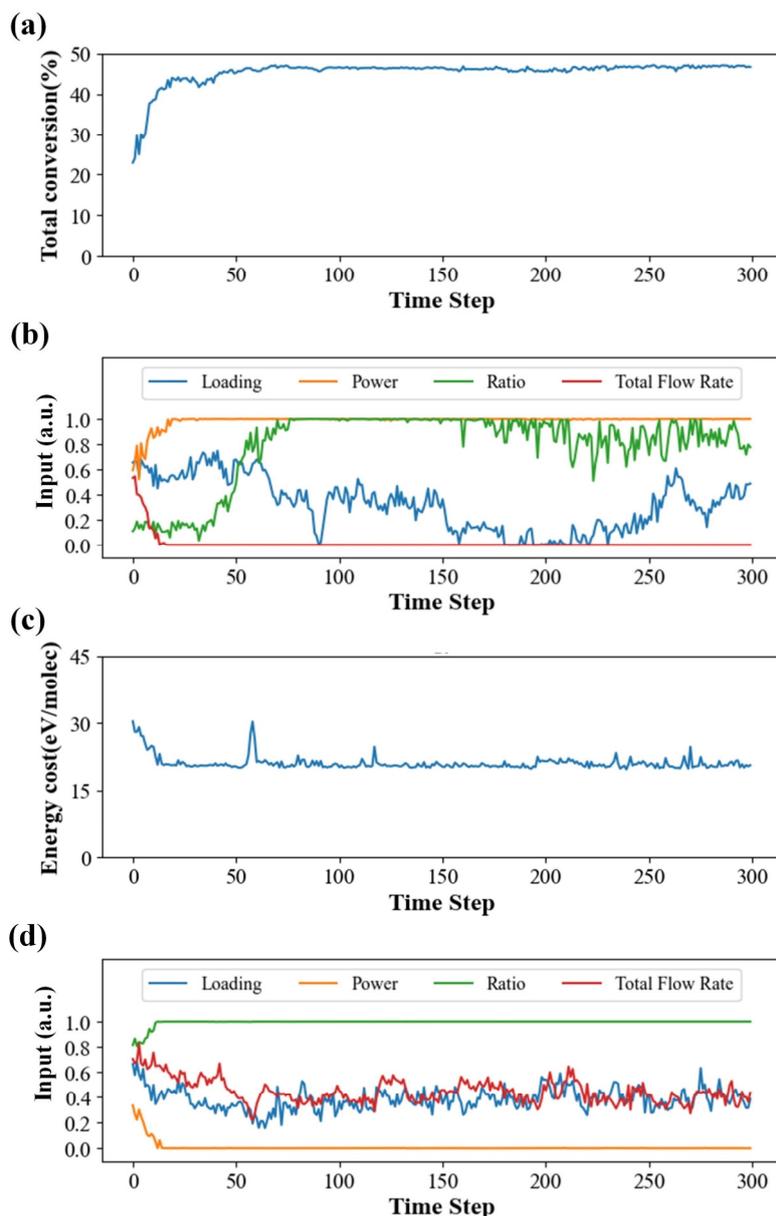


Fig. 7. Testing curve of the RL models of (a) total conversion and (c) energy cost, by plotting them as a function of time step, as well as the corresponding actions (b) and (d), within the investigated range. The y-axis representing the input parameters (for (b) and (d)) shows the normalized values.

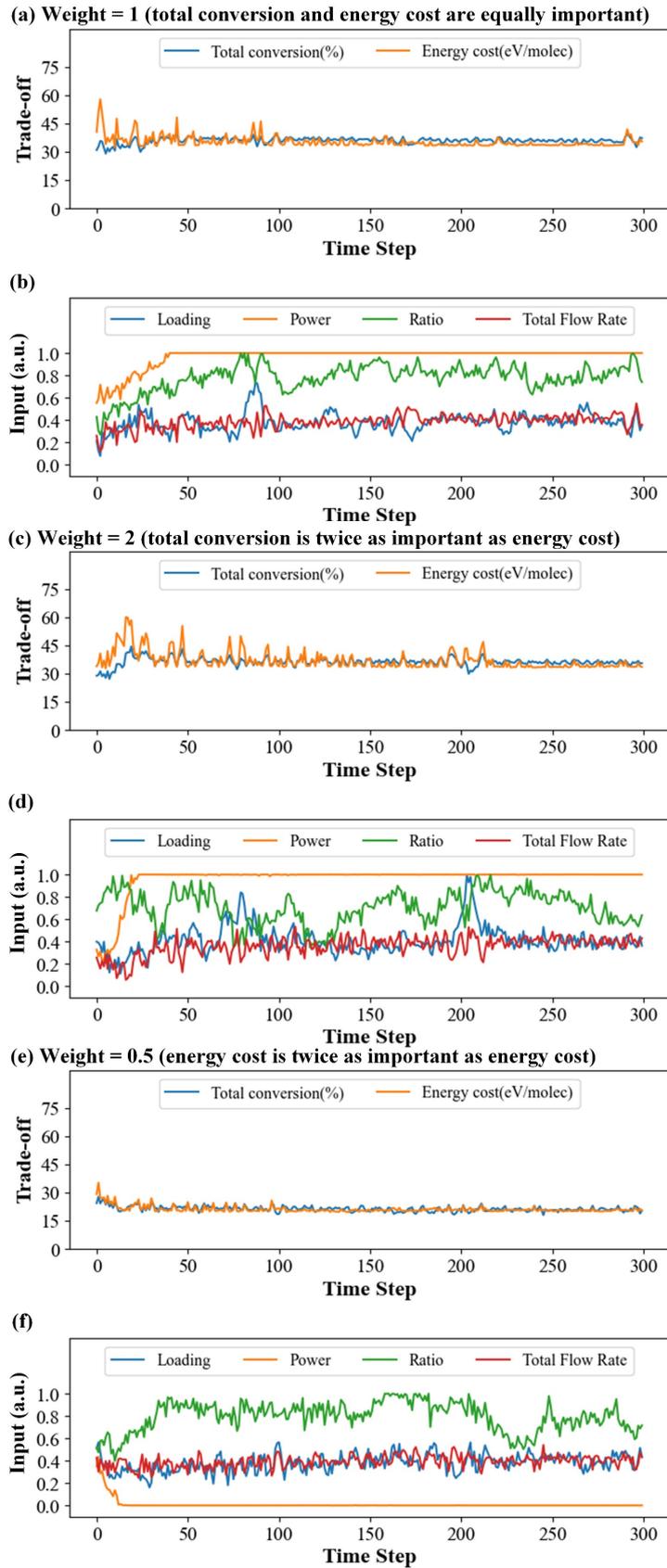


Fig. 8. Superposition of total conversion and energy cost, showing the best trade-offs (a,c,e), as well as the corresponding actions of the four input parameters within the investigated range (b,d,f), for different weights, i.e., weight = 1: (a) and (b), weight = 2: (c) and (d), weight = 0.5: (e) and (f).

The testing curve of the RL models for the total conversion and EC, with the corresponding actions within the investigated range, is presented in Fig. 7. From Fig. 7 (a) and Fig. 7 (b), we can see that the total conversion can reach its maximum value of 47% when the total flow rate first reaches its lower boundary (i.e. 25 mL/min), followed by the discharge power reaching its upper boundary (i.e. 60 W), while the Ni loading and CO<sub>2</sub>/CH<sub>4</sub> ratio have their optimal value at around 7.5 wt% and 1.3, respectively. The EC can reach its minimum value of 21 eV/molec when the power first reaches its lower boundary (i.e. 20 W), followed by the CO<sub>2</sub>/CH<sub>4</sub> ratio reaching its upper boundary (i.e. 1.5). It is worth to point out that the discharge power yields a trade-off between total conversion and EC. Notably, eq. 3 suggests there is an inversely linear dependence between the total flow rate and the EC, but the total flow rate does not reach the upper boundary. The reason will be further clarified in section 4.2.

In addition to separately investigating the total conversion and EC, we need a comprehensive understanding of the effect of all operating parameters for optimizing both performance metrics of the plasma-catalytic DRM process. Similar as in section 3.3.1, we present in Fig. 8 the superposition of total conversion and EC for different weights, with the corresponding actions. When the total conversion is more or equally important than the EC, the discharge power is a positive factor and is near-linearly proportional to the weighted superposition. Especially when the total conversion is equally important to the EC, the total conversion reaches 36%, while the EC reaches 34 eV/molec, when the discharge power reaches its upper boundary (i.e. 60W), the Ni loading is 9.5% and total flow rate is 74 mL/min. When the EC is more important than the total conversion, the power is a negative factor and is near-linearly but inversely proportional to the weighted superposition. Furthermore, the CO<sub>2</sub>/CH<sub>4</sub> ratio has a positive correlation with the weighted superposition and fluctuates within the range to seek its optimal value, while the total flow rate and Ni loading seems at their fixed value.

## 4. Discussion

### 4.1 Effect of investigated range on the RL model

The investigated range of the training dataset has a noticeable effect on the RL model's performance. By comparing the two RL models of total conversion and energy cost outside and within the range in Fig. 3 and Fig. 7 (b and d), the total flow rate and discharge power converge to negative values and reach their lower boundary for total conversion and energy cost, respectively. We notice that the actions of agents reach negative values (i.e., -50 mL/min of total flow rate and -20 W of power), which is not in line with scientific knowledge, and impossible to implement as their value should be higher than 0 in reality. This phenomenon can be explained as follows: the total flow rate and discharge power are the negative factors and exhibit a near-linear decrease with total conversion and EC within the investigated range. The RL agent is trained on the current dataset and learned such near-linear patterns within the investigated range. However, it is not trained by the dataset outside the investigated range. When this very same model is subjected to predictions outside the investigated range, the patterns learned within the investigated range are not effective, which induces an inaccurate RL model. Thus, the agent decreases the negative factor as much as possible, even to a

negative value, to obtain the maximum return. This result demonstrates that the RL model can only be effective on patterns prediction and policy optimization within the investigated range.

A similar optimization policy can also be applied to the positive factors. In Fig. 5 and Fig. 6 (b and d), the power is the positive factor for the gas conversion and product yield, and the actions of agent are to reach their upper boundary, i.e. at the maximum value of the investigated range. Although the actions of agent are not to reach negative value this time, the maximum value of input parameters depends on the capability of the equipment (e.g. the power supply). Specifically, the RL model might predict that a discharge power of 1 MW is super-efficient for the conversion, but it is impossible to implement. In addition, the safety boundary to implement in experiments should also be considered (e.g. explosion limit when considering CO<sub>2</sub>/CH<sub>4</sub>/O<sub>2</sub> gas mixture for future study). This means that the RL model can only optimize the input parameters within the range of experimental capability.

## 4.2 Optimal policy developed by the RL model

Table 3. Regulation sequence of various input parameters within their range

Objectives	1 <sup>st</sup> order	2 <sup>nd</sup> order	3 <sup>rd</sup> order	4 <sup>th</sup> order
CO <sub>2</sub> conversion ↑	Flow (↓)	Power (↑)	Ratio (↓)	Loading (optimal)
CH <sub>4</sub> conversion ↑	Flow (↓)	Ratio (↑)	Power (↑)	Loading (optimal)
CO yield ↑	Flow (↓)	Ratio (↑)	Power (↑)	Loading (optimal)
H <sub>2</sub> yield ↑	Flow (↓)	Ratio (↑)	Power (↑)	Loading (optimal)
Total conversion ↑	Flow (↓)	Power (↑)	Ratio (optimal); Loading (optimal)	
Energy cost ↓	Power (↓)	Ratio (↑)	Flow (optimal); Loading (optimal)	
Total conversion + energy cost (1+1) ↑	Power (↑)	Ratio (optimal)	Flow (near stable)	Loading (near stable)

List of the abbreviations included in the table: Total flow rate (Flow), Discharge power (Power), CO<sub>2</sub>/CH<sub>4</sub> ratio (Ratio) and Ni loading (Loading). When the factors reach their upper and lower bounds, it is represented by (↑) and (↓), respectively. The expected objective is to maximize the reaction performance and minimize the energy cost, which is represented by (↑) and (↓), respectively.

Comparing the relative significance of the input parameters in Table 2, we can summarize two regulation policies in Table 3: 1) the agent’s actions on the input parameters are in line with the concluded trend toward the output parameters. To maximize the reaction performance, the positive factors should reach upper bounds and the negative factors should reach lower bounds. When the factors first increase (decrease) and then decrease (increase), they would have an optimal value. To minimize the EC, the power should reach lower bounds and the CO<sub>2</sub>/CH<sub>4</sub> ratio should reach upper bounds, since the EC decreases with decreasing power and increasing CO<sub>2</sub>/CH<sub>4</sub> ratio. 2) To build the bridge between the agent’s actions on the operating parameters and the reaction performance within the investigated range, the optimal sequence can be concluded as follows: the agent ideally starts with a coarse tuning of the more influential parameters, followed by fine-tuning of the less important parameters, to obtain the maximum rewards. In other words, the regulation sequence exhibits good agreement with the predicted significance results. This indicates that the RL agent

mimics human experience in the decision-making process. However, the optimization sequence on EC (alone or in combination with total conversion) seems different from the summarized policy, due to the limited training dataset. The most important factor, i.e., total flow rate, which is inversely proportional to energy cost, is not effective, as it is fixed at 75 mL/min in the training dataset near the optimal value for energy cost. If more data could be used for training with different flow rates, the agent can learn a more comprehensive regulation in the current dataset, which is interesting for further investigation.

### **4.3 Comparison between SL model and RL model**

To make a clear comparison between the SL and RL models in terms of performance and application, we list here the advantages and disadvantages of these two approaches:

- 1) Training efficiency: Training a SL model can be relatively efficient, especially with large datasets and powerful computing resources, while training a RL model is generally less efficient, due to the need for exploration and the potentially long sequences of actions required to learn an optimal policy.
- 2) Interactivity and adaptability: SL models are generally static, meaning they do not adapt to changes in the environment after training, while RL models are inherently interactive and adaptive, as they learn from the consequences of their actions, so they are well-suited for dynamic environments.

In conclusion, SL models are suitable for tasks that require high precision and accuracy in predictions from static data, while RL models are more appropriate for tasks that involve complex sequences of actions and interactions with a dynamic environment. In many real-world scenarios, a hybrid approach, combining elements of both SL and RL, may offer the best performance and adaptability. Hence, it is interesting to compare the predictions of the SL model (static) with those of the RL model (dynamic) when combining SL and RL models in this work.

To evaluate the RL model's effectiveness, we firstly compare the minimum value of the EC obtained by our RL model with the real experimental dataset, considering that the EC is the most important factor for industrial-scale DRM [45]. According to our RL model, the EC can reach values as low as 21 eV/molec when the discharge power is 20 W and the CO<sub>2</sub>/CH<sub>4</sub> ratio is 1.5. This outcome aligns with prior work on the optimal energy yield (EY), which was predicted at same conditions by SL model, and also consistent with the experimental result on EC, reaching 21.3 eV/molec at the same conditions [32].

Similarly, our RL model reveals that the total conversion reaches 36%, while the EC reaches 34 eV/molec, when the discharge power reaches its upper boundary (i.e. 60 W), the Ni loading is 9.5 wt% and the total flow rate is 74 mL/min. This result is also validated by the experimental dataset on total conversion of 34.0 % and EC of 35.2 eV/molec, at a near same conditions [32]. Therefore, the final results of the actions by the agent and the desired outcome (both maximum total conversion and minimum EC) are in line with the real experimental dataset.

### **4.4 DBD reactor used for DRM**

The Ni loading, the only non-linear factor among these input parameters, has an optimal value at approximately 7.5 % for the performance of plasma-catalytic DRM. This is because the catalyst's specific surface area is larger at lower Ni loading [46], but it can be reduced at higher Ni loading, which in turn diminishes the performance [47], as indicated by studies based on similar preparation method in literature [46,48]. Despite its significant effect, Ni loading is not the dominant factor on the performance, indicated by the relative significance analysis in our work.

Besides Ni loading, the other three factors have much impact on the performance, particularly affecting the reaction kinetics, thermodynamics, and mechanisms, as revealed from literature [7]. As indicated in Fig. 5 to Fig. 7, the reaction performance and EC exhibit a near-linear increase with discharge power. In general, higher discharge power leads to more micro-discharges, thereby producing more reaction channels and reactive species, which play an active role in both catalytic surface and gas-phase reactions to boost the CO and H<sub>2</sub> yield [49]. On the contrary, only the EC (alone or in combination with total conversion when the EC weights more than total conversion) hope to reach lower discharge power as the conversion rises less than linearly with rising power, as can be deduced from eq. (3) above [50].

The CO<sub>2</sub>/CH<sub>4</sub> ratio positively influences the (absolute) CH<sub>4</sub> conversion and product yield, but it reduces the CO<sub>2</sub> conversion because there are less CH<sub>x</sub> radicals that can contribute to the CO<sub>2</sub> conversion, as indicated in Table 2 [50,51]. With higher CO<sub>2</sub> contents, more oxygen atoms generated by CO<sub>2</sub> dissociation can efficiently react with the H atoms produced from CH<sub>4</sub> dissociation through electron impact, thus limiting the backward reaction, CH<sub>3</sub> + H → CH<sub>4</sub>, enabling higher CH<sub>4</sub> conversion [51]. In the meanwhile, the total conversion reaches its maximum for a CO<sub>2</sub>/CH<sub>4</sub> ratio of around 1.3. Indeed, the CH<sub>4</sub> conversion is typically higher than the CO<sub>2</sub> conversion (cf. Fig. 5 (a) and (c)), suggesting that a CO<sub>2</sub>/CH<sub>4</sub> ratio below 1 would be beneficial for the total conversion, but on the other hand, prior works have also demonstrated that CO<sub>2</sub>/CH<sub>4</sub> ratios below 1 lead to carbon deposition and catalyst deactivation [52]. By integrating these findings with our model predictions, it is suggested to set a CO<sub>2</sub>/CH<sub>4</sub> ratio between 1 and 1.5 for plasma-catalytic DRM applications and it also corresponds with the common composition of biogas [53].

In our work, the EC can reach a minimum value of 12 eV/molec without any limitation (i.e., when we vary the parameters outside the investigated range) and 21 eV/molec when the parameters are kept within the investigated range. These values are however much higher than the EC target, i.e. 4.3 eV/molecule defined by Snoeckx and Bogaerts [7], to be contentious with traditional DRM, and other promising technologies. The same conclusion about too high energy cost for DBD plasmas was also made in [7], and is generally found in literature (see details in [6]). It should be noted that the task of RL is to aid scientists to optimize the energy cost based on the patterns learned within the investigated range, but they cannot lead to a breakthrough of the physical limitation in DBD reactors. Therefore, we should consider other plasma reactors to enhance the cost-effectiveness of DRM, i.e., warm plasmas like gliding arc, microwave, spark discharge and atmospheric pressure glow discharge (APGD). Indeed, these plasmas can combine energy costs below this target of 4.3 eV/molec with relatively

high conversions. For instance, a recent paper from PLASMANT shows that a confined APGD (cAPGD) reactor can reach an EC of 3.5 eV/molec, for a total conversion of 62% (and at other conditions, an EC of 4.2 eV/molec, for a total conversion of 74%) [13], while O<sub>2</sub> addition could even reach a lower EC of 1.98 eV/molec for a total conversion of 67% [54]. These values are clearly lower than the EC target of 4.3 eV/molecul, showing the superior performance of an APGD, and other warm plasmas (yielding similar performance) for DRM. On the other hand, although a DBD reactor cannot reach such excellent performance, it is the most suitable reactor to apply ML, due to the much larger dataset available, because of the much more extensive literature. That is the reason why we applied our ML model to the DBD results. In the future, we also plan to investigate the performance of ML for the other plasma reactor types.

#### 4.5 ML applied to plasma catalysis

ML is crucial in advancing plasma applications that require an accurate description and control of complex plasma-surface interactions, e.g., plasma catalysis [10]. While our data-driven ML model shines at prediction, it cannot explain the fundamental reaction mechanisms, where plasma chemical kinetics models is required, e.g., in Ref. [11] and the specific role of the catalyst for the synergy effect was not considered. Moreover, the results of ML models are mostly validated based on their effectiveness by the experimental dataset and they cannot optimize the reactor performance exceeding its potential [55]. This suggests that the interpretability and potential of ML model can be limited, emphasizing the necessity of prior knowledge in evaluating ML results.

Nevertheless, our ML model demonstrates significant potential for predicting and optimizing other chemical processes, for instance, thermal catalytic DRM. Since catalyst compositions and reactor operation in thermal catalysis have been well-explored through various modeling approaches, our ML methods offer beneficial insights from existing experimental data, achieving enhanced reliability and consistency, to further optimize industrial-scale DRM process.

### 5. Conclusion

We developed a ML (SL and RL) model to both predict and optimize the plasma-catalytic DRM process, respectively. The SL model utilizes a typical ANN algorithm for reaction performance and cost-effectiveness prediction with a strong connection to the experimental data, indicated by R<sup>2</sup> values close to 1 for all output. However, we go one step further, by also developing a RL model for process optimization. For optimal regulation policy on single output revealed by our RL model, the agent starts with a coarse tuning of the more influential parameters, followed by fine-tuning of the less important parameters. Furthermore, our RL model effectively pinpointed the ideal conditions for achieving the lowest energy cost of 21 eV/molec at lowest bond of discharge power (i.e., 20 W) but highest bond of CO<sub>2</sub>/CH<sub>4</sub> ratio (i.e., 1.5), which is in line with our SL prediction and the experimental dataset. For the combined optimization of total conversion and energy cost, our RL model reveals that the discharge power yields a trade-off between both performance metrics. However, an optimal discharge power of 60 W, Ni loading of 9.5 wt% and total flow rate of 74 mL/min resulted in both

maximum total conversion and minimum energy cost. Overall, our ML model excels at deriving new insights to facilitate the optimization of intricate nonlinear and dynamic systems, like in plasma-based gas conversion process.

### **CRedit authorship contribution statement**

Jiayin Li: Conceptualization, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing review & editing. Jing Xu: Formal analysis, Investigation, Writing – review & editing. Evgeny Rebrov: Writing – review & editing. Annemie Bogaerts: Funding acquisition, Project administration, Supervision, Writing – review & editing.

### **Declaration of competing interest**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### **Acknowledgments**

This project received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no.810182-SCOPE ERC Synergy project), and the University of Antwerp BOF-FWO postdoc seal of excellence mandate (grant agreement no.41/FA070200/FFB240228).

### **References**

- [1] D. Pakhare, J. Spivey, A review of dry (CO<sub>2</sub>) reforming of methane over noble metal catalysts, *Chem. Soc. Rev.* 43 (2014) 7813–7837. <https://doi.org/10.1039/C3CS60395D>.
- [2] Y. Kathiraser, U. Oemar, E.T. Saw, Z. Li, S. Kawi, Kinetic and mechanistic aspects for CO<sub>2</sub> reforming of methane over Ni based catalysts, *Chem. Eng. J.* 278 (2015) 62–78. <https://doi.org/10.1016/j.cej.2014.11.143>.
- [3] H. Zhao, R. Yu, S. Ma, K. Xu, Y. Chen, K. Jiang, Y. Fang, C. Zhu, X. Liu, Y. Tang, L. Wu, Y. Wu, Q. Jiang, P. He, Z. Liu, L. Tan, The role of Cu<sub>1</sub>–O<sub>3</sub> species in single-atom Cu/ZrO<sub>2</sub> catalyst for CO<sub>2</sub> hydrogenation, *Nat. Catal.* 5 (2022) 818–831. <https://doi.org/10.1038/s41929-022-00840-0>.
- [4] Y. Song, E. Ozdemir, S. Ramesh, A. Adishev, S. Subramanian, A. Harale, M. Albuali, B.A. Fadhel, A. Jamal, D. Moon, S.H. Choi, C.T. Yavuz, Dry reforming of methane by stable Ni–Mo nanocatalysts on single-crystalline MgO, *Science* 367 (2020) 777–781. <https://doi.org/10.1126/science.aav2412>.
- [5] W. Lu, Q. Cao, B. Xu, H. Adidharma, K. Gasem, M. Argyle, F. Zhang, Y. Zhang, M. Fan, A new approach of reduction of carbon dioxide emission and optimal use of carbon and hydrogen content for the desired syngas production from coal, *J. Clean. Prod.* 265 (2020) 121786. <https://doi.org/10.1016/j.jclepro.2020.121786>.
- [6] A. George, B. Shen, M. Craven, Y. Wang, D. Kang, C. Wu, X. Tu, A Review of Non-Thermal Plasma Technology: A novel solution for CO<sub>2</sub> conversion and utilization, *Renew. Sustain. Energy Rev.* 135 (2021) 109702. <https://doi.org/10.1016/j.rser.2020.109702>.
- [7] R. Snoeckx, A. Bogaerts, Plasma technology – a novel solution for CO<sub>2</sub> conversion?, *Chem. Soc. Rev.* 46 (2017) 5805–5863. <https://doi.org/10.1039/C6CS00066E>.

- [8] D.B. Nguyen, S. Saud, Q.T. Trinh, H. An, N.-T. Nguyen, Q.H. Trinh, H.T. Do, Y.S. Mok, W.G. Lee, Generation of Multiple Jet Capillaries in Advanced Dielectric Barrier Discharge for Large-Scale Plasma Jets, *Plasma Chem. Plasma P.* 43 (2023) 1475–1488. <https://doi.org/10.1007/s11090-023-10404-0>.
- [9] J. Osorio-Tejada, M. Escriba-Gelonch, R. Vertongen, A. Bogaerts, V. Hessel, CO<sub>2</sub> conversion to CO via plasma and electrolysis: a techno-economic and energy cost analysis, *Energy Environ. Sci.* 17 (2024) 5833–5853. <https://doi.org/10.1039/D4EE00164H>.
- [10] A. Bogaerts, X. Tu, J.C. Whitehead, G. Centi, L. Lefferts, O. Guaitella, F. Azzolina-Jury, H.-H. Kim, A.B. Murphy, W.F. Schneider, T. Nozaki, J.C. Hicks, A. Rousseau, F. Thevenet, A. Khacef, M. Carreon, The 2020 plasma catalysis roadmap, *J. Phys. D: Appl. Phys.* 53 (2020) 443001. <https://doi.org/10.1088/1361-6463/ab9048>.
- [11] B. Loenders, R. Michiels, A. Bogaerts, Is a catalyst always beneficial in plasma catalysis? Insights from the many physical and chemical interactions, *J. Energy Chem.* 85 (2023) 501–533. <https://doi.org/10.1016/j.jechem.2023.06.016>.
- [12] A. Bogaerts, E.C. Neyts, Plasma Technology: An Emerging Technology for Energy Storage, *ACS Energy Lett.* 3 (2018) 1013–1027. <https://doi.org/10.1021/acsenerylett.8b00184>.
- [13] B. Wanten, S. Maerivoet, C. Vantomme, J. Slaets, G. Trenchev, A. Bogaerts, Dry reforming of methane in an atmospheric pressure glow discharge: Confining the plasma to expand the performance, *J. CO<sub>2</sub> Util.* 56 (2022) 101869. <https://doi.org/10.1016/j.jcou.2021.101869>.
- [14] A. Aziznia, H.R. Bozorgzadeh, N. Seyed-Matin, M. Baghalha, A. Mohamadlizeh, Comparison of dry reforming of methane in low temperature hybrid plasma-catalytic corona with thermal catalytic reactor over Ni/ $\gamma$ -Al<sub>2</sub>O<sub>3</sub>, *J. Nat. Gas Chem.* 21 (2012) 466–475. [https://doi.org/10.1016/S1003-9953\(11\)60392-7](https://doi.org/10.1016/S1003-9953(11)60392-7).
- [15] J. Martin-del-Campo, M. Uceda, S. Coulombe, J. Kopyscinski, Plasma-catalytic dry reforming of methane over Ni-supported catalysts in a rotating gliding arc – Spouted bed reactor, *J. CO<sub>2</sub> Util.* 46 (2021) 101474. <https://doi.org/10.1016/j.jcou.2021.101474>.
- [16] O. Biondo, C.F.A.M. van Deursen, A. Hughes, A. van de Steeg, W. Bongers, M.C.M. van de Sanden, G. van Rooij, A. Bogaerts, Avoiding solid carbon deposition in plasma-based dry reforming of methane, *Green Chem.* 25 (2023) 10485–10497. <https://doi.org/10.1039/D3GC03595F>.
- [17] S. Li, J. Sun, Y. Gorbanev, K. van't Veer, B. Loenders, Y. Yi, T. Kenis, Q. Chen, A. Bogaerts, Plasma-Assisted Dry Reforming of CH<sub>4</sub>: How Small Amounts of O<sub>2</sub> Addition Can Drastically Enhance the Oxygenate Production—Experiments and Insights from Plasma Chemical Kinetics Modeling, *ACS Sustainable Chem. Eng.* 11 (2023) 15373–15384. <https://doi.org/10.1021/acssuschemeng.3c04352>.
- [18] M. Ronda-Lloret, Y. Wang, P. Oulego, G. Rothenberg, X. Tu, N.R. Shiju, CO<sub>2</sub> Hydrogenation at Atmospheric Pressure and Low Temperature Using Plasma-Enhanced Catalysis over Supported Cobalt Oxide Catalysts, *ACS Sustainable Chem. Eng.* 8 (2020) 17397–17407. <https://doi.org/10.1021/acssuschemeng.0c05565>.
- [19] B. Ashford, Y. Wang, C.-K. Poh, L. Chen, X. Tu, Plasma-catalytic conversion of CO<sub>2</sub> to CO over binary metal oxide catalysts at low temperatures, *Appl. Catal. B Environ.* 276 (2020) 119110. <https://doi.org/10.1016/j.apcatb.2020.119110>.
- [20] D. Mei, B. Ashford, Y.-L. He, X. Tu, Plasma-catalytic reforming of biogas over supported Ni catalysts in a dielectric barrier discharge reactor: Effect of catalyst supports, *Plasma Process.*

- Polym. 14 (2017) 1600076. <https://doi.org/10.1002/ppap.201600076>.
- [21] N. Lu, D. Sun, Y. Xia, K. Shang, B. Wang, N. Jiang, J. Li, Y. Wu, Dry reforming of CH<sub>4</sub>CO<sub>2</sub> in AC rotating gliding arc discharge: Effect of electrode structure and gas parameters, *Int. J. Hydrogen Energy* 43 (2018) 13098–13109. <https://doi.org/10.1016/j.ijhydene.2018.05.053>.
- [22] R.S. Abiev, D.A. Sladkovskiy, K.V. Semikin, D.Y. Murzin, E.V. Rebrov, Non-Thermal Plasma for Process and Energy Intensification in Dry Reforming of Methane, *Catalysts* 10 (2020) 1358. <https://doi.org/10.3390/catal10111358>.
- [23] M. Witman, D. Gidon, D.B. Graves, B. Smit, A. Mesbah, Sim-to-real transfer reinforcement learning for control of thermal effects of an atmospheric pressure plasma jet, *Plasma Sources Sci. Technol.* 28 (2019) 095019. <https://doi.org/10.1088/1361-6595/ab3c15>.
- [24] A.D. Bonzanini, K. Shao, D.B. Graves, S. Hamaguchi, A. Mesbah, Foundations of machine learning for low-temperature plasmas: methods and case studies, *Plasma Sources Sci. Technol.* 32 (2023) 024003. <https://doi.org/10.1088/1361-6595/acb28c>.
- [25] U.K. Ercan, G.D. Özdemir, M.A. Özdemir, O. Güren, Plasma medicine: The era of artificial intelligence, *Plasma Process. Polym.* 20 (2023) e2300066. <https://doi.org/10.1002/ppap.202300066>.
- [26] M. Suvarna, T.P. Araújo, J. Pérez-Ramírez, A generalized machine learning framework to predict the space-time yield of methanol from thermocatalytic CO<sub>2</sub> hydrogenation, *Appl. Catal. B Environ.* 315 (2022) 121530. <https://doi.org/10.1016/j.apcatb.2022.121530>.
- [27] A. Bhardwaj, A.S. Ahluwalia, K.K. Pant, S. Upadhyayula, A principal component analysis assisted machine learning modeling and validation of methanol formation over Cu-based catalysts in direct CO<sub>2</sub> hydrogenation, *Sep. Purif. Technol.* 324 (2023) 124576. <https://doi.org/10.1016/j.seppur.2023.124576>.
- [28] Y. Wang, Z. Liao, S. Mathieu, F. Bin, X. Tu, Prediction and evaluation of plasma arc reforming of naphthalene using a hybrid machine learning model, *J. Hazard. Mater.* 404 (2021) 123965. <https://doi.org/10.1016/j.jhazmat.2020.123965>.
- [29] S.Y. Liu, D.H. Mei, Z. Shen, X. Tu, Nonoxidative Conversion of Methane in a Dielectric Barrier Discharge Reactor: Prediction of Reaction Performance Based on Neural Network Model, *J. Phys. Chem. C* 118 (2014) 10686–10693. <https://doi.org/10.1021/jp502557s>.
- [30] X. Zhu, S. Liu, Y. Cai, X. Gao, J. Zhou, C. Zheng, X. Tu, Post-plasma catalytic removal of methanol over Mn–Ce catalysts in an atmospheric dielectric barrier discharge, *Appl. Catal. B Environ.* 183 (2016) 124–132. <https://doi.org/10.1016/j.apcatb.2015.10.013>.
- [31] Y. Wang, Y. Chen, J. Harding, H. He, A. Bogaerts, X. Tu, Catalyst-free single-step plasma reforming of CH<sub>4</sub> and CO<sub>2</sub> to higher value oxygenates under ambient conditions, *Chem. Eng. J.* 450 (2022) 137860. <https://doi.org/10.1016/j.cej.2022.137860>.
- [32] Y. Cai, D. Mei, Y. Chen, A. Bogaerts, X. Tu, Machine learning-driven optimization of plasma-catalytic dry reforming of methane, *J. Energy Chem.* 96 (2024) 153–163. <https://doi.org/10.1016/j.jechem.2024.04.022>.
- [33] A. Mesbah, D.B. Graves, Machine learning for modeling, diagnostics, and control of non-equilibrium plasmas, *J. Phys. D: Appl. Phys.* 52 (2019) 30LT02. <https://doi.org/10.1088/1361-6463/ab1f3f>.
- [34] R.S. Sutton, A.G. Barto, Reinforcement Learning: An Introduction, Cambridge, MA, USA: 624 MIT Press (2018).
- [35] Z. Hou, T. Lee, M. Keidar, Reinforcement Learning With Safe Exploration for Adaptive Plasma

- Cancer Treatment, *IEEE Trans. Radiat. Plasma Med. Sci.* 6 (2022) 482–492. <https://doi.org/10.1109/TRPMS.2021.3094874>.
- [36] B. Wanten, R. Vertongen, R.D. Meyer, A. Bogaerts, Plasma-based CO<sub>2</sub> conversion: How to correctly analyze the performance?, *J. Energy Chem.* 86 (2023) 180. <https://doi.org/10.1016/j.jechem.2023.07.005>.
- [37] T. Nishida, [Data transformation and normalization], *Rinsho Byori* 58 (2010) 990–997.
- [38] X. Wang, X. Du, K. Chen, Z. Zheng, Y. Liu, X. Shen, C. Hu, Predicting the Ammonia Synthesis Performance of Plasma Catalysis Using an Artificial Neural Network Model, *ACS Sustainable Chem. Eng.* 11 (2023) 4543–4554. <https://doi.org/10.1021/acssuschemeng.2c04715>.
- [39] J. Li, L. Pan, M. Suvarna, X. Wang, Machine learning aided supercritical water gasification for H<sub>2</sub>-rich syngas production with process optimization and catalyst screening, *Chem. Eng. J.* 426 (2021) 131285. <https://doi.org/10.1016/j.cej.2021.131285>.
- [40] Y. Shen, C. Fu, W. Luo, Z. Liang, Z.-R. Wang, Q. Huang, Machine learning for CO<sub>2</sub> conversion driven by dielectric barrier discharge plasma and Cs<sub>2</sub>TeCl<sub>6</sub> photocatalysts, *Green Chem.* 25 (2023) 7605–7611. <https://doi.org/10.1039/D3GC02354K>.
- [41] D. Tan, M. Suvarna, Y. Shee Tan, J. Li, X. Wang, A three-step machine learning framework for energy profiling, activity state prediction and production estimation in smart process manufacturing, *Appl. Energy* 291 (2021) 116808. <https://doi.org/10.1016/j.apenergy.2021.116808>.
- [42] X. Yuan, M. Suvarna, S. Low, P.D. Dissanayake, K.B. Lee, J. Li, X. Wang, Y.S. Ok, Applied Machine Learning for Prediction of CO<sub>2</sub> Adsorption on Biomass Waste-Derived Porous Carbons, *Environ. Sci. Technol.* 55 (2021) 11925–11936. <https://doi.org/10.1021/acs.est.1c01849>.
- [43] Y. Gao, Z. Saedi, H. Shi, B. Zeng, B. Zhang, X. Zhang, Machine Learning-Assisted Optimization of Microbubble-Enhanced Cold Plasma Activation for Water Treatment, *ACS EST Water* 4 (2024) 735–750. <https://doi.org/10.1021/acsestwater.3c00783>.
- [44] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, O. Klimov, Proximal Policy Optimization Algorithms, (2017). <https://doi.org/10.48550/arXiv.1707.06347>.
- [45] A.H. Khoja, A. Mazhar, F. Saleem, M.T. Mehran, S.R. Naqvi, M. Anwar, S. Shakir, N.A. Saidina Amin, M.B. Sajid, Recent developments in catalyst synthesis using DBD plasma for reforming applications, *Int. J. Hydrogen Energy* 46 (2021) 15367–15388. <https://doi.org/10.1016/j.ijhydene.2021.02.043>.
- [46] Sk. Mahammadunnisa, P. Manoj Kumar Reddy, B. Ramaraju, Ch. Subrahmanyam, Catalytic Nonthermal Plasma Reactor for Dry Reforming of Methane, *Energy Fuels* 27 (2013) 4441–4447. <https://doi.org/10.1021/ef302193e>.
- [47] N. Sahli, C. Petit, A.C. Roger, A. Kiennemann, S. Libs, M.M. Bettahar, Ni catalysts from NiAl<sub>2</sub>O<sub>4</sub> spinel for CO<sub>2</sub> reforming of methane, *Catal. Today* 113 (2006) 187–193. <https://doi.org/10.1016/j.cattod.2005.11.065>.
- [48] Q. Wang, B.-H. Yan, Y. Jin, Y. Cheng, Dry Reforming of Methane in a Dielectric Barrier Discharge Reactor with Ni/Al<sub>2</sub>O<sub>3</sub> Catalyst: Interaction of Catalyst and Plasma, *Energy Fuels* 23 (2009) 4196–4201. <https://doi.org/10.1021/ef900286j>.
- [49] A. Bogaerts, T. Kozák, K. van Laer, R. Snoeckx, Plasma-based conversion of CO<sub>2</sub>: current status and future challenges, *Faraday Discuss.* 183 (2015) 217–232. <https://doi.org/10.1039/C5FD00053J>.

- [50] R. Snoeckx, Y.X. Zeng, X. Tu, A. Bogaerts, Plasma-based dry reforming: improving the conversion and energy efficiency in a dielectric barrier discharge, *RSC Adv.* 5 (2015) 29799–29808. <https://doi.org/10.1039/C5RA01100K>.
- [51] R. Snoeckx, R. Aerts, X. Tu, A. Bogaerts, Plasma-Based Dry Reforming: A Computational Study Ranging from the Nanoseconds to Seconds Time Scale, *J. Phys. Chem. C* 117 (2013) 4957–4970. <https://doi.org/10.1021/jp311912b>.
- [52] J.-M. Lavoie, Review on dry reforming of methane, a potentially more environmentally-friendly approach to the increasing natural gas exploitation, *Front. Chem.* 2 (2014). <https://doi.org/10.3389/fchem.2014.00081>.
- [53] D. Mei, X. Shen, S. Liu, R. Zhou, X. Yuan, Z. Rao, Y. Sun, Z. Fang, X. Du, Y. Zhou, X. Tu, Plasma-catalytic reforming of biogas into syngas over Ni-based bimetallic catalysts, *Chem. Eng. J.* 462 (2023) 142044. <https://doi.org/10.1016/j.cej.2023.142044>.
- [54] S. Maerivoet, B. Wanten, R. De Meyer, M. Van Hove, S. Van Alphen, A. Bogaerts, Effect of O<sub>2</sub> on Plasma-Based Dry Reforming of Methane: Revealing the Optimal Gas Composition via Experiments and Modeling of an Atmospheric Pressure Glow Discharge, *ACS Sustainable Chem. Eng.* (2024). <https://doi.org/10.1021/acssuschemeng.4c04283>.
- [55] T. Toyao, Z. Maeno, S. Takakusagi, T. Kamachi, I. Takigawa, K. Shimizu, Machine Learning for Catalysis Informatics: Recent Applications and Prospects, *ACS Catal.* 10 (2020) 2260–2297. <https://doi.org/10.1021/acscatal.9b04186>.

## Supplementary Information

### **Machine Learning-based prediction and optimization of plasma-catalytic dry reforming of methane in a dielectric barrier discharge reactor**

Jiayin Li<sup>1</sup>, Jing Xu<sup>2</sup>, Evgeny Rebrov<sup>3,4</sup> and Annemie Bogaerts<sup>1,\*</sup>

<sup>1</sup> Research Group PLASMANT, Department of Chemistry, University of Antwerp, Antwerp, 2610, Belgium

<sup>2</sup> School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan, Hubei, 430074, People's Republic of China

<sup>3</sup> School of Engineering, University of Warwick, Coventry, CV4 7AL, the United Kingdom

<sup>4</sup> Department of Chemical Engineering and Chemistry, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, the Netherlands

\*Corresponding author: [jiayin.li@uantwerpen.be](mailto:jiayin.li@uantwerpen.be) (J. Li),  
[annemie.bogaerts@uantwerpen.be](mailto:annemie.bogaerts@uantwerpen.be) (A. Bogaerts).

## Contents

1. List of all abbreviations in the paper .....	3
2. Experimental setup .....	4
3. The database for the ML model development .....	4
4. Data processing.....	9
5. Hyperparameters optimization.....	9
6. Artificial neural network model.....	10
7. Relative significance of different parameters .....	10
8. Reinforcement learning model .....	11
8.1 Basic concepts.....	11
8.2 States, actions, and rewards .....	12
8.3 Actor-critic framework .....	12
8.4 Proximal policy optimization algorithm.....	13
8.5 Network structure for RL agent training .....	13
9. Comparative analysis of the predicted and experimental results .....	14
10. RL agents training results .....	16
11. A visualization figure of the input distribution for the SL model .....	16

## 1. List of all abbreviations in the paper

Table S1. List of all abbreviations

<b>Original description</b>	<b>Abbreviations</b>
Dry reforming of methane	DRM
Microwave discharge	MW
Dielectric barrier discharge	DBD
Artificial intelligence	AI
Machine learning	ML
Supervised learning	SL
Reinforcement learning	RL
Artificial neural network	ANN
Mean square error	MSE
Coefficient of determination	$R^2$
Reinforcement learning controllers	RLC
Energy cost	EC
Backpropagation	BP
Pearson's Correlation Coefficient	PCC
Proximal Policy Optimization	PPO
Actor-Critic	AC
Energy yield	EY
Atmospheric pressure glow discharge	APGD
Temporal difference	TD

## 2. Experimental setup

Fig. S1 shows the schematic overview of the experimental setup, which was detailedly described in Ref. [1]. The catalyst preparation and reaction performance diagnostics were described in detail in Ref. [2].

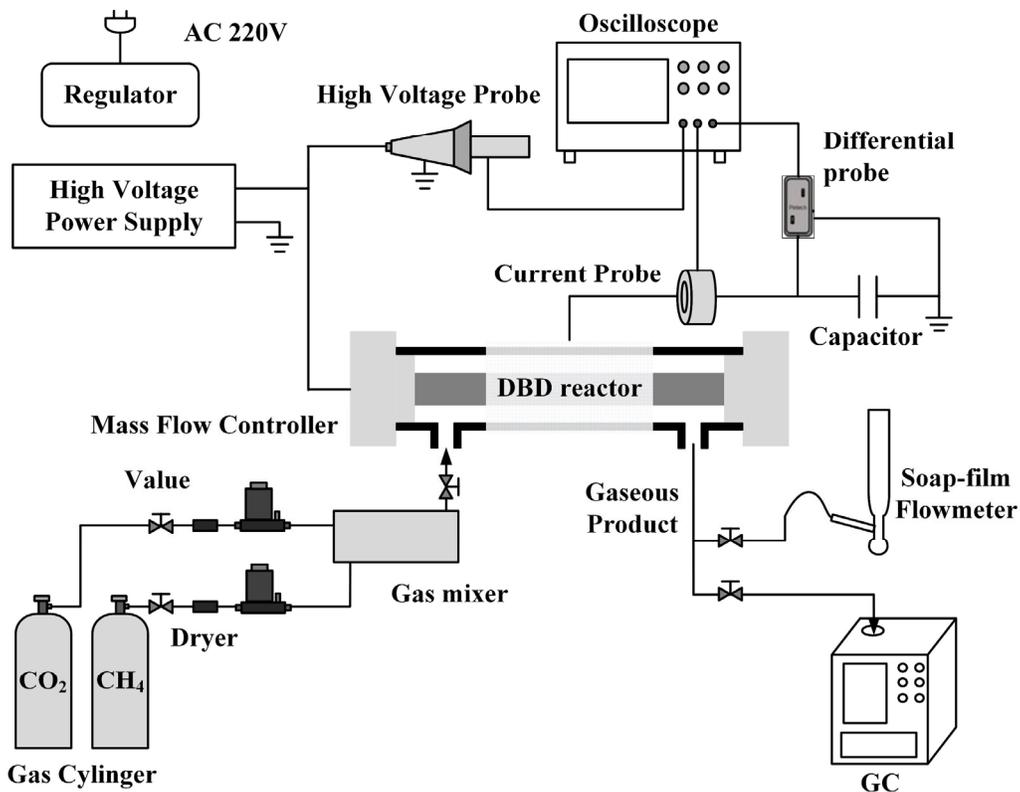


Fig. S1. Schematic overview of the experimental setup [2].

## 3. The database for the ML model development

Table S2. Operating parameters for the plasma-catalytic DRM process

No.	Ni loading (wt%)	Discharge Power (W)	CO <sub>2</sub> /CH <sub>4</sub> molar ratio	Total flow rate (mL/min)
1	5	20	1	75
2	7.5	20	1	75
3	10	20	1	75
4	12.5	20	1	75
5	15	20	1	75
6	5	20	1.25	75
7	7.5	20	1.25	75
8	10	20	1.25	75
9	12.5	20	1.25	75
10	15	20	1.25	75
11	5	20	1.5	75
12	7.5	20	1.5	75
13	10	20	1.5	75

14	12.5	20	1.5	75
15	15	20	1.5	75
16	5	20	0.75	75
17	7.5	20	0.75	75
18	10	20	0.75	75
19	12.5	20	0.75	75
20	15	20	0.75	75
21	5	20	0.5	75
22	7.5	20	0.5	75
23	10	20	0.5	75
24	12.5	20	0.5	75
25	15	20	0.5	75
26	10	20	1	75
27	7.5	30	1.25	100
28	7.5	30	0.75	100
29	12.5	30	0.75	100
30	12.5	30	1.25	100
31	7.5	30	0.75	50
32	7.5	30	1.25	50
33	12.5	30	0.75	50
34	10	40	1.5	75
35	10	40	0.5	75
36	10	40	1	75
37	10	40	1	75
38	15	40	1	75
39	5	40	0.5	25
40	7.5	40	0.5	25
41	10	40	0.5	25
42	12.5	40	0.5	25
43	15	40	0.5	25
44	5	40	0.75	25
45	7.5	40	0.75	25
46	10	40	0.75	25
47	12.5	40	0.75	25
48	15	40	0.75	25
49	5	40	1.5	25
50	7.5	40	1.5	25
51	10	40	1.5	25
52	12.5	40	1.5	25
53	15	40	1.5	25
54	5	40	1.25	25
55	7.5	40	1.25	25
56	10	40	1.25	25

57	12.5	40	1.25	25
58	15	40	1.25	25
59	5	40	1	25
60	7.5	40	1	25
61	10	40	1	25
62	12.5	40	1	25
63	15	40	1	25
64	7.5	40	0.5	125
65	7.5	40	0.75	125
66	7.5	40	1	125
67	7.5	40	1.25	125
68	7.5	40	1.5	125
69	15	40	0.5	125
70	15	40	0.75	125
71	15	40	1	125
72	15	40	1.25	125
73	15	40	1.5	125
74	5	40	0.5	125
75	5	40	0.75	125
76	5	40	1	125
77	5	40	1.25	125
78	5	40	1.5	125
79	12.5	40	0.5	125
80	12.5	40	0.75	125
81	12.5	40	1	125
82	12.5	40	1.25	125
83	12.5	40	1.5	125
84	10	40	0.5	125
85	10	40	0.75	125
86	10	40	1	125
87	10	40	1.25	125
88	10	40	1.5	125
89	15	40	1	75
90	7.5	50	1.25	50
91	12.5	50	0.75	100
92	7.5	50	0.75	100
93	7.5	50	0.75	50
94	7.5	50	1.25	100
95	12.5	50	1.25	100
96	12.5	50	1.25	50
97	7.5	60	1	75
98	7.5	60	1.5	75
99	10	60	1	75

100	10	60	0.5	75
-----	----	----	-----	----

Table S3. Experimental results for the plasma-catalytic DRM process

No.	CO yield (%)	H <sub>2</sub> yield (%)	CO <sub>2</sub> conversion (%)	CH <sub>4</sub> conversion (%)	Total conversion (%)	Energy cost (eV/molec)
1	8.4	6.5	13.3	20.1	16.7	23.87604731
2	9.5	7.2	14.8	21.4	18.1	22.02928122
3	9.9	7.4	15.2	21	18.1	22.02928122
4	9.3	7	14.5	19	16.75	23.80477552
5	7.9	6.1	12.8	15.4	14.1	28.2787227
6	9.8	7.9	11.7	25	17.552	22.71706871
7	11	8.6	13.3	25.7	18.756	21.25879665
8	11.3	8.7	13.7	24.9	18.628	21.40487385
9	10.7	8.3	13.1	22.4	17.192	23.1927635
10	9.3	7.4	11.3	18.4	14.424	27.64351012
11	11	9.1	9.7	29.5	17.62	22.62939785
12	12.1	9.7	11.3	29.8	18.7	21.32245936
13	12.3	9.9	11.8	28.5	18.48	21.57629816
14	11.7	9.4	11.2	25.6	16.96	23.510023
15	10.3	8.5	9.4	21	14.04	28.39957194
16	6.5	4.9	14.4	15	14.742	27.04721137
17	7.7	5.6	15.9	16.6	16.299	24.4634634
18	8.1	5.8	16.3	16.7	16.528	24.12451537
19	7.6	5.5	15.6	15.2	15.372	25.9387191
20	6.2	4.6	13.8	12	12.774	31.21418428
21	4.3	3.1	15.1	9.4	11.281	35.34526993
22	5.6	3.9	16.6	11.6	13.25	30.09282944
23	6	4.1	16.9	12.1	13.684	29.13840909
24	5.5	3.8	16.2	11	12.716	31.35655788
25	4.1	3	14.3	8.4	10.347	38.53580652
26	9.8	7.3	15.5	21	18.25	21.84821863
27	9.8	7.4	11.4	22.5	16.284	27.54674765
28	6.7	4.7	14	14.2	14.114	31.78200643
29	6.6	4.6	13.7	14.2	13.985	32.07516902
30	9.5	7.2	11.2	20.6	15.336	29.24955913
31	10.9	9.3	23.5	26.2	25.039	35.82980461
32	14.7	12.5	19.7	37.1	27.356	32.79508984
33	10.7	9	23	24.3	23.741	37.78874005
34	15.4	12.8	15.8	37.8	24.6	32.41707236
35	8.4	6.9	23.3	19.4	20.687	38.54884614
36	12.6	10.2	20.5	29.3	24.9	32.02650522
37	12.6	10.2	20.5	29.3	24.9	32.02650522
38	10.6	8.9	17.8	24.8	21.3	37.43943568

39	12.8	12	33.7	29.8	31.087	76.95756877
40	13.9	12.6	34.9	31.6	32.689	73.18608523
41	14.2	12.7	35.1	31.7	32.822	72.8895235
42	13.6	12.2	34.1	30.3	31.554	75.8185948
43	12.1	11.3	32	27.2	28.784	83.11492288
44	15.4	14.1	31.8	37.2	34.878	68.59280751
45	16.5	14.7	33.1	38.5	36.178	66.12803196
46	16.8	14.7	33.2	38.2	36.05	66.36282775
47	16.2	14.3	32.3	36.3	34.58	69.18391961
48	14.7	13.2	30.2	32.8	31.682	75.51227638
49	21.3	19.3	23.5	57.3	37.02	64.62398542
50	22.3	19.8	24.8	57.2	37.76	63.3575196
51	22.5	19.8	25.1	55.5	37.26	64.20772786
52	21.8	19.2	24.2	52.2	35.4	67.58135424
53	20.2	18.1	22.3	47.3	32.3	74.06749041
54	19.7	17.7	26.7	50.9	37.348	64.05644051
55	20.8	18.3	28	51.3	38.252	62.54261059
56	20.9	18.3	28.2	50.1	37.836	63.23025532
57	20.3	17.7	27.3	47.3	36.1	66.27091247
58	18.7	16.6	25.4	42.8	33.056	72.37354611
59	17.8	16	29.5	44.2	36.85	64.92211507
60	18.8	16.6	30.8	45.1	37.95	63.04031463
61	19	16.6	31	44.3	37.65	63.54262789
62	18.4	16.1	30	41.9	35.95	66.54742532
63	16.9	15	28	38	33	72.49636182
64	2.6	1.5	11.6	7.9	9.121	52.45872032
65	4.6	3.1	10.9	12.1	11.584	41.30490228
66	6.2	4.4	9.9	15.9	12.9	37.09116186
67	7.5	5.6	8.4	19.4	13.24	36.13866979
68	8.5	6.5	6.4	22.6	12.88	37.14875684
69	1.3	0.8	9.5	9	9.165	52.20687267
70	3.2	2.2	9	11.8	10.596	45.15628426
71	4.8	3.5	8	14.3	11.15	42.91264467
72	6	4.5	6.6	16.5	10.956	43.67250712
73	6.8	5.4	4.7	18.3	10.14	47.18698107
74	1.3	0.7	10.1	4.3	6.214	76.99967622
75	3.3	2.3	9.4	8.9	9.115	52.49325157
76	5	3.6	8.3	13.2	10.75	44.50939424
77	6.3	4.8	6.8	17.2	11.376	42.06012553
78	7.3	5.8	4.8	20.9	11.24	42.56903808
79	2.6	1.6	11.3	10.2	10.563	45.29735757
80	4.5	3.1	10.7	13.5	12.296	38.91314151
81	6.1	4.3	9.7	16.5	13.1	36.52488458

82	7.3	5.4	8.3	19.1	13.052	36.6592084
83	8.2	6.3	6.4	21.3	12.36	38.71164952
84	3	1.8	12	9.8	10.526	45.45658256
85	5	3.3	11.4	13.6	12.654	37.81223234
86	6.6	4.6	10.4	17	13.7	34.9252546
87	7.9	5.7	8.9	20	13.784	34.71241933
88	8.8	6.7	7	22.8	13.32	35.92162072
89	11	9	18.1	25.4	21.75	36.66482667
90	20.5	17.4	26.5	48.4	36.136	41.37805686
91	8.6	6.8	18.6	23	21.108	35.41873846
92	8.8	7	19	21.9	20.653	36.19903798
93	16.5	14	31.6	36.4	34.336	43.5472234
94	12.2	9.8	15.2	31.1	22.196	33.68258836
95	11.9	9.5	14.9	30.4	21.72	34.4207519
96	20.1	16.9	25.9	45.8	34.656	43.14512531
97	17.5	14.4	26.7	40.3	33.5	35.70716329
98	20.7	17.2	20.8	50.7	32.76	36.51373535
99	17.8	14.4	27	41	34	35.18205794
100	13.2	11	31	30.1	30.397	39.35223772

#### 4. Data processing

To normalize the input and output data, the linear function method was used to transform the sample data into the interval [0,1]:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

where  $X_{norm}$  denotes the normalized input parameters and predicted output;  $X$  denotes the original input parameters and actual output;  $X_{max}$  and  $X_{min}$  denotes the maximum and minimum value of the data, respectively. The inverse normalization formula for the predicted output is:

$$Y = Y_{min} + Y_{norm} (Y_{max} - Y_{min}) \quad (2)$$

where  $Y_{norm}$  denotes the normalized predicted value;  $Y$  denotes the inverse normalized predicted value; and  $Y_{max}$  and  $Y_{min}$  denote the maximum and minimum values of the data, respectively.

#### 5. Hyperparameters optimization

Hyperparameters are parameters that are set before the actual learning process begins. For an artificial neural network (ANN) model, the number of hidden layers and neurons determine the capacity of the network. More layers and neurons can increase the model's ability to learn complex patterns but may also increase the risk of overfitting. Therefore, it was optimized by the grid search method to find the optimal combination within the specified ranges. For the reinforcement learning (RL) model, the

hyperparameter selection was optimized by a random search method: 1) Define a search space for each hyperparameter; 2) Generate multiple configurations by randomly sampling from the defined search space; 3) Train the RL model with each configuration, and then evaluate the performance of each configuration using a validation metric (e.g., accumulative reward); 4) Choose the configuration that performs best based on the evaluation metric, and consider running additional simulations with slight variations around the best configuration to refine the results.

## 6. Artificial neural network model

An ANN is a common SL model [3,4]. Indeed, it can be viewed as a sophisticated function [5]. The goal of ANN training process is to find optimum weight and bias values to reduce the discrepancy between predicted and actual values. The loss function at the output serves as an adjustment signal to constantly optimize the weights in the direction of the input and will reduce to an acceptable level or within a specified number of training epochs (where epoch is defined as one complete cycle of training data being processed through the algorithm). Fig. S2 presented the example of this process. The MSE of the network decreases continuously during the training and converges at about 60 epochs. Once the loss function is computed, the partial derivatives of the weights are determined by using the chain rule, then we use gradient descent method to update the network parameters.

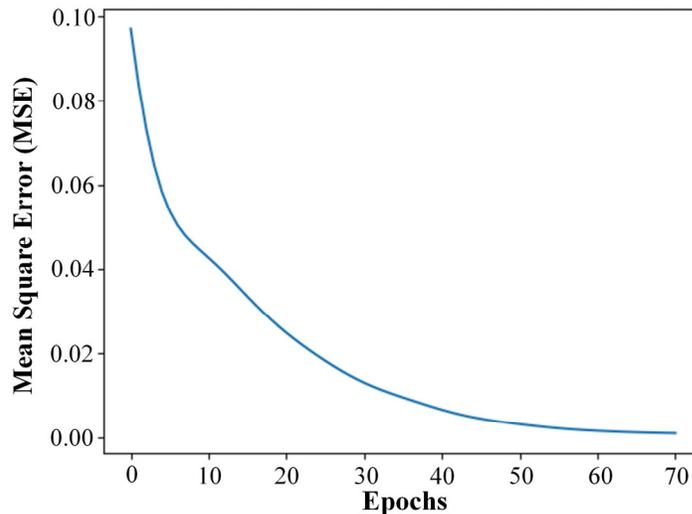


Fig. S2. MSE of the best fitness value in each epoch for ANN

## 7. Relative significance of different parameters

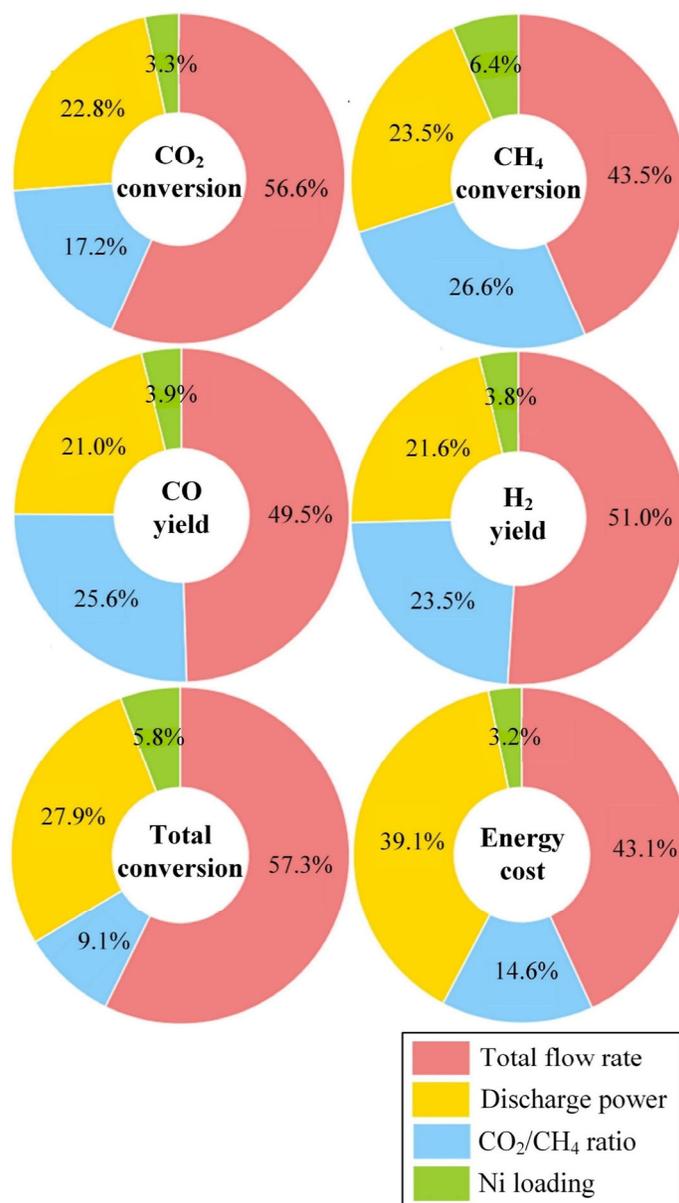


Fig. S3. Importance analysis between reaction performance and four different operating parameters.

## 8. Reinforcement learning model

### 8.1 Basic concepts

RL aims to formalize the decision-process based on experience through interaction with the world. The decision-maker or learner is called as *agent* while the *environment* encompasses everything the agent interacts with. At any time step  $t$ , the agent observes the environment, represented as state  $s_t$ , and it must choose an action  $a(t)$ , and it also receives some rewards,  $r(t)$ . Through this action, it reaches new state ( $s_{t+1}$ ). The objective is to develop a policy  $\pi(a|s)$ , in order to maximize the expected cumulative future reward. Thus, the reward signal determines which actions are good or bad, guiding the agent's subsequent actions. Fig. S4 shows such a closed-loop operation. In the next, we will present fundamental RL concepts and explain how to train the RL

controller (RLC) for plasma-catalytic DRM process discussed in this paper.

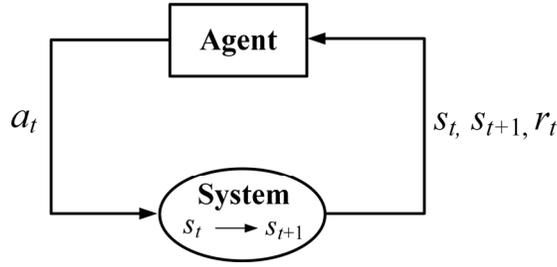


Fig. S4. Schematic principle of RL.

## 8.2 States, actions, and rewards

The states and actions are defined by the four operating parameters described in the main paper. It should be noted that the agent's actions must remain within the defined limits when considering the model within investigated range. The reward function is employed to guide optimization toward the best possible outcome. In our case, we aim to maximize the reaction performance (gas conversion and product yield) while minimizing the EC. Therefore, the reward function is determined by the value difference between current time step and previous time step during the iteration. Specifically, the higher the reaction performance and lower the energy cost, the higher is the reward.

## 8.3 Actor-critic framework

The Actor-critic (AC) algorithm is an extension of the idea of gradient bandit methods [6]. The *actor* is responsible for choosing actions, represented by the policy and the *critic* is used to evaluate the quality of actions made by the *actor*. The working process of AC framework is presented in Fig. S7:

- (1) The actor selects an action  $a_t$  by using its policy  $\pi_{\theta}(a_t|s_t)$
- (2) According to the current state  $s_t$ , leading to a new state  $s_{t+1}$  and a reward  $r_t$ ;
- (3) Based on the reward, the critic compares the value of the new state  $V(s_{t+1})$  with the previous state  $V(s_t)$  by evaluating the temporal difference (TD) error:

$$\delta_t = r_t + \gamma V_{\phi}(s_{t+1}) - V_{\phi}(s_t) \quad (3)$$

where  $\gamma$  denotes the discount factor (0,1).

- (4) The critic's value function is updated by gradient descent to minimize TD error:

$$\theta_c \leftarrow \theta_c + \alpha_c \delta_t \nabla_{\theta_c} V(s_t) \quad (4)$$

where  $\alpha_c$  and  $\theta_c$  are the learning rate and parameters for the critic network, respectively.

- (5) Based on the feedback from the critic, the actor's policy is updated by policy gradient [7]:

$$\nabla_{\theta} J(\theta) = \nabla_{\theta} \log \pi(a_t|s_t; \theta) \cdot \delta_t \quad (5)$$

The actor updates the policy according to the critic's evaluation, to favor actions that are more likely to yield higher rewards:

$$\theta \leftarrow \theta + \alpha_a \nabla_{\theta} J(\theta) \quad (6)$$

where  $\alpha_a$  and  $\theta$  are the learning rate and parameters for the actor network, respectively.

The process is repeated for step (1) to (5) in the episode, continuously updating the actor and critic until convergence or until the end of the episode.

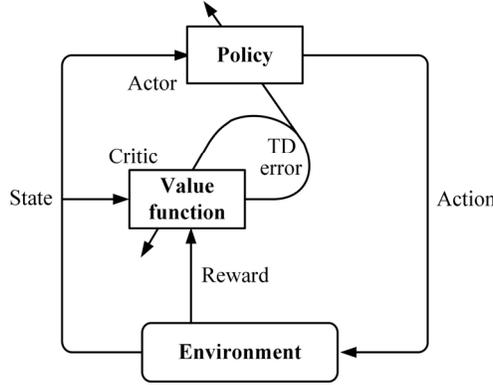


Fig. S5: The actor-critic architecture.

#### 8.4 Proximal policy optimization algorithm

Proximal policy optimization (PPO) is a specific algorithm built on the AC framework, designed to improve the stability and performance by constraining policy updates through clipping. The PPO algorithm updates its critic network similarly to the AC algorithm, but it features two types of policies in its actor network, which are called the target policy and the current policy, respectively. The current policy is used to generate the batch of trajectories (sequences of states, actions, and rewards), and the target policy updates the gradient according to these data and updates the current strategy at the end of each iteration cycle. The probability ratio  $r_t(\theta)$  measures the divergence between the new policy  $\pi_{\theta}(a_t|s_t)$  and the old one  $\pi_{\theta_{old}}(a_t|s_t)$  for a given action by importance sampling:

$$r_t(\theta) = \frac{\pi_{\theta}(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)} \quad (7)$$

The TD error is then multiplied by the probability ratio to update the target policy, thus transforming the AC algorithm from an On policy to Off policy. The PPO's objective function is designed to maximize the expected reward, while keeping the target policy close to the current policy. Therefore, the loss function is clipped in the PPO algorithm as follows:

$$L^{Clip}(\theta) = Clip(r_t(\theta), 1 - \varepsilon, 1 + \varepsilon) \cdot \delta_t \quad (8)$$

where  $\varepsilon$  denotes a hyperparameter that determines the clipping range, restricting  $r_t(\theta)$  to the interval  $[1-\varepsilon, 1+\varepsilon]$ .

#### 8.5 Network structure for RL agent training

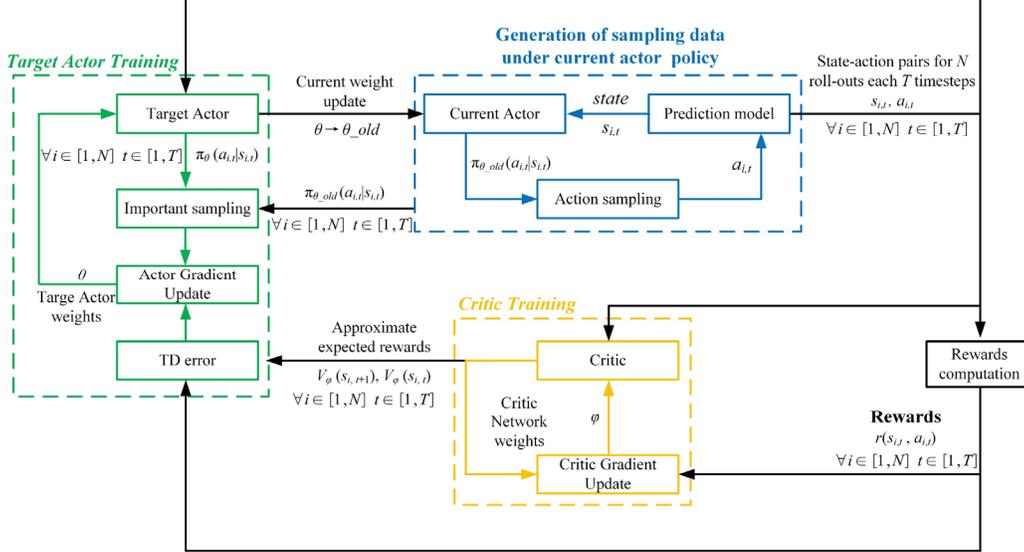


Fig. S6. Block diagram representation of the network structure.

The block diagram of the PPO algorithm is shown in Fig. S8. Each iteration cycle needs to generate a batch of training data ( $N = 200$  roll-outs and  $T = 300$  time steps) based on the current actor policy to update current weight. At the start of each roll-out, we randomly select a new setpoint from the uniform distribution,  $\chi_{sp} \sim u[7,8]$ . In each actor training step, we perform 10 times updates of the critic target values, with the critic network being updated 1 gradient steps per target update. Based on the optimised critic network, the target actor network is updated with 1 gradient steps.

Table S4. Detailed parameters of the RL controller

Parameter	Actor network	Critic network
Number of input layers	4	4
Number of hidden layer	16	16
Number of output layers	4	1
Activation function	tanh	softplus
Learning rate $\alpha$	1e-3	1e-2
Discount $\gamma$	0.98	
Scaling factor	0.95	
Clipped factor $\epsilon$	0.2	

The parameters of the RLC are shown in Table S4. The input layer of the network contains one hidden layer with 16 neurons. The output layer of the actor network contains 4 nodes, represented by the probability distribution of the four action parameters. The activation function for the mean parameter is a tanh function. The sampled action values are controlled by clipping in the range of  $[0,1]$ . The softplus activation function is used for the critic network.

## 9. Comparative analysis of the predicted and experimental results

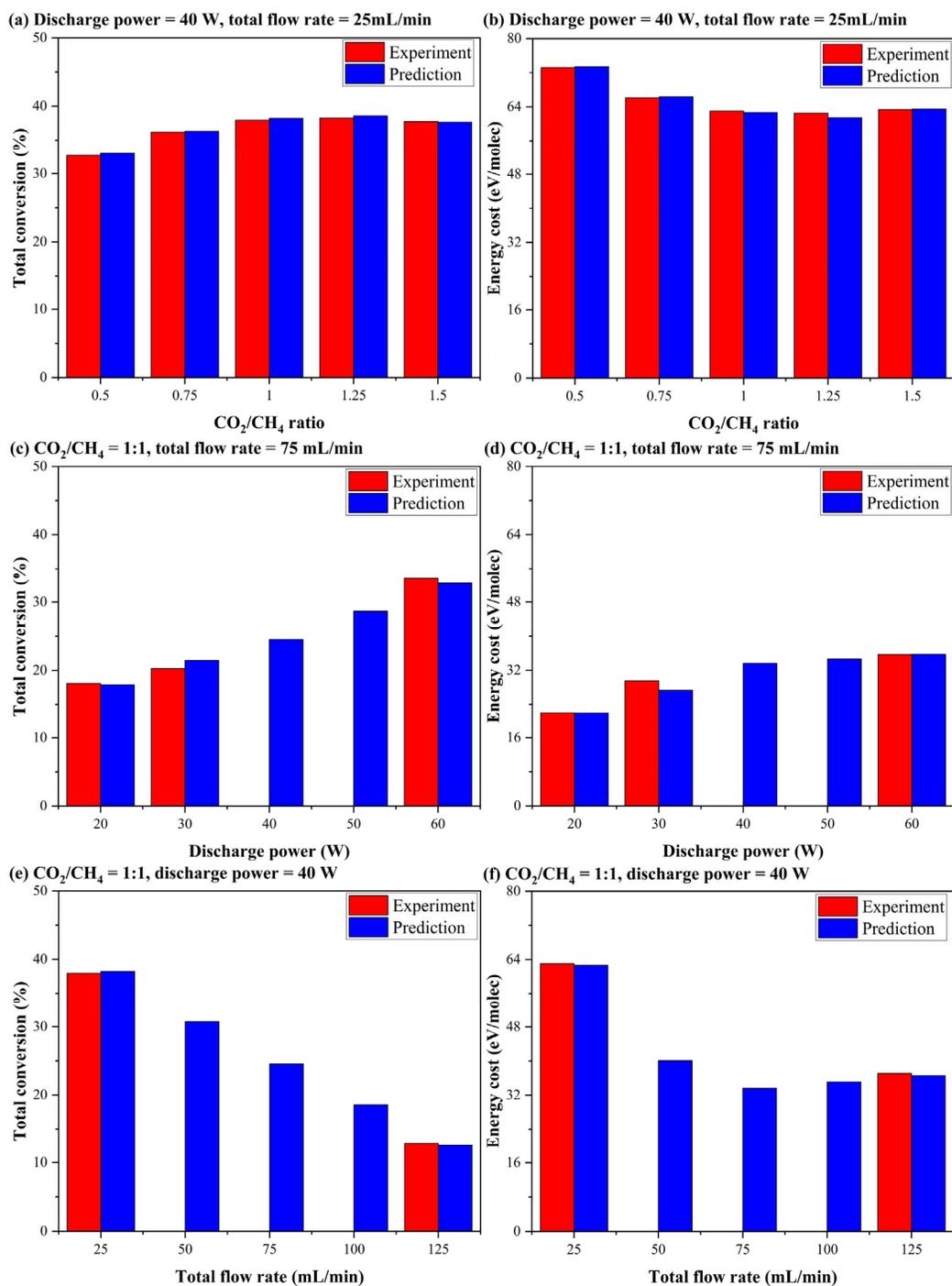


Fig. S7. Comparison of predicted values with available experimental data, using 7.5 wt% Ni/Al<sub>2</sub>O<sub>3</sub> for total conversion (a, c, e) and energy cost (b, d, f).

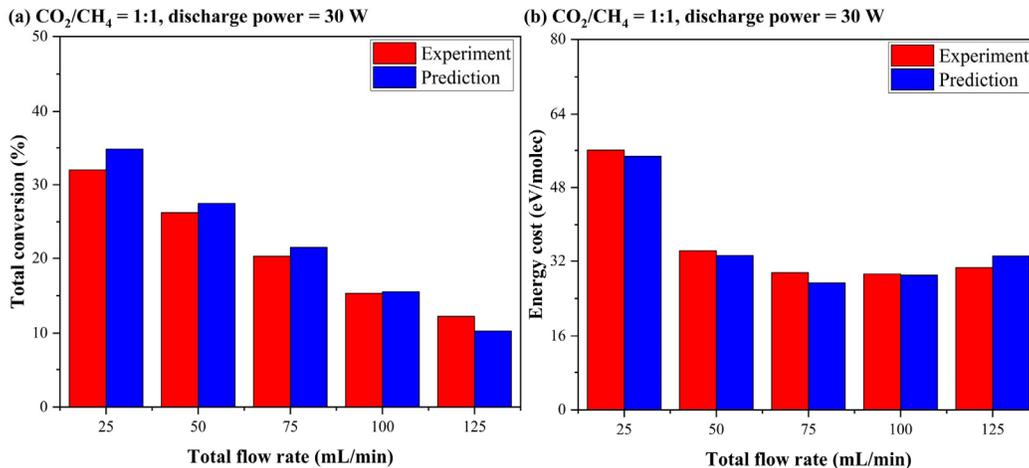


Fig. S8. Comparison of predicted values and unseen experimental data, using 7.5 wt% Ni/Al<sub>2</sub>O<sub>3</sub> for model generalization evaluation: (a) Total conversion; (b) Energy cost.

## 10. RL agents training results

The maximum training iterations for the agent is 300. Before training, the two RLC will generate 1 random number uniformly distributed within their limited interval, and each random number will be trained for 200 rounds. Therefore, in each iteration cycle, 200 data are used to train the critic network. The training results of CO<sub>2</sub> conversion and total conversion within investigated range are shown in Figure S9. The average return of both agents in each iteration converges to the maximum value 23 after about 25 iterations.

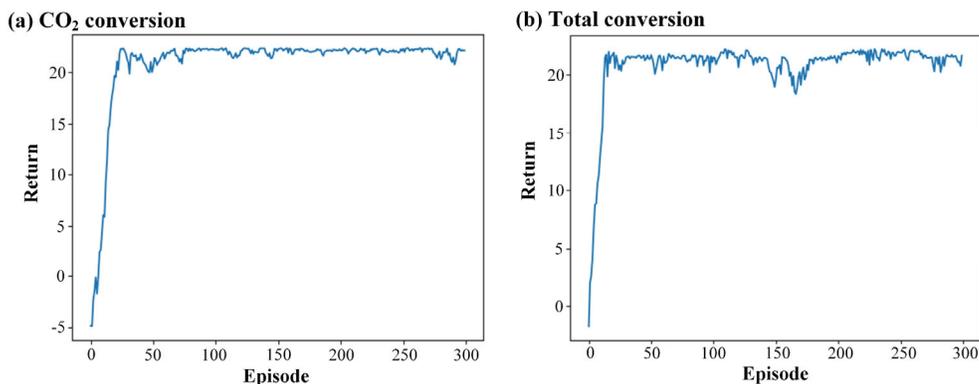


Fig. S9. Training curve of CO<sub>2</sub> conversion and total conversion RL agent.

## 11. A visualization figure of the input distribution for the SL model

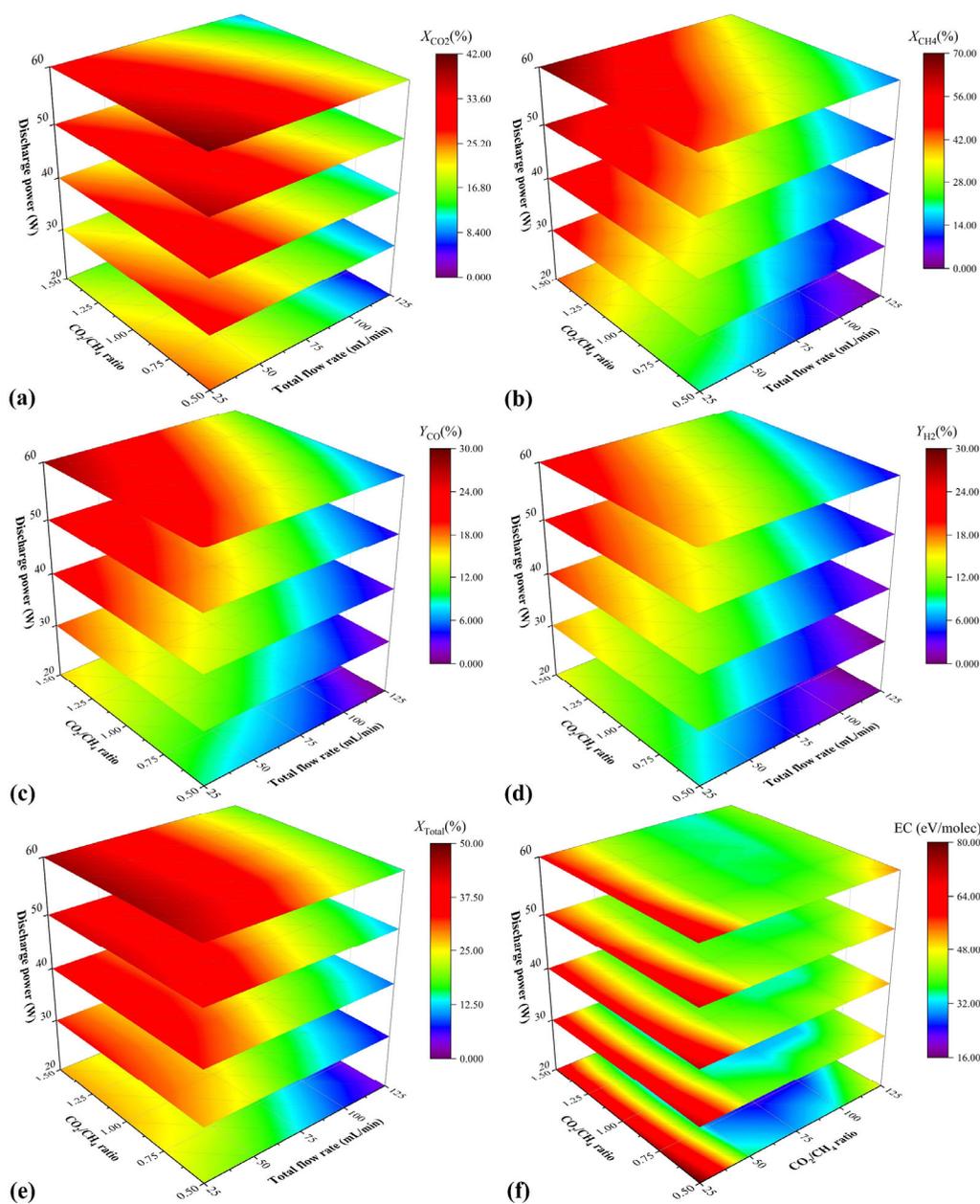


Fig. S10. Prediction results of input distribution including discharge power,  $\text{CO}_2/\text{CH}_4$  ratio, and total flow rate on the performance of plasma-catalytic DRM process using 7.5 wt%  $\text{Ni}/\text{Al}_2\text{O}_3$ . (a)  $\text{CO}_2$  conversion; (b)  $\text{CH}_4$  conversion; (c) CO yield; (d)  $\text{H}_2$  yield; (e) Total conversion; (f) EC.

## References

- [1] D. Mei, B. Ashford, Y.-L. He, X. Tu, Plasma-catalytic reforming of biogas over supported Ni catalysts in a dielectric barrier discharge reactor: Effect of catalyst supports, *Plasma Process. Polym.* 14 (2017) 1600076. <https://doi.org/10.1002/ppap.201600076>.
- [2] Y. Cai, D. Mei, Y. Chen, A. Bogaerts, X. Tu, Machine learning-driven optimization of plasma-catalytic dry reforming of methane, *J. Energy Chem.* 96 (2024) 153–163.

<https://doi.org/10.1016/j.jechem.2024.04.022>.

[3] M.A.N. Dewapriya, R.K.N.D. Rajapakse, W.P.S. Dias, Characterizing fracture stress of defective graphene samples using shallow and deep artificial neural networks, *Carbon* 163 (2020) 425–440. <https://doi.org/10.1016/j.carbon.2020.03.038>.

[4] Y. Wang, C. Ling, H. Yin, W. Liu, Z. Tang, Z. Li, Thermophysical properties of KCl-NaF reciprocal eutectic by artificial neural network prediction and experimental measurements, *Solar Energy* 204 (2020) 667–672. <https://doi.org/10.1016/j.solener.2020.05.029>.

[5] Y. Wang, Y. Chen, J. Harding, H. He, A. Bogaerts, X. Tu, Catalyst-free single-step plasma reforming of CH<sub>4</sub> and CO<sub>2</sub> to higher value oxygenates under ambient conditions, *Chem. Eng. J.* 450 (2022) 137860. <https://doi.org/10.1016/j.cej.2022.137860>.

[6] R.S. Sutton, D. McAllester, S. Singh, Y. Mansour, Policy Gradient Methods for Reinforcement Learning with Function Approximation, in: *Advances in Neural Information Processing Systems*, MIT Press, 1999. [https://proceedings.neurips.cc/paper\\_files/paper/1999/hash/464d828b85b0bed98e80ade0a5c43b0f-Abstract.html](https://proceedings.neurips.cc/paper_files/paper/1999/hash/464d828b85b0bed98e80ade0a5c43b0f-Abstract.html).

[7] J. Schulman, P. Moritz, S. Levine, M. Jordan, P. Abbeel, High-Dimensional Continuous Control Using Generalized Advantage Estimation, (2018). <https://doi.org/10.48550/arXiv.1506.02438>.

[8] J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba, P. Abbeel, Domain Randomization for Transferring Deep Neural Networks from Simulation to the Real World, (2017). <https://doi.org/10.48550/arXiv.1703.06907>.