ARTÍCULO DE INVESTIGACIÓN / RESEARCH ARTICLE
https://dx.doi.org/10.14482/inde.44.01.103.911

A Reinforcement Learning-Based Approach for Retail Electricity Pricing Strategy for Multi-Microgrid Distribution Systems

Estrategia de asignación de precios minoristas de electricidad basada en aprendizaje por refuerzo para sistemas de distribución con múltiples microrredes

Juan M. Rey*
Camilo Carrillo-Valera**
Jhon Sandoval-Manrique***
Andrés Luna****
Iván D. Serna-Suárez*****

* Associate Professor, Universidad Industrial de Santander (UIS), Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones (E3T), Bucaramanga (Colombia). PhD in Electronic Engineering. Orcid-ID: https://orcid.org/0000-0002-5465-4769. juanmrey@uis.edu.co

** Electrical Engineering Master's Student, Universidad Industrial de Santander (UIS), Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones (E3T), Bucaramanga (Colombia). Electronics Engineer. Orcid-ID: https://orcid.org/0009-0004-7848-5009. camilo2248093@correo.uis.edu.co

*** Electrical Engineering Master's Student, Universidad Industrial de Santander (UIS), Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones (E3T), Bucaramanga (Colombia). Electronics Engineer. Orcid-ID: https://orcid.org/0009-0008-1895-9398. jhon2248094@correo.uis.edu.co

**** Specialist in Electrical Power Distribution Systems, Universidad Industrial de Santander (UIS), Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones (E3T), Bucaramanga (Colombia). Electrical Engineer. Orcid-ID: https://orcid.org/0009-0005-4273-703X. andres.luna@correo.uis.edu.co

***** Assistant Professor, Universidad Industrial de Santander (UIS), Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones (E3T), Bucaramanga (Colombia). PhD in Electrical Engineering. Orcid-ID: https://orcid.org/0000-0003-4079-2252. idsersua@uis.edu.co

Correspondencia: Juan M. Rey. LP 204, Carrera 27 calle 9. Bucaramanga (Colombia). Teléfono: +57 (607) 6344000, Ext. 1525.

Abstract

This paper presents a retail electricity pricing strategy for multi-microgrid distribution systems based on a policy-driven reinforcement learning algorithm. Developed from the perspective of the distribution system operator (DSO), the approach enables the practical derivation of a set of hourly electricity prices that simultaneously maximize profit from energy exchanges and minimize the system's peak-to-average load ratio, thereby flattening the aggregate load profile. To address the absence of a complete system model from the DSO's viewpoint, the training process employs a Monte Carlo-based method that generates synthetic data from representative base profiles, enabling extensive interaction between the DSO and its environment. Simulations are conducted to validate the effectiveness of the proposed method. Additionally, a sensitivity analysis is presented to evaluate the influence of key parameters on the training performance and the strategy's effectiveness.

Keywords: distribution systems, microgrids, reinforcement learning, retail electricity pricing.

Resumen

Este artículo presenta una estrategia de asignación de precios minoristas de electricidad para sistemas de distribución con múltiples microrredes, basada en un algoritmo de aprendizaje por refuerzo orientado por políticas. Desarrollado desde la perspectiva del operador del sistema de distribución, este enfoque permite determinar de forma práctica un conjunto de precios horarios de electricidad que, simultáneamente, maximiza el beneficio del intercambio de energía y minimiza la relación pico-promedio de la carga del sistema, aplanando así el perfil de carga agregado. Para abordar la ausencia de un modelo completo del sistema desde el punto de vista del operador del sistema de distribución, el proceso de entrenamiento emplea un método basado en Monte Carlo que genera datos sintéticos a partir de perfiles base representativos, lo que permite una amplia interacción entre el operador y su entorno. Se realizaron simulaciones para validar la efectividad del método propuesto. Adicionalmente, se presenta un análisis de sensibilidad que permite evaluar la influencia de parámetros claves en el desempeño del entrenamiento y la eficiencia de la estrategia.

Palabras clave: aprendizaje por refuerzo, asignación de precios minoristas de electricidad, microrredes, sistemas de distribución.

Fecha de recepción: 28 *de marzo de* 2025
Fecha de aceptación: 18 *de julio de* 2025

INTRODUCTION

Microgrids are electrical networks comprising distributed energy resources (DERs), energy storage systems (ESSs), and clusters of local loads that function as a single controllable entity [1], [2]. With the increasing penetration of renewable energy, microgrids have gained popularity for enhancing the reliability and flexibility of distribution systems through improved controllability [3], [4]. However, multi-microgrid distribution systems (MMDS), i.e., distribution systems with multiple microgrids connected at different nodes, introduce complex challenges for managing energy exchange [5]-[8]. Factors such as the intermittency of renewable sources, the stochastic nature of load demand, and the privacy constraints of microgrids make retail electricity pricing a difficult task for distribution system operators (DSOs) [9], [10].

Advanced metering infrastructure (AMI) in smart grids provides users with real-time electricity pricing and detailed insights into their power consumption, enabling strategic decisions to reduce energy costs [11], [12]. A well-designed retail electricity pricing strategy serves as a flexible tool to achieve specific operational goals within microgrids. For example, it can reduce load consumption during peak hours, flattening the load curve and reducing stress on distribution systems. In multi-microgrid distribution systems (MMDS), it can also encourage self-consumption based on resource availability. Additionally, AMI data enables DSOs to analyze load patterns and user consumption behaviors, improving system awareness and mitigating failures from unexpected events [13]. In this context, AMI-based retail electricity pricing plays a crucial role in ensuring the reliable and secure operation of MMDS [14], [15].

Due to data privacy and ownership concerns, DSOs typically lack access to detailed information about the topology, configuration, or models beyond the Point of Common Coupling (PCC) of each microgrid. This limitation restricts the use of conventional model-based techniques that rely on precise system data and predictions. In contrast, data-driven techniques such as Reinforcement Learning (RL) have emerged as a promising alternative, enabling effective grid operation without requiring comprehensive modeling of all distribution system components [16].

In recent years, the application of Reinforcement Learning (RL) has gained traction in addressing various control and decision-making challenges in distribution systems with high renewable energy penetration. Notable contributions can be found in areas such as online microgrid scheduling [17], real-time energy management [18], [19], power management of networked microgrids [20], and the operation of EV charging stations [21], among others.

Several studies have proposed RL-based dynamic pricing strategies that rely on a service provider acting as an intermediary between the utility company and customers

[22]-[25]. In [22], a service provider is trained to manage the energy exchange of multiple microgrids, optimizing their collective power transactions with the main grid. Similarly, [23] introduces a reinforcement and imitation learning approach to train a multi-objective retail broker, aiming to maximize economic benefits while balancing energy supply and demand under operational constraints. Additionally, in [24] presents an RL-based dynamic pricing model where a service provider adjusts retail electricity prices based on customers' load demand levels. While these approaches leverage RL, they are not explicitly designed from the DSO's perspective to formulate an MMDS pricing strategy that maximizes total profit while enhancing key distribution network operational objectives.

Some studies formulate a bi-level problem consisting of a high and a low level. For instance, in [26], [27], the DSO utilizes RL at the higher level to set retail electricity prices. In [26], training uncertainties are mitigated through interval predictions. Similarly, [27] employs an interactive mechanism based on a leader-multi-follower Stackelberg game, where the DSO acts as the leader. However, both approaches primarily focus on cost considerations while overlooking key operational objectives, such as flattening consumption curves.

Many studies use simulation-based models to train the DSO in managing MMDS responses. For example, [28] proposes a retail electricity pricing strategy that employs a deep neural network-based model to predict microgrid responses, ensuring customer privacy during RL agent training. Similarly, [29] presents an interactive MMDS pricing method aimed at minimizing deviations between real-time and day-ahead load curves. In this approach, the DSO optimizes profits using estimations from a deep learning-based interaction model. However, these methods may face significant uncertainties due to the unpredictable expansion of generation systems and the lack of customer-side information.

Finally, in [15] a pricing scheme is proposed to encourage consumer participation in demand response by offering a selection of pricing plan options. Users are categorized using a classification algorithm, allowing for tailored pricing strategies. While this work primarily focuses on load response, it stands out for its practical implementation.

As observed in the review presented above, modeling and predicting microgrids behavior within multi-microgrid distribution systems (MMDS) remains a challenging and active area of research. Various methods continue to be developed to enhance prediction accuracy, particularly under conditions of limited or uncertain data. Therefore, inspired by previous research and seeking to contribute to the state of the art, this paper proposes a retail electricity pricing strategy for MMDS based on a policy-driven reinforcement learning (RL) algorithm.

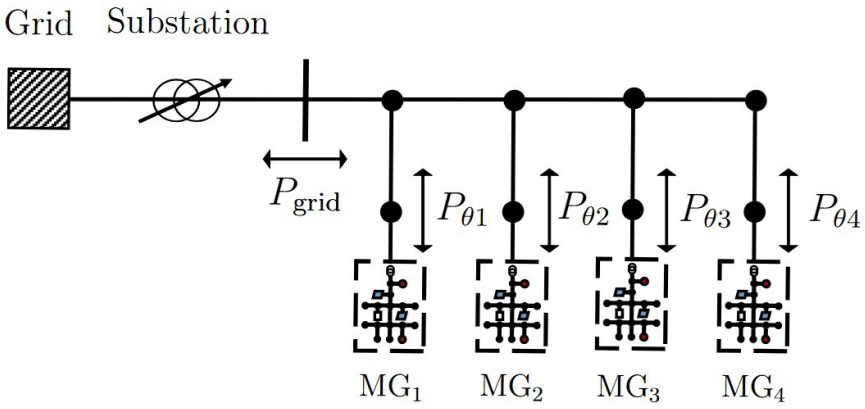
The proposed strategy is designed from the perspective of the Distribution System Operator (DSO), who seeks to dynamically determine a sequence of 24-hourly electricity prices. The reward function is designed to maximize energy exchange profits while simultaneously minimizing the system's peak-to-average load ratio (PAR), a critical metric for reducing operational stress and improving grid reliability.

A key advantage of the proposed approach lies in its practical application, as it can be integrated with any simulation framework capable of capturing microgrid dynamics within MMDS. Additionally, the resulting pricing policy can serve as a basis for other dynamic operational controls and retraining methods in response to changing system conditions.

The rest of the paper is organized as follows: Section 2 presents the modeling of the MMDS. Section 3 describes in detail the formulation of the RL problem, including the reward and the pricing policy. Then, Section 4 describes the training considerations. The simulations conducted to validate the proposed method and an analysis of the impact of some key parameters are presented and discussed in Section 5. Finally, the main conclusions are drawn in Section 6.

MODELING OF THE MULTI-MICROGRID DISTRIBUTION SYSTEM

This section presents the modeling of the MMDS considered in this study. Fig. 1 illustrates a schematic of a radial distribution system comprising four interconnected microgrids. The thick dots represent nodes, connected by lines that denote line impedances. It is assumed that a communication channel exists, enabling microgrids to receive real-time retail price updates from the DSO.



Source: own elaboration.

Figure 1. Scheme of a radial multi-microgrid distribution system

Microgrid Modeling

For a given microgrid, the dispatch problem can be stated as:

$$\min_{P_D,\, P_\theta} \{C_D + C_\theta\} \tag{1}$$

s.t.

$$P_D + P_\theta + p_R = p_L \tag{2}$$

$$P_D^{min.} \leq P_D \leq P_D^{max.} \tag{3}$$

where C_D is the cost of operation of the dispatchable sources (diesel generation) and C_θ is the cost of the exchanged power. Regarding the constraints in (2) and (3), P_D is the hourly energy production of the dispatchable sources of the microgrid (diesel generation), P_θ is the hourly energy imported from the grid, p_R is the hourly renewable energy production and p_L is the hourly load consumption. Note that constraint (2) corresponds to the power balance of the microgrid. Also, constraint (3) models the operative limits of the dispatchable diesel generator, where $P_D^{min.}$ and $P_D^{max.}$ are the minimum and maximum values, respectively.

The following assumptions are made in the proposed formulation:

- ***Negligible internal power losses***: Power losses within each microgrid are assumed to be negligible. Although it would be possible to account for these losses by solving detailed internal power flow equations, a simplified mathematical model is employed during training to estimate P_θ. This simplification is justified by the fact that microgrids typically exhibit lower internal losses compared to conventional power grids.
- ***Neglect of reactive power***: While reactive power management plays an important role in the safe and efficient operation of distribution systems, its exclusion allows the analysis to remain focused on the global impact of the pricing mechanism on active power. This simplification enables a clearer interpretation of the results without undermining the generality of the conclusions.
- ***Exclusion of energy storage systems***: For the training of the proposed strategy, microgrids are modeled as prosumers connected at specific nodes, from which hourly exchanged power data are derived. From the grid operator's perspective, increasing the internal complexity of microgrids, such as by including storage, does not significantly alter the operational logic of the strategy.

In this dispatch problem, the energy management system (EMS) of each microgrid seeks to minimize the total cost, defined as the sum of C_D and C_θ. These terms are calculated as follows: C_D is modeled as a quadratic function of the dispatched power P_D, as:

$$C_D = a \cdot P_D^2 + b \cdot P_D + c \tag{4}$$

where a, b and c are constant coefficients which can be estimated experimentally [30]-[32]. Whereas C_θ is calculated using the hourly retail price π and the traded power P_θ as follows:

$$C_\theta = \pi \cdot P_\theta \tag{5}$$

The renewable power generation p_R is the addition of the power of the PV system p_{pv} and the power of the wind system p_w:

$$p_R = p_{pv} + p_w \tag{6}$$

For the PV system, the following simplified model is used:

$$p_{pv} = n_{pv} \cdot A_{pv} \cdot \eta \cdot I \tag{7}$$

where n_pv is the number of PV panels of the microgrid, Apv is the superficial area of each PV panel, η is the conversion efficiency and I is the solar irradiation [33], [34]; while for the wind power, the following simplified model is used

$$p_w = n_w \cdot \frac{1}{2} \cdot \rho \cdot A_w \cdot C_{pw} \cdot v^3 \tag{8}$$

where n_w is the number of wind generators of the microgrid, ρ is the density of air, A_w is the swept area of the wind turbine blades, C_{pw} is a power conversion coefficient, and v is the wind speed, which is used in a cubic form [35].

Notice that the retail price π is determined by the DSO. Thus, once this value is assigned, the traded power P_θ for each microgrid m corresponds to $P_{\theta m}$ from the DSO's perspective.

The complete MMDS consists of multiple microgrids, each with distinct generation and load configurations. Table 1 summarizes the key characteristics of these configurations for the numerical example of four interconnected microgrids used in the following sections. The nominal power of each microgrid load will be used to generate a per-unit load profile for the training stage, as detailed in Section 4.

Table 1. Microgrid configurations

MG	PV panels[1] (n_{pv})	Wind generators[2] (n_w)	Diesel Gen.[3] (n_d)	Nominal power of MG load [kW]
1	15	11	1	25
2	28	19	1	45
3	39	26	1	65
4	24	17	1	40

Note
1 Nominal power of each PV panel: 0.306 [kW].
2 Nominal power of each wind generator: 0.303 [kW].
3 Nominal power of each diesel generator: 30 [kW].
Source: own elaboration.

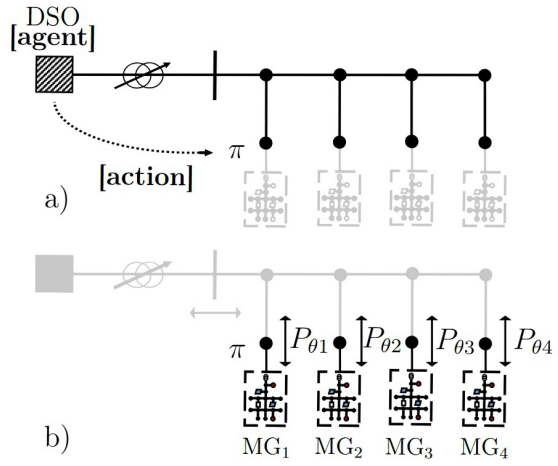
Distribution System Modeling

The distribution system is modeled as a balanced radial three-phase grid. Once the exchanged power for each microgrid is determined, the distribution system's power flow is solved to obtain the total power exchanged at the substation, P_{grid} (see Fig. 1). This value is then used to compute the reward, as described in the next section.

REINFORCEMENT LEARNING FORMULATION

To formulate retail electricity pricing as a reinforcement learning problem, it is essential to define its key elements. Figure 2 illustrates this framework:

- The **agent** is the DSO, whose **action** is to set the hourly retail electricity price π for the microgrids connected to the distribution system (the **environment**), as shown in Fig. 2(a).
- Given this price, each microgrid's EMS determines its traded power by solving the optimization problem described in subsection 2.1, as depicted in Fig. 2(b).



Source: own elaboration.

Figure 2. Reinforcement learning formulation: (a) action of the agent, (b) environment

Due to privacy constraints, the DSO lacks knowledge of the internal topology and operational details of each microgrid (represented in light gray in the figure). However, for evaluation purposes in this study, each microgrid is modeled to obtain its respective output $P_{\theta m}$. The total traded power P_{grid} is then computed by solving the distribution system's power flow equations using the power injections $P_{\theta m}$ from all microgrids.

Reward

The **reward** function is formulated to maximize profit from power exchanges with microgrids while simultaneously smoothing the load profile by minimizing the peak-to-average ratio (PAR) over a daily cycle:

$$R = \alpha \cdot B - (1-\alpha) \cdot PAR \tag{9}$$

where B is the normalized profit of the DSO, PAR is the normalized peak-to-average ratio, and α is a weight introduced to maintain a balance between both objectives. Note that α can be adjusted according to system requirements.

In (9), the profit B is defined as follows:

$$B_\tau = \sum_{t=\tau}^{\tau+23} \lambda^{t-\tau} \cdot \pi_t \cdot \frac{P_{\text{grid},t}}{B_{\text{base}}} \tag{10}$$

where λ corresponds to a discount factor in range of [0, 1], and B_base is a regularization term. A window of 24 hours is used to compute the benefits for t=τ. The effect of λ is discussed in the next section.

On the other hand, the PAR, which is the relationship between the maximum and the average value of the hourly total power exchanged in a daily cycle, is defined as:

$$PAR = \frac{\max P_{grid}}{\text{mean } P_{grid}} \tag{11}$$

The PAR must be over an entire daily cycle rather than on an hourly basis. Consequently, the training process will consist of 24-hour episodes. Additionally, PAR is not explicitly expressed in terms of the retail price π, which serves as the DSO's decision variable. This key aspect complicates solving the problem using traditional linear techniques.

Pricing Policy

The proposed formulation aims to obtain a daily pricing policy **Π** consisting of 24-hourly prices. This policy maximizes the total profit from power sales while minimizing the PAR, aligning with the defined objective function. These values are calculated for a daily cycle rather than individually for each hour. For this reason, the training algorithm must account for long-term effects, where the discount factor λ plays a crucial role. Additionally, the pricing policy **Π** can be updated through a retraining algorithm or used as an initial policy for a dynamic operational control strategy.

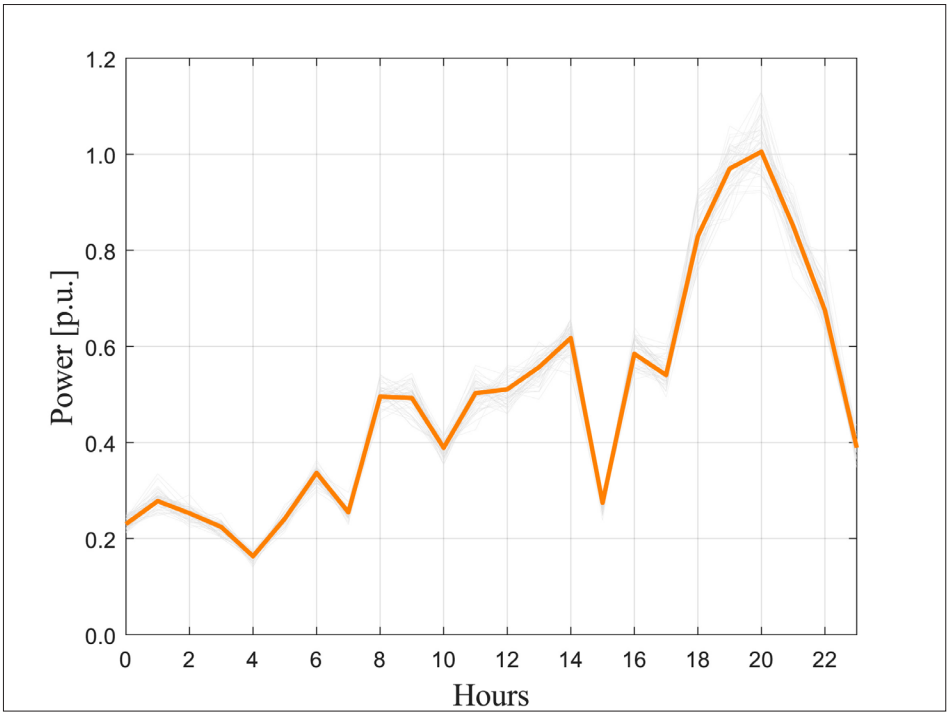
TRAINING METHOD

The lack of a complete model from the DSO perspective poses challenges in formulating an explicit transition probability function. To overcome this, a Monte Carlo-based training approach is employed, enabling extensive interaction between the DSO and its environment. In order to capture the stochastic nature of energy resources and loads, representative base profiles or *seed profiles* are defined, from which *synthetic* data are generated by introducing random variations using a normal Gaussian distribution.

For example, in the case of load profiles, Fig. 3 presents the per-unit seed day (solid line), derived from a typical residential consumption curve in Colombia. To simulate realistic day-to-day variability, more than 50 synthetic daily profiles are generated by multiplying each point on the *seed profile* by a random factor drawn from

a normal distribution with a mean (μ) of 1 and a standard deviation (σ) of 0.05. In this context, μ= 1 ensures that the average of the generated profiles remains centered around the original seed curve, while σ= 0.05 introduces a 5% variability around that average, capturing the natural fluctuations observed in real-world consumption.

These per-unit synthetic profiles are then scaled by the nominal power of each microgrid's load (see Table 1) for use in the training process. The same methodology is applied to generate synthetic data for solar irradiation and wind speed: in those cases, the random factors are drawn from normal distributions with μ=1 and σ=0.035; and μ=1 and σ=0.005, respectively.



Source: own elaboration.

Figure 3. Daily seed of the per-unit load profile and synthetic days obtained using a normal distribution function

The training method is described in Algorithm 1. Initially, a search space of prices π_s, is defined. These prices can be generated randomly or proportionally based on historical market data relevant to the specific application. In this study, a set of five prices was selected:

π_s={0.027,0.0285,0.030,0.0315,0.033}.

Algorithm 1: Monte Carlo training method

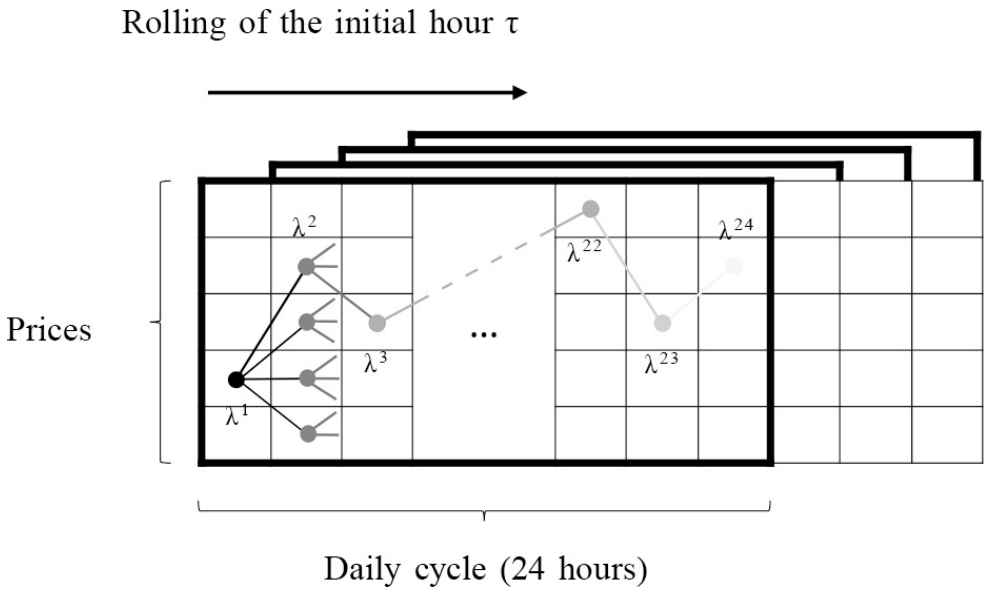
1: Set the search space of prices π_s
2: Define the number of episodes N_e
3: **for** τ = 1 to 24 (rolling initial hour) **do**
4: **for** N_e episodes **do**
5: **for** t = τ to τ+23 (daily cycle) **do**
6: Select a random price
7: Run the MG modeling to obtain $P_{\theta m}$
8: Solve the DS power flows to obtain $P_{\theta m}$
9: **end for**
10: Calculate the **reward** R for the episode
11: **end for**
12: Select the price for the hour τ and store in Π
13: **end for**

For this formulation, an episode corresponds to a run of a daily cycle from an initial hour τ to a terminal hour τ+23. The sequence in each episode is the following:

- A random price of π_s is selected.
- The DSO sets this hourly retail price and the powers $P_{\theta m}$ are obtained.
- The total traded power P_{grid} is obtained solving the distribution system power flows.

This process is repeated N_e times, calculating the reward R for each episode. The average reward of all the episodes that started in each price of π_s is obtained, and the one with the highest value is selected as the price for the hour τ in Π. For example, if for hour 1, the episodes that started with a price of 0.030 yield the highest average reward compared to the other prices, then 0.030 is assigned to the first position in the final policy Π. This procedure is repeated iteratively, rolling the initial hour to obtain the 24 prices of **Π**.

Although the hourly price selection for policy Π is made only for the initial hour of the set of episodes, the long-term effects of a price choice on daily performance can be farsighted due to the use of the discount factor λ. Fig. 4 seeks to illustrate the training method, with the gradual fading of the line's color representing the diminishing influence of future prices over time. For this work, λ is set to 0.9, a commonly used value for this type of problem.



Source: own elaboration.

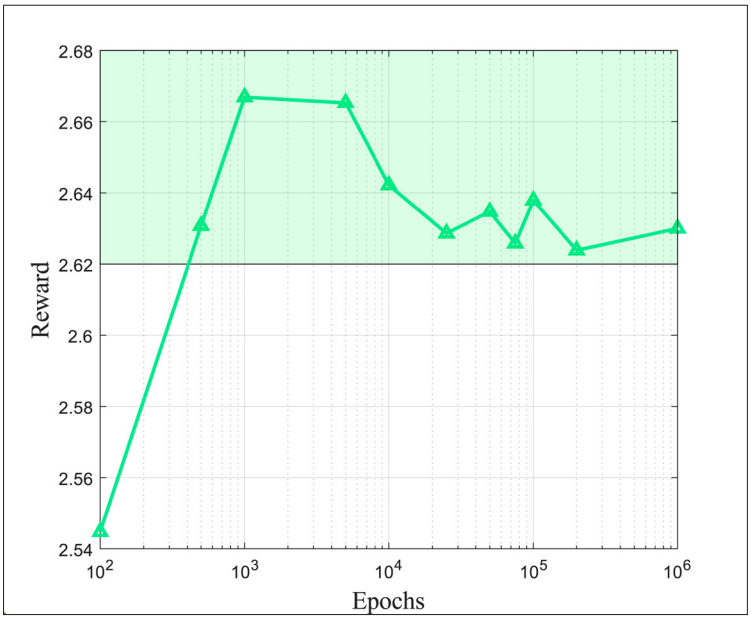
Figure 4. Illustrative diagram of the training method

RESULTS AND DISCUSSION

This section presents the results and analyzes the impact of key variables on the training process and the resulting pricing policy. For this, two synthetically generated 50-days datasets were generated: a training dataset X_{train} and a testing dataset X_{test}. These were obtained following the indications presented in the previous section.

Impact of the Selection of the Numbers of Episodes N_e over the Reward

Initially, the impact of varying the number of episodes N_e over the pricing policy selection is analyzed. The training process was repeated 50 times for each value of N_e using X_{train}, and the averages results are presented in Fig. 5.



Source: own elaboration.

Figure 5. Impact of the number of episodes simulated over the reward of price policy obtained

The results show that policies trained with a low N_e present lower rewards, as an insufficient number of episodes limits effective exploration in a large search space. Conversely, increasing N_e leads to higher rewards, with the results stabilizing within a certain range. However, excessively increasing N_e results in computationally expensive and time-consuming simulations without significant reward improvement. Based on these findings, an N_e of 5000 is used for the remaining simulations.

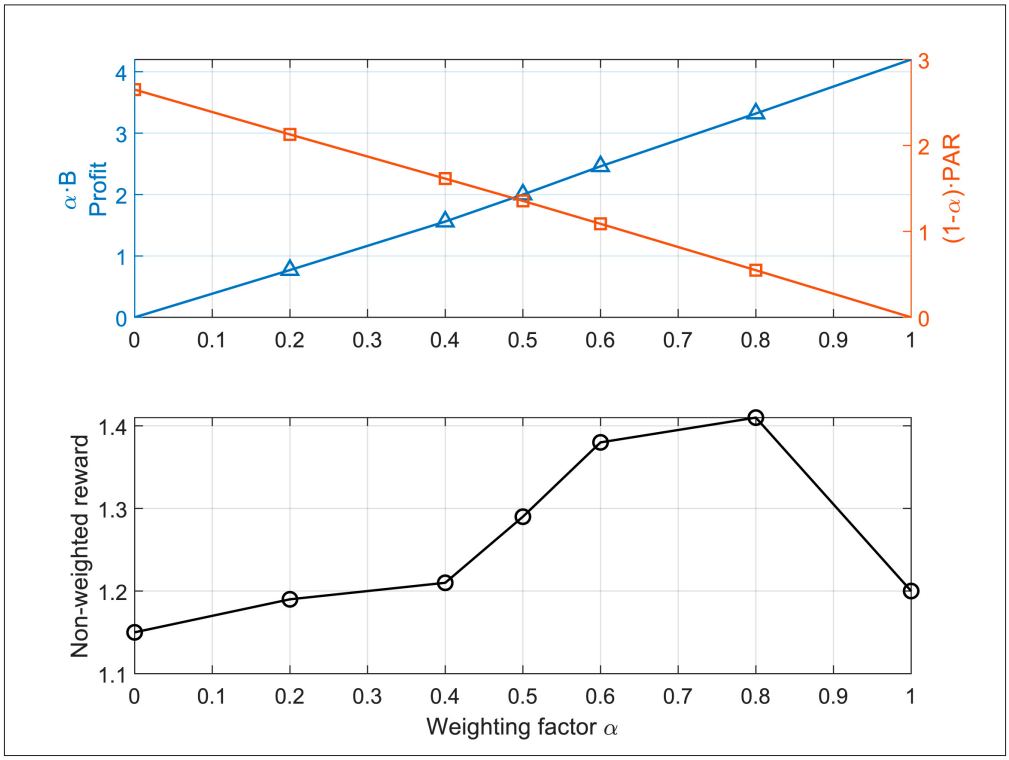
Impact of the Selection of the Weighting Factor over the Reward

Next, the relevance of the weighting factor was evaluated. To do so, the value of α was varied from 0 to 1 by steps of 0.2 (including 0.5), training for X_{train}. The results are presented in Fig. 6.

In this formulation, the objective is to maximize profit while minimizing PAR (which is always a positive number). When α = 1, the reward is entirely based on profit, maximizing the term α · B. Conversely, when α = 0, the reward depends solely on PAR, which has a negative sign in (9), leading to the smallest possible PAR value.

Fig. 6(a) illustrates these effects, where the x-axis represents α values. The left y-axis (blue) corresponds to the reward term α · B, while the right y-axis (orange) represents the reward term (1 - α)PAR.

To better visualize the influence of α, Fig. 6(b) presents non-weighted rewards. The results indicate that α = 0.8 achieves a well-balanced trade-off between the desired objectives, maximizing the non-weighted reward. Therefore, this value is used for the remaining simulations.



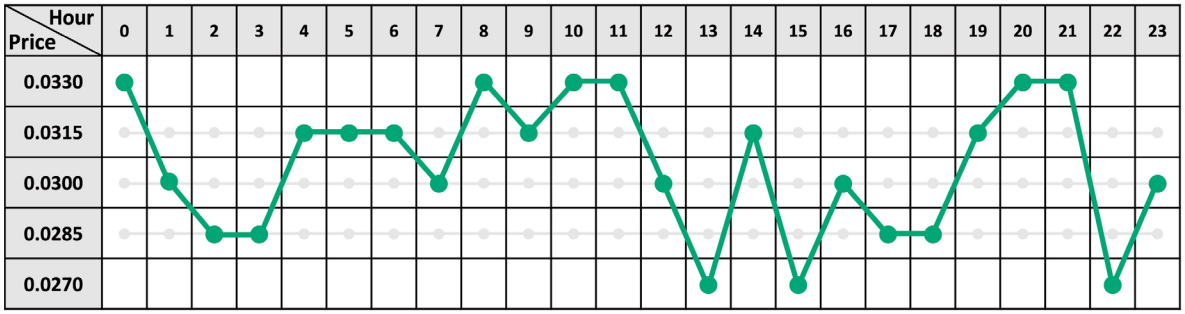
Source: own elaboration.

Figure 6. Impact of the weight factor α over: (a) the reward terms αB and (1-α)PAR, (b) the non-weighted reward

Pricing Policy Selection

Finally, a pricing policy Π was obtained using α = 0.8 and N_e = 5000 in a training using X_{train}. Fig. 7 displays the hourly prices of Π as a green line. Additionally, three constant price policies are depicted in gray lines, $\Pi_{0.0315}$, $\Pi_{0.0300}$ and $\Pi_{0.0285}$, each representing a constant price applied throughout the day.

The obtained pricing policy **Π** fluctuates between different values in the set π_s, without following an obvious pattern that could have been easily identified without the formulation and training of the proposed strategy.



Source: own elaboration.

Figure 7. Pricing policies: in green Π obtained using α = 0.8 and N_e=5000 trained with X_{train}; In gray, three constant price policies

The four pricing policies were evaluated using X_{test}, and the results are presented in Table 2. As can be seen, the pricing policy Π achieved the highest reward, outperforming some of the constant price policies by up to 8%. While these results may vary slightly with different test datasets, the overall trend confirms that Π consistently outperforms the constant pricing policies in the vast majority of cases, demonstrating the effectiveness of the proposed strategy.

Table 2. Rewards for different pricing policies

Price policy	Type	Reward R
Π	Variable	2.665
$\Pi_{0.0315}$	Constant	2.657
$\Pi_{0.0300}$	Constant	2.556
$\Pi_{0.0285}$	Constant	2.457

Source: own elaboration.

CONCLUSIONS

This work presented a retail electricity pricing strategy for MMDS based on a policy-driven RL algorithm. Developed from the perspective of the distribution system operator, the proposed strategy determines a sequence of hourly electricity prices. The policy is derived through a reward function that seeks to maximize energy exchange profits while minimizing system stress, quantified via the PAR. The obtained policy can be easily applied as a starting policy for dynamic operational control and retraining methods, in response to changing system conditions.

The effectiveness of the proposed strategy has been validated through a numerical case study. The results demonstrate that the RL agent generates a set of hourly prices for a given operating day that outperforms constant price strategy in terms of accumulated reward. Moreover, the obtained price set does not exhibit a clear trend that could have been identified without the formulation and training process.

It is important to note that certain simplifying assumptions were made in this case study. These include neglecting internal power losses within microgrids, disregarding the effects of reactive power flows, and omitting the presence of energy storage systems. However, these elements could be incorporated into a more detailed case study without affecting the findings obtained in this work.

As future work, the derived pricing policy will be further explored as a baseline for other dynamic operational control and policy retraining algorithms. This will support the broader objective of investigating how retail electricity pricing can be systematically linked to other operational aspects of distribution networks.

REFERENCES

[1] Ton, D.T., Smith, M.A., 2012. The U.S. department of energy's microgrid initiative. The Electricity Journal 25, 84–94. doi:https://doi.org/10.1016/j.tej.2012.09.013.

[2] Farrokhabadi, M., Cañizares, C.A., Simpson-Porco, J.W., Nasr, E., Fan, L., Mendoza-Araya, P.A., Tonkoski, R., Tamrakar, U., Hatziargyriou, N., Lagos, D., Wies, R.W., Paolone, M., Liserre, M., Meegahapola, L., Kabalan, M., Hajimiragha, A.H., Peralta, D., Elizondo, M.A., Schneider, K.P., Tuffner, F.K., Reilly, J., 2020. Microgrid stability definitions, analysis, and examples. IEEE Trans. Power Systems 35, 13–29. doi:10.1109/TPWRS.2019.2925703.

[3] Olivares, D.E., Mehrizi-Sani, A., Etemadi, A.H., Cañizares, C.A., Iravani, R., Kazerani, M., Hajimiragha, A.H., Gomis-Bellmunt, O., Saeedifard, M., Palma-Behnke, R., Jiménez-Estévez, G.A., Hatziargyriou, N.D., 2014. Trends in microgrid control. IEEE Trans. Smart Grid 5, 1905– 1919. doi:10.1109/TSG.2013.2295514.

[4] Asmus, P., 2014. Why microgrids are moving into the mainstream: Improving the efficiency of the larger power grid. IEEE Electr. Mag. 2, 12–19. doi:10.1109/MELE.2013.2297021.

[5] Kumar Nunna, H.S.V.S., Doolla, S., 2013. Multiagent-based distributed-energy-resource man- agement for intelligent microgrids. IEEE Trans. Industrial Electronics 60, 1678–1687. doi:10.1109/TIE.2012.2193857.

[6] Mao, M., Jin, P., Hatziargyriou, N.D., Chang, L., 2014. Multiagent-based hybrid energy management system for microgrids. IEEE Trans. Sustainable Energy 5, 938–946. doi:10.1109/TSTE.2014.2313882.

[7] Arefifar, S.A., Ordonez, M., Mohamed, Y.A.R.I., 2017. Energy management in multi-microgrid systems—development and assessment. IEEE Trans. on Power Systems 32, 910–922. doi:10.1109/TPWRS.2016.2568858.

[8] Jiang, W., Yang, K., Yang, J., Mao, R., Xue, N., Zhuo, Z., 2019. A multiagent-based hierarchical energy management strategy for maximization of renewable energy consumption in interconnected multi-microgrids. IEEE Access 7, 169931–169945. doi:10.1109/ACCESS.2019.2955552.

[9] Yang, J., Zhao, J., Luo, F., Wen, F., Dong, Z.Y., 2018b. Decision-making for electricity retailers: A brief survey. IEEE Trans. Smart Grid 9, 4140–4153. doi:10.1109/TSG.2017.2651499.

[10] Braithwait, S.D., 2018. Retail pricing responses to the challenge of distributed energy resources. The Electricity Journal 31, 38–43. doi:https://doi.org/10.1016/j.tej.2018.09.001.

[11] Wang, B., Guo, Q., Yu, Y., 2022. Mechanism design for data sharing: An electricity retail perspective. Applied Energy 314, 118871. doi:https://doi.org/10.1016/j.apenergy.2022.118871.

[12] Thirugnanam, K., Moursi, M.S.E., Khadkikar, V., Zeineldin, H.H., Al Hosani, M., 2021. Energy management of grid interconnected multi-microgrids based on p2p energy ex- change: A data driven approach. IEEE Trans. Power Systems 36, 1546–1562. doi:10.1109/TPWRS.2020.3025113.

[13] Qiu, D., Wang, Y., Wang, J., Jiang, C., Strbac, G., 2023. Personalized retail pricing design for smart metering consumers in electricity market. Applied Energy 348, 121545. doi:https://doi.org/10.1016/j.apenergy.2023.121545.

[14] Trujillo-Baute, E., del Río, P., Mir-Artigues, P., 2018. Analysing the impact of renewable energy regulation on retail electricity prices. Energy Policy 114, 153–164. doi:https://doi.org/10.1016/j.enpol.2017.11.042.

[15] Yang, H., Zhang, J., Qiu, J., Zhang, S., Lai, M., Dong, Z.Y., 2018a. A practical pricing approach to smart grid demand response based on load classification. IEEE Trans. Smart Grid 9, 179–190. doi:10.1109/TSG.2016.2547883.

[16] Liu, W., Zhuang, P., Liang, H., Peng, J., Huang, Z., 2018. Distributed economic dispatch in microgrids based on cooperative reinforcement learning. IEEE Trans. Neural Networks and Learning Systems 29, 2192–2203. doi:10.1109/TNNLS.2018.2801880.

[17] Shuai, H., He, H., 2021. Online scheduling of a residential microgrid via monte-carlo tree search and a learned model. IEEE Trans. Smart Grid 12, 1073–1087. doi:10.1109/TSG.2020.3035127.

[18] Kuznetsova, E., Li, Y.F., Ruiz, C., Zio, E., Ault, G., Bell, K., 2013. Reinforcement learning for microgrid energy management. Energy 59, 133–146. doi:https://doi.org/10.1016/j.energy.2013.05.060.

[19] Ye, Y., Qiu, D., Wu, X., Strbac, G., Ward, J., 2020. Model-free real-time autonomous control for a residential multi-energy system using deep reinforcement learning. IEEE Trans. Smart Grid 11, 3068–3082. doi:10.1109/TSG.2020.2976771.

[20] Zhang, Q., Dehghanpour, K., Wang, Z., Huang, Q., 2020. A learning-based power management method for networked microgrids under incomplete information. IEEE Trans. Smart Grid 11, 1193–1204. doi:10.1109/TSG.2019.2933502.

[21] Lee, S., Choi, D.H., 2021. Dynamic pricing and energy management for profit maximization in multiple smart electric vehicle charging stations: A privacy-preserving deep reinforcement learn- ing approach. Applied Energy 304, 117754. doi:https://doi.org/10.1016/j.apenergy.2021.117754.

[22] Ghanbari, S., Bahramara, S., Golpîra, H., 2024. Modeling market trading strategies of the intermediary entity for microgrids: A reinforcement learning-based approach. Electric Power Systems Research 227, 109989. doi:https://doi.org/10.1016/j.epsr.2023.109989.

[23] Zhang, Y., Yang, Q., Li, D., An, D., 2022. A reinforcement and imitation learning method for pricing strategy of electricity retailer with customers' flexibility. Applied Energy 323, 119543. doi:https://doi.org/10.1016/j.apenergy.2022.119543.

[24] Kim, B.G., Zhang, Y., van der Schaar, M., Lee, J.W., 2016. Dynamic pricing and energy con- sumption scheduling with reinforcement learning. IEEE Trans. Smart Grid 7, 2187–2198. doi:10.1109/TSG.2015.2495145.

[25] Lu, R., Hong, S.H., Zhang, X., 2018. A dynamic pricing demand response algo- rithm for smart grid: Reinforcement learning approach. Applied Energy 220, 220–230. doi:https://doi.org/10.1016/j.apenergy.2018.03.072.

[26] Xiong, L., Tang, Y., Mao, S., Liu, H., Meng, K., Dong, Z., Qian, F., 2022. A two-level en- ergy management strategy for multi-microgrid systems with interval prediction and reinforce- ment learning. IEEE Trans. Circuits and Systems I: Regular Papers 69, 1788–1799. doi:10.1109/TCSI.2022.3141229.

[27] Guo, C., Wang, X., Zheng, Y., Zhang, F., 2021. Optimal energy management of multi-microgrids connected to distribution system based on deep reinforcement learning. International Journal of Electrical Power Energy Systems 131, 107048. doi:https://doi.org/10.1016/j.ijepes.2021.107048.

[28] Du, Y., Li, F., 2020. Intelligent multi-microgrid energy management based on deep neural network and model-free reinforcement learning. IEEE Trans. Smart Grid 11, 1066– 1076. doi:10.1109/TSG.2019.2930299.

[29] Peng, D., Xiao, H., Pei, W., Kong, L., 2021. Interactive pricing optimization of multi-microgrid based on deep learning, in: 2021 IEEE 1st Int. Conf. on Digital Twins and Parallel Intelligence (DTPI), pp. 82–85. doi:10.1109/DTPI52967.2021.9540113.

[30] Bukar, A.L., Tan, C.W., Lau, K.Y., 2019. Optimal sizing of an autonomous photo- voltaic/wind/battery/diesel generator microgrid using grasshopper optimization algorithm. Solar Energy 188, 685–696. doi:https://doi.org/10.1016/j.solener.2019.06.050.

[31] Jiménez-Vargas, I., Rey, J.M., Osma-Pinto, G., 2023. Sizing of hybrid microgrids considering life cycle assessment. Renewable Energy 202, 554–565. doi:10.1016/j.renene.2022.11.103.

[32] Lujano-Rojas, J.M., Monteiro, C., Dufo-López, R., Bernal-Agustín, J.L., 2012. Optimum load management strategy for wind/diesel/battery hybrid power systems. Renewable Energy 44, 288– 295. doi:https://doi.org/10.1016/j.renene.2012.01.097.

[33] Rey, J.M., Jiménez-Vargas, I., Vergara, P.P., Osma-Pinto, G., Solano, J., 2022. Sizing of an autonomous microgrid considering droop control. International Journal of Electrical Power Energy Systems 136, 107634. doi:https://doi.org/10.1016/j.ijepes.2021.107634.

[34] Ayop, R., Tan, C.W., 2017. A comprehensive review on photovoltaic emulator. Ren. and Sustainable Energy Reviews 80, 430–452. doi:https://doi.org/10.1016/j.rser.2017.05.217.

[35] Rey, J.M., Vergara, P.P., Solano, J., Ordóñez, G., 2019. Design and optimal sizing of microgrids, in: Microgrids Design and Implementation. Springer, pp. 337–367.