

Article

Reinforcement Learning-Based Energy Storage Management for Microgrid Power Exchanges

Federico Perquoti ¹, Davide Milillo ¹, Lorenzo Sabino ¹, Michele Quercio ¹, Francesco Riganti Fulginei ^{1,*}, George Cristian Lazaroiu ² and Fabio Crescimbin ¹

¹ Department of Industrial, Electronic and Mechanical Engineering, Roma Tre University, 00146 Rome, Italy; fed.perquoti@stud.uniroma3.it (F.P.); davide.milillo@uniroma3.it (D.M.); lorenzo.sabino@uniroma3.it (L.S.); michele.quercio@uniroma3.it (M.Q.); fabio.crescimbin@uniroma3.it (F.C.)

² Department of Power Systems, National University of Science and Technology POLITEHNICA Bucharest, 060042 Bucharest, Romania

* Correspondence: francesco.rigantifulginei@uniroma3.it

Abstract

Intelligent energy management systems are increasingly necessary for integrating renewable energy sources within microgrids. This paper investigates the application of a reinforcement learning (RL) neural network to optimize the operation of an electrochemical storage system in an environment composed of residential loads, commercial loads, and a photovoltaic plant, all connected to the grid. A dataset combining market purchase prices, photovoltaic generation, and residential and commercial load profiles was generated and used to train a Twin Delayed Deep Deterministic Policy Gradient (TD3) agent with the primary goal of deriving a reliable and adaptive post-training policy capable of maximizing photovoltaic self-consumption, minimizing operational costs through intelligent price arbitrage, and ensuring strict compliance with battery physical constraints. The system state includes battery state of charge, load demand, PV generation, and normalized market purchase prices, whereas the action represents the battery's charge/discharge power, which is restricted from exporting energy to the grid. Results show that the agent learns to effectively store surplus PV energy and minimize grid dependency through dynamic charge management. The proposed approach outperforms strategies based solely on storing surplus self-generated energy and maintains the battery within safe operational limits. Tests with previously unseen data demonstrate robust, adaptive, and economically efficient energy management, highlighting the potential of reinforcement learning in intelligent energy systems.

Keywords: renewable; microgrids; reinforcement learning; storage system



Academic Editors: Yichen Shen,
Alexis Pengfei Zhao and Shuangqi Li

Received: 12 January 2026

Revised: 24 February 2026

Accepted: 3 March 2026

Published: 9 March 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

The increasing penetration of renewable energy sources in power systems, particularly solar and wind, has introduced new challenges due to their intrinsic intermittency and variability, as energy generation strongly depends on weather conditions that can change significantly within minutes. In this context, energy storage systems (ESSs) play a fundamental role, enabling temporal shifting of renewable energy to periods of demand and improving grid stability. Their main benefits include peak shaving, increased self-consumption, and enhanced integration of renewable resources. However, their deployment also presents challenges, including high investment costs and progressive degradation of electrochemical components over time [1].

Microgrids represent an effective solution for locally managing distributed energy resources, storage systems, and loads. They can operate both isolated and grid-connected, providing flexibility, resilience, and improved utilization of renewable energy. A microgrid capable of intelligently coordinating its components enhances self-consumption, reduces grid stress, and supports ancillary services such as demand response [2].

Although machine learning techniques such as Random Forest and LSTM networks have shown good performance in forecasting energy loads and renewable generation, they present limitations when applied to highly volatile energy exchanges. These models are generally based on historical data and assume that past patterns can repeat, whereas energy markets are inherently nonlinear, unpredictable, and subject to unforeseen events such as geopolitical crises, sudden production outages, or regulatory changes. Such conditions can cause rapid and extreme price fluctuations, making predictions from purely data-driven models often unreliable. Moreover, LSTM models can suffer from overfitting to past time series, and Random Forests, while robust to noise, are not designed to capture market dynamics beyond the observed data. These challenges motivate the use of adaptive approaches capable of continuous interaction with the environment, such as reinforcement learning [3].

Deep Reinforcement Learning (DRL) is an advanced machine learning approach in which an agent learns optimal decision-making strategies through continuous interaction with the environment, maximizing a reward function defined during design [4–10]. Unlike traditional predictive models, DRL develops adaptive strategies capable of responding in real time to highly unpredictable, nonlinear, and sometimes seemingly irrational conditions typical of energy markets [11–31]. In microgrids, the model state represents the current state of the local energy system, including the photovoltaic generation, total load, electricity buy prices, and the battery state of charge (SoC). Since SoC is the main state variable in the framework, its accurate estimation is crucial; recent literature highlights the importance of advanced model-based and physics-informed machine learning SoC estimation techniques [32–38]. The agent's action represents the power of charging or discharging the battery, while the reward function evaluates its performance to maximize self-consumption within operational constraints. This approach enables robust, adaptive and real-time optimized energy management strategies, overcoming the limitations of models based solely on historical data.

In summary, this study proposes a DRL framework to optimize the energy management of a grid-connected microgrid. The main contributions and novelties of this work are as follows: (i) the development of a realistic microgrid simulation that includes PV generation, aggregated residential and commercial loads, and an electrochemical lithium-ion battery system; (ii) the training of a TD3 agent to optimize battery operation for self-consumption and price-based energy arbitrage; (iii) the demonstration of the agent's "zero-shot" robustness by testing the learned policy on entirely unseen, future market data from 2025; and (iv) the validation of the agent's intrinsic capability to automatically adapt to drastic hardware component resizings (e.g., increased PV or battery capacity) without the need for any retraining. The findings provide insights into the practical potential of reinforcement learning for scalable microgrid energy management and highlight its suitability for real-time deployment.

2. Materials and Methods

2.1. System Description

The investigated microgrid, whose schematic architecture is depicted in Figure 1, consists of a set of residential and commercial loads, a photovoltaic (PV) system, and an electrochemical energy storage system. It can operate either in an isolated mode or

connected to the main distribution grid. The objective is to intelligently manage the power exchanges of the storage system through a reinforcement learning-based controller.

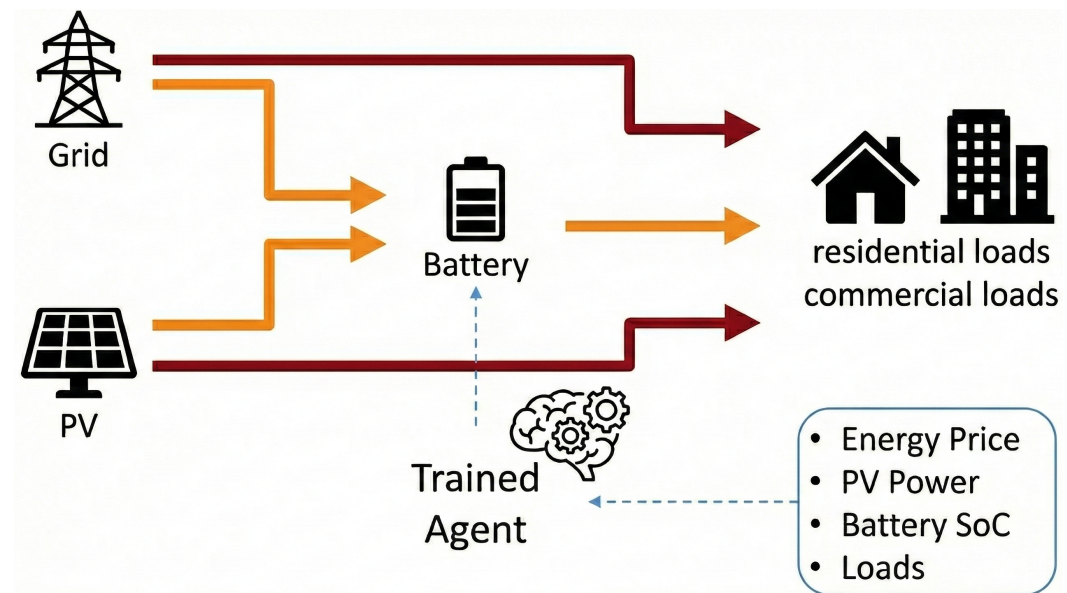


Figure 1. Schematic representation of the microgrid architecture.

2.2. Dataset Generation

The dataset used for training the RL agent was constructed by integrating multiple sources of information that describe the main physical and economic drivers of microgrid operation.

2.2.1. Market Prices

The hourly zonal purchase prices for Central-Southern Italy were downloaded from the Italian Electricity Market Operator (GME) for the period 1 January 2014 to 31 December 2024. This extended time frame was intentionally chosen to expose the learning agent to a wide range of market conditions and price dynamics. Covering more than a decade allows the model to capture both regular seasonal patterns and exceptional events that have significantly affected electricity prices. The moving average is calculated over the 12 h preceding the considered instant to help the TD3 agent discriminate between transient spikes and sustained price movements, enabling more informed decisions such as opportunistic grid charging during prolonged low-price periods while avoiding overreactions to isolated single-hour anomalies.

2.2.2. Photovoltaic Generation

Photovoltaic generation was modeled on the basis of normalized solar irradiance profiles corresponding to the installed photovoltaic system's peak power. Random noise was added to replicate different daily scenarios: sunny, partially cloudy, or fully cloudy days. Seasonal variability was also considered, adjusting irradiance levels according to the time of year, to create a more realistic and comprehensive data set. While the current implementation applies a uniform random noise to introduce daily variability, future developments will explore the use of coloured noises (e.g., to simulate the low-frequency dynamics of passing clouds) to further enhance the realism of the simulator.

2.2.3. Load Profiles

Residential loads: Aggregated into a single profile with random noise and typical morning and evening peaks to represent main household consumption periods.

Commercial loads: Modeled with a more constant pattern during working hours (9:00–18:00), including random noise to simulate daily operational variations.

2.2.4. Energy Storage System

The microgrid includes an electrochemical Energy Storage System (ESS) characterized by a nominal energy capacity E_{batt} and a maximum charging and discharging power P_{batt} . These parameters define the operational limits of the storage system and directly influence its ability to shift energy over time and interact with both the photovoltaic generator and the electrical load.

The ESS operation is subject to a set of physical constraints. Charging and discharging powers are bounded within their respective minimum and maximum limits, and simultaneous charging and discharging is not allowed. Moreover, battery discharging is permitted exclusively to supply the local load, preventing direct power injection into the grid. The rationale behind this no-export constraint lies in common grid regulations, which frequently prohibit energy exporting from storage systems to prevent market manipulation (such as pure price arbitrage without local renewable generation) and to maintain overall grid stability. The state of charge (SoC) is constrained between predefined minimum and maximum thresholds, and no charging or discharging is allowed once these limits are reached.

Finally, the dependence of the effective storage capacity on the charging or discharging rate (C-rate) is not considered in the adopted model; for consistency, fixed and identical efficiency values are therefore assumed for both the charging and discharging phases, which represents a limitation of the proposed approach.

Figure 2 visualizes the interaction between these profiles over a randomly selected week from the training set (1 June 2019–7 June 2019), highlighting the variability faced by the agent.

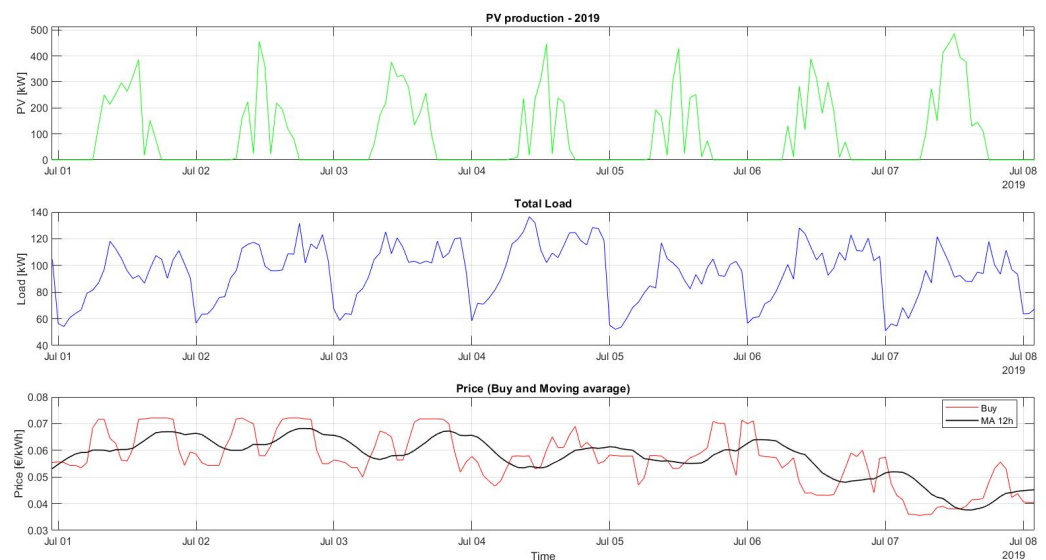


Figure 2. Example load, PV generation and electricity price profiles used in the dataset.

The integration of these three components resulted in a realistic dataset suitable for training DRL agents capable of optimally managing the balance between self-consumption, storage and grid exchange.

2.3. TD-3 Agent and Network Architecture

The agent designed for microgrid energy management is based on the Twin Delayed Deep Deterministic Policy Gradient (TD3) algorithm, a state-of-the-art method for continuous control problems. The decision-making process is formulated as a Markov Decision

Process (MDP), providing a rigorous mathematical foundation where each state transition depends exclusively on the current state and action [5].

The TD3 agent comprises an Actor network (μ_θ), responsible for generating continuous control actions, and two parallel Critic networks (Q_{ϕ_1}, Q_{ϕ_2}), which independently estimate the value of state–action pairs. The use of twin Critics is crucial to reduce the overestimation bias typical of standard DDPG. Both Actor and Critics are implemented as fully connected neural networks with ReLU activations.

To ensure robust learning in the nonlinear microgrid environment, training stability is enhanced through specific mechanisms visualized in Figure 3. These include Target Policy Smoothing, which adds noise to target actions to prevent exploiting Q-function peaks, and Delayed Policy Updates, where the actor and target networks are updated less frequently than the critics. Soft updates of the target networks further stabilize the training trajectory.

In this specific application, the Actor employs a final tanh-scaled output layer to enforce physical constraints. The agent's action represents the battery's charge/discharge power; to comply with grid regulations, the control logic ensures that any exported energy (E_{export}) originates exclusively from surplus photovoltaic generation, effectively masking the raw neural network output when necessary.

Figure 3 summarizes the complete architecture and training loop. While a detailed mathematical derivation is beyond the scope of this section, the diagram highlights the interaction between the primary and target networks and the flow of gradients that allows the agent to learn an optimal policy.

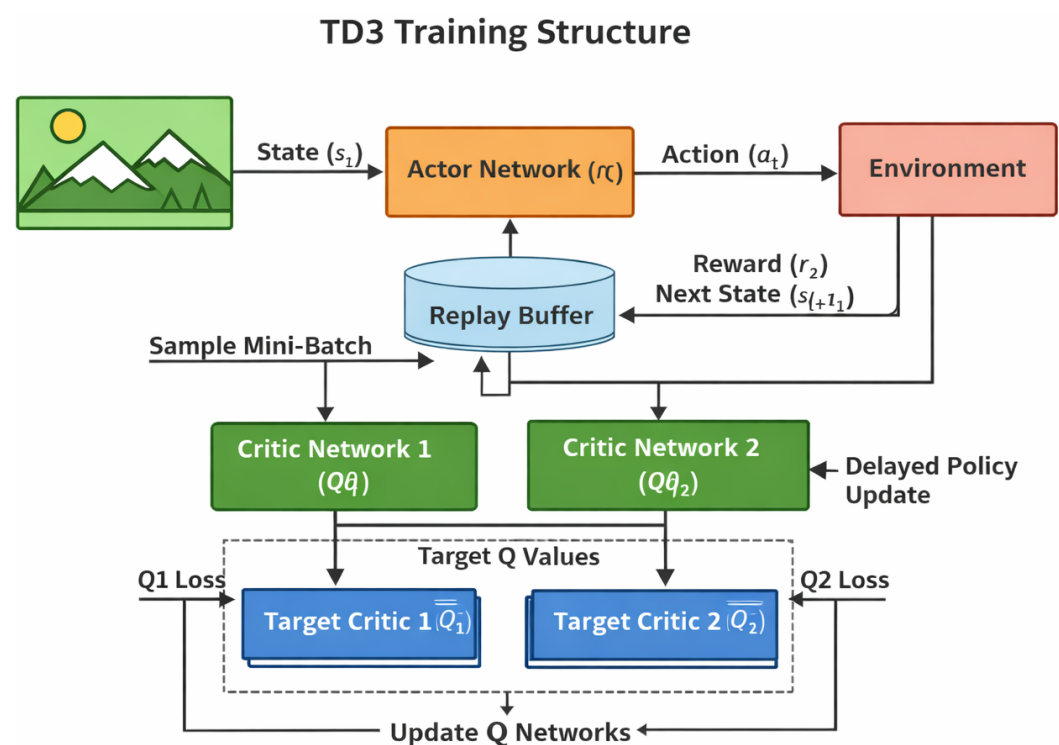


Figure 3. Schematic representation of the TD3 architecture and training loop.

2.3.1. Reinforcement Learning Training Environment

The training environment models the dynamic behaviour of a grid-connected microgrid equipped with a battery energy storage system. The system state encompasses data on electrical load, photovoltaic generation, electricity prices, and battery state of charge

(SoC), providing the agent with a comprehensive representation of the operating conditions. Formally, at each time step t , the agent observes the state vector S_t defined as:

$$S_t = [P_{\text{load},t}, P_{\text{pv},t}, C_{f,t}, \text{SoC}_t]$$

A crucial preprocessing phase involves the normalization of state variables. Given that input features exhibit heterogeneous magnitudes (e.g., power in kW versus prices in €/kWh), the data is scaled to standardized ranges. This procedure mitigates numerical disparities, thereby enhancing the stability of gradient descent and accelerating the convergence of the neural networks during the training process.

To provide a comprehensive understanding of the simulation environment's internal mechanics and how the agent's actions translate into physical energy flows, Figure 4 details the algorithmic workflow executed at each time step t . The process initiates by reading the current state vector and receiving the continuous action $a \in [-1, 1]$ output by the TD3 agent.

As illustrated in the flowchart, the decision logic branches based on the sign of a . If $a \geq 0$ (Charging Mode), the system calculates the commanded charge power as a fraction of the nominal power P_{max} , strictly limited by the available capacity (SoC room). The algorithm prioritizes self-consumption by directing PV surplus to the battery; if the commanded power exceeds the available PV surplus, the difference is drawn from the grid. Conversely, if $a < 0$ (Discharging Mode), the system interprets the magnitude $|a|$ as the requested discharge power. Crucially, the discharge is physically constrained to cover only the residual load demand ($\min(P_{\text{req}}, \text{Residual Load})$), thereby enforcing a strict “no-export” policy for stored energy.

Following the physical update of the battery SoC and the resolution of net grid interactions, the Total Reward r_t is computed. This cycle concludes by returning the next state observation S_{t+1} , the computed reward r_t , and the episode termination flag.

Following the operational logic detailed in Figure 4, the immediate reward r_t serves as the numerical feedback to guide the agent's learning process. Formally, it is computed as:

$$r_t = \underbrace{C_t (E_t^{\text{pv} \rightarrow \text{load}} + E_t^{\text{pv} \rightarrow \text{batt} \rightarrow \text{load}})}_{\text{value of self-consumption}} - \underbrace{\lambda_{\text{exp}} C_t E_t^{\text{pv} \rightarrow \text{grid}}}_{\text{export penalty}} + \underbrace{k_{\text{store}} E_t^{\text{grid} \rightarrow \text{batt/load}} \max\left(\frac{C_{\text{ref}} - C_t}{C_{\text{ref}}}, 0\right)}_{\text{bonus for low-price grid import}}$$

where C_t denotes the electricity price at time t , and $E_t^{(\cdot)}$ represents the energy flows between the PV system, load, grid, and battery. The coefficients λ_{exp} and k_{store} are hyperparameters tuned to balance the penalty for exports and the incentive for opportunistic grid arbitrage, respectively.

2.3.2. Algorithmic Updates with Adam Optimizer

In TD3, two Critic networks with parameters θ^{Q_1} and θ^{Q_2} are trained using the Adam optimizer [7]. For each transition (s_t, a_t, r_t, s_{t+1}) , the target action is computed via the target Actor with clipped noise (policy smoothing):

$$a_{t+1}^{\text{tar}} = \mu_{\text{tar}}(s_{t+1}) + \text{clip}(\epsilon, -c, c), \quad \epsilon \sim \mathcal{N}(0, \sigma^2).$$

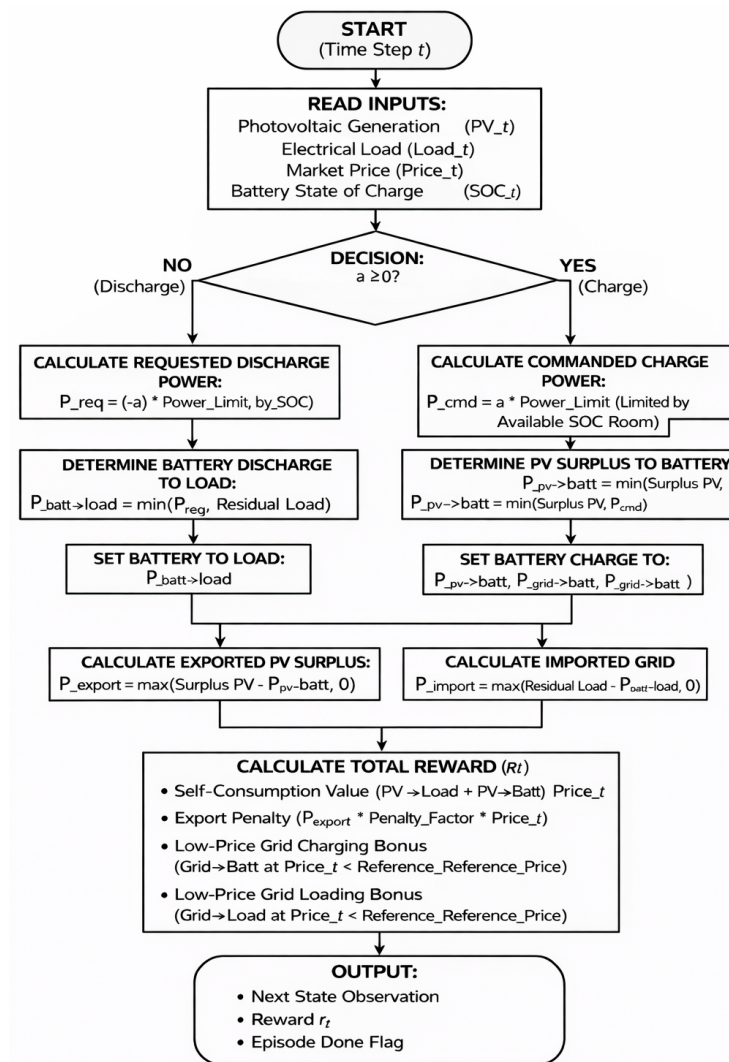


Figure 4. Detailed operational logic flowchart of the microgrid environment at each time step t . The diagram illustrates how the continuous action a is translated into physical power flows, enforcing system constraints and calculating the reward. The asterisk (*) denotes the multiplication operator.

The target value is obtained from the minimum of the two target Critics:

$$y_t = r_t + \gamma \min_{i=1,2} Q_{i,\text{tar}}(s_{t+1}, a_{t+1}^{\text{tar}}).$$

Each Critic is updated by minimizing the TD loss:

$$L(\theta^{Q_i}) = \mathbb{E} \left[\left(Q_i(s_t, a_t \mid \theta^{Q_i}) - y_t \right)^2 \right], \quad i = 1, 2,$$

with parameter update:

$$\theta^{Q_i} \leftarrow \theta^{Q_i} - \eta_Q \cdot \text{Adam}(\nabla_{\theta^{Q_i}} L).$$

The Actor parameters θ^μ are updated less frequently (delayed update) using the deterministic policy gradient:

$$\nabla_{\theta^\mu} J = \mathbb{E}_{s_t} \left[\nabla_a Q_1(s_t, a) \Big|_{a=\mu(s_t)} \cdot \nabla_{\theta^\mu} \mu(s_t) \right],$$

$$\theta^\mu \leftarrow \theta^\mu + \eta_\mu \cdot \text{Adam}(\nabla_{\theta^\mu} J).$$

Target networks are updated via soft updates:

$$\theta_{\text{tar}} \leftarrow \tau \theta + (1 - \tau) \theta_{\text{tar}}.$$

Here, η_Q and η_μ denote the learning rates, while c and σ specify the noise clipping and variance. The Adam optimizer is employed for both Actor and Critic networks due to its adaptive learning rates, ensuring stable and efficient convergence by minimizing the temporal-difference error in Critics and updating the Actor's policy via the Critics' gradients.

3. Results

To evaluate the performance and generalization capability of the trained TD3 agent, a dedicated test set was generated following the same structural logic of the photovoltaic and load power profiles used during training. However, a stochastic perturbation was applied to these profiles to ensure that their temporal evolution differs from the trajectories used during training.

In addition, the electricity buy prices used for the test were taken from real GME market data covering the period from 1 January 2025 to 2 November 2025. These price signals were never included in the training set, therefore allowing a fully unbiased evaluation of the agent's ability to operate the storage system under unseen economic conditions.

This setup enables a rigorous assessment of the robustness, adaptability, and overall operational behavior of the proposed control strategy. Tests were conducted using different values for the nominal sizing parameters of the microgrid components (PV capacity in kWp, battery capacity, and battery power), while keeping the electricity price profile unchanged. The corresponding results were then collected and analyzed.

3.1. Baseline Configuration: Same PV and Battery Parameters as in Training

The baseline test scenario represents the first test case, adopting the same nominal sizing parameters used during the training phase of the TD3 agent. These values serve as a reference configuration for the subsequent sensitivity analyses. Figure 5 illustrates the electricity purchase price profile used throughout the test set.

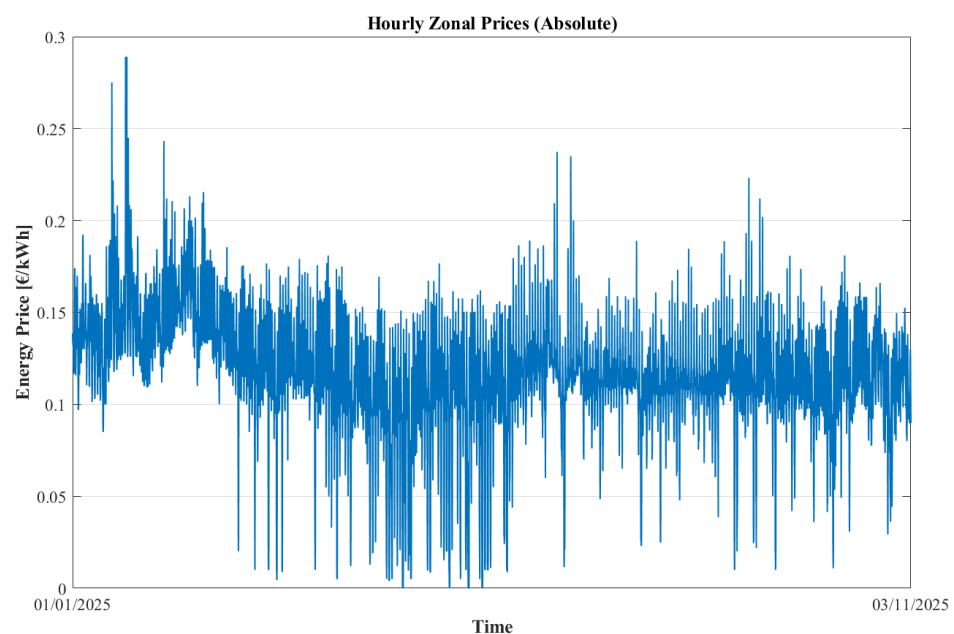


Figure 5. Electricity purchase price profile used in the test set.

The baseline sizing parameters are summarized in Table 1.

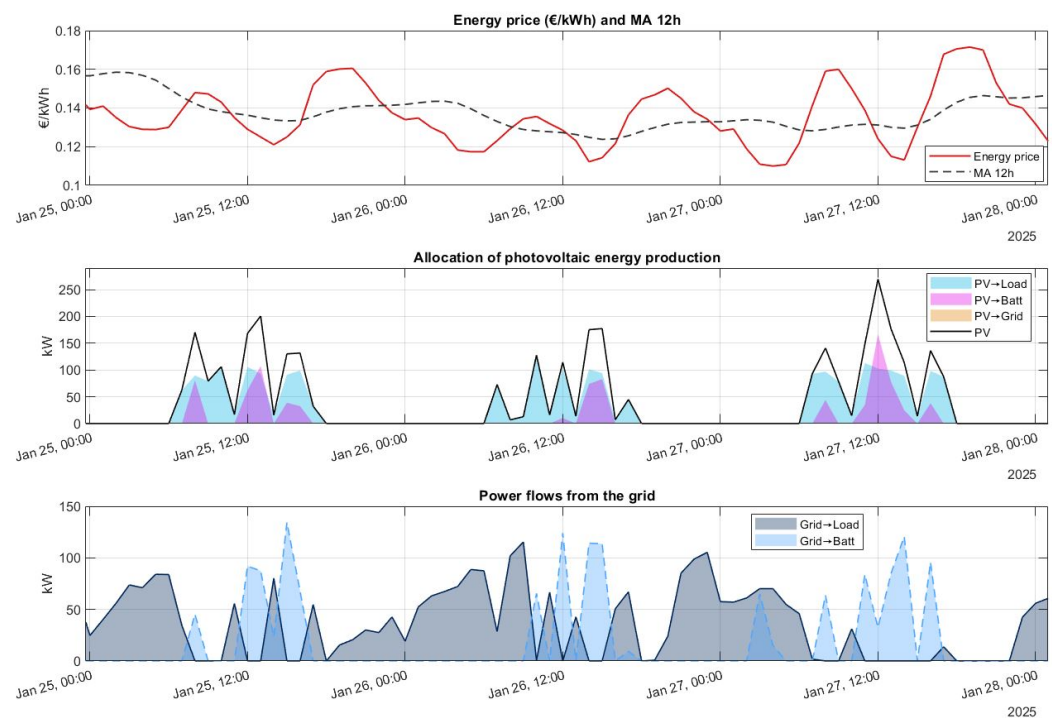
Table 1. Nominal sizing parameters of the microgrid components in the baseline configuration.

PV Capacity [kWp]	Battery Capacity [kWh]	Battery Power [kW]
600	800	200

To illustrate the operational behavior of the proposed control strategy under different photovoltaic availability conditions, three representative three-day periods were selected from the test dataset. These intervals correspond to: (i) a day with low PV generation, (ii) a day with medium PV generation, and (iii) a day with high PV generation. For each case, two figures are provided. The first figure reports the electricity purchase price together with its moving average, the time evolution of the PV production, and the grid power exchange. The second figure illustrates the response of the control agent, represented by the battery charging and discharging power and the resulting evolution of the state of charge, shown in relation to the load profile. These plots highlight how the agent adapts its decisions based on the instantaneous system state, modifying the storage operation according to price variations, PV availability, and load demand.

3.1.1. Low Photovoltaic Production Days

As shown in Figures 6 and 7, the temporal evolution of the microgrid variables illustrates the agent's response to this specific scenario.

**Figure 6.** Three-day profiles for the low-PV-production scenario, showing the electricity purchase price and its moving average, the photovoltaic power generation and the power imported from the grid.

3.1.2. Medium Photovoltaic Production Days

As shown in Figures 8 and 9, the temporal evolution of the microgrid variables illustrates the agent's response to this specific scenario.

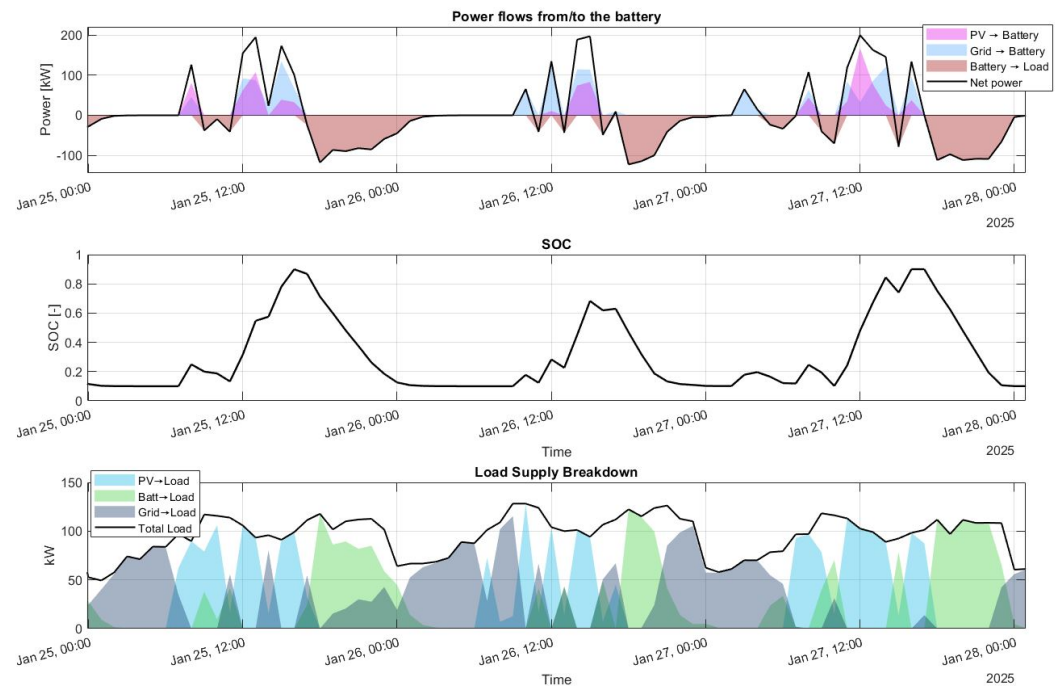


Figure 7. Three-day profiles for the low-PV-production scenario, showing battery charge and discharge power, the battery state of charge (SoC) and the load demand, highlighting the agent's reliance on grid energy during low-price night hours to compensate for the solar deficit.

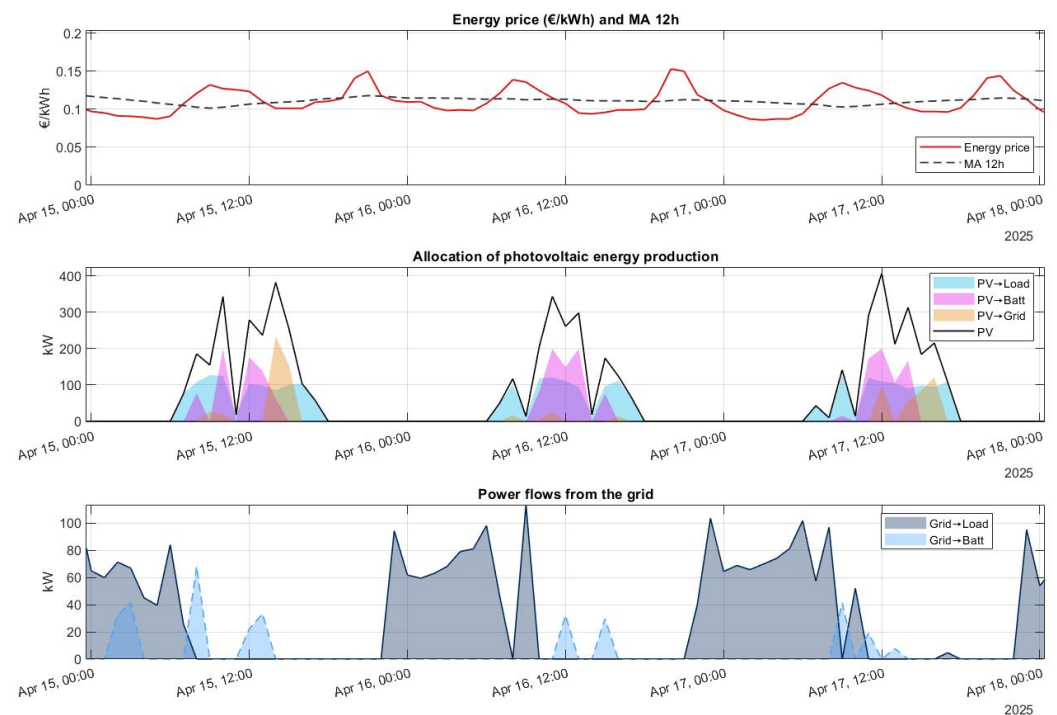


Figure 8. Three-day profiles for the medium-PV-production scenario, showing the electricity purchase price and its moving average, the photovoltaic power generation, and the power imported from the grid.

3.1.3. High Photovoltaic Production Days

Table 2 summarizes the cumulative energy flows and the associated economic indicators computed over the entire test period. These values provide an overall picture of the microgrid operation under the tested configuration, aggregating the contributions of PV generation, battery usage, grid exchanges, and the resulting economic balance.

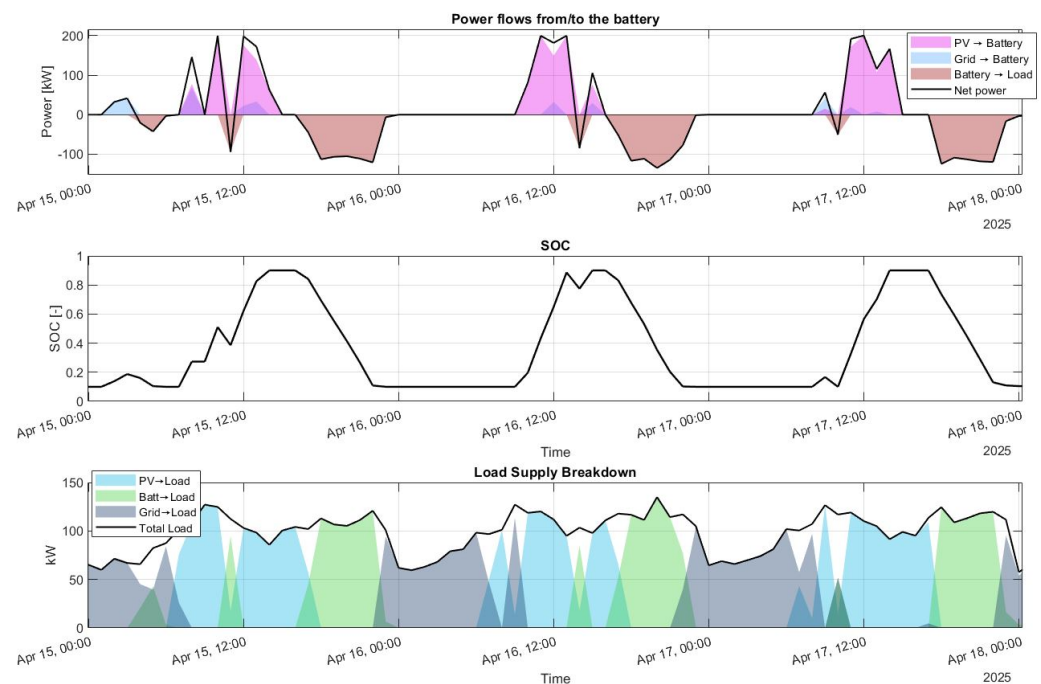


Figure 9. Three-day profiles for the medium-PV-production scenario, showing battery charge and discharge power, the battery state of charge (SoC), and the load demand, showing the battery effectively absorbing the midday solar peak and discharging during evening load peaks.

Table 2. Summary of cumulative energy flows and economic performance over the full test period (1 January 2025–2 November 2025).

E_load	711,233 kWh
E_pv	560,267 kWh
E_grid_import	273,323 kWh
E_pv→Load	276,908 kWh
E_batt→Load	215,985 kWh
E_pv→Batt	184,096 kWh
E_grid→Batt	54,983 kWh
E_export (PV)	99,262 kWh
Self-consumption (kWh)	64.8% of load
PV self-consumption	82.3%
Import energy cost	30,835.22 €
Export energy gain	8612.69 €
Self-consumption value	58,652.22 €
Net cost (import–export)	22,222.53 €
Battery equivalent cycles	≈270 cycles

As shown in Figures 10 and 11, the temporal evolution of the microgrid variables illustrates the agent’s response to this specific scenario.

3.2. Rule-Based Baseline Strategy

To establish a benchmark for evaluating the performance of the proposed DRL approach, a standard industry rule-based (RB) control strategy was simulated over the same unseen 2025 test dataset, utilizing the baseline nominal sizing parameters (600 kWp PV, 800 kWh battery, 200 kW power). The RB controller follows a greedy logic commonly implemented in commercial microgrid inverters: it strictly prioritizes self-consumption by charging the battery whenever there is a PV surplus and discharging it to cover the load whenever PV generation is insufficient. Unlike the TD3 agent, the RB controller does not

perform price arbitrage and only imports energy from the main grid when both the local PV generation and the battery storage are unable to meet the load demand.

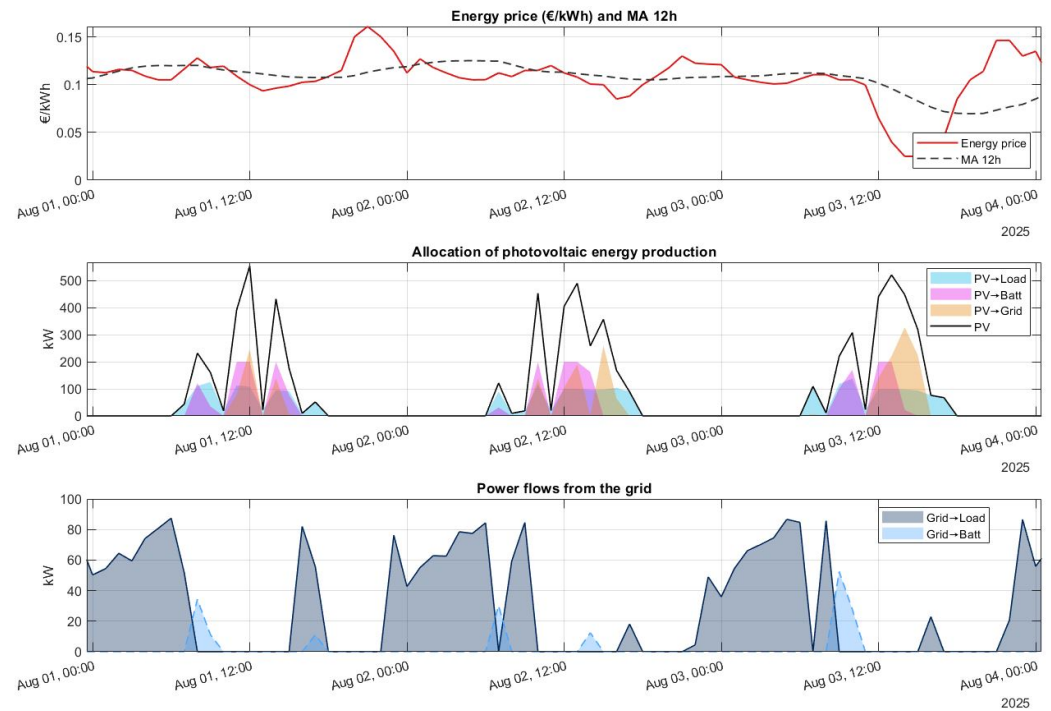


Figure 10. Three-day profiles for the high-PV-production scenario, showing the electricity purchase price and its moving average, the photovoltaic power generation, and the power imported from the grid.

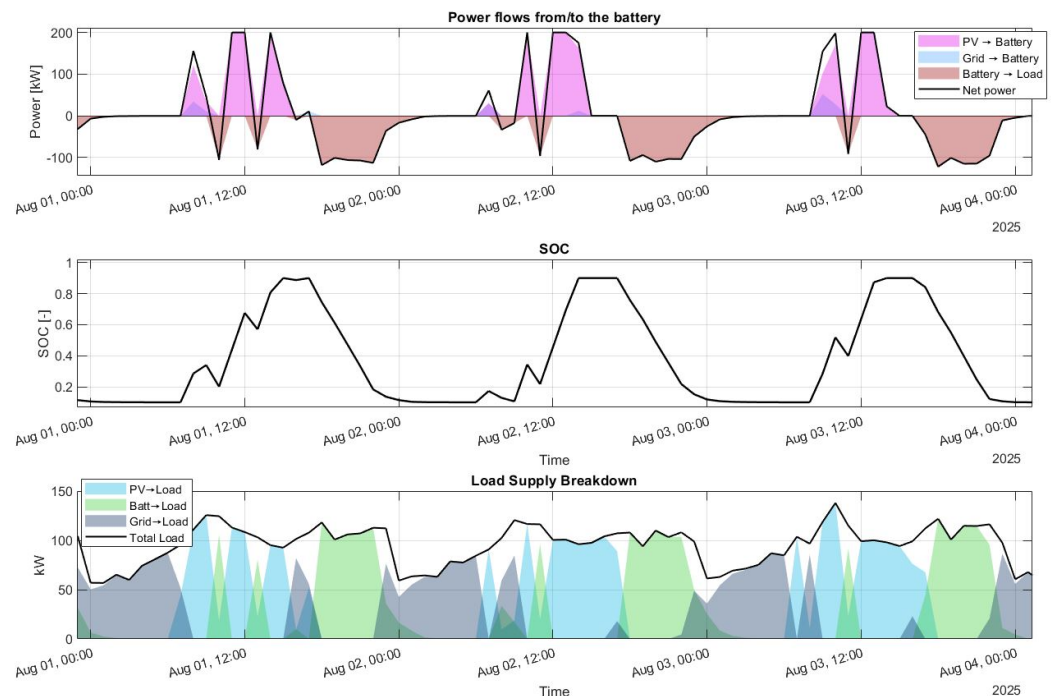


Figure 11. Three-day profiles for the high-PV-production scenario, showing battery charge and discharge power, the battery state of charge (SoC), and the load demand, illustrating the battery reaching full capacity early in the day and the subsequent massive export of PV surplus.

Table 3 summarizes the cumulative energy flows and the resulting economic indicators for the Rule-Based strategy. These metrics provide a standard reference point for the comparative discussion in the following section.

Table 3. Summary of cumulative energy flows and economic performance for the Rule-Based (RB) controller over the full test period.

E_load	711,233 kWh
E_pv	560,267 kWh
E_grid_import	247,972 kWh
E_pv→Load	276,908 kWh
E_batt→Load	186,353 kWh
E_pv→Batt	206,149 kWh
E_grid→Batt	0 kWh
E_export (PV)	77,209 kWh
Self-consumption (kWh)	67.9% of load
PV self-consumption	86.2%
Import energy cost	29,730.83 €
Export energy gain	6711.25 €
Self-consumption value	54,008.63 €
Net cost (import–export)	23,019.58 €
Battery equivalent cycles	≈233 cycles

3.3. Test Scenario with Increased Battery Capacity

This test scenario evaluates the performance of the TD3 agent when the battery capacity is increased relative to the baseline configuration, while keeping PV capacity and battery power unchanged. The configuration parameters are detailed in Table 4.

Table 4. Nominal sizing parameters of the microgrid components for the test scenario with increased battery capacity.

PV Capacity [kWp]	Battery Capacity [kWh]	Battery Power [kW]
600	1600	200

To illustrate the strategic behavior of the TD3 agent, a representative one-week period was randomly selected from the test dataset. As shown in Figures 12 and 13, the system dynamics over this interval highlight how the agent adapts its decisions based on the evolving state of the microgrid.

Table 5 summarizes the cumulative energy flows and the corresponding economic indicators for this test scenario, providing an overall view of the system operation with the increased battery capacity.

3.4. Test Scenario with Increased Battery Power

This test scenario evaluates the performance of the TD3 agent when the battery power is increased relative to the baseline configuration, while maintaining the same PV capacity and battery capacity as in the training setup. The configuration parameters are detailed in Table 6.

To demonstrate the decision-making strategy of the TD3 agent, a one-week period was randomly chosen from the test dataset. As shown in Figures 14 and 15, the temporal evolution of the system illustrates how the agent adjusts its actions in response to the changing state of the microgrid.

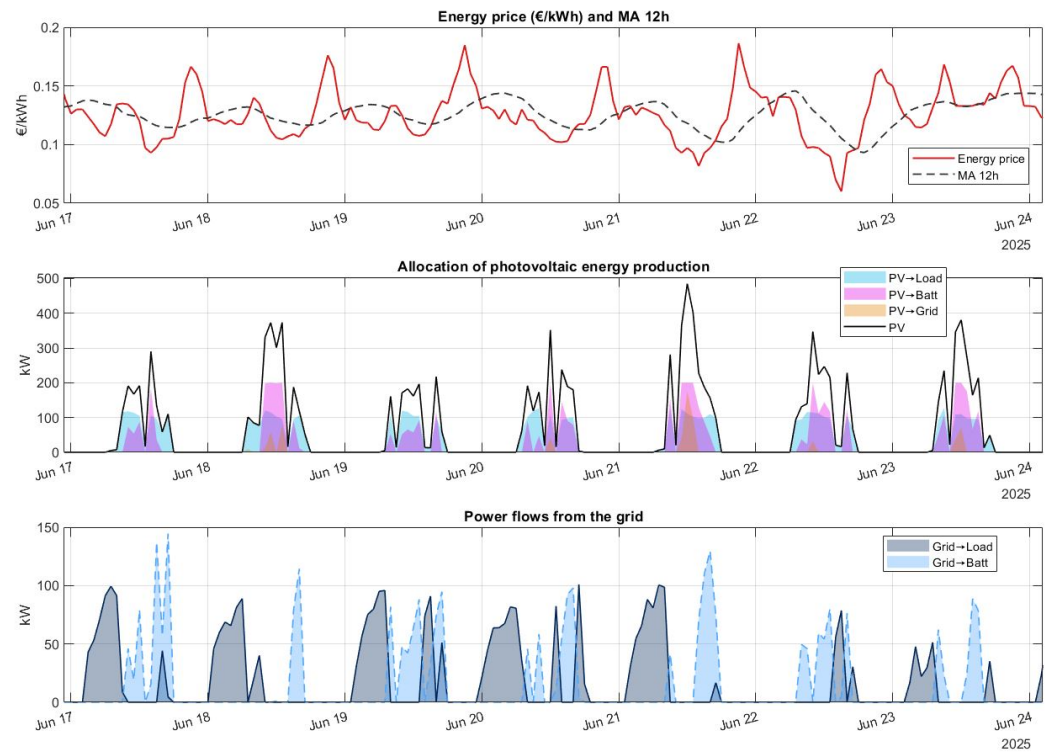


Figure 12. One-week profiles for a randomly selected period from the test dataset, showing the electricity purchase price and its moving average, the photovoltaic power generation, and the power imported from the grid.

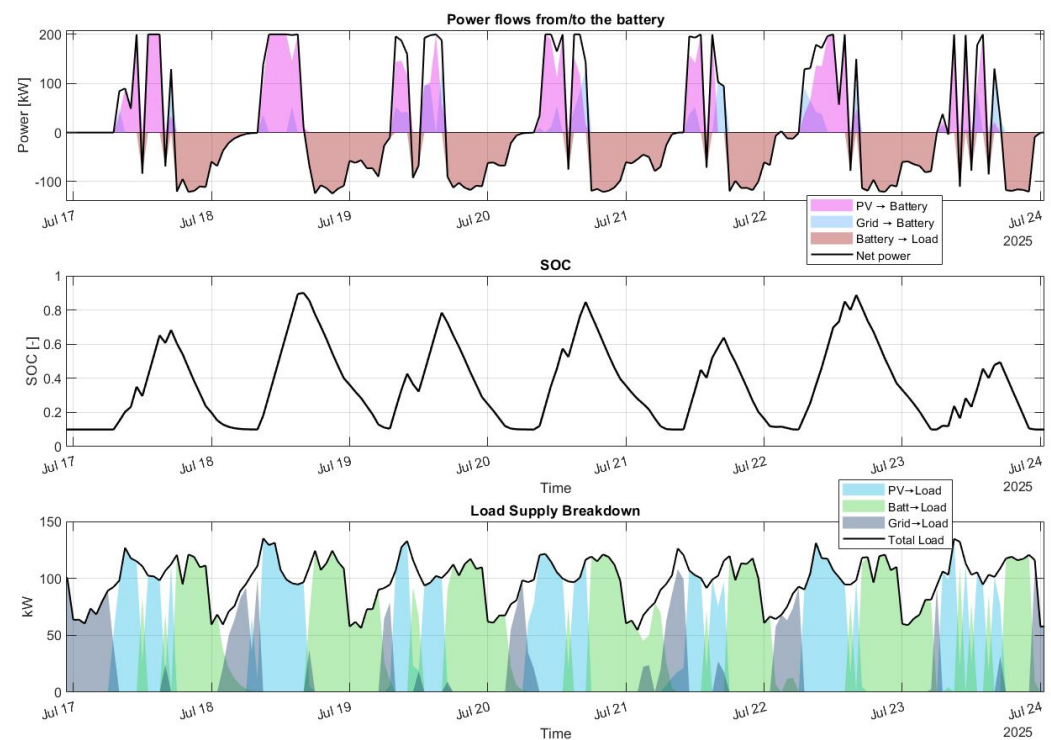


Figure 13. One-week profiles for a randomly selected period from the test dataset, showing battery charge and discharge power, the battery state of charge (SoC), and the load demand.

Table 5. Summary of cumulative energy flows and economic performance over the second full test period (1 January 2025–2 November 2025).

E_load	711,233 kWh
E_pv	560,267 kWh
E_grid_import	226,004 kWh
E_pv→Load	276,908 kWh
E_batt→Load	317,677 kWh
E_pv→Batt	242,353 kWh
E_grid→Batt	109,356 kWh
E_export (PV)	41,005 kWh
Self-consumption (kWh)	73.0% of load
PV self-consumption	92.7%
Import energy cost	23,745.78 €
Export energy gain	3902.40 €
Self-consumption value (€/h)	70,519.66 €
Net cost (import–export)	19,843.38 €
Battery equivalent cycles	≈199 cycles

Table 6. Nominal sizing parameters of the microgrid components for the test scenario with increased battery power.

PV Capacity [kWp]	Battery Capacity [kWh]	Battery Power [kW]
600	800	400

Table 7 summarizes the cumulative energy flows and the corresponding economic indicators for this test scenario, providing an overall view of the system operation with the increased battery power.

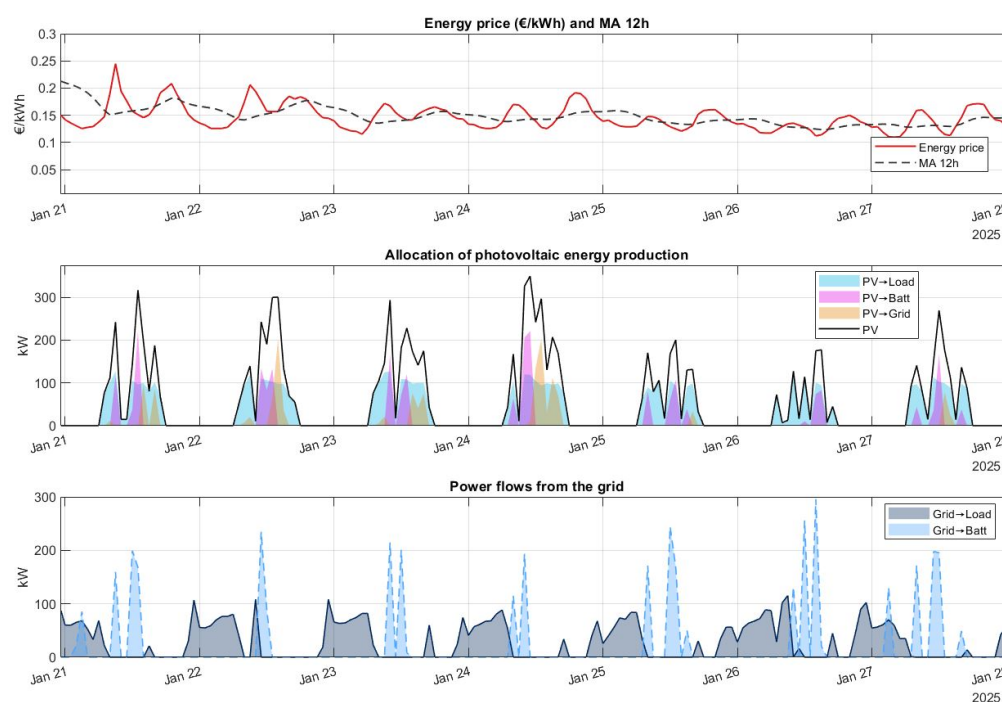
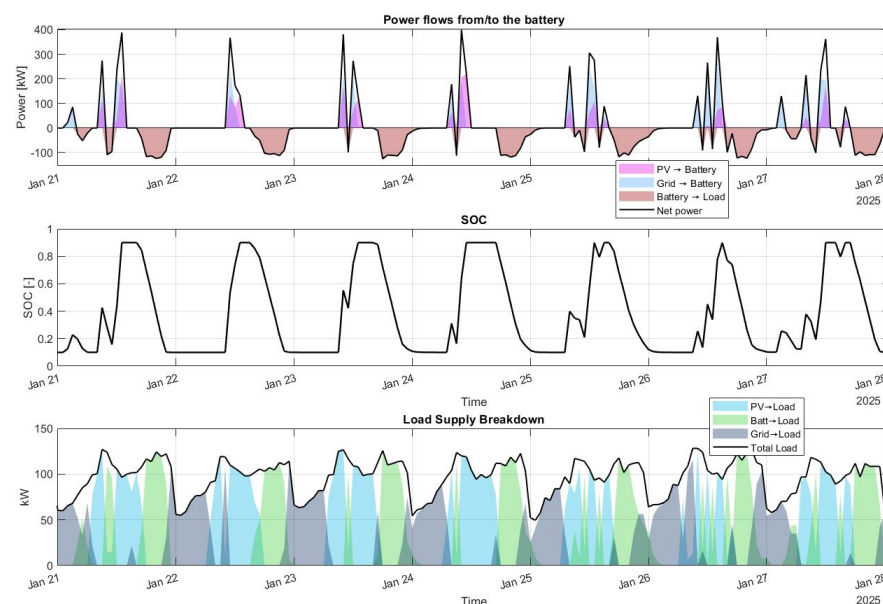
**Figure 14.** One-week profiles for a randomly selected period from the test dataset, showing the electricity purchase price and its moving average, the photovoltaic power generation, and the power imported from the grid.

Table 7. Summary of cumulative energy flows and economic performance over the third full test period (1 January 2025–2 November 2025).

E_load	711,233 kWh
E_pv	560,267 kWh
E_grid_import	330,689 kWh
E_pv→Load	276,908 kWh
E_batt→Load	226,430 kWh
E_pv→Batt	127,859 kWh
E_grid→Batt	122,794 kWh
E_export (PV)	155,499 kWh
Self-consumption (kWh)	56.9% of load
PV self-consumption	72.2%
Import energy cost	38,365.69 €
Export energy gain	13,749.16 €
Self-consumption value (€/h)	59,878.82 €
Net cost (import–export)	24,616.52 €
Battery equivalent cycles	≈283 cycles

**Figure 15.** One-week profiles for a randomly selected period from the test dataset, showing battery charge and discharge power, the battery state of charge (SoC), and the load demand.

3.5. Test Scenario with Increased PV Capacity

In this test configuration, the photovoltaic peak capacity is increased with respect to the baseline scenario, while the battery capacity and power rating remain unchanged. This setup allows assessing how the TD3 agent responds to a substantially higher availability of renewable generation and how it reallocates charging, discharging, and grid exchange decisions under enhanced PV conditions. The configuration parameters are detailed in Table 8.

Table 8. Nominal sizing parameters of the microgrid components for the test scenario with increased photovoltaic capacity.

PV Capacity [kWp]	Battery Capacity [kWh]	Battery Power [kW]
1200	800	200

To examine how the control policy behaves under this expanded PV potential, a randomly selected week from the test dataset was analyzed. As shown in Figures 16 and 17, the temporal

evolution of the microgrid variables over this interval illustrates how the agent reorganizes its operational strategy in response to the increased renewable power availability.

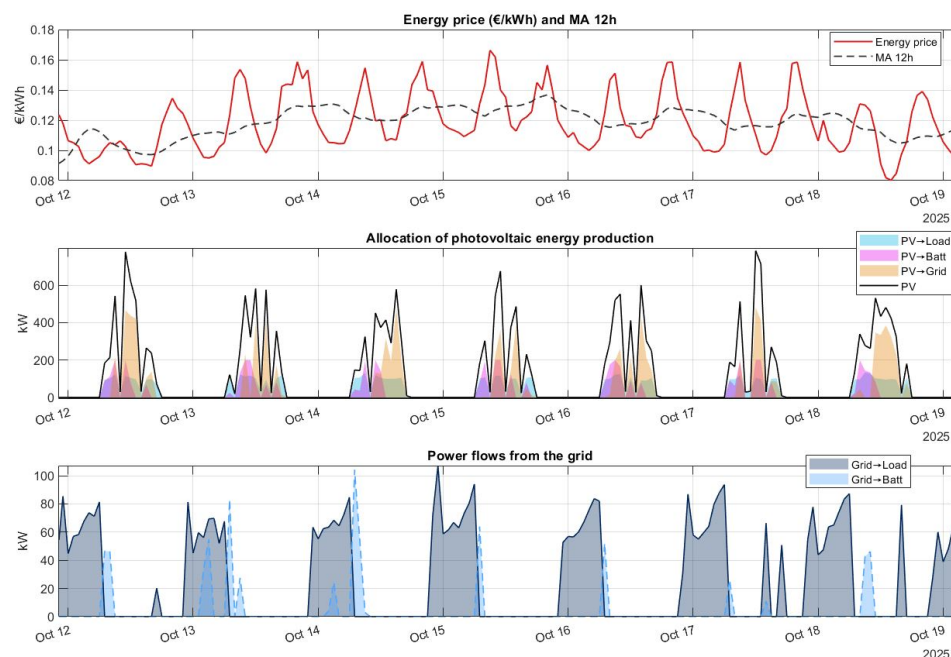


Figure 16. One-week profiles for a randomly selected period from the test dataset, showing the electricity purchase price and its moving average, the photovoltaic power generation, and the power imported from the grid.

Table 9 reports the aggregated energy balances and economic metrics computed over the full test horizon. These results provide an integrated view of the system behavior when the PV capacity is doubled relative to the baseline, highlighting the resulting shifts in energy flows, self-consumption, and economic performance.

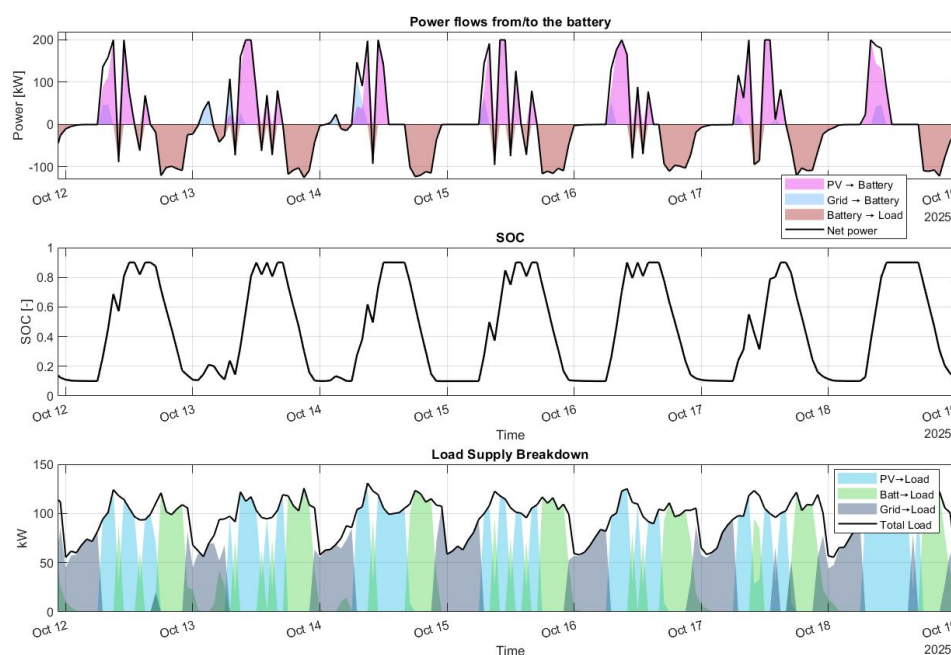


Figure 17. One-week profiles for a randomly selected period from the test dataset, illustrating the battery charge and discharge power, the battery state of charge (SoC), and the load demand.

Table 9. Summary of cumulative energy flows and economic performance over the test period (1 January 2025–2 November 2025) with increased PV capacity.

E_load	711,233 kWh
E_pv	1,120,533 kWh
E_grid_import	211,469 kWh
E_pv→Load	301,047 kWh
E_batt→Load	221,563 kWh
E_pv→Batt	222,419 kWh
E_grid→Batt	22,846 kWh
E_export (PV)	597,068 kWh
Self-consumption (kWh)	73.6% of load
PV self-consumption	46.7%
Import energy cost	23,846.83 €
Export energy gain	57,162.04 €
Self-consumption value (€/h)	62,489.67 €
Net cost (import–export)	–33,315.21 €
Battery equivalent cycles	≈277 cycles

4. Discussion

In the following section, the results are discussed from both qualitative and quantitative perspectives. On the qualitative side, the analysis focuses on how the trained agent adapts its control strategy under varying photovoltaic availability. This includes examining the temporal evolution of the main operational variables to gain insight into the agent's decision-making behaviour.

On the quantitative side, the discussion evaluates how key techno-economic indicators—such as energy flows, battery usage, self-consumption performance, and operational costs—vary as a function of the adopted sizing parameters. The analysis places particular emphasis on the impact of design choices on the overall operation of the microgrid, as well as on the verification of the correct functioning of the control strategy previously examined.

4.1. Behaviour of the Agent Under Different PV Production Scenarios

The TD3 agent directly determines the battery's charging and discharging power and indirectly influences the power exchanged with the grid. By modulating the battery operation, the agent shapes the balance between photovoltaic generation, load demand, and interaction with the external network. Across the different photovoltaic availability conditions, two distinct operating regimes clearly emerge. Before analysing the differences across photovoltaic availability scenarios, it is useful to outline a set of behaviours consistently exhibited by the trained agent throughout all test conditions. First, the electrical load is always fully supplied, indicating that the control strategy never jeopardizes supply security. Second, all physical constraints of the microgrid components are strictly satisfied, including battery power limits and state-of-charge (SoC) bounds. Third, whenever PV generation drops unexpectedly—for instance, due to cloud coverage—the agent promptly compensates by discharging the battery, ensuring seamless load support. The agent also demonstrates a stable hierarchy in its decision-making, reflecting the structure of the reward function:

- (i) PV power is used to supply the load whenever available;
- (ii) Any surplus PV generation is stored in the battery;
- (iii) The battery is discharged to supply the load when PV is insufficient;
- (iv) Energy is imported from the grid only when economically favourable.

As a consequence, the agent naturally performs price-driven arbitrage, concentrating grid imports during hours in which the electricity price is below its moving average reference.

Finally, regarding battery degradation, the learned policy proves to be highly conservative. Although the thermodynamic model adopted in this study assumes a fixed efficiency and neglects C-rate-dependent capacity degradation, the agent's behaviour intrinsically preserves battery health. Over the 11-month testing period, the agent rarely executes more than one full equivalent charge-discharge cycle per day (e.g., approximately 270 cycles in the baseline scenario, as summarized in the results tables), despite no explicit constraint enforcing this limitation. This conservative, emergent outcome effectively limits cycle aging and mitigates the immediate need for a complex degradation model, highlighting the ability of the reward structure to promote battery health preservation alongside cost minimization. Nonetheless, explicitly incorporating C-rate effects and a degradation penalty in the reward formulation represents a priority for future investigations aiming at multi-year techno-economic optimization.

4.1.1. Low PV Production

Under low or very low PV generation (see Figures 6 and 7) the agent adopts a distinctly different strategy. Because the reduced solar input does not allow for completing a full battery cycle within the day, the agent increases the use of grid charging, taking advantage of favourable price spreads. Grid imports become larger in volume and more concentrated during low-price hours, consistent with price-arbitrage behaviour. In this scenario, the battery is primarily used to shift energy from low-price periods to high-price periods, rather than to absorb PV surplus. The grid-import curves therefore exhibit more pronounced peaks, but still aligned with the downward deviations of the electricity price relative to its moving average reference.

4.1.2. Medium and High PV Production

In conditions of medium or high photovoltaic availability (see Figures 8–11) the agent behaves in a manner resembling classical rule-based controllers. The battery is charged whenever PV production exceeds the instantaneous load and discharged when PV alone cannot cover demand. The grid-import profile exhibits moderate peaks and reduced imported energy, with most grid purchases occurring during intervals in which the electricity price falls below its moving average. Thanks to the adequate sizing of the storage system, the battery often reaches its minimum SoC as the price curve enters its typical night-time decline, allowing the agent to avoid unnecessary charging from the grid at inefficient times. In this operating regime, the agent prioritizes direct PV-to-load usage and uses the battery primarily as a buffer for intra-day balancing.

4.2. Comparison with the Rule-Based Strategy

To accurately evaluate the added value of the proposed DRL methodology, the performance of the TD3 agent can be directly compared against the standard Rule-Based (RB) controller presented in Section 3.2. Table 10 highlights the key differences in operational and economic metrics between the two strategies under the baseline configuration.

The comparison clearly highlights the economic superiority of the RL approach, with a Net Cost reduction of approximately 3.5% over the 11-month period. While the RB controller is highly efficient at maximizing purely physical self-consumption (67.9% vs. 64.8%) by blindly storing all available PV surplus, it is inherently oblivious to market price fluctuations. In contrast, the TD3 agent learns to strategically shift grid imports to low-price hours and leverages the battery for price arbitrage. This is evidenced by the higher grid import volume (273,323 kWh vs. 247,972 kWh) which, however, results in a disproportional

tionately low increase in cost due to the lower average purchase price (≈ 0.112 €/kWh for TD3 vs. ≈ 0.120 €/kWh for RB). Furthermore, the agent generates significantly higher revenues from exports (+1900 €), indicating that it intelligently discharges or exports energy during high-price peak hours. This price-driven behavior—learned autonomously without predefined rules—drastically reduces the net operational cost of the microgrid, constituting the primary advantage of the DRL control over standard logic.

Table 10. Performance comparison between the TD3 agent and the standard Rule-Based (RB) controller (Baseline configuration).

Metric	Rule-Based Controller	TD3 Agent
Self-consumption (% of load)	67.9%	64.8%
Total grid import (kWh)	247,972	273,323
Import energy cost (€)	29,730.83	30,835.22
Export energy gain (€)	6711.25	8612.69
Net operational cost (€)	23,019.58	22,222.53

4.3. Impact of Component Sizing on Microgrid Performance

From a quantitative perspective, the analysis clearly shows that the agent is able to adapt remarkably well to variations in the nominal sizes of the microgrid components. Despite significant modifications to the photovoltaic capacity, battery storage capacity, and maximum exchange power compared to the configuration used in the training set, the agent consistently applies the same optimal operational strategy observed in the baseline test.

This behaviour indicates that the learned policy has not been memorized as a rigid response tailored to a specific sizing configuration. Instead, the agent has assimilated a scalable model of the relationships between production, demand, storage capability, and electricity prices. As a result, it recognizes the same energy-management patterns even when the power scale changes, adjusting only the magnitude of its actions. In other words, the agent preserves the same decision hierarchy—maximizing self-consumption, minimizing costly grid imports, and ensuring an efficient use of the BESS—while applying it to larger or smaller quantities depending on the configuration. Such robustness demonstrates strong generalization of the learned policy, suggesting that training enabled the agent to capture the functional relationships within the environment rather than relying on the absolute values of the parameters.

What does change significantly, however, are the overall performance metrics of the microgrid, evaluated under a fixed load profile. Since the agent applies identical strategies across different energy scales, quantitative indicators—such as self-consumption, exported energy, storage cycling, grid imports, and economic performance—become highly sensitive to component sizing. In essence, while the policy remains structurally unchanged, the energetic and economic efficiency of the microgrid is strongly influenced by the nominal characteristics of its components: higher PV capacity modifies the balance between surplus and deficit, larger storage enhances arbitrage opportunities and load coverage, and different inverter power limits affect the hourly dynamics of power flows.

These differences impact not only the absolute energy volumes but also the interaction quality among components, producing direct effects on operating costs, the share of renewable energy effectively utilised, and the overall profitability of the system. It is therefore essential to analyse each configuration individually to understand how design variations alter the balance between renewable production, storage capability, and grid exchange, bearing in mind that all results are derived under forecast-based operation.

The following subsections examine in detail the effects of modifying each component, illustrating how each sizing choice influences the energetic and economic performance of the microgrid.

4.3.1. Effect of Increasing the Battery Energy Capacity

Doubling the battery capacity significantly alters the distribution of energy flows within the microgrid. Table 4 shows that this increased capacity enables the battery to absorb a much larger share of the photovoltaic surplus. This results in a marked rise in the energy transferred from the PV system to the battery, also reflected in the cumulative values reported in Table 5. Consequently, the energy supplied from the battery to the load increases, since the system can draw from a larger energy reservoir during periods of low irradiance or higher demand.

At the same time, the need to export surplus photovoltaic energy decreases, as part of the energy that could not be used locally in the baseline configuration is now stored instead. This reduction in exported energy, together with a decrease in grid imports (see Table 5), enhances the local exploitation of renewable generation.

These dynamics improve both the overall self-consumption and the PV self-consumption indicators. From an economic standpoint, the net cost decreases because the reduction in energy purchases outweighs the loss of revenue due to the smaller amount of exported energy.

4.3.2. Effect of Increasing the Battery Power Rating

Increasing the nominal power of the storage system—as specified in Table 6—significantly affects the charging dynamics, while leaving the discharging behaviour unchanged. Since the battery can discharge only towards the load, the discharge power never exceeds the instantaneous demand, which remains below the nominal limit even in the most demanding conditions.

The higher charging power leads to a notable increase in the energy imported from the grid, as confirmed by the cumulative values in Table 7. The agent exploits more aggressively the possibility of charging the battery when the purchase price is below the moving average, which reduces the share of load directly supplied by photovoltaic generation. As a consequence, overall self-consumption, PV self-consumption, and the energy delivered from the battery to the load all decrease.

4.3.3. Effect of Increasing the PV Peak Power

A higher nominal installed photovoltaic capacity—doubled in the configuration reported in Table 8—leads to a substantial rise in energy production during daylight hours. As reflected in the cumulative balances of Table 9, both the energy directly supplied to the load and the energy exported to the grid increase noticeably, since the grid becomes the main sink for the large surplus that exceeds local consumption.

The amount of PV energy stored in the battery increases only marginally because the storage capacity remains unchanged and thus becomes a bottleneck with respect to the additional production. A significant portion of the extra PV generation cannot be stored and is therefore exported.

Regarding performance, the overall self-consumption improves due to the greater direct contribution of PV to covering the load. However, the PV self-consumption metric worsens because the exported fraction grows more rapidly than the locally consumed share. Economically, import costs decrease thanks to reduced grid dependency, while revenues from exported energy increase significantly due to the higher availability of surplus PV production.

5. Conclusions

The study demonstrated that the TD3 agent effectively manages a microgrid equipped with a photovoltaic (PV) plant and an energy storage system (ESS), adapting its control policy to diverse operational conditions without requiring explicit reconfiguration. Specifically, the agent achieved self-consumption rates up to 73.6% of the total load and significantly reduced net operational costs through intelligent price arbitrage, outperforming standard rule-based strategies.

Qualitative analysis highlighted that the controller's behavior remains consistent even as PV availability varies: the agent maintains a stable decision-making hierarchy, ensures the load supply, adheres to the physical constraints of the components, and reacts promptly to sudden variations in solar generation. Simultaneously, it correctly integrates the economic dimension, modulating energy flows as a function of price dynamics.

From a quantitative perspective, the results indicate that the agent generalizes well with respect to component sizing: while maintaining the same decision-making policy, it successfully adapts the intensity of its actions automatically when photovoltaic power, battery capacity, or grid exchange power limits vary. Nevertheless, the energy and economic performance of the microgrid remain strongly dependent on these design choices:

- Higher storage capacity reduces exports and increases self-consumption;
- An increase in charging power favors price arbitrage but reduces the direct contribution of photovoltaics;
- High installed power raises the level of local production but also increases the energy injected into the grid.

In all scenarios, the learned policy reproduces dynamics known in the literature, confirming that the agent has correctly internalized the relationships between generation, demand, storage, and price signals.

While the present study focuses on the adaptive behavior of the TD3 agent, the advantages of RL-based strategies over traditional approaches are well-documented. For instance, Foruzan et al. [3] demonstrated that multi-agent RL allows distributed entities to converge to optimal profit strategies (Nash equilibrium) even under high uncertainty, a capability often lacking in deterministic rule-based controllers. Furthermore, Oliveira da Silva et al. [29] showed that RL agents can achieve near-optimal performance comparable to complex Model Predictive Control (MPC) schemes but with significantly lower computational complexity, rendering them superior for real-time microgrid applications. While the offline training phase requires adequate computational resources (e.g., GPUs), the real-time inference consists solely of a few matrix multiplications. This minimal computational footprint allows the model to be executed in fractions of a millisecond, making it highly suitable for real-time deployment on lightweight embedded systems (such as advanced Battery Management Systems) or via cloud-based APIs.

Beyond these quantitative benefits, the employment of a controller based on Reinforcement Learning (RL) offers significant operational advantages. The neural network representing the policy is capable of capturing non-linear relationships between operational variables and adapting to different plant configurations without the need for manual re-tuning. This renders the approach particularly attractive for real-world Energy Management Systems (EMS), whether integrated directly into storage devices or implemented via cloud platforms for residential, commercial, or industrial microgrids.

A particularly relevant aspect concerns the use of this methodology to support the sizing of microgrid components. Through simulation tests based on the learned policy, it is possible to explore various configurations of photovoltaic capacity, storage, and grid exchange power, imposing specific operational objectives such as maximizing self-consumption, reducing energy procurement costs, or increasing economic returns from exported energy.

Consequently, the controller not only manages the system in real time but also serves as a decision-making tool for system design, enabling the identification of optimal configurations that balance energy efficiency, economic performance, and component lifespan.

However, certain limitations persist: control reliability depends on the quality of forecasts, operation under rare or unseen conditions requires further verification, and real-world implementation must account for measurement noise, component faults, and the need for periodic policy updates. In perspective, natural extensions of this work include the use of multi-agent algorithms, the integration of more complex dynamic pricing schemes, and the evaluation of control robustness under non-ideal conditions. Furthermore, real-world implementation must account for measurement uncertainties inherent in physical sensors. As highlighted in recent literature [33], noise in Battery Management System (BMS) readings can affect the state observed by the RL agent. Additionally, recent studies have further explored advanced modelling frameworks, battery parameter estimation under aging uncertainties, and the integration of novel photovoltaic architectures [35–38], highlighting the continuous need for robust, adaptive analytical strategies in modern microgrids.

Overall, the obtained results indicate that an approach based on Deep Reinforcement Learning (DRL) and neural networks constitutes a promising solution for the optimization of next-generation microgrids, owing to its ability to react in real time, valorize local renewable energy, and adapt to different plant configurations. This paves the way for the practical application of such techniques in future distributed energy management systems.

Author Contributions: Conceptualization, F.P.; Methodology, D.M.; Validation, D.M. and L.S.; Writing—original draft, F.P.; Writing—review & editing, D.M.; Supervision, L.S., M.Q., F.R.F., G.C.L. and F.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Xu, Y.; Zhang, W.; Hug, G.; Kar, S.; Li, Z. Cooperative control of distributed energy storage systems in a microgrid. *IEEE Trans. Smart Grid* **2015**, *6*, 238–248. [\[CrossRef\]](#)
2. Jiang, Q.; Xue, M.; Geng, G. Energy management of microgrid in grid-connected and stand-alone modes. *IEEE Trans. Power Syst.* **2013**, *28*, 3380–3389. [\[CrossRef\]](#)
3. Foruzan, E.; Soh, L.K.; Asgarpour, S. Reinforcement learning approach for optimal distributed energy management in a microgrid. *IEEE Trans. Power Syst.* **2018**, *33*, 5749–5758. [\[CrossRef\]](#)
4. Lillicrap, T.P.; Hunt, J.J.; Pritzel, A.; Heess, N.; Erez, T.; Tassa, Y.; Silver, D.; Wierstra, D. Continuous control with deep reinforcement learning. In *4th International Conference on Learning Representations (ICLR 2016)—Conference Track Proceedings*; ICLR: San Juan, Puerto Rico, 2016.
5. Littman, M.L. Markov Decision Processes. In *International Encyclopedia of the Social & Behavioral Sciences*; Elsevier: Oxford, UK, 2001; pp. 9240–9242.
6. Fujimoto, S.; van Hoof, H.; Meger, D. Addressing function approximation error in actor–critic methods. In *Proceedings of the 35th International Conference on Machine Learning (ICML)*; PMLR: Stockholm, Sweden, 2018; pp. 1587–1596.
7. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.
8. Lund, H.; Østergaard, P.A.; Connolly, D.; Mathiesen, B.V. Smart energy and smart energy systems. *Energy* **2017**, *137*, 556–565. [\[CrossRef\]](#)
9. Laureti, L.; Massaro, A.; Costantiello, A.; Leogrande, A. The impact of renewable electricity output on sustainability in the context of circular economy: A global perspective. *Sustainability* **2023**, *15*, 2160. [\[CrossRef\]](#)

10. Abdelwanis, M.I.; Elmezain, M.I. A comprehensive review of hybrid AC/DC networks: Insights into system planning, energy management, control, and protection. *Neural Comput. Appl.* **2024**, *36*, 17961–17977. [\[CrossRef\]](#)
11. Dragicevic, T.; Lu, X.; Vasquez, J.C.; Guerrero, J.M. DC microgrids—Part I: A review of control strategies and stabilization techniques. *IEEE Trans. Power Electron.* **2016**, *31*, 4876–4891. [\[CrossRef\]](#)
12. Dragicevic, T.; Lu, X.; Vasquez, J.C.; Guerrero, J.M. DC microgrids—Part II: A review of power architectures, applications, and standardization issues. *IEEE Trans. Power Electron.* **2016**, *31*, 3528–3549. [\[CrossRef\]](#)
13. Hassan, S.J.U.; Ali, M.; Abbas, G.; Hussain, M.R. Towards medium voltage hybrid AC/DC distribution systems: Architectural topologies, planning and operation. *Int. J. Electr. Power Energy Syst.* **2024**, *159*, 110003. [\[CrossRef\]](#)
14. Elsayed, A.T.; Mohamed, A.A.; Mohammed, O.A. DC microgrids and distribution systems: An overview. *Electr. Power Syst. Res.* **2015**, *119*, 407–417. [\[CrossRef\]](#)
15. Liu, X.; Wang, P.; Loh, P.C. A hybrid AC/DC microgrid and its coordination control. *IEEE Trans. Smart Grid* **2011**, *2*, 278–286. [\[CrossRef\]](#)
16. Dragičević, T.; Guerrero, J.M.; Vasquez, J.C.; Škrlec, D. Supervisory control of an adaptive-droop regulated DC microgrid with battery management capability. *IEEE Trans. Power Electron.* **2014**, *29*, 695–706. [\[CrossRef\]](#)
17. Guerrero, J.M.; Chandorkar, M.; Lee, T.L.; Loh, P.C. Advanced control architectures for intelligent microgrids—Part I: Decentralized and hierarchical control. *IEEE Trans. Ind. Electron.* **2013**, *60*, 1254–1262. [\[CrossRef\]](#)
18. Adegbeyega, A.W.; Sepasi, S.; Howlader, H.O.R.; Griswold, B.; Matsuura, M.; Roose, L.R. DC microgrid deployments and challenges: A comprehensive review of academic and corporate implementations. *Energies* **2025**, *18*, 1064. [\[CrossRef\]](#)
19. Lasseter, R.H. Smart distribution: Coupled microgrids. *Proc. IEEE* **2011**, *99*, 1074–1082. [\[CrossRef\]](#)
20. Abhishek, A.; Ranjan, A.; Devassy, S.; Verma, B.K.; Ram, S.K.; Dhakar, A.K. Review of hierarchical control strategies for DC microgrid. *IET Renew. Power Gener.* **2020**, *14*, 1631–1640. [\[CrossRef\]](#)
21. Meng, L.; Shafiee, Q.; Trecate, G.F.; Karimi, H.; Fulwani, D.; Lu, X.; Guerrero, J.M. Review on control of DC microgrids and multiple microgrid clusters. *IEEE J. Emerg. Sel. Top. Power Electron.* **2017**, *5*, 928–948. [\[CrossRef\]](#)
22. Osorio, E.; Saavedra-Montes, A.J.; Ramos-Paja, C.A.; Herrera-Murcia, J.G.; Bastidas-Rodríguez, J.D. Power Converters for Microgrids and Distributed Generation Systems. *Tecnológicas* **2023**, *26*, e2498. [\[CrossRef\]](#)
23. Ahmed, I.; Rehan, M.; Basit, A.; Ahmad, H.; Ahmed, W.; Ullah, N.; Piecha, M.; Blazek, V.; Prokop, L. Review on microgrids design and monitoring approaches for sustainable green energy networks. *Sci. Rep.* **2023**, *13*, 21663. [\[CrossRef\]](#)
24. Zhang, Y.; Wang, J.; Wang, X.; Zhao, J. Review on probabilistic forecasting of wind power generation. *Renew. Sustain. Energy Rev.* **2014**, *32*, 255–270. [\[CrossRef\]](#)
25. Biju, N.; Fang, H. BattX: An equivalent circuit model for lithium-ion batteries over broad current ranges. *Appl. Energy* **2023**, *339*, 120905. [\[CrossRef\]](#)
26. Li, Y.; He, S.; Li, Y.; Shi, Y.; Zeng, Z. Federated multiagent deep reinforcement learning approach via physics-informed reward for multimicrogrid energy management. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**, *35*, 5902–5914. [\[CrossRef\]](#)
27. Naidji, I.; Choucha, C.E.; Ramdani, M. Decentralized Federated Learning Architecture for Networked Microgrids. In *Proceedings of the 20th International Conference on Informatics in Control, Automation and Robotics (ICINCO)*; SciTePress: Setubal, Portugal, 2023; Volume 1, pp. 291–294, ISBN: 978-989-758-670-5. [\[CrossRef\]](#)
28. Franco, A.; Salza, P. Strategies for optimal penetration of intermittent renewables in complex energy systems based on techno-operational objectives. *Renew. Energy* **2011**, *36*, 743–753. [\[CrossRef\]](#)
29. Oliveira da Silva, C.F.; Dabiri, A.; De Schutter, B. Integrating Reinforcement Learning and Model Predictive Control for Mixed-Logical Dynamical Systems. *IEEE Open J. Control. Syst.* **2025**, *4*, 316–331. [\[CrossRef\]](#)
30. Mohammadi, M.; Noorollahi, Y.; Mohammadi-Ivatloo, B.; Yousefi, H. Energy hub: From a model to a concept—A review. *Renew. Sustain. Energy Rev.* **2017**, *80*, 1512–1527. [\[CrossRef\]](#)
31. Hadi, M.; Elbouchikhi, E.; Zhou, Z.; Saim, A.; Shafie-khah, M.; Siano, P.; Rahbarimaghani, H.; Colom, P.M. Artificial intelligence for microgrids design, control, and maintenance: A comprehensive review and prospects. *Energy Convers. Manag.* **2025**, *27*, 101056. [\[CrossRef\]](#)
32. Saeed, M.; Khalatbarisoltani, A.; Deng, Z.; Liu, W.; Altaf, F.; Lu, S.; Hu, X. Comparative analysis of control observer-based methods for state estimation of lithium-ion batteries in practical scenarios. *IEEE/ASME Trans. Mechatron.* **2025**, *30*, 3697–3709. [\[CrossRef\]](#)
33. Saeed, M.; Khalatbarisoltani, A.; Deng, Z.; Hu, X. Integrated framework for accurate state estimation of lithium-ion batteries subject to measurement uncertainties. *IEEE Trans. Power Electron.* **2024**, *39*, 8813–8823. [\[CrossRef\]](#)
34. Saeed, M.; Khalatbarisoltani, A.; Deng, Z.; Liu, W.; Lu, S.; Liu, C.; Hu, X. A PID framework with intelligent derivative control and model bias for robust state estimation of large format lithium-ion batteries. *IEEE/ASME Trans. Mechatron.* **2025**, *30*, 8038–8050. [\[CrossRef\]](#)

35. Pannala, S.; Valecha, P.; Mohtat, P.; Siegel, J.B.; Stefanopoulou, A.G. Electrochemical Battery State Estimation Under Parameter Uncertainty Caused by Aging Using Expansion Measurements. In *2021 American Control Conference (ACC)*; IEEE: New Orleans, LA, USA, 2021; pp. 3088–3093. [[CrossRef](#)]
36. Dahmardeh, M. Reliable State Estimation of Lithium-Ion Batteries Under Model and Parameter Uncertainties. Ph.D. Thesis, The State University of New Jersey, New Brunswick, NJ, USA, 2021.
37. Guo, F.; Couto, L.D.; Mulder, G.; Trad, K.; Hu, G.; Capron, O.; Haghverdi, K. A systematic review of electrochemical model-based lithium-ion battery state estimation and management systems. *J. Energy Storage* **2024**, *101*, 113850. [[CrossRef](#)]
38. Rahman, S. Estimating battery state of charge using Kalman filters. *Nat. Rev. Clean Technol.* 2026, *in press*. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.