Investigating Deep Reinforcement Learning Advantages and Response to Different Forecast Error Distribution in Energy Applications

Lorenzo Giannuzzo
Polytechnic of Turin Energy Center Lab, Polytechnic of Turin, via Paolo Borsellino 38/16, 10152
Turin, Italy
lorenzo.giannuzzo@polito.it

Francesco Demetrio Minuto
Polytechnic of Turin Energy Center Lab, Polytechnic of Turin, via Paolo Borsellino 38/16, 10152
Turin, Italy
francesco.minuto@polito.it

Daniele Salvatore Schiera
Polytechnic of Turin Energy Center Lab, Polytechnic of Turin, via Paolo Borsellino 38/16, 10152
Turin, Italy
daniele.schiera@polito.it

Andrea Lanzini
Polytechnic of Turin Energy Center Lab, Polytechnic of Turin, via Paolo Borsellino 38/16, 10152
Turin, Italy
andrea.lanzini@polito.it

Abstract*—Deep reinforcement learning (DRL) algorithms have demonstrated significant advantages over traditional optimization methods for energy management systems operating under conditions of uncertainty. However, the variability of forecast errors in electricity markets raises critical questions about the generalizability of DRL performance across different uncertainty conditions, as DRL responses may vary substantially depending on the nature and distribution of forecasting error inputs. This study compares the effectiveness of Proximal Policy Optimization (PPO) and Particle Swarm Optimization (PSO) for battery energy storage system arbitrage under three distinct forecast error distributions: random, normal, and Ornstein-Uhlenbeck. These distributions represent diverse real-world forecasting scenarios with varying error structures and temporal correlations. Using a one-year dataset of 8,760 hourly time steps from a 4 MWh/2 MW battery system, we demonstrate that DRL consistently outperformed PSO across all tested scenarios. PPO achieved profit improvements ranging from 7.2% under time-correlated errors to 55.4% under random errors. This variation in performance across error distributions shows that DRL behavior significantly adapts to the characteristics of input uncertainty, with the greatest advantages emerging under the most challenging random error conditions. These findings validate that DRL advantages extend across diverse uncertainty regimes and quantify how algorithm performance responds to realistic variations in forecast error behavior.

***Keywords*—Deep Reinforcement Learning; Forecast Uncertainties; Forecast Error Distribution; Energy Applications.**

I. INTRODUCTION

Deep Reinforcement Learning (DRL) algorithms have been widely used in many scientific and commercial fields [1]. Specifically, in the field of Energy Management Systems (EMS) optimization, DRL algorithms have largely been proven to be able to handle input variable uncertainties more effectively than heuristic optimization algorithms, such as Particle Swarm Optimization (PSO) and Genetic Algorithms (GAs), as well as mathematical optimization algorithms, such as Mixed-Integer Linear Programming (MILP) [2]. However, due to the unpredictable behavior of decision variables in energy system optimization, such as volatility of electricity prices, load demand, and energy production from Renewable Energy Sources (RESs), DRL algorithms may not always respond the same way over time. Forecast algorithms are typically subject to significant errors, particularly in energy applications [3]. The variability of input variables used for forecasting (e.g., weather, solar radiance, and energy usage), combined with the various ways in which forecasting algorithms can be built and trained, may lead to variations in algorithms performance over time based on different forecasting behaviors and responses to variations in input data [4-6]. Therefore, DRL algorithm performance may vary a lot, especially in energy applications. For this reason, a more detailed study is needed to investigate how such models respond to different error distributions in order to predict performance based on changes and uncertainties in input variables and the different behavior of forecast algorithms. Many papers suggest that using DRL models for energy system optimization is highly beneficial. For example, Liu et al. proposed a home energy management optimization strategy based on Deep Q-learning (DQN) and Double Deep Q-learning (DDQN) for scheduling home energy appliances in [7]. They demonstrated that their algorithm could help customers reduce electricity consumption by responding to a dynamic environment with a series of actions. They compared their results with PSO results, which validated the performance of the proposed algorithm and demonstrated the advantages of DRL by using a real-world database combined with an energy storage model for a single household. Similarly, Guo et al. used real-time dynamic optimal energy management based on the Proximal Policy Optimization (PPO) DRL algorithm. They demonstrated that their proposed approach is more computationally efficient and accurate than PSO and MILP under input variable uncertainties [8]. Furthermore, a DRL-based approach for day-ahead scheduling of generation resources under demand and wind power uncertainties was presented in [9]. The proposed approach yielded a causal policy that relies only on historical uncertainty realizations and forecast data. This policy is trained with an actor-critic (AC)-based reinforcement learning algorithm. The results obtained showed that the proposed approach had superior performance in terms of computational efficiency and operational costs. It also significantly reduced penalty costs caused by insufficient net load supply compared to the deterministic approach. Moreover, [10] proposed a real-time economic energy management strategy based on DRL to optimize microgrid scheduling. Unlike traditional model-based approaches, the adopted strategy was learning-based and did not require an explicit model of uncertainty. Considering the uncertainties in RESs, load demand, and electricity prices, they formulated a Markov Decision Process (MDP) to solve the real-time economic energy management

979-8-3315-8138-1

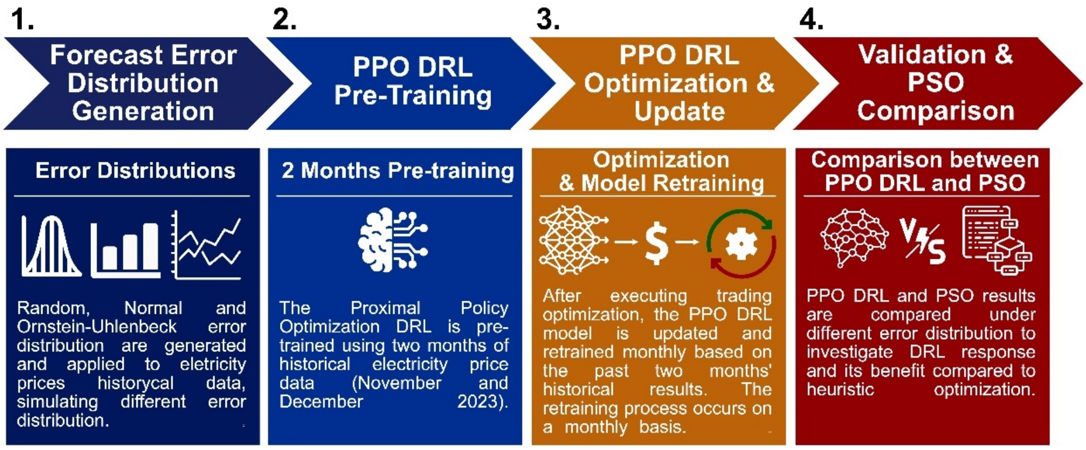

Fig. 1. Proposed methodological framework.

problem of microgrids. The objective was to minimize the system's daily operating cost by scheduling controllable distributed generators and Energy Storage Systems (ESS). Specifically, they used a Deep Deterministic Policy Gradient (DDPG) to solve the MDP. They validated the effectiveness of their approach by comparing their results to those obtained using different DRL approaches, such as DQN, Soft AC, and Model Predictive Control (MPC) methods. These comparisons demonstrated that the DDPG approach reduced operating costs by 29.59%, 17.39%, and 9.55%, respectively. In [11], a behavioral decision model for energy storage based on the SAC DRL algorithm for energy storage system dispatch was tested to address uncertainties in wind farm energy production. This demonstrates the advantages of DRL, as shown in previous analyses. Similar promising results were obtained in many different fields [12]. In these cases, the algorithms' predictions in response to input variables can change significantly, especially for the prediction of parameters like electricity prices. This may significantly change the magnitude of the error during prediction and the overall error distribution, potentially leading to different DRL operational results compared to those achieved with a limited amount of simulation time. Additionally, the error magnitude and distribution can vary based on the algorithm selection choice and the way the algorithm is trained and tested due to the different nature and structure of different forecast algorithms and the diversity of the input variables required to predict. Thus, results may not be generalizable. For these reasons, this paper proposes investigating the response of DRL algorithms to different forecast error distributions and validating that response by comparing the results obtained using DRL to those obtained using heuristic algorithms.

II. Methodology

A. Methodological Framework

Figure 1 shows the proposed methodological framework. The main goal is to study the response of DRL to different forecast errors in energy applications. For simplicity, in this paper, DRL response is evaluated in the context of energy arbitrage/trading. A stand-alone li-ion Battery Energy Storage System (BESS) is dispatched to optimize revenues generated by purchasing energy during low-price periods and selling energy during high-price periods. The methodological framework consists mainly of four phases. In the first phase different forecast error distributions are generated and applied to the system input variables (e.g., electricity prices). The different forecasts error distributions are used to simulate forecast errors that could occur in real-world scenarios. Next, a PPO DRL algorithm is created and pre-trained using two months of historical electricity price data (i.e., November and December 2023). Then, the PPO-DRL is employed to optimize the energy system and maximize revenues. The DRL model is then constantly updated through monthly retraining using the last two months of historical data. This allows the algorithm to capture the dynamic evolution of the input variables and adapt the system operation accordingly. The results obtained with the PPO DRL are then validated and compared to those obtained using a PSO within the same optimization context.

B. Case Study and Optimization Problem

This paper analyzes a case study of energy trading provided by a utility-scale BESS. Both PPO DRL and PSO are used with a rolling-horizon optimization window of 24 h, with a time resolution of 1 h. The modeled battery has a capacity of 4 MWh and a rated power of 2 MW (P_{max}). To simulate a realistic scenario, a Depth-of-Discharge (DoD) of 0.8 is used, with lower and upper State-of-Charge (SOC) limits of 0.1 (SOC_{min}) and 0.9 (SOC_{max}), respectively. Additionally, a C-rate limit of 0.5 (C_{rate}) is used, as well as a battery degradation model based on the number of cycles described in [13]. The optimization problem is modeled according to Eq (1).

$$OF = max(Revenues) = \max\left(\left(\sum_i^T\left(E_{discharged,i} - E_{charged,i}\right)\right) * c_{electricity,i} - \sum_i^T\left(E_{tot,i} * \frac{c_{aging}}{2*EOL}\right)\right) \quad (1)$$

where: $E_{discharged,i}$ is the energy discharged from the battery to the grid at the timestep i; $E_{charged,i}$ is the energy charged into the battery from the grid at the timestep i; $c_{electricity,i}$ is the electricity cost at the timestep i; $E_{tot,i}$ is the sum of the energy charged and discharged at the timestep i; c_{aging} is the unit cost of aging, used to consider degradation effects due to purchasing/selling energy over time, which is assumed to be fixed at 400 €/MW; EOL is the End-of-Life number of cycles, which is assumed to be 6000; and T is the rolling-horizon time window extension, which is 24h. BESS SOC is updated using Coulomb Counting, following Eq. (2) and (3), where a charge efficiency (η_{charge}) and a discharge efficiency (η_{charge}) of 0.95 are considered. Finally, the optimization problem constraints are summarized in Eq. (4), (5), and (6).

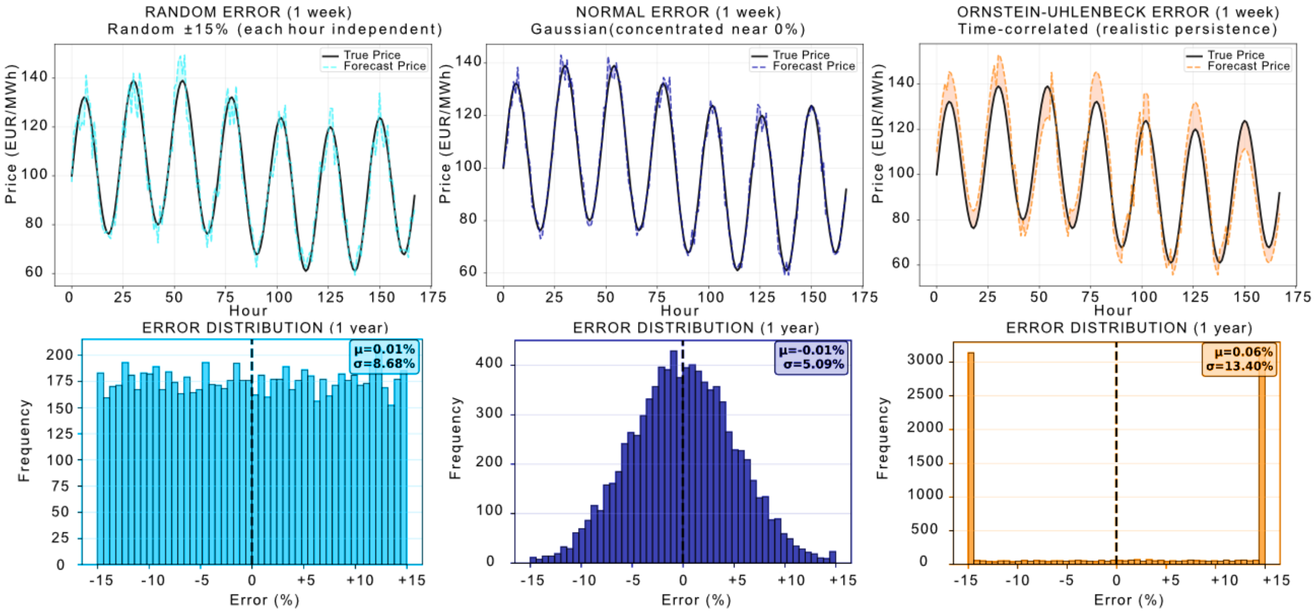

Fig. 2. Forecast error distribution applied to electricity prices. It shows a one-week comparison of three different error distributions applied to the true electricity price value, together with the overall one-year error distribution.

$$SOC_{(i+\Delta t)} = SOC_i + \frac{E_{charged,i}\,\eta_{charge}}{C_{BESS,i}} \tag{2}$$

$$SOC_{(i+\Delta t)} = SOC_i - \frac{E_{discharged,i}}{C_{BESS,i} * \eta_{discharge}} \tag{3}$$

$$SOC_{min} \leq SOC_i \leq SOC_{max} \qquad \forall\, i \in T \tag{4}$$

$$P\,-_{max} \leq P_i \leq P_{max}) \qquad \forall\, i \in T \tag{5}$$

$$E_{charged,i} * E_{discharged,i} = 0 \qquad \forall\, i \in T \tag{6}$$

C. *Proximal Policy Deep Reinforcement Learning*

This paper employs the PPO DRL algorithm. It is a DRL algorithm with an actor-critic architecture, as detailed described in [14]. The actor network is a deep neural network (DNN) whose primary function is to learn and execute a policy learned from data. The critic network, also called the "evaluator," estimates the value function to evaluate state quality. The DNN critic network measures how good the current state is under the current policy and provides a baseline for computing the advantage function. The critic network also helps stabilize actor training by providing low-variance estimates. In summary, it distinguishes between good and bad actions by providing state value baselines, which accelerates learning and improves the efficiency of the algorithms. The main parameters of the employed PPO DRL algorithms are summarized in Table I. In the proposed model, the actor and the critic DNN share the same architecture. The pre-training phase of the PPO agent consists of 500,000 timesteps, during which the neural network learns optimal battery management strategies. The architecture consists of two hidden layers, with 1,024 and 512 neurons, respectively, for the actor (policy) and critic (value function) networks. The input layer receives 26 features, including the current state of charge, a 24-hour forecast of electricity prices, and the current real-time price. During training, the agent gathers experience through 2048 step rollouts before updating, balancing exploration and computational efficiency. Each update involves ten epochs of optimization using mini batches of sixty-four samples, with a learning rate of 10^{-4} controlling the step size of gradient descent. The discount factor, gamma, is set to 0.99, meaning the agent values future rewards at 99% of their original value per time step. This encourages long-term strategic planning. The generalized advantage estimation lambda parameter of 0.95 balances the bias and variance in the advantage calculations, resulting in more stable learning. The clip range of 0.2 prevents destructively large policy updates that could destabilize training. Regularization is achieved through an entropy coefficient of 0.01, which rewards policy diversity and encourages exploration, as well as a value function coefficient of 0.5, which balances the importance of

TABLE I. PPO PARAMETERS

Parameter	Value	Parameter	Value
Pre-training		***Hidden Layers (Actor & Critic)***	
Learning Rate	1e-4	Number of neurons first layer	1024
Total timesteps	500 000	Number of neurons second layer	512
Steps per rollout	2048	Input Features	26 (SOC (1), forecast electricity prices (24), Real price (1))
Batch size	64	***Regularization***	
Epochs per update	10	Entropy Coefficient	0.01
Discount & GAE		Clip Range	0.2
Gamma	0.99	Value Function Coefficient	0.5
GAE Lambda	0.95	Max Gradient norm	0.5

979-8-3315-8138-1

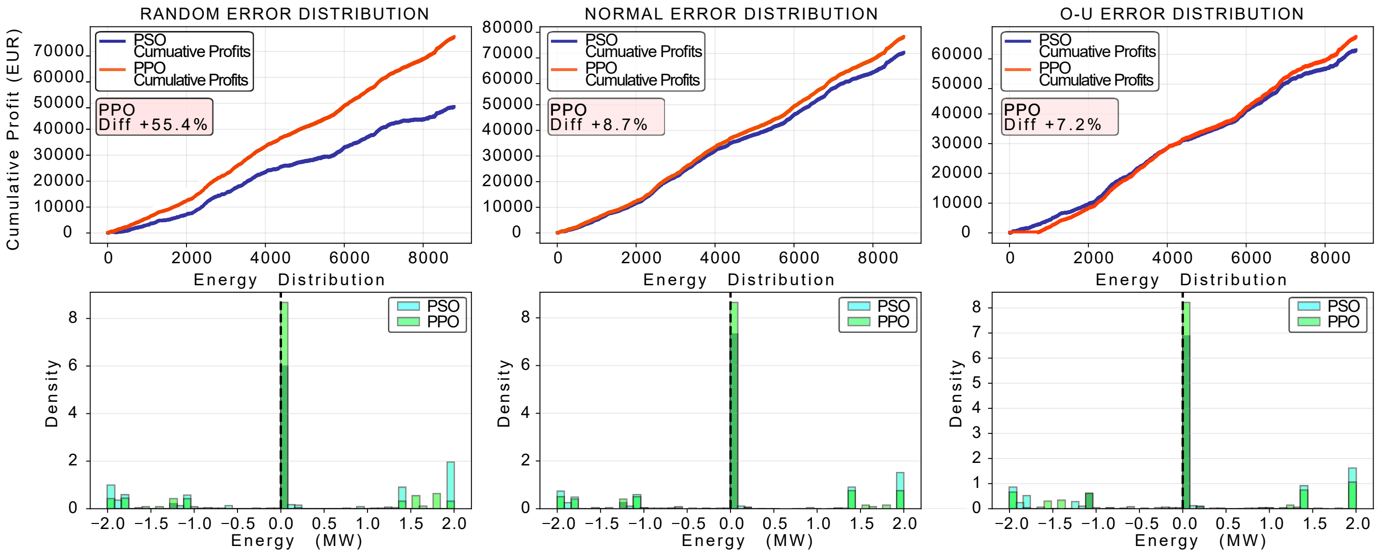

Fig. 2. PPO DRL and PSO profit and BESS energy density distribution comparison over one year.

value function accuracy relative to policy improvement. Finally, the maximum gradient norm of 0.5 prevents exploding gradients by clipping them and ensures stable convergence throughout the training process. The reward (r) and profit functions are similar to Eq. (1), and are defined according to Eq. (7), and (8).

$$r(t) = \pi(t) - C_{deg}(t) \tag{7}$$

$$\pi(t) = -P(t) * c_{electricity}(t) * \Delta t \tag{8}$$

where: $\pi(t)$ is the profit at the timestep t; $P(t)$ is the power at the timestep t; and C_{deg} is the degradation cost, which is evaluated as stated in Eq.(1).

D. Particle Swarm Optimization

In this study, a PSO is used to validate and compare the results obtained by the proposed DRL model. The objective function is the same the one used in PPO (Eq. (1) and (7)), using the same optimization constraints expressed in Eq. (4), (5), and (6). The PSO algorithm employed 50 particles over 150 iterations with a 24-hour time horizon at 1-hour resolution. The inertia weight decreased linearly from 0.95 to 0.2, and both acceleration coefficients (c1 and c2) were set to 2.0.

E. Forecast Error Distribution Generation

Forecast error is added to the real value based on three different error distributions: random, normal, and Ornstein-Uhlenbeck. According to different papers, like [15-17], the maximum forecast error on electricity prices is assumed to be 15%. For testing their impact on DRL response for energy trading, error distributions are applied on historical electricity prices as shown in Figure 2. As can be seen, the normal distribution has most of its errors concentrated near 0%, with lower frequencies close to the extremes. In contrast, random distribution has almost equal frequencies across the entire error range. Ornstein-Uhlenbeck distribution, on the other hand, has most of its errors concentrated near the extremes.

F. Data

The electricity prices are obtained through the Italian Electricity Market Operator (GME) for the year 2023 (2 months pre-training) and 2024 [18].

III. Results and Discussion

The results presented in Figure 3 were obtained by executing the energy trading optimization problem with both PPO-based DRL and PSO over a comprehensive one-year dataset comprising 8,760 hourly time steps. The comparative analysis encompasses three distinct forecast error distributions: random, normal, and Ornstein-Uhlenbeck, each representing different real-world forecasting uncertainty scenarios that BESS operators may encounter in electricity markets. Across all three tested error distributions, PPO consistently demonstrates superior cumulative profitability compared to PSO, establishing the general advantage of employing DRL methodologies for optimizing energy systems operating under forecasting uncertainties. This systematic outperformance validates the hypothesis that reinforcement learning algorithms can effectively learn adaptive trading strategies that account for stochastic price dynamics. The cumulative profit trajectories, visualized in the upper panels of Figure 3, reveal that both algorithms generate positive returns throughout the annual simulation period, indicating that arbitrage opportunities exist and can be exploited even under imperfect price forecasting conditions. However, the divergence between the PPO and PSO profit curves becomes increasingly pronounced as the simulation progresses, particularly beyond the 4,000-hour mark (approximately mid-year), suggesting that PPO's learned policy exhibits superior long-term strategic planning capabilities and better risk-adjusted decision-making when compounding effects of trading decisions accumulate over extended time horizons. The best results emerge under the random error distribution scenario, in which forecast errors span the full range of ±15% around the true electricity price. PPO achieves a remarkable cumulative annual profit of approximately €73,000, compared to PSO's €47,000, representing a substantial 55.4% improvement in profit. This is the largest performance gap among the three scenarios and is particularly significant given that uniform random errors constitute the most challenging and unpredictable forecasting environment. PPO's superior performance stems from its neural network architecture, which enables complex nonlinear mappings between the 26-dimensional state space and optimal actions. Combined with monthly retraining, this architecture continuously adapts the policy based on the most recent 60 days of observed forecast errors. Energy distribution density

979-8-3315-8138-1

plots reveal that PPO has a distinctly more conservative action distribution with a pronounced peak at zero power and significantly reduced density at maximum charge/discharge extremes. PPO maintains idle operation approximately 35-40% of the time, whereas PSO maintains idle operation 25-30% of the time. Under normally distributed forecast errors with $\sigma \approx 5\%$, both algorithms demonstrate improved absolute profitability. PPO achieves approximately €70,000, and PSO reaches €64,500 in annual cumulative profits. This represents an 8.7% profit advantage for PPO. This improved performance is due to the concentration of forecast errors near zero. The Gaussian distribution allocates approximately 68% of errors within ±5% and 95% within ±10%. This creates a more favorable environment than the uniform distribution. PPO's action distribution exhibits a sharper central peak and lighter tails. This indicates a preference for moderate power levels, balancing profit maximization with prudent risk management to avoid excessive degradation costs and SOC constraint violations. The Ornstein-Uhlenbeck error distribution presents the most challenging operational environment, with both algorithms achieving their lowest absolute profitability. PPO generates approximately €62,000, compared to €58,000 for PSO, corresponding to a 7.2% advantage for PPO. PPO's ability to identify patterns in which consecutive forecast errors are correlated gives it a measurable advantage over PSO's myopic optimization, which treats each hour's forecast independently. The analysis of the energy distribution reveals that both algorithms adopt more conservative strategies under these conditions. PPO demonstrates greater conservatism with an even more pronounced concentration of actions near zero power. It rationally responds to the higher effective uncertainty created by error persistence by reducing trading frequency.

IV. Conclusion

This study presents a comprehensive comparative analysis of DRL and PSO for BESS arbitrage under three distinct forecast error distributions: uniform random, normal, and Ornstein-Uhlenbeck. The results demonstrate that DRL consistently outperforms PSO across all tested scenarios, achieving profit improvements ranging from 7.2% to 55.4%, addressing the research gap regarding DRL performance generalizability under varying forecast uncertainty conditions. The fundamental advantage of DRL, specifically the employed PPO-based algorithm, stems from its capacity to learn adaptive risk-adjusted policies through extensive environment interaction and monthly retraining. Our analysis reveals systematically more conservative trading strategies, with PPO maintaining 35-40% idle frequency compared to PSO's 25-30%, reducing adverse selection risk while preserving operational flexibility. These findings validate that DRL's advantages in energy system optimization extend across diverse uncertainty regimes, from favorable low-error environments to challenging time-correlated conditions. Future research could investigate alternative DRL algorithms, incorporate explicit uncertainty characterization into state representations, extend to multi-market participation with ancillary services, and develop enhanced interpretability tools for regulatory compliance.

Acknowledgment

The authors carried out this publication within the project “Network 4 Energy Sustainable Transition – NEST”, Project code PE0000021, Concession Decree No. 1561 of 11.10.2022 adopted by Ministero dell’Università e della Ricerca (MUR), CUP E13C22001890001. Project funded under the National Recovery and Resilience Plan (NRRP), Mission 4 Component 2 Investment 1.3 - Call for tender No. 341 of 15.03.2022 of MUR; funded by the European Union – NextGenerationEU.

References

[1] Li, S. E. (2023). Deep reinforcement learning. In *Reinforcement learning for sequential decision and optimal control* (pp. 365-402). Singapore: Springer Nature Singapore.

[2] Zhang, Z., Zhang, D., & Qiu, R. C. (2019). Deep reinforcement learning for power system applications: An overview. *CSEE Journal of Power and Energy Systems*, *6*(1), 213-225.

[3] Sanders, N. R., & Graman, G. A. (2009). Quantifying costs of forecast errors: A case study of the warehouse environment. *Omega*, *37*(1), 116-125.

[4] Bezdek, R. H., & Wendling, R. M. (2002). A half century of long-range energy forecasts: errors made, lessons learned, and implications for forecasting. *Journal of Fusion Energy*, *21*(3), 155-172.

[5] Wu, J., Zhang, B., Li, H., Li, Z., Chen, Y., & Miao, X. (2014). Statistical distribution for wind power forecast error and its application to determine optimal size of energy storage system. *International Journal of Electrical Power & Energy Systems*, *55*, 100-107.

[6] Alessandrini, S., & Sperati, S. (2017). Characterization of forecast errors and benchmarking of renewable energy forecasts. In *Renewable Energy Forecasting* (pp. 235-256). Woodhead Publishing.

[7] Liu, D., Zang, C., Zeng, P., Li, W., Wang, X., Liu, Y., & Xu, S. (2023). Deep reinforcement learning for real-time economic energy management of microgrid system considering uncertainties. *Frontiers in Energy Research*, *11*, 1163053.

[8] Guo, C., Wang, X., Zheng, Y., & Zhang, F. (2022). Real-time optimal energy management of microgrid with uncertainties based on deep reinforcement learning. *Energy*, *238*, 121873.

[9] Ajagekar, A., & You, F. (2022). Deep reinforcement learning based unit commitment scheduling under load and wind power uncertainty. *IEEE Transactions on Sustainable Energy*, *14*(2), 803-812.

[10] Cui, Y., Xu, Y., Li, Y., Wang, Y., & Zou, X. (2024). Deep reinforcement learning based optimal energy management of multi-energy microgrids with uncertainties. *CSEE Journal of Power and Energy Systems*.

[11] Li, Z., Xiang, Y., & Liu, J. (2025). Forecasting error-aware optimal dispatch of wind-storage integrated power systems: A soft-actor-critic deep reinforcement learning approach. *Energy*, *318*, 134798.

[12] Ramkumar, M. S., Subramani, J., Sivaramkrishnan, M., Munimathan, A., Michael, G. K. O., & Alam, M. M. (2025). Optimal energy management for multi-energy microgrids using hybrid solutions to address renewable energy source uncertainty. *Scientific Reports*, *15*(1), 7755.

[13] Giannuzzo, L., Massano, M., Schiera, D. S., Minuto, F. D., Papurello, D., Meneghetti, F., & Lanzini, A. (2025, July). Benchmarking Genetic Algorithms for Short-Term Battery Energy Storage Systems Optimization. In *2025 IEEE 49th Annual Computers, Software, and Applications Conference (COMPSAC)* (pp. 2053-2059). IEEe.

[14] Guo, C., Wang, X., Zheng, Y., & Zhang, F. (2022). Real-time optimal energy management of microgrid with uncertainties based on deep reinforcement learning. *Energy*, *238*, 121873.

[15] Jiang, P., Nie, Y., Wang, J., & Huang, X. (2023). Multivariable short-term electricity price forecasting using artificial intelligence and multi-input multi-output scheme. *Energy Economics*, *117*, 106471.

[16] Yang, W., Sun, S., Hao, Y., & Wang, S. (2022). A novel machine learning-based electricity price forecasting model based on optimal model selection strategy. *Energy*, *238*, 121989.

[17] Belenguer, E., Segarra-Tamarit, J., Pérez, E., & Vidal-Albalate, R. (2025). Short-term electricity price forecasting through demand and renewable generation prediction. Mathematics and Computers in Simulation, 229, 350-361.

[18] Home. Prezzi Zonali. (n.d.). https://gme.mercatoelettrico.org/it-it/Home/Esiti/Elettricita/MI/MI-A1/Esiti/PrezziZonali