

Intelligent Energy Management of a Microgrid Using Reinforcement Learning

Mohammad Safayet Hossain

Department of Electrical and Computer Engineering
University of Central Florida
Orlando, FL, USA
mohammad.safayet.hossain@ucf.edu.

Hisham Mahmood

Pacific Northwest National Laboratory
Richland, WA, USA
hisham.mahmood@pnnl.gov.

Abstract—This paper presents an intelligent energy management system (EMS) for microgrids. The intelligent EMS employs load and photovoltaic (PV) generation forecasting schemes and a double deep Q-learning network (DDQN) based reinforcement learning (RL) scheme. The trained RL agent policy utilizes the 24-h ahead forecast to schedule the generation and the battery charge/discharge operation to minimize the microgrid operating cost and meet the battery, fuel, and the grid feeder constraints. Simulation results show that the trained EMS agent is capable of producing a near-optimal scheduling policy when subjected to new test scenarios. The agent shows a consistent trend of adjusting the microgrid operation based on the type of the day to meet the policy constraints. Moreover, the performance of the proposed EMS is compared with that of deep Q-learning and proximal policy optimization algorithms to explore the superiority of double deep Q-learning.

Index Terms—energy management system, microgrid, reinforcement learning, double deep, Q-learning.

I. INTRODUCTION

According to the IEEE Standard 2030.7–2017, a microgrid is defined as a group of interconnected electric loads and distributed energy resources in a defined electrical boundary. It can be operated in two modes, grid-connected and islanded. Renewable energy sources (RESs) such as solar and wind are expected to dominate. These energy sources are intermittent and stochastic in nature. The electric load profile is also considered as a stochastic time series. Consequently, generation from these renewable sources is not always aligned with the load demand which makes it a challenging task to operate the microgrid efficiently and reliably. Energy storage units are required in microgrids to mitigate this challenge, among other tasks. An energy storage system (ESS) stores/supplies energy when needed to efficiently use the available resources while considering the variable energy prices and the electric network limitations. This efficient operation of the microgrid is achieved through optimizing the power set points of the energy storage and dispatchable units by the EMS to reduce the operational costs. Hence, the operation of the microgrid requires an efficient and adaptable generation scheduling policy. The main challenge for any EMS is the stochastic nature of the intermittent RESs and the load demand.

Several approaches for microgrid EMSs are available in the literature to solve day-ahead and real-time scheduling

problems of microgrids. Day ahead scheduling approaches alone are not able to manage the real-time uncertainty of load and RES generation as the scheduling policy is predetermined and fixed for the whole day [1]. The decision tree based EMS is one of the popular conventional methods to operate a microgrid dynamically in real-time. However, the decision tree method may cause forced outages as reported in [2]. An energy system is proposed in [1] to optimize the operational cost of a microgrid using a deep neural network (DNN) and dynamic programming (DP). An intelligent EMS is designed in [2] using adaptive DP and RL to supply critical load continuously by leveraging renewable energy. However, the forecasting knowledge of load and renewable energy are not considered in the strategies proposed in [1], [2]. Recently, prediction models for load demand and renewable energy generation are exploited to improve the robustness of energy management algorithms [3]–[7]. The forecasting data of PV generation and load demand are utilized in a model predictive control (MPC) algorithm in [3], [4] with mixed integer linear programming (MILP) to optimize the battery operation in a microgrid. Forecasts of electric load and renewable energy generation are employed in an MILP based EMS in [5] to optimize the operation of distributed energy resources (DERs). A predictive EMS is proposed in [6] and [7] based on utilizing forecasts of PV generation and load, to reduce load shedding in the considered islanded microgrid. However, the proposed strategies in [6] and [7] are considered deterministic and do not consider the forecast uncertainty. Reinforcement learning (RL) is one of the promising optimization methods reported in the literature for solving energy system problems [2], [8]–[12]. A Q-learning based EMS is proposed in [8] to optimize a battery operation while considering wind forecasts. However, the Q-learning algorithm struggles to solve a large state space problem [10]. Multiple microgrids are controlled by multiple agents using proximal policy optimization (PPO) algorithm in [13]. However, the proposed EMS can be improved by integrating PV power forecasting and load forecasting data.

Inspired by the previous works, an intelligent predictive EMS is proposed in this paper. The trained agent considers the day ahead forecast, battery state-of-charge (SOC), microturbine state-of-fuel (SOF), and the energy tariff to schedule the battery and microturbine operation. The 24-hour forecast

data of the electrical load and PV generation are produced using a long short term memory (LSTM) neural network (NN) based forecasting models. A near-optimal policy is obtained by training a double deep Q network (DDQN) agent to optimize the battery and the microturbine operation. The paper provides a complete framework to design and verify predictive EMS for microgrids. Simulation results show that agent realizes a consistent trend of adjusting the microgrid operation based on the type of the day to meet the policy constraints and achieve a near optimal operation. Finally, the performance of the DDQN is compared with that of a deep Q network (DQN) and a proximal policy optimization (PPO) agent. After the brief introduction, Section II describes the problem statement. Section III presents the microgrid control scheme using double deep Q -learning. Section IV presents and illustrates simulation results. The paper is concluded in Section V.

II. PROBLEM STATEMENT

A microgrid comprising a photovoltaic plant (1200kW), a battery (2000kWh), a microturbine gas generator (150kW) and residential loads is considered in this paper, as shown in Fig. 1. The microgrid operation is formulated for a finite horizon of time. The measurements throughout a full day are indexed as, $T = \{t \in \mathbb{N} : 1 \leq t \leq 24\}$ with $\Delta t = 1$ hour time interval. The battery operation is formulated as in equations (1) - (4). The battery power¹, P_t^{bat} and the charging/discharging efficiencies (η_{ch}, η_{dis}) are utilized in the equation of the state of charge, SOC [7], [14]. The battery is operated by controlling the SOC within predefined constraints as in equation (3). Meanwhile, it is desired to maintain a certain SOC level at the end of the day to cover the early time period of the next day operation.

$$-P_t^{bat,max} \leq P_t^{bat} \leq P_t^{bat,max} \quad (1)$$

$$SOC_{t+1} = \begin{cases} SOC_t - (\frac{\eta_{ch} \cdot P_t^{bat} \cdot \Delta t \cdot 100}{E_{bat}}); P_t^{bat} < 0, \forall t \in T \\ SOC_t - (\frac{P_t^{bat} \cdot \Delta t \cdot 100}{E_{bat} \cdot \eta_{dis}}); P_t^{bat} > 0, \forall t \in T \end{cases} \quad (2)$$

$$SOC^{min} \leq SOC_t \leq SOC^{max}, \forall t \in T \quad (3)$$

$$SOC^{end,min} \leq SOC_t \leq SOC^{end,max}, t = 24 \quad (4)$$

The microturbine model is derived from Gottwalt *et al.* [15] and Simons *et al.* [16]. The model includes the fuel consumption (P^{fuel}), utilization factor (ρ^{gen}), microturbine capacity (E_{gen}), efficiency (η^{gen}), cost of fuel (c^{fuel}), gas price² (c_{gas}), and the state of fuel (SOF). The active power supplied by the microturbine (P^{gen}) and the SOF are constrained as in equations (6) and (12), respectively. The deficit in generation in the microgrid is supplied by the microturbine, as in (5). The fuel cost has two parts, the variable cost and the ramping cost, as formulated in (7). The ramping cost is considered at the first hour of the microturbine operation only, which is indicated by the binary variable, b^r .

¹ $P_t^{bat,max} = 400kW, \eta_{ch} = 90\%, \eta_{dis} = 90\%$

² 1 MMBTU=293.071 kWh, the gas price is 4.09 \$/MMBTU=0.014 \$/kWh (<https://www.eia.gov/dnav/ng/hist/rngwhhdm.htm>)

$$P_t^{gen} = P_t^L - (P_t^{PV} + P_t^{bat}), \forall t \in T \quad (5)$$

$$P_t^{gen,min} \leq P_t^{gen} \leq P_t^{gen,max}, \forall t \in T \quad (6)$$

$$c_t^{fuel} = c_{gas} \cdot \underbrace{((1.4306 \cdot P_t^{gen} + 0.2788 \cdot E_{gen}))}_{\text{variable fuel cost}} + \underbrace{2 \cdot \Delta t \cdot (1.4306 \cdot 0.4 \cdot E_{gen} + 0.2788 \cdot E_{gen}) \cdot b_t^r}_{\text{ramping cost}}, \forall t \in T \quad (7)$$

$$\rho_t^{gen} = \frac{P_t^{gen}}{E_{gen}}, \forall t \in T \quad (8)$$

$$\eta_t^{gen} = 0.1917 \cdot \rho_t^{gen} + 0.3933, \forall t \in T \quad (9)$$

$$P_t^{fuel} = \left(\frac{P_t^{gen} \cdot 3412}{\eta_t^{gen}} \right), \forall t \in T \quad (10)$$

$$SOF_{t+1} = SOF_t - \left(\frac{P_t^{fuel} \cdot \Delta t \cdot 100}{E_{fuel}} \right), \forall t \in T \quad (11)$$

$$SOF_t \geq SOF^{min}, \forall t \in T \quad (12)$$

The microgrid exports power to the utility grid to participate in the electricity market. A grid limit is considered for the power exchange as in equation (13). The main grid power exchange (P^G) is determined using the microgrid power balance as in equation (14) [11]. The real-time electricity price (R^{tariff}) and feed-in tariff (R^{FIT}) rates are used for the energy import and export respectively, as in equation (15).

$$-P_t^{G,max} \leq P_t^G \leq P_t^{G,max}, \forall t \in T \quad (13)$$

$$P_t^G = P_t^L - (P_t^{PV} + P_t^{bat} + P_t^{gen}), \forall t \in T \quad (14)$$

$$R_t = \begin{cases} R^{tariff}; P_t^G > 0, \forall t \in T \\ R^{FIT}; P_t^G < 0, \forall t \in T \end{cases} \quad (15)$$

The day ahead forecasting scheme as of PV generation and load demand are integrated within the EMS. The PV power forecast ($P^{PV,f}$) and the load forecast ($P^{L,f}$) are utilized to estimate the day ahead grid power profile ($P^{G,f}$). The day ahead hourly PV power and load forecasting model is developed using an LSTM neural network as in [17], [18]. The predicted day ahead grid power is calculated as the difference between the PV power forecast and the load forecast as in (16). Whereas, the expected grid power profile of a day, $P_d^{G,f} \in \mathbb{R}^{24}$ is calculated from the PV power forecasting profile, $P_d^{PV,f} \in \mathbb{R}^{24}$ and load forecasting profile, $P_d^{L,f} \in \mathbb{R}^{24}$ of a day. Where, $P_d^{G,f} \in \mathbb{R}^{24}$ contains the predicted power differences of all time instants of a day. The expected grid power of a day can be estimated as

$$P_d^{G,f} = P_d^{L,f} - P_d^{PV,f} \quad (16)$$

The microgrid operational cost for 24 hours (C_d^{opt}) is calculated as follows [1].

$$C_d^{opt} = \sum_{t=1}^{24} (R_t \cdot P_t^G \cdot \Delta t + c_t^{fuel}) \quad (17)$$

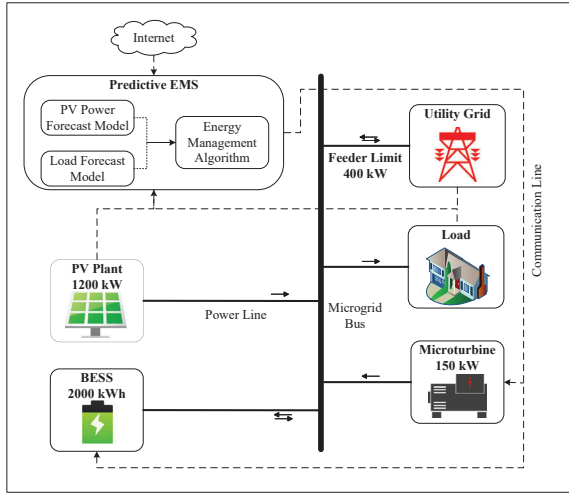


Fig. 1. Microgrid structure with the predictive EMS.

III. DOUBLE DEEP Q -LEARNING BASED ENERGY MANAGEMENT STRATEGY

A. Basics of Double Deep Q -Learning

The double deep Q -Learning algorithm is introduced by Hasselt et al. [19]. In deep Q -learning, two deep neural networks (DNN) are used to approximate the action values. The first DNN, so called the online network, is used to approximate the action values for a given state. This network is implemented using a set of parameters, θ , where θ is updated using gradient decent method as follows:

$$\theta_{t+1} = \theta_t + \alpha(Y_t^Q - Q(s_t, a_t; \theta_t)) \cdot \nabla_{\theta_t} Q(s_t, a_t; \theta_t) \quad (18)$$

where α is the training step size. The DNN maps the state vector, $s \in \mathbb{R}^n$ into the action vector, $a \in \mathbb{R}^m$ and defined as, $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$. Another DNN, so called the target network, with another set of parameters, θ^- , is used to evaluate the actions and serves as a target to train the online network. The network parameter set θ^- is duplicated from the online network regularly once every certain number of time steps, i.e., $\theta_t = \theta_t^-$. The target value, Y_t can be expressed by,

$$Y_t^{DDQN} \equiv r_{t+1} + \gamma \cdot \max_a Q(s_{t+1}, a; \theta_t^-) \quad (19)$$

where γ is the discount factor, $\gamma \in (0, 1)$ and r_{t+1} is the reward of the next time step. The new feature of the deep Q -learning algorithm is the inclusion of an experience buffer, also called the replay memory. The agent stores its experiences in a replay memory with a defined sample size, and draws a sample batch randomly to train the network.

One drawback of this approach is that the approximated action values can be overestimated as the actions are selected and evaluated by same values in Q -learning and Deep Q -learning as shown by equation (20) which is the decoupled form of equation (19).

$$Y_t^{DDQN} \equiv r_{t+1} + \gamma \cdot Q(s_{t+1}, \arg \max_a Q(s_{t+1}, a; \theta_t); \theta_t) \quad (20)$$

On the other hand, in a double deep Q network (DDQN), the machine learns two value functions using two different weight vectors θ and θ^- to overcome the overestimation of deep Q -learning algorithm.

$$Y_t^{DDQN} \equiv r_{t+1} + \gamma \cdot Q(s_{t+1}, \arg \max_a Q(s_{t+1}, a; \theta_t); \theta_t^-) \quad (21)$$

B. Predictive Control Strategy

The PV power output (P^{PV}), the load demand (P^L), the state-of-charge (SOC) of the battery, the utility grid tariff (R), time of the day index (I), and the predicted grid power profile of a day ($P_d^{G,f} \in \mathbb{R}^{24}$) are included in the state space, \mathcal{S} where the state vector is defined as, $s_t = \{SOC_t, SOF_t, P_t^{PV}, P_t^L, R_t, I_t, P_d^{G,f}\}^T \in \mathbb{R}^{53} \subset \mathcal{S}$. The numerical states are normalized using the wellknown min-max formula. Whereas, the one hot encoder technique is used to normalize the categorical state, $I_t \in \mathbb{R}^{24}$. The actions of the DDQN agent comprises the control commands of the BESS and microturbine $a_t = \{a_t^{bat}, a_t^{gen}\}^T \in \mathbb{R}^2 \subset \mathcal{A}$, where the action space, $\mathcal{A} = \{(\frac{x}{10}, y) : x \in [-10..10], y \in [0, 1]\}$ includes all possible action pairs. Then the battery power is formulated as, $P_t^{bat} = a_t^{bat} \cdot P^{bat,max}$ and the generator power can be written as, $P_t^{gen} = a_t^{gen} \cdot (P_t^L - (P_t^{PV} + P_t^{bat}))$. The RL agent observes, $s_{t+1} \in \mathbb{R}^{53} \subset \mathcal{S}$. The reward function considers the penalties, $R^{SOC}, R^{SOC,end}, R^{SOF}, R^G$ associated with the constraints of SOC_t, SOF_t , and P_t^G . Where, $R^{SOC} = R^{SOF} = R^G = 100, R^{SOC,end} = 500$. The objective of the DDQN agent is to reduce operating cost by considering the generation cost, grid tariff, and the feed-in tariff. An episode comprises a finite horizon of a day with one hour interval. In each time step, the reward, r_t is computed as follows:

$$r_t = -(R_t \cdot P_t^G \cdot \Delta t + c_t^{fuel}) - (R^{SOC} + R^{SOC,end} + R^{SOF} + R^G), \forall t \in T \quad (22)$$

The problem is formulated using a Markov decision process and solved by the DDQN RL algorithm. A simplified block diagram of the system environment and the DDQN agent is shown in Fig. 2.

IV. RESULTS AND DISCUSSION

A data set including 12 months (January 2018 - December 2018) hourly samples of PV generation and residential load, for the city Sarasota in Florida, is prepared for the RL agent training and testing [20], [21]. The RL agent is trained on first 10 months data, whereas the rest of the data set is used to test the performance. The RL algorithm and environment are implemented using the RL toolbox in MATLAB 2022b [22]. The EMS test scenarios of a sunny day and a normal day are shown in Fig. 3 and Fig. 4, respectively. The figures include 25 data samples of SOF, SOC, P^{PV}, P^L and R^{tariff} states where the last sample represents the observation in response to the action of the last hour of the day. The 24 data samples of actions and grid power represent the 24-hour operation of

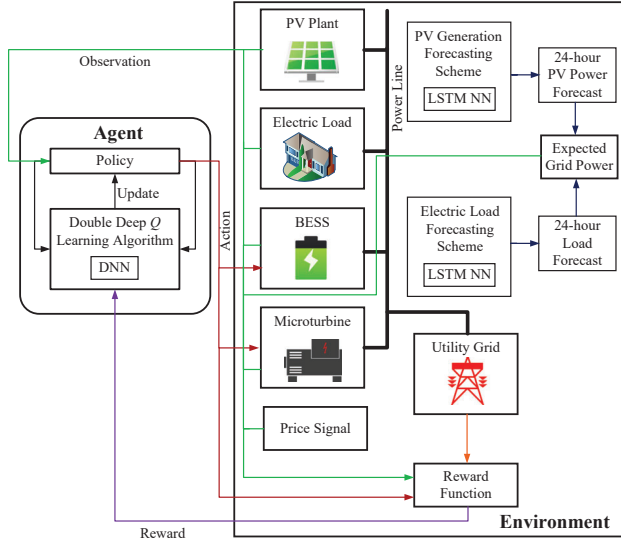


Fig. 2. RL algorithm framework.

the microgrid. The agent has optimized the battery actions to secure the SOC trajectory within the defined boundary ($SOC^{max} = 90\%$, $SOC^{min} = 25\%$) using the trained policy. The results show that the agent has learned to utilize surplus energy and the off-peak demand hours of the day to charge the battery. Consequently, the power consumption of peak hours has been reduced using the battery stored energy. The EMS stores more charge in the battery during sunny days than that in the cloudy days which shows the systematic approach of the agent.

The agent consistently stores energy at the end hour of the day to satisfy the defined limit ($SOC^{end,min} = 40\%$, $SOC^{end,max} = 50\%$) regardless of the type of the day. The results also show the consistent agent trend to operate the microturbine at the beginning and the end of the day to support the peak demand hours and charge the battery as needed. It controls the fuel consumption above the minimum fuel limit ($SOF^{min} = 5\%$). Fig. 5 shows that during a cloudy day the agent utilizes the microturbine to compensate for the low PV generation.

The probability distributions of the constrained states are shown in Fig. 6 for the operation of the trained DDQN agent. The microturbine fuel consumption depends on the type of the day which makes the last hour fuel state, SOF^{end} , more random. The performance of the DDQN agent is compared with a DQN agent and a PPO agent, in Fig. 7, using the daily operational cost of the microgrid. Table I presents the total operation cost achieved by the three different agent throughout the testing period. The optimal cost is obtained by training the DDQN on the same test data, which is an ideal case used to set the benchmark for comparison. Both Fig. 7 and Table I show the ability of the DDQN agent to achieve a near optimal operation.

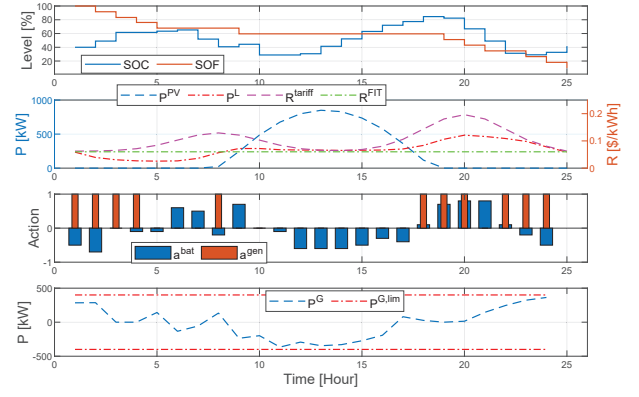


Fig. 3. Performance of the DDQN agent during a sunny day (11/14/2018).

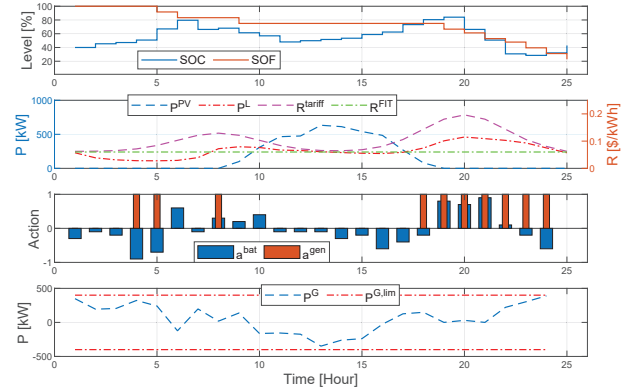


Fig. 4. Performance of the DDQN agent during a normal day (12/26/2018).

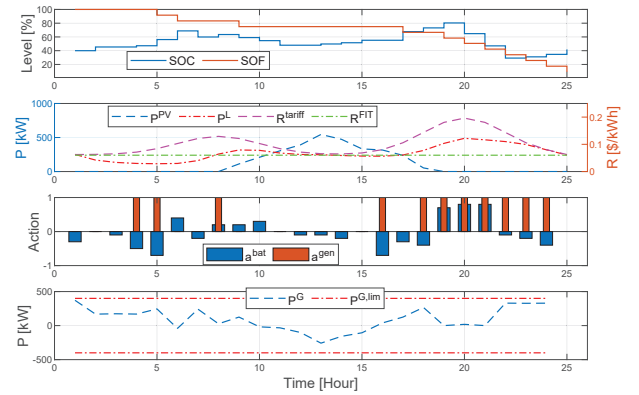


Fig. 5. Performance of the DDQN agent during a cloudy day (12/11/2018).

V. CONCLUSION

The load and PV generation forecasting schemes are integrated with a double-deep Q -learning algorithm to construct an intelligent predictive EMS. Simulation results show that the agent consistently stores energy at the end hour of the day to satisfy the defined limit ($SOC^{end,min}$, $SOC^{end,max}$).

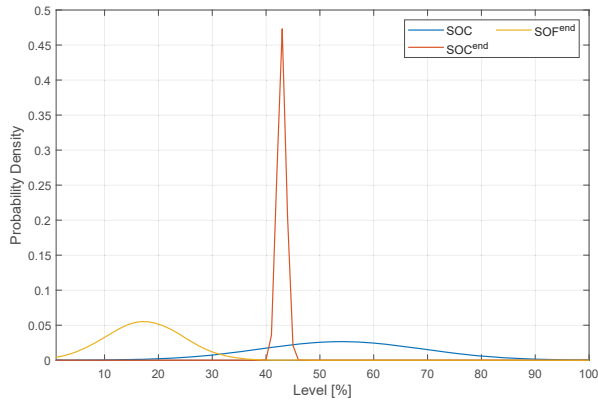


Fig. 6. Probability distribution of constrained states achieved by the trained DDQN agent in the test period.

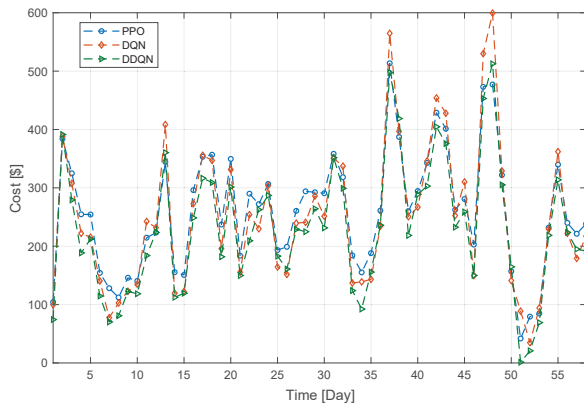


Fig. 7. Performance comparison of the DDQN, DQN, and PPO agents daily operational cost.

TABLE I
COMPARISON OF VARIOUS RL AGENTS BASED ON OPERATIONAL COST
(NOVEMBER, 2018 - DECEMBER, 2018).

Parameter	Optimal Cost (\$)	DDQN Cost (\$)	DQN Cost (\$)	PPO Cost (\$)
Total	13197	13290	14504	15015
Daily Avg.	227.53	229.13	250.06	258.87
Std. Dev.	115.88	112.99	123.29	103.12
Variance	13429	12766	15201	10635

regardless of the type of the day. The results also show the consistent agent trend to operate the microturbine at the beginning and the end of the day to support the peak demand hours and charge the battery as needed. The probability distribution of the constrained states illustrates the learning capability of DDQN agent. Total operating cost of the test period shows that the DDQN agent can achieve lower cost than other RL agents.

REFERENCES

- [1] P. Zeng, H. Li, H. He and S. Li, "Dynamic Energy Management of a Microgrid Using Approximate Dynamic Programming and Deep

- Recurrent Neural Network Learning," *IEEE Transactions on Smart Grid*, vol. 10, no. 4, pp. 4435-4445, July 2019.
- [2] G. K. Venayagamoorthy, R. K. Sharma, P. K. Gautam and A. Ahmadi, "Dynamic Energy Management System for a Smart Microgrid," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 8, pp. 1643-1656, Aug. 2016.
- [3] J. Lee et al., "Optimal Operation of an Energy Management System Using Model Predictive Control and Gaussian Process Time-Series Modeling," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 6, no. 4, pp. 1783-1795, Dec. 2018.
- [4] I. Novickij and G. Joós, "Model Predictive Control Based Approach for Microgrid Energy Management," *2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE)*, Edmonton, AB, Canada, 2019, pp. 1-4.
- [5] R. Palma-Behnke et al., "A Microgrid Energy Management System Based on the Rolling Horizon Strategy," *IEEE Transactions on Smart Grid*, vol. 4, no. 2, pp. 996-1006, June 2013.
- [6] D. Michaelson, H. Mahmood and J. Jiang, "A Predictive Energy Management System Using Pre-Emptive Load Shedding for Islanded Photovoltaic Microgrids," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 7, pp. 5440-5448, July 2017.
- [7] D. Michaelson, H. Mahmood and J. Jiang, "Reduction of Forced Outages in Islanded Microgrids by Compensating Model Uncertainties in PV Rating and Battery Capacity," *IEEE Power and Energy Technology Systems Journal*, vol. 5, no. 4, pp. 129-138, Dec. 2018.
- [8] B. Jiang and Y. Fei, "Smart Home in Smart Microgrid: A Cost-Effective Energy Ecosystem With Intelligent Hierarchical Agents," *IEEE Transactions on Smart Grid*, vol. 6, no. 1, pp. 3-13, Jan. 2015.
- [9] V. Bui, A. Hussain and H. Kim, "Q-Learning-Based Operation Strategy for Community Battery Energy Storage System (CBESS) in Microgrid System," *Energies*, 2019, 12, 1789.
- [10] V. Bui, A. Hussain and H. Kim, "Double Deep Q -Learning-Based Distributed Operation of Battery Energy Storage System Considering Uncertainties," *IEEE Transactions on Smart Grid*, vol. 11, no. 1, pp. 457-469, Jan. 2020.
- [11] Y. Lv, Z. Wu and X. Zhao, "Data-Based Optimal Microgrid Management for Energy Trading With Integral Q-Learning Scheme," in *IEEE Internet of Things Journal*, vol. 10, no. 18, pp. 16183-16193, 15 Sept. 15, 2023.
- [12] S. -H. Hong and H. -S. Lee, "Robust Energy Management System With Safe Reinforcement Learning Using Short-Horizon Forecasts," in *IEEE Transactions on Smart Grid*, vol. 14, no. 3, pp. 2485-2488, May 2023.
- [13] M. S. Hossain and C. Enyioha, "Multi-Agent Energy Management Strategy for Multi-Microgrids Using Reinforcement Learning," *2023 IEEE Texas Power and Energy Conference (TPEC)*, College Station, TX, USA, 2023, pp. 1-6.
- [14] J. Lee et al., "Optimal Operation of an Energy Management System Using Model Predictive Control and Gaussian Process Time-Series Modeling," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 6, no. 4, pp. 1783-1795, Dec. 2018.
- [15] S. Gottwalt, J. Gärtner, H. Schmeck and C. Weinhardt, "Modeling and Valuation of Residential Demand Flexibility for Renewable Energy Integration," *IEEE Transactions on Smart Grid*, vol. 8, no. 6, pp. 2565-2574, Nov. 2017.
- [16] G. Simons and S. Barsun, "Chapter 23: Combined Heat and Power Evaluation Protocol," Aug. 2017. [Online]. Available: <https://www.nrel.gov/docs/fy17osti/68579.pdf>
- [17] M. S. Hossain and H. Mahmood, "Short-Term Photovoltaic Power Forecasting Using an LSTM Neural Network and Synthetic Weather Forecast," *IEEE Access*, vol. 8, pp. 172524-172533, 2020.
- [18] M. S. Hossain and H. Mahmood, "Short-Term Load Forecasting Using an LSTM Neural Network," *2020 IEEE Power and Energy Conference at Illinois (PECI)*, 2020, pp. 1-6.
- [19] H. V. Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double q-learning," *Proceedings of the AAAI conference on artificial intelligence*. Vol. 30. No. 1. 2016.
- [20] National Renewable Energy Laboratory (NREL). Accessed: Nov. 27, 2020. [Online]. Available: <https://www.nrel.gov/grid/solar-power-data.html>.
- [21] Open Energy Information. Accessed: Dec. 6, 2020. [Online]. Available: <https://openei.org/datasets/dataset/commercial-and-residential-hourly-load-profiles-for-all-tmy3-locations-in-the-united-states>.
- [22] Reinforcement Learning Toolbox User's Guide. [Online]. Available: <https://www.mathworks.com>