University of Sindh Journal of Information and Communication Technology (USJICT)
Volume 6, Issue 3, September 2022
ISSN-E: 2523-1235, ISSN-P: 2521-5582
Website: http://sujo.usindh.edu.pk/index.php/USJICT/

Sensitivity Analysis of Reinforcement Learning to Schedule the Battery in Grid-tied Microgrid

Khawaja Haider Ali[1], Hasnain Hyder[1], Muhammad Asif Khan[2]
[1]Department of Electrical Engineering, Sukkur IBA University, Sukkur, Pakistan
[2]Department of Computer Science, Sukkur IBA University, Sukkur, Pakistan
haiderali@iba-suk.edu.pk, hasnain.be16@iba-suk.edu.pk, and asif.khan@iba-suk.edu.pk

Abstract: This research paper explores the application of offline reinforcement learning (RL) in controlling battery operation in a grid-connected microgrid. The study investigates the impact of different parameters on the performance of the RL algorithm, such as the number of discretization levels, gamma, and alpha values. The results show that the convergence time and optimality of the RL algorithm are affected by the choice of these parameters. The research concludes that carefully selecting the discretization levels of state-action spaces and RL hyperparameters is crucial for optimal RL algorithm performance. The benchmark offline sensitivity analysis can be compared in the future with other RL approaches, such as function approximation or DRL methods.

Keywords: Reinforcement learning, offline RL, microgrid, battery operation, state action spaces, discretisation, hyperparameters, convergence, optimal solution, cost savings, gamma, learning rate, sensitivity analysis.

I. INTRODUCTION

Renewable energy sources are gaining popularity worldwide, with countries adopting policies and regulations to encourage their use in power generation [1]. Grid-connected microgrids are an innovative solution to manage energy from renewable sources effectively, and battery energy storage systems (BESS) play a crucial role in this process [2]. BESS help to store excess energy during periods of low demand, which can be used to meet demand during peak periods or when renewable energy generation is low [3]. However, managing BESS in grid-connected microgrids can be challenging, especially when dealing with uncertain energy generation and demand patterns [4].

Reinforcement learning (RL) algorithms have shown great potential in optimizing the performance of BESS in grid-connected microgrids [5]. RL is a type of machine learning algorithm that uses trial-and-error learning to develop an optimal policy for a given problem [6]. The RL algorithm learns by interacting with the environment and receiving feedback in the form of rewards or penalties, enabling it to find the best actions to take in different situations [7]. However, the performance of the RL algorithm is influenced by various system variables and hyperparameters, which can have a significant impact on its effectiveness [8].

In this regard, sensitivity analysis is a crucial step in evaluating the effectiveness of the RL algorithm [9]. Sensitivity analysis involves testing the algorithm's sensitivity to changes in system variables and hyperparameters, allowing for the identification of the optimal settings for achieving maximum performance [10].

This research aims to investigate the sensitivity of RL algorithms to changes in system variables and hyperparameters, with a specific focus on managing BESS in a grid-connected

microgrid environment. This study will use MATLAB 2019 to simulate and evaluate different combinations of system states and hyperparameters to determine their impact on the performance of RL algorithms in grid-connected microgrids. This research will also investigate the effects of different discretisation levels of states and actions, and the modification of tuning parameters such as discount factor and learning rate on the performance of the RL algorithm. The findings of this research will provide insights into the best practices for implementing RL algorithms in energy management systems and will have significant implications for the development of future energy storage technologies.

In this paper is structured as follows. Section 2 outlines the research methodology used in this study, including the simulation framework and the sensitivity analysis approach. Section 3 presents the results of the simulation experiments and provides a discussion and analysis of the findings. Finally, section 4 concludes the paper with a summary of the research findings.

II. METHODOLOGY

This section provide an overview of the methodology used to perform sensitivity analysis. The forecasted PV and load data are input to RL at the beginning of the day. Q-learning is then run using the same input data until convergence is achieved. A policy is developed at the end of this phase that is used to generate commands for battery: charging, discharging, or remain idle. In addition, a backup controller monitors battery control commands output from RL and make any necessary modifications before applying them to the physical system. The backup controller ensures that all physical constraints and limitations are met after applying the control actions from Q-learning.

Each episode of one day consists of 24 steps (1-hour interval). The RL continues to use the same data until convergence is achieved. A total of 10,000 iterations were employed. A convergence may occur before 10000 iterations, but the objective of using a high number of iterations is to observe how convergence behaves when different levels of discretized system states are used. The optimized battery commands are dispatched at the start of the following day for the battery to operate in real-time using real load and PV profiles. However, this work assumes, the forecasted and real net demands are equal, i.e., forecast error is zero. Although this case does not exist in practice, it will serve as a useful comparison when different levels of system states and actions are analyzed.

A. STATE SPACE

The state space (S) is discretized at Δt = 1hr, which suggests that the learning agent captures the information related to the dynamics of the microgrid after the time interval of an hour. In Equation 1, t represents the time period, which has 24 states in 24h of a day due to its discretization at every hour of the day.

$$s_t = [SOC, t] \in S, \tag{1}$$

where SOC, battery state of charge.

The SOC should be bounded by maximum and minimum limits such that:

$$SOC_{\min} \le SOC(t) \le SOC_{\max}. \tag{2}$$

We discretize the state space as shown in Equation 3 below in which the i, j indices represent the SOC and t, respectively as:

$$S_{discrete} = \{S_{i,j}\}. \tag{3}$$

Each index in the state space can be selected using different levels. For example: i = 3,5,7,8 levels, j = 24, 48, 72 levels. The total number of states depend on number of level selected, for example if i and j are 3 and 24 respectively then 3×24=72.

In this work we use different levels of SOC only to see the system performance, therefore, j is 24 in every case.

B. ACTION SPACE

The action space consists of the charge, discharge, and idle command of the battery such as:

$$A = \{a | (\text{Discharge, Idle, Charge})\}. \quad (4)$$

At each time step t, one action is selected from the action space A. If the action "Discharge" is chosen, the battery discharges into the main grid, supplies the load, or both. In case of the action "Idle", the load demand is fulfilled by the PV source, main grid, or both. If the "Charge" action is selected, the battery is charged from the PV, the grid, or both. The actions levels or battery power can be determined using below equation 5.

$$P_{batt} = (Soc_{\max} - Soc_{\min}) * A(1,2). \quad (5)$$

A1 and A2 are mentioned in below case 1 and 2. In this work two cases are considered for simulation.

a. CASE 1.

$$A2 = \{a | 0, \pm 20\%\}. \quad (6)$$

Hence, total number of actions =3

b. CASE 2

$$A1 = \{a | 0, \pm 10\%, \pm 20\%, \pm 30\% \ldots\ldots \pm 100\%\}, \quad (7)$$

where the sign ± means charging or discharging of the battery while zero means battery is Idle. The actions in percentages are calculated on the basis of total capacity of the battery as shown in above equation 5. Therefore, total number of actions =21.

c. EXPLORATION VS. EXPLOITATION

Epsilon-greedy decision-making methods use a trade-off between exploration and exploitation in order to determine a suitable action. The greedy method refers to a strategy that promotes the actions the agent believes will produce the highest rewards. The ϵ -greedy methods include some non-greedy (exploratory) decisions with some degree of probability $1-\epsilon$. Some problems remain, though ϵ - greedy methods are that they choose the action to explore without discrimination. This means that there is no weighting or schedule for selecting which actions to explore. These methods are, however, robust and efficient, and are included in state-of-the-art techniques despite some limitations. In cases where random exploration is not ideal, the exploration strategy itself may be inadequate. Figure 1 shows that in the beginning of the episodes RL agent is exploring so the sum of rewards are diverse in different episodes. However, as the agent starts exploiting the sum of rewards in each episode give similar values.

$$r(s_t, a_t) = \begin{cases} -P_t^{grid} \times \Delta t \times \text{Tariff}_{\text{imp}}, & P_{grid} \geq 0 \\ P_t^{grid} \times \Delta t \times \text{Tariff}_{\text{exp}}, & P_{grid} < 0 \end{cases},$$

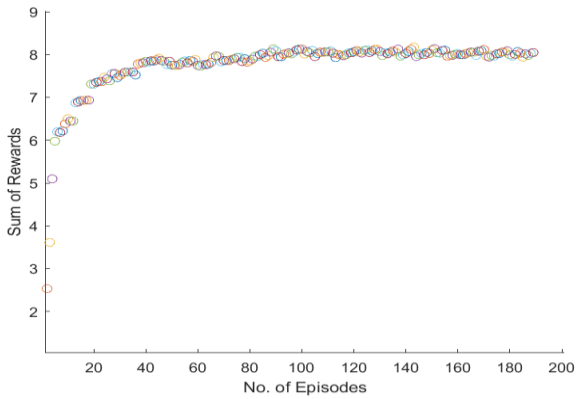

Figure 1 Exploration verses Exploitation

C. REWARD FUNCTION

The reward function (s_t,a_t) is the negative of the cost of imported energy or the cost of exported energy. The cost is calculated every 1hr (as Δt =

1hr) by multiplying the respective tariff rates, as mentioned in equation 8.

Hence, the reward function used in this work can be formulated as follows:

(8)

where $Tariff_{imp}$ and $Tarrif_{exp}$ are the import and export tariffs, respectively. P_t^{grid} is the grid power and is given by equation 9:

$$P_t^{grid} = e_t^{Net} + P_t^{batt}, \quad (9)$$

where P_t^{batt} is the power used to charge the battery.

D. TARIFF

This work used following imported tariff rates:

$$Tariff_{imp} = \begin{Bmatrix} 0.05£/\text{kWh} & \text{low peak, 21:00 to 8:00} \\ 0.08£/\text{kWh} & \text{medium peak, 9:00 to 12:00 ; 19 to 20} \\ 0.171£/\text{kWh} & \text{high peak, 13:00 to 18:00} \end{Bmatrix}. \quad (10)$$

The export tariff does not vary and it is $Tariff_{\exp} = 0.033£/\text{kWh}$.

III. SIMULATION RESULTS

In this work, 10 kW of installed PV power is used to drive large residential load. The battery capacity is 12 kWh. The constraints Soc_{max} and Soc_{min} are 100% and 40%, respectively. Open-source data website has been used to retrieve data profiles for one day. For a better comprehension of the findings, one month of data sets are simulated on a daily basis to analyze convergence and optimality. Since the predicted load and PV profiles are the same as the real profiles. Hence, this case study presents the results of an offline RL ideal scenario. Simulated results for the standard one-day are presented in below section. In this regard, figure 2, shows PV, load, and tariff rates with time intervals of 1hr for a complete day.

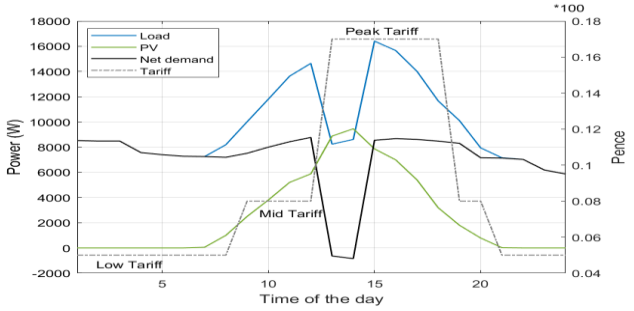

Figure 2 Data Profiles (24 hours).

During the optimization of the battery in a grid connected microgrid, the effects of System Space variables are studied in the following first subsection.

In second subsection, the effect of hyper parameters related to RL are inspected by altering their values during simulation.

A. EFFECT OF SYSTEM SPACES

In this section, two important state space variables related to RL algorithm to manage BESS in a grid connected microgrid are varied to see the effect on EMS performance.

1) State space
2) Action Space

a. VARY DISCRETISATION LEVELS OF STATE SPACE (SOC)

The discretization levels of Soc as in equation 5.1 are changed and the respective simulation results are shown in figure 3 below. The values of hyper parameters **γ**, α, **ϵ**, chosen after tuning are 0.85, 0.9, and 0.8, respectively. The total number of discretization actions is 21 in this case.

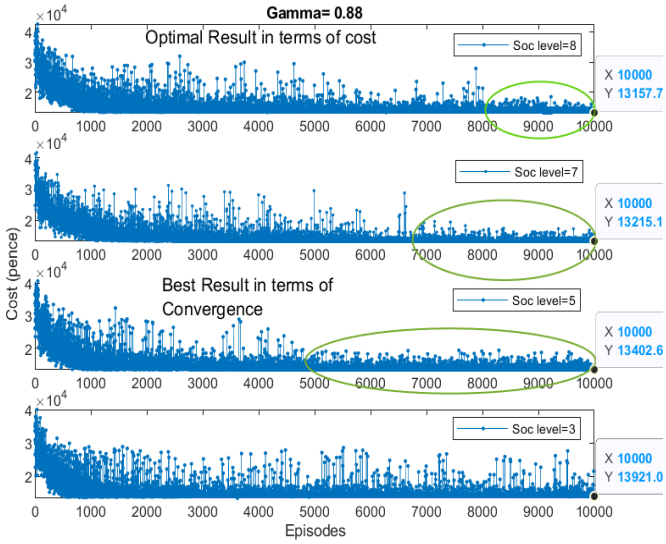

Figure 3 Effect of SOC level's on convergence and optimal cost.

Figure 3 shows, when Soc has the higher discretized level, it gives the best optimal cost as compared to a lower level of discretization. After convergence, the total cost achieved is also shown in the boxes at the end of each subplot, where X and Y depict the episode number and total cost respectively. Contrary to this, convergence achieved at higher discretization levels (Soc) takes more time than at lower discretization levels (Soc). In addition, when the number of descriptions is too low e.g. 3, the convergence is not achieved or is suboptimal. As a result, the total cost achieved for one day is random or sparse. As a result, the number of description levels during the selection process is crucial to achieving an optimal cost.

b. VARIATION IN ACTION SPACE

RL simulations are presented in this section to demonstrate different levels of action for controlling the operation of BESS in a grid-connected microgrid. For instance, in case 1, the BESS in grid-connected microgrid will be configured using three discretization levels, while in case 2, it is configured using twenty one discretization levels.

a) CASE 1

The number of action levels considered in this case is 3. Hence, there is only one level available for each of battery charging, discharging, and idle mode as shown in figure 4. The power that can be used at any time interval is equal to 1440W for each charging and discharging cycle.

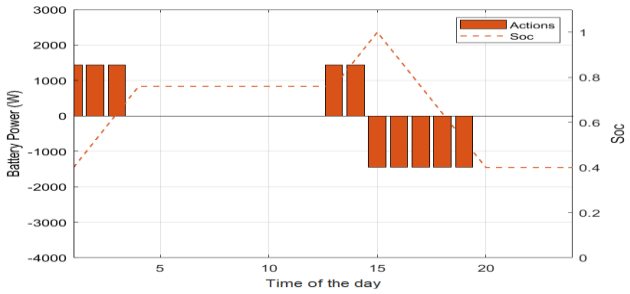

Figure 4 **Battery actions at total discretisation level =3 (Charging level=1; discharging level=1; Idle level=1).**

When the total number of battery actions are 3, convergence is achieved after 4000 iterations. For this case 1, the cost is £134.6 (13461 pence) as described in figure 5.

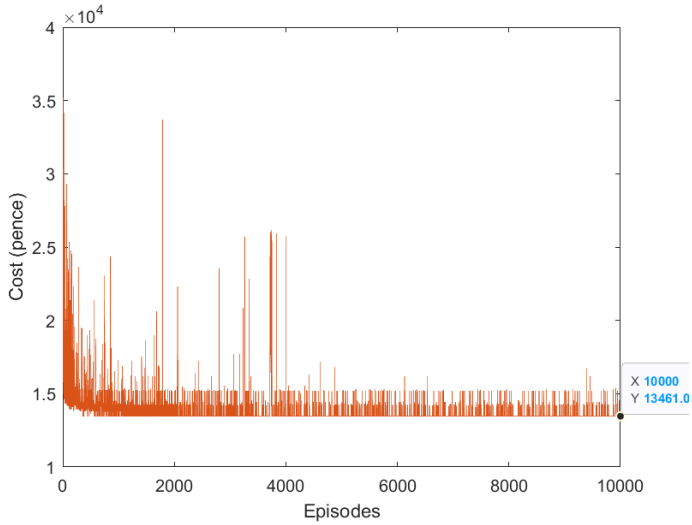

Figure 5 Performance in terms of convergence at total discretization of actions=3.

b) CASE 2

As per equation 7, there are 21 total levels of action in this case: 10 charging, 10 discharging, and 1 idle. The simulation result after the convergence of RL algorithm are mentioned in below figure 6.

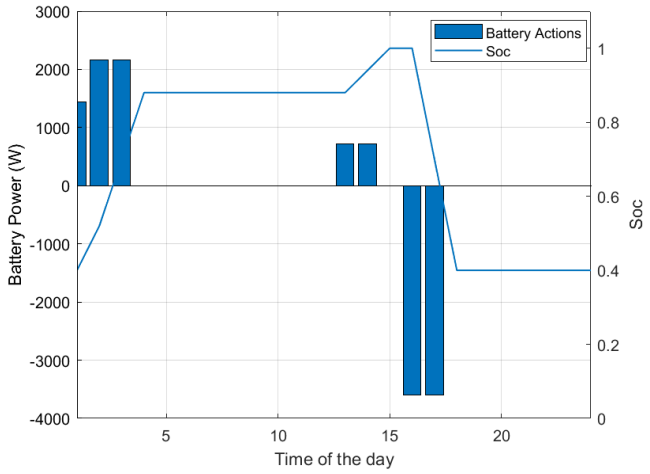

Figure 6 Battery actions at total discretisation level =21 (charging=10; discharging =10; Idle=1).

Simulation results in figure 7 shows, convergence in case 2 occurs at approximately 8000 iterations/episodes. Accordingly, the optimal cost achieved is 131.4 £ or 13157.3 pence.

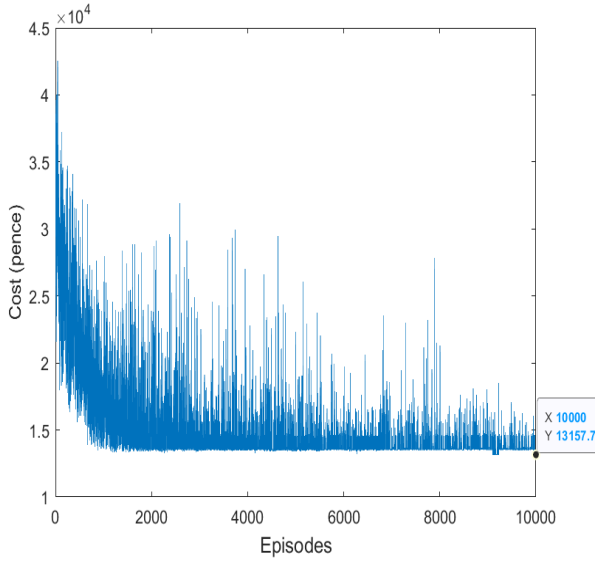

Figure 7 Performance in terms of convergence at total discretisation of actions=21.

c. COMPARISON OF CASE 1 AND 2

RL performance for battery optimization in a grid-connected microgrid is affected by the discretization levels of action. It gives better cost optimization if the actions are flexible, such as different levels of options available to the RL agent regarding charging and discharging the battery. Conversely, convergence occurs sooner at lower levels of discretization (e.g. 3) but with suboptimal results than at higher levels (e.g. 21).

B. EFFECT OF VARIATIONS IN HYPER-PARAMETERS

In this section the hyper parameters are varied to show the behaviour and performance of RL during the optimization of BESS connected to grid-tied microgrid. The benchmark results achieved and shown below when total state discretization levels (Soc) and total discretization actions levels are 8 and 21 respectively. Both used as a constant in following simulation section.

a. VARIATION IN GAMMA

In below simulation, figure 8, total number of discretization levels for Soc and actions are 8 and 21 respectively. The values of hyper parameters ϵ, α are 0.9 and 0.8 respectively. While the γ values changes to see the effect of discount factor in this section.

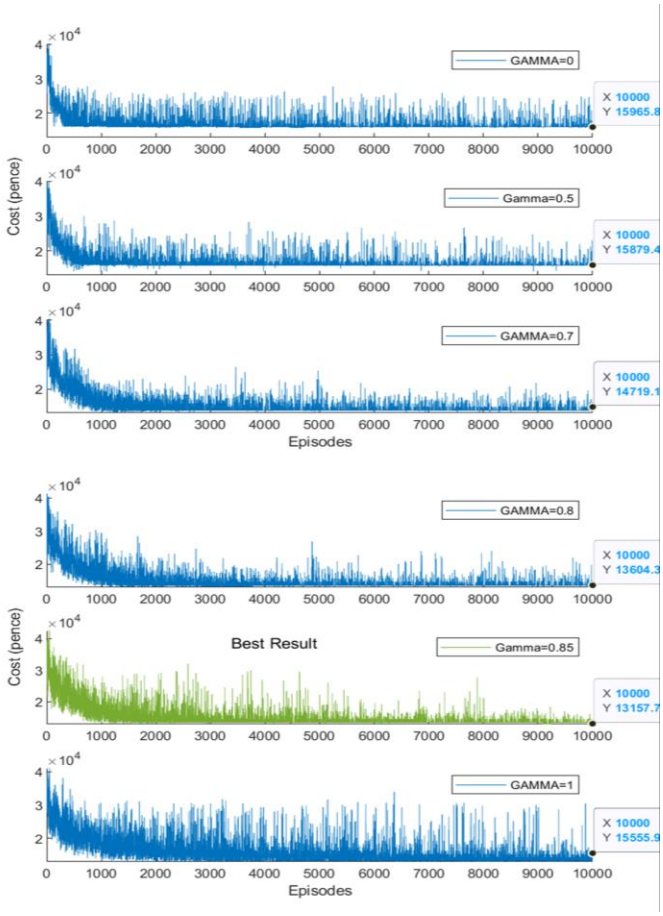

Figure 8 Effect of gamma on convergence and optimal cost

In terms of cost, the best result is achieved when γ =0.85. Convergence occurs after approximately 8000 episodes. Although convergence can be achieved with γ values lower than 0.85 but the cost and convergence are suboptimal. When γ =1, the RL algorithm fails to converge within 10000 iterations or steps. If γ is equal to one, the agent will consider future rewards with greater weight. It means that if an agent does something good in tenth action, it is just as valuable as doing it directly. So learning doesn't work at that well at γ value equal to 1. Despite achieving convergence at γ =0, the sum of daily cost is not optimal because of the myopic (short sighted) behaviour of the γ at 0 prevents it being optimal.

b. VARIATION IN LEARNING RATE

In this simulation figure 9 shows the total number of discretization levels for Soc and actions are 8 and 21 respectively. The values of hyper parameters **ϵ**, **γ** are 0.9 and 0.85 respectively. This section displays the effect of learning rate (**α**) while changing its values.

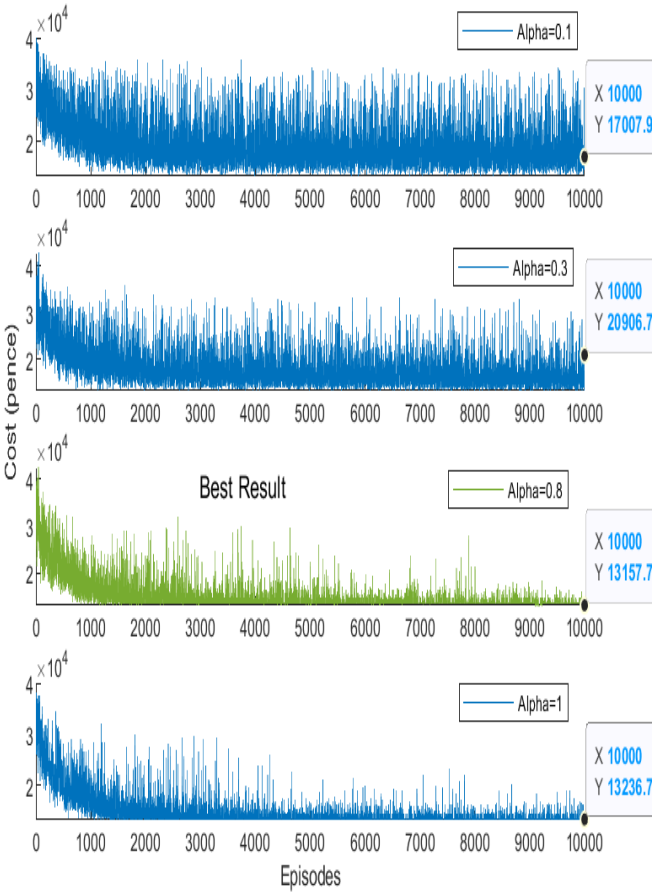

Figure 9 Effect of learning rate on RL performance.

In simulation, the benchmark learning rate in order to tune RL is 0.8 out of different alpha values used in this work. In the case of alpha=0.8, the algorithm achieved convergence after approximately 8000 iterations, while the cost was the lowest. If the alpha value is very low, for example, 0.1, the agent will not learn, therefore convergence may not occur. That is why the total cost per day is high compared to the benchmark alpha value. Furthermore, upon maximization of alpha=1, convergence was achieved, but cost per day was suboptimal.

IV. CONCLUSION

RL algorithm performance is examined in this study by adjusting the different parameters of the offline RL approach to control the battery operation in grid-connected microgrid. The impact of different parameters is clearly visible when this ideal offline case is used. Although variations of these parameters will affect this ideal scenario, it is used as a benchmark when examining the sensitivity of other types of RL algorithms and in real world scenarios when forecasted profiles differ from real RES and load profiles.

Following are the key findings of this work:

- When there are high number of discretization levels of state action spaces, the convergence time will be longer but the outcome in terms of cost saving (optimality) will be higher.
- Convergence and optimal solutions are adversely affected by choosing the discretization level of state-action spaces too high or too low.
- When gamma is 0 or 1, convergence is sparse. Values very close to 1 have a convergence problem, so they produce less optimal results. If gamma is closer to zero, the agent will tend to consider only immediate rewards. If gamma is closer to one, the agent will consider future rewards with greater weight, willing to delay the reward. In this work optimal results are achieved when gamma is between 0.8 and 0.9. Specifically, 0.85 gives best results in terms of cost savings,

Low alpha, such as 0.1, gives divergent results, causing non-optimality. If agent's learning rate is too low, it takes a very long time to learn causing non optimal solution especially in real time applications when computational time is very important. When alpha values are very high such as 1, convergence is achieved, but the results are suboptimal due to inadequate learning. In this work, the best results in terms of cost and convergence are achieved when alpha is 0.8.

Therefore, this work recognizes and concludes that it is important to carefully select the discretization levels of state action spaces and RL hyper parameters. A small change in these variables and parameters can have a large impact on the performance of the RL algorithm. The benchmark offline sensitivity analysis can be compared in the future with other RL approaches, such as function approximation or DRL methods.

V. REFRENCES

[1] A. R. Mišić, Y. Ma, and D. Kammen, "Renewable Energy Policies and Their Implications for Energy Poverty Alleviation: A Global Review," Energies, vol. 15, no. 3, pp. 1-18, Feb. 2022, Art. no. 713.

[2] Q. Zhu and M. Shahidehpour, "Optimal Sizing of Battery Energy Storage System in Grid-Connected Microgrids: A Review of Modeling Techniques and Control Strategies," in IEEE Transactions on Sustainable Energy, vol. 12, no. 1, pp. 252-264, Jan. 2021, doi: 10.1109/TSTE.2020.2986251.

[3] J. Zhang, M. R. Hesamzadeh, T. Ding, and Z. Wang, "Battery energy storage systems in power systems: Technologies, applications, and future outlooks," in Applied Energy, vol. 292, p. 116889, Dec. 2021, doi: 10.1016/j.apenergy.2021.116889.

[4] R. Hou, C. Li, X. Wang, H. Liang, S. Li, Y. Zhang, and W. Li, "Model Predictive Control-Based Energy Management Strategy for a Microgrid with Battery Energy Storage System," in IEEE Access, vol. 9, pp. 118054-118066, 2021, doi: 10.1109/ACCESS.2021.3104017.

[5] J. Zhang, Y. Gao, X. Li, and Y. Li, "A Reinforcement Learning-Based Optimal Control Strategy for Grid-Connected Microgrid Energy

Management System with Electric Vehicles," in IEEE Transactions on Transportation Electrification, vol. 7, no. 1, pp. 245-255, March 2021, doi: 10.1109/TTE.2020.3046106.

[6] K. El-Metwally, A. A. Moustafa and K. F. Hussain, "Deep Reinforcement Learning: An Overview," in IEEE Computational Intelligence Magazine, vol. 16, no. 2, pp. 22-35, May 2021, doi: 10.1109/MCI.2021.3060973.

[7] Y. Wang, J. Wen and Y. Yang, "Recent Advances in Deep Reinforcement Learning: A Survey," in International Journal of Machine Learning and Cybernetics, vol. 12, no. 3, pp. 559-577, Mar. 2021, doi: 10.1007/s13042-021-01229-w.

[8] A. Ali-Eldin, A. H. El-Banna and M. M. Gaber, "Optimization of Hyperparameters in Deep Reinforcement Learning: A Comprehensive Survey," in IEEE Access, vol. 9, pp. 72525-72554, 2021, doi: 10.1109/ACCESS.2021.3086419.

[9] R. Cardell-Oliver, L. Li and C. J. Leckie, "Sensitivity Analysis of Reinforcement Learning Hyperparameters for Optimal Control of HVAC Systems," in IEEE Transactions on Smart Grid, vol. 12, no. 2, pp. 1403-1413, March 2021, doi: 10.1109/TSG.2020.3031399.

[10] S. P. Mohanty, R. R. Negenborn and K. Turitsyn, "Sensitivity analysis for reinforcement learning in power systems," Electric Power Systems Research, vol. 198, 106375, 2021.