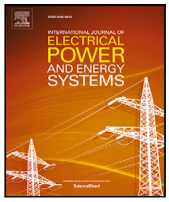
Electrical Power and Energy Systems 152 (2023) 109210

Contents lists available at ScienceDirect

International Journal of Electrical Power and Energy Systems

journal homepage: www.elsevier.com/locate/ijepes



Microgrid energy scheduling under uncertain extreme weather: Adaptation from parallelized reinforcement learning agents☆

Avijit Das [a], Zhen Ni [b,*], Xiangnan Zhong [b]

[a] Pacific Northwest National Laboratory, Richland, WA 99352, USA
[b] Florida Atlantic University, Boca Raton, FL 33431, USA

ARTICLE INFO

Keywords:
Aggregating parallelized agents
Energy optimization
Extreme weather events
Microgrid energy scheduling
Reinforcement learning
Q learning

ABSTRACT

Microgrids are useful solutions for integrating renewable energy resources and providing seamless green electricity to minimize carbon footprint. In recent years, extreme weather events happened often worldwide and caused significant economic and societal losses. Such events bring uncertainties to the microgrid energy scheduling problems and increase the challenges of microgrid operation. Traditional optimization approaches suffer from the inaccuracy of the uncertain microgrid model and the unseen events. Existing reinforcement learning (RL) - based approaches are also hampered by the limited generalization and the increasing computational burden when stochastic formulations are required to accommodate the uncertainties. This paper proposes a new parallelized reinforcement learning (PRL) method based on the probabilistic events to handle the microgrid energy uncertainties. Specifically, several local learning agents are employed to interact with pertinent microgrid environments in a distributed manner and report outcomes to the global agent, which will optimize microgrid energy resources online during extreme events. The stochastic microgrid energy optimization problem is reformulated to include all possible scenarios with probabilities. The advantage estimate functions of learning agents are designed with a backward sweep to transfer the outcomes to the value function updating process. Two simulation studies, stochastic optimization and online testing, are performed to compare with several existing RL approaches. Results substantiate that the proposed PRL method can achieve up to 20% optimization performance improvement with 4 and 28 times less computation cost than Q-learning with experience replay and multi-agent Q-learning approaches, respectively.

1. Introduction

Trends and impacts of recent extreme weather events and natural disasters alarm us with the urgency of improving power infrastructure resilience and the smart grid technologies. These extreme weather events have been identified as one of the main causes of power outages and blackouts in the U.S. [1]. As shown in Fig. 1, Climate Central reports on U.S. power outages due to weather-related and non weather-related incidents [2]. Note that the plot shows only the number of outages affecting more than 50k customers, and the number of weather-related outages is significantly higher than others.

More than 10 million customers have experienced the weather-related power intermittency between 2003 and 2012 with 58% of total power grid outages [3]. After 2012, more than 17 million customers have been affected by power outages due to severe weather events [4]. One study shows that the annual economic impact of weather-related blackout costs between $20 to $75 billion in the U.S., and such losses keep increasing every year [5,6]. Thus, research on improving power infrastructure resiliency has become one of the top priorities for the U.S. electric utilities [1,7,8].

Microgrids attract researchers' attention worldwide as a viable solution for improving power infrastructure resiliency [9]. In recent years, many researchers have devoted on the optimization and control for the microgrid operation. In [10], the authors proposed a mixed-integer linear programming (MILP)-based optimization model for determining the microgrid spinning reserve requirements by analyzing the characteristics of unit outage events. A model-based centralized microgrid design considering multi-period islanding constraints is proposed in [11] to ensure microgrid resiliency. In [12,13], the authors proposed

☆ This work was supported in part by the National Science Foundation, United States under Grant 2047064, 2047010, 1947418, and 1949921. Avijit Das is with the Energy and Environment Directorate, Pacific Northwest National Laboratory, Richland, WA, 99354. This work was started when Avijit Das was a Ph.D. student at the Florida Atlantic University. Zhen Ni and Xiangnan Zhong are with the Department of Electrical Engineering and Computer Science, Florida Atlantic University, Boca Raton, FL, 33431.

* Corresponding author.
E-mail addresses: avijit.das@pnnl.gov (A. Das), zhenni@fau.edu (Z. Ni), xzhong@fau.edu (X. Zhong).

https://doi.org/10.1016/j.ijepes.2023.109210
Received 6 October 2022; Received in revised form 10 January 2023; Accepted 27 April 2023
Available online 20 May 2023
0142-0615/© 2023 Elsevier Ltd. All rights reserved.

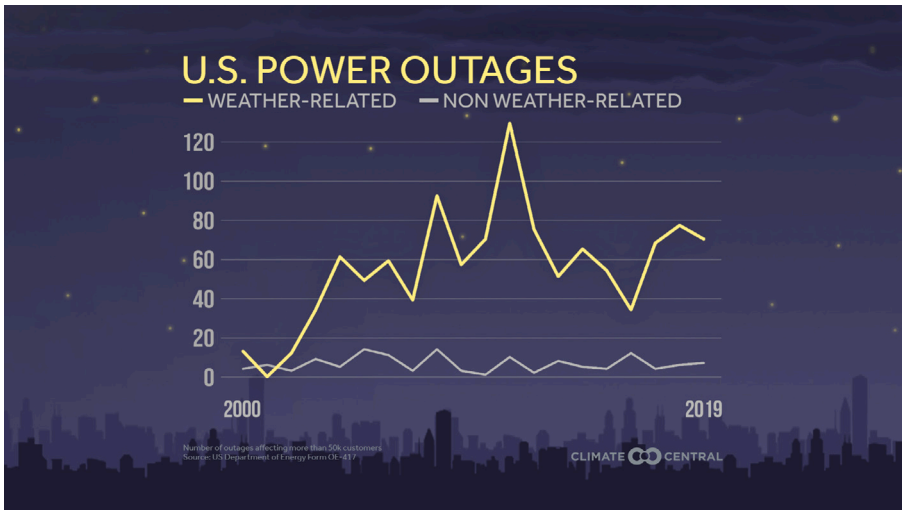


Fig. 1. U.S. power outages due to weather-related and non weather-related incidents [2].

MILP formulations to solve the optimal service restoration problems and determine operation strategy to avoid the possible power shortage considering the control actions of coordinating switches, distributed generators, and controllable loads. Model predictive control based optimization strategies have been proposed for outage management and resilient scheduling with the goal to minimize the load curtailment of microgrid [14,15]. Stochastic optimization frameworks have also been designed to solve microgrid scheduling problem for normal and emergency situations, considering the challenges of generation-demand balance [9,16] and renewable generation (RG) and load uncertainties [17]. These model-based approaches require accurate model information, which is not a trivial task. They are usually limited to certain operation plans and are hardly adaptive to unseen instances. Since the microgrid is a small distribution system with diverse distributed energy resources (DERs), the scheduling and dispatch are crucial to utilize the DERs in a reliable way, especially during extreme weather events. Stochastic formulation is recommended in microgrid scheduling to consider the uncertain nature of the extreme weather events. In stochastic formulations, the size of the microgrid state space increases exponentially with the scenarios. The occurring complexity also challenges the aforementioned optimization approaches.

During the past few years, model-free reinforcement learning (RL) approaches have been recognized and applied in the various engineering applications for online decision-making and control [18–21]. A RL framework for autonomous multi-state and multi-criteria decision-making of energy storage management in a microgrid system was presented in [22]. In [23], the authors used RL to allow the service provider to learn the behaviors of customers and the change of electricity cost to make an optimal pricing decision. Multi-agent RL approaches were explored to solve microgrid energy management problems [24,25], where the agents interacted with the environment in a cooperative manner. The size of the state space increases significantly with the increment of sub-environments, and this incurred considerable computation costs. In [26,27], the authors adopted a conventional Q-learning method that is computationally expensive for solving stochastic optimization problems and outputted optimization policy with extra operating costs. In order to improve the microgrid post-disaster resiliency, a multi-agent RL technique was proposed with the goal to minimize the outage duration [6]. In the post-disaster resilience study, microgrid spinning reserves were used to minimize the outage duration and restore the loads that may output extra-operating costs. Effective stochastic planning and efficient DER utilization during the extreme weather events can be a cost-effective solution with strengthening reliability and resiliency. Stochastic planning requires considering different microgrid operating scenarios based on extreme event uncertainties. This will introduce state space with a massive size and an intensive computational burden for the existing RL approaches.

The concept of asynchronous and synchronous RL approaches have been reported to train neural network controllers on a single multi-core CPU for continuous motor control problems [28–30]. They are similar with the existing multi-task RL approaches from certain aspects [31,32]. Most of these approaches claimed to improve the data efficiency by transferring knowledge (e.g., parameters) to the related tasks. However, the gradients from various *random* tasks could cause the unstable learning results and sometimes downgrade the performance [32]. These approaches have mainly been demonstrated to computer video games and are not yet readily applicable to complex engineering applications, which have many domain constraints and safety rules to satisfy. To the best of the authors' knowledge, none of the aforementioned techniques have been applied to address the uncertain weather events in the microgrid. The authors' recent results [33] proved the concept of the learning combination from both the normal and emergency operations could benefit the operator's decision-making process.

In this paper, we design a new parallelized RL (PRL) method to accommodate the uncertain extreme events systematically and compare it with several existing RL approaches on the microgrid stochastic optimization problem. Specifically, in the proposed design, we reformulate the microgrid energy scheduling problem with the stochastic operation scenarios. The event probability matrix is incorporated in the ultimate objective function so that the proposed PRL method could adapt to the uncertainties. The proposed PRL method has two key steps: employing local learning agents to interact with pertinent microgrid events in a distributed way and aggregating state–action pairs together with the value functions for the learning of the global agent. That being said, the uncertain events are represented by the local state–action information and are incorporated together with the event probabilities to the ultimate optimization objective function. The knowledge aggregation procedure assembles the state and action from the local agents and builds the global state and action vectors for online microgrid decision making process. This helps the global agent to adapt to the potential extreme weather in a timely manner. In addition, the proposed PRL method is computed through an effective double-pass iterative process. The local learning agents explore the designated microgrid

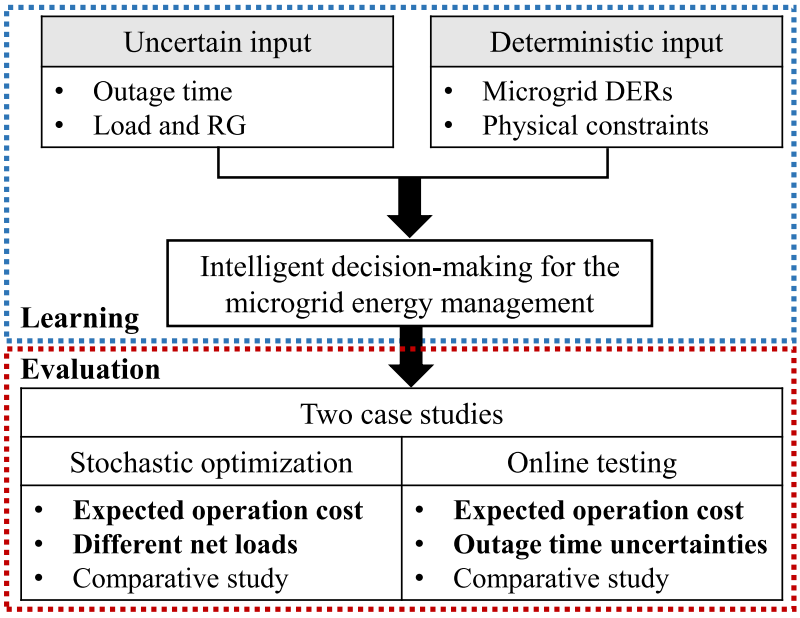


Fig. 2. A schematic overview of the learning and testing procedures of the proposed approach.

environments in the forward pass, and the global agent's value function is updated in backward pass through the advantage estimates. Furthermore, we conduct extensive simulation studies with various event scenarios as shown in Fig. 2. The stochastic scheduling evaluates learning performance of time uncertainties of the extreme weather. The online testing assesses the proposed approach's adaptability considering RG and load uncertainties during these events. Results show that the proposed PRL method can achieve up to 20% improvement of optimization performance and is much faster than existing approaches.

The rest of this paper is organized as follows. The model description and problem formulation are presented in Section 2. In Section 3, we introduce solutions of the basic Q learning, Q learning with experience replay, multi-agent reinforcement learning, and the proposed PRL methods. Simulation results and analysis are carried out in Section 4. Finally, the conclusion is presented in Section 5.

2. Model description and problem formulation

Uncertain extreme weather could significantly impact the renewable generations, load demands and others. For example, the time uncertainties of the weather events play an important factor for the operation of microgrid energy system. To this end, stochastic optimization formulation is recommended to incorporate the uncertain scenarios. This is quite different and challenging than many existing microgrid optimization problems. Our paper reformulates the microgrid stochastic optimization to capture these probabilistic events. Thus, the state and action spaces are much larger [34]. The traditional RL approaches are not feasible to solve it in a timely manner, and this motivates our proposed design.

Specifically, we consider a grid-connected microgrid with multiple DERs, including RG, battery energy storage system (BESS), and dispatchable distributed generator (DG). The RG unit contains photovoltaics and wind turbines. The residential community loads are used as the microgrid load demand. In the grid-connected microgrid, the main grid connection gives the flexibility to the microgrid to export/import power to/from the utility network and maintain the reference voltage and frequency of the system. During the islanded mode, microgrid can use its DERs for the generation-demand balance and maintain the ancillary services in accordance with the microgrid operation strategy.

2.1. Input and decision variables

The multi-time period microgrid decision-making problem with stochasticity is formulated following the Markov decision process (MDP). The probabilities of the outage scenarios are integrated while defining the objective function. So that microgrid generation units can be scheduled accordingly and the total expected operational cost will be minimized. This is different from the existing works. We are working to solve this optimization problem for a day with an hour interval. While the microgrid scheduling problem usually uses DER input information as the state variables regardless of uncertainties, this paper defines the state variables with the microgrid input information considering the extreme weather event scenarios in a vector form. The microgrid state is defined as

$$S_t = (s_{t,1}, \dots, s_{t,k}, \dots, s_{t,K}),\ 1 \le k \le K, \tag{1}$$

$$s_{t,k} = (SOC_{t,k}, k_{t,k}^{DG}, R_{t,k}, G_{t,k}, D_{t,k}), \tag{2}$$

where t is the time index, k is the scenario index, $1 \le k \le K$, and K is the total number of scenarios. $SOC_{t,k}$ is the state of charge (SOC) of the BESS. $k_{t,k}^{DG}$ is a binary variable representing the ON/OFF status of the DG. $R_{t,k}$ is the available RG output. $G_{t,k}$ is the grid price. $D_{t,k}$ is the microgrid load demand.

The decision/action variables are represented as the power output of different microgrid units for the corresponding extreme weather event scenarios. Microgrid decision vector at time t is

$$a_t = (a_{t,1}, \dots, a_{t,k}, \dots, a_{t,K}),\ a_t \in \chi_t, \tag{3}$$

$$a_{t,k} = (a_{t,k}^{B,c}, a_{t,k}^{B,d}, a_{t,k}^{DG}, a_{t,k}^{m,G}, a_{t,k}^{G,m}, a_{t,k}^{dl}), \tag{4}$$

where χ_t is the feasible action space constrained by the microgrid operational constraints. $a_{t,k}^{B,c}$ and $a_{t,k}^{B,d}$ represent charging and discharging

power of the BESS, respectively. $a_{t,k}^{DG}$ is the DG power output. $a_{t,k}^{m,G}$ and $a_{t,k}^{G,m}$ represent the export and import powers to and from the main grid, respectively. $a_{t,k}^{dl}$ is the dumped or unserved load.

2.2. Objective and cost functions

The proposed objective function in (5) is to minimize the expected operation cost of microgrid, considering the defined extreme weather event uncertainties and the cost function in (6) for each scenario.

$$\min_{a_{t,1},\ldots,a_{t,K}} \sum_{k=1}^{K} p_k \sum_{t=0}^{T} C(s_{t,k}, a_{t,k}), \tag{5}$$

$$\begin{aligned} C(s_{t,k}, a_{t,k}) = (a_{t,k}^{G,m} - a_{t,k}^{m,G}) G_{t,k} \Delta t \\ + k_{t,k}^{DG}(x(a_{t,k}^{DG})^2 + y a_{t,k}^{DG} + z), \end{aligned} \tag{6}$$

where, p_k represents the probability of extreme weather event k, and Δt is the time interval. In the cost function, as shown in (6), the first part implies the cost of having energy exchange to/from the grid, and the second part represents the fuel cost of the dispatchable DG unit. The proposed cost function is scenario-sensitive and calculated using the state–action pairs of the corresponding scenarios. The DG quadratic cost function depends on the ON/OFF status ($k_{t,k}^{DG}$) and the given x, y and z are the DG fuel cost-curve coefficients. The proposed microgrid stochastic optimization problem subjects to the operational constraints, as shown in the next subsection.

2.3. Operational constraints

Operational constraints encompass the economic and technical aspects of the microgrid scheduling problem. Since stochastic optimization formulations are used to consider the extreme event uncertainties, the operational constraints are applied for each scenario so that the uncertainties can be captured by the constraints and a healthy microgrid operation can be maintained.

2.3.1. Power balance

Power balance is a crucial microgrid operational constraint, as shown in (7).

$$a_{t,k}^{DG} + a_{t,k}^{G,m} + a_{t,k}^{B,d} - a_{t,k}^{B,c} - a_{t,k}^{m,G} + a_{t,k}^{dl} + R_{t,k} = D_{t,k}. \tag{7}$$

Here, the power balance is an equality constraint that balances the microgrid's generation and demand and helps to maintain the required system ancillary service.

2.3.2. Battery constraints

BESS is one of the major DER units, and its charging and discharging processes require to be maintained within a certain limit. The BESS charging and discharging constraints are presented in (8) and (9), respectively.

$$0 \le a_{t,k}^{B,c} \le (1 - b_t)\psi^C, \tag{8}$$

$$0 \le a_{t,k}^{B,d} \le b_t \psi^D, \tag{9}$$

where ψ^C and ψ^D represent the maximum charging and discharging battery power limit, respectively. b_t is a binary variable introduced to maintain the BESS charging and discharging operation.

According to [35], the SOC of the BESS plays a vital role in the battery lifetime, and a healthy operation can be achieved by keeping the SOC within a certain range. Therefore, we also introduce the battery SOC constraint, as shown in (10).

$$SOC_{\min} \le SOC_{t,k} \le SOC_{\max}, \tag{10}$$

where, $SOC_{\min}$ and $SOC_{\max}$ are the defined minimum and maximum SOC of the BESS, respectively.

After determining the battery charging/discharging operation, a transition function is used to determine the change of SOC for the taken action, as shown in (11).

$$SOC_{t+1,k} = SOC_{t,k} + \frac{1}{B_{\text{cap}}}\left(\phi^C a_{t,k}^{B,c} - \frac{a_{t,k}^{B,d}}{\phi^D}\right), \tag{11}$$

where ϕ^C and ϕ^D are the BESS charging and discharging efficiency, and B_{cap} represents the energy capacity of the BESS.

2.3.3. DG constraints

The DG requires to be operated in a certain range. Therefore, the DG power output is constrained as follows

$$k^{\text{gen}} p^{\text{rated}} k_{t,k}^{DG} \le a_{t,k}^{DG} \le p^{\text{rated}} k_{t,k}^{DG}, \tag{12}$$

where k^{gen} is defined as a percentage of the DG rated power p^{rated}. The value k^{gen} can be obtained from the manufacturer's dataset and can be applied in the constraint so that the DG output can be determined based on the operating requirement.

In the given context of stochastic formulation, the number of input and output variables increases significantly with scenarios, adding complexity in finding the solution. In this paper, we investigate this challenge, propose the parallelized reinforcement learning agents, and compare the performance with the existing RL approaches as follows.

3. RL approaches for microgrid optimization

In this section, we introduce several existing RL approaches, i.e., Q-learning, Q-learning with experience replay (ER), and multi-agent RL, as the comparable solutions to the microgrid application. We will also introduce the unique designs of the proposed PRL approach to address the aforementioned challenges.

3.1. Existing approaches

3.1.1. Q-learning approach

RL can be defined as an agent and environment (system model) interactive system where the RL agent observes the environment state, selects an action, receives a feedback reward, and learns the optimization policy through the sequential decision-making process [36]. At each time step t, the RL agent selects an action a_t according to its policy after receiving the state information S_t. In return, the environment sends next-state S_{t+1} and a reward/cost feedback r_t. This process continues until the agent reaches to the terminal state. At every iteration, the agent receives the return as $R_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k}$ which is the total cumulative from time step t with discount factor γ. The agent's goal is to minimize/maximize the expected return from each state S_t. In the Q-learning approach, a Q-value function $Q(s, a)$ is used to map the relationship between state s and action a. The Q-value $Q^{\pi}(s, a) = \mathbb{E}[R_t | s_t = s, a]$ represents the expected return for taking action a from state s following the policy π. The optimal value function can be obtained as $Q^*(s, a) = max_{\pi} Q^{\pi}(s, a)$ that returns the maximum action value for the given state–action pair following the policy π.

In this paper, the reward (r_t) is replaced with the microgrid cost function $C(S_t, a_t)$. The Q-value function will be calculated recursively using the Bellman equation. In terms of solving the given microgrid stochastic optimization problem, multiple scenarios are combined to define the state information. Therefore, the state space increases significantly with the increment of scenarios. In this case, the Q-learning agent will need intensive exploration to find the proper policy.

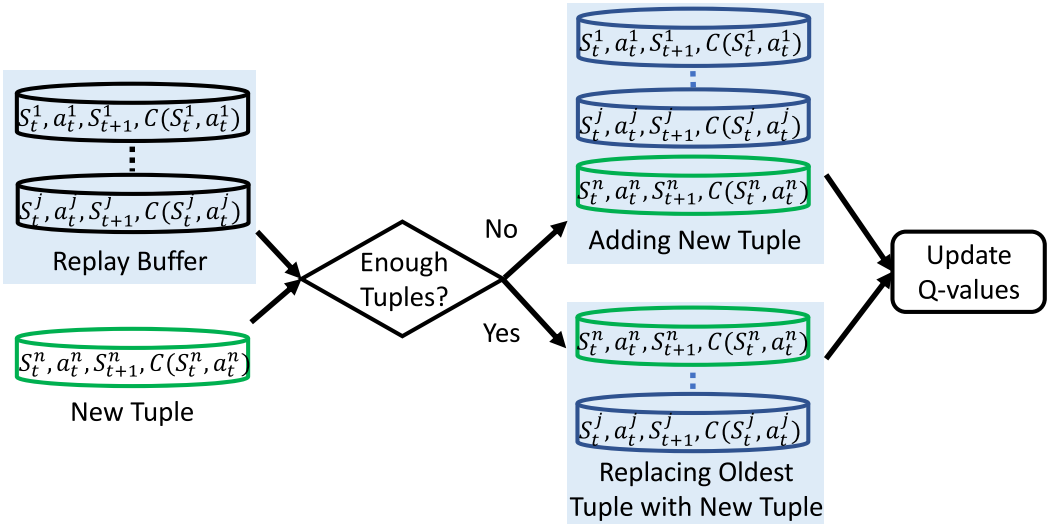


Fig. 3. Replay buffer and its circulation (adding and removing tuples) process.

3.1.2. Q-learning with experience replay approach

In recent years, Q-learning with ER approach attracts researcher's attention due to its skill of improving learning efficiency. In this approach, a replay buffer is used to store agent-environment interaction experiences as tuples. They are used for off-policy training in future episodes [37,38]. The replay buffer is usually defined as a circular buffer, where the oldest transition is replaced by a new transition. A graphical representation of the replay buffer is illustrated in Fig. 3. In the figure, j represents the index of tuples in the replay buffer. When a new tuple is available, the replay buffer checks its size. If the size exceeds the defined capacity, the new tuple is added to the buffer to replace the oldest tuple.

There are different sampling strategies to use tuples in the replay buffer. This process could improve the sample efficiency by enabling data to be reused multiple times for training and also improve the training stability. For the problems with low variance in immediate outcomes, the inclusion of ER in Q-learning also helps to find the proper policy.

In this approach, at every iteration, the RL agent provides a new tuple from interacting with the microgrid environment. The replay buffer is updated as shown in Fig. 3, and the Q-values are calculated using the state–action pairs of each tuple [37]. This procedure continues until the maximum iteration is reached, and the trained value functions are used to determine microgrid operations. Since Q-learning with ER leverages the past experiences, this approach may struggle to reach to the optimal solution for the stochastic optimization problem if there are exploration gaps or unvisited states due to lack of iterations. Also, training with ER adds the computation cost, which is proportional to the experience replay buffer size.

3.1.3. Multi-agent RL approach

Multi-agent RL approach uses multiple RL agents to solve sequential decision-making problem operating in a common environment. In this approach, the agents aim to optimize their own objective by interacting with the environment and other agents. Based on the problem requirement, the agents can be designed to employ in cooperative, independent and mix manner. In this approach, inspired by [24], we use a multi-agent Q-learning approach where the agents interact with the environment independently. The Q-table is updated based on the experiences observed through the agent-environment interactions. At every iteration, the environment sends the state information to all agents, which will take decision independently and update its Q-table accordingly. While the traditional Q-learning approach interacts with the environment once in an iteration, the M learning agents can interact M times and thus can improve the exploration capacity per iteration. The Q-value update can be expressed as

$$\begin{aligned}\begin{bmatrix} Q(S_t^1, a_t^1) \\ Q(S_t^2, a_t^2) \\ . \\ . \\ Q(S_t^M, a_t^M) \end{bmatrix} &= (1-\alpha)\begin{bmatrix} Q(S_t^1, a_t^1) \\ Q(S_t^2, a_t^2) \\ . \\ . \\ Q(S_t^M, a_t^M) \end{bmatrix} \\ &+ \alpha\begin{bmatrix} C(S_t^1, a_t^1) \\ C(S_t^2, a_t^2) \\ . \\ . \\ C(S_t^M, a_t^M) \end{bmatrix} \\ &+ \alpha\gamma\begin{bmatrix} \min\limits_{a_{t+1}^1} Q(S_{t+1}^1, a_{t+1}^1) \\ \min\limits_{a_{t+1}^2} Q(S_{t+1}^2, a_{t+1}^2) \\ . \\ . \\ \min_{a_{t+1}^M} Q(S_{t+1}^M, a_{t+1}^M) \end{bmatrix}. \end{aligned} \tag{13}$$

where M is the total number of agents employed to interact with the microgrid environment, and α is the learning rate.

3.2. Proposed parallelized reinforcement learning approach

We propose a new PRL approach with multiple local agents for solving stochastic microgrid scheduling problem. The proposed approach solves the given stochastic optimization problem following two key steps: distributed learning for local agents and knowledge aggregation for the global agent. In the distributed learning, the local RL agents are employed to interact with pertinent microgrid environments and obtain their own learned knowledge. Next, we aggregate state–action information and build the value function of the global agent with a probabilistic cost. The proposed approach is illustrated in the grid-connected microgrid application in Fig. 4.

As shown in this figure, our proposed approach receives system parameters and scenario environments from the microgrid, and employs local agents to interact with the environments in a parallel manner. Note that we employ a local RL agent for dealing with a certain problem scenario. Therefore, the number of agents should be equal to the number of scenarios, and we use k as the scenario/local agent index. In our proposed approach, we use double-pass action value updating process. In forward pass, local RL agents interact with different environment scenarios using $\epsilon - greedy$ technique. In this technique, at any state $s_{t,k}$,

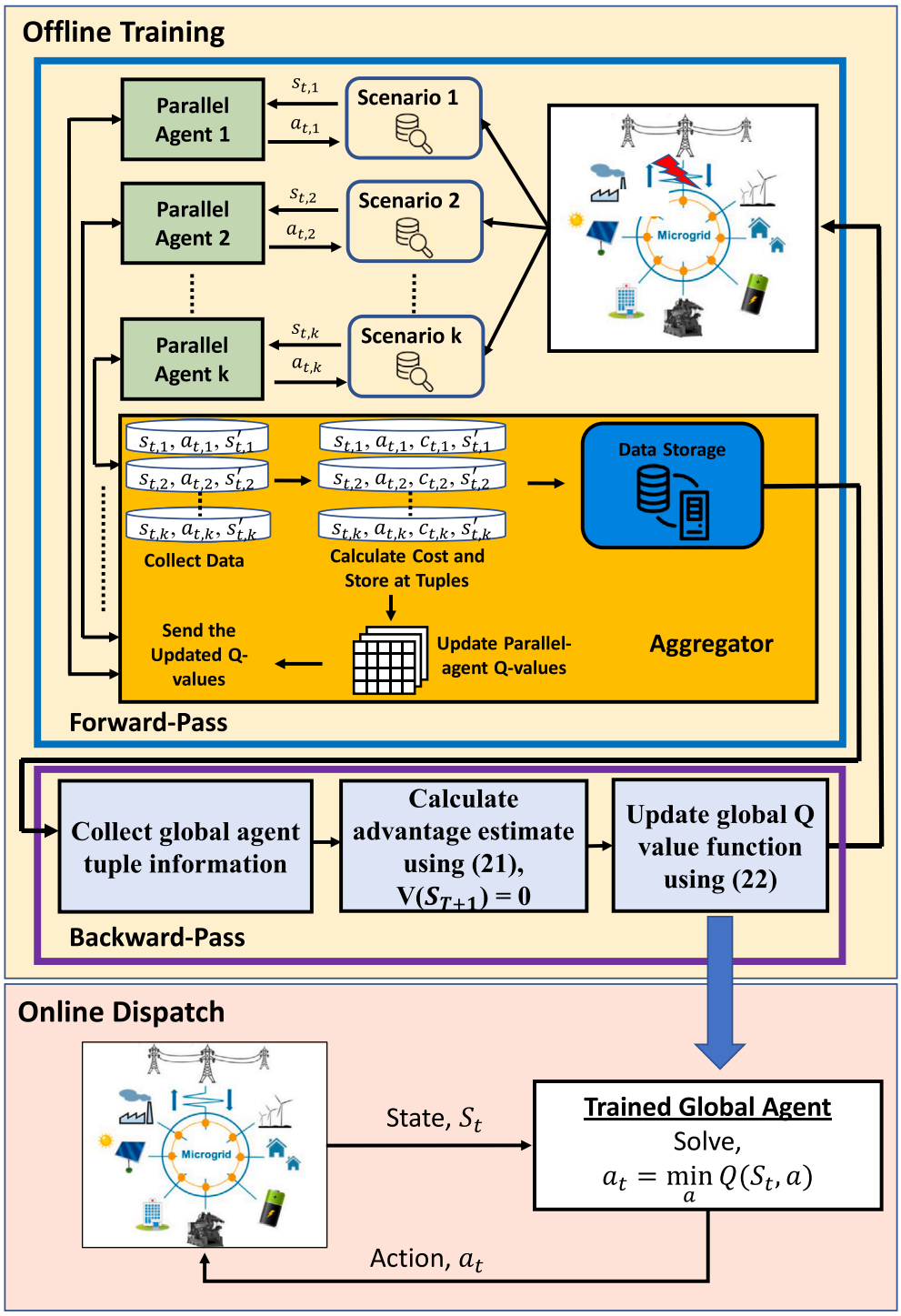


Fig. 4. The diagram of the proposed PRL approach with a double-pass structure of information computation. The parallelized agents interact with scenarios and the aggregator combines the learned knowledge. The global agent's value function is updated in the backward pass and will be used for the microgrid stochastic energy scheduling online.

the action $a_{t,k}$ is determined either selecting a random action from the feasible actions or using the greedy formula as

$$a_{t,k} = \min_{a} Q_k(s_{t,k}, a). \tag{14}$$

This paper uses the physical state variables such as discretized battery SOC and DG ON/OFF status to define the states since other information state variables are the same as the forecast information and do not vary over iteration. For any local learning agent, to determine the action variables, the main idea is to first find all feasible BESS power solutions based on the current SOC and feasible SOC at the next time step. Next, the power output of DG and grid's export and import powers are determined using a rule-based dispatch strategy for each feasible BESS power solution [39]. Note that the proposed decision-making strategy is applicable for microgrids with DGs, and it guarantees the dispatch solution. The steps for determining the action variables are detailed as follows.

1. At any time t, we first find all feasible BESS SOCs at the next time step from the current BESS SOC $SOC_{t,k}$, which satisfy (10). A feasible SOC must also needs to satisfy (15), which can be

Algorithm 1 Decision-making strategy.

1: Set the required power: $p^a = D_{t,k} - R_{t,k} - a^{B,d}_{t,k} + a^{B,c}_{t,k}$
2: Determine power output of DG:
$a^{DG}_{t,k} = \min(\max(k^{DG}_{t,k} p^{\text{rated}}, p^a_t), p^{\text{rated}})$
3: Update the required power: $p^a \leftarrow p^a - a^{DG}_{t,k}$
4: **if** $p^a \geq 0$ **then**
$a^{G,m}_{t,k} = p^a$; $a^{m,G}_{t,k} = 0$
5: **else**
$a^{G,m}_{t,k} = 0$; $a^{m,G}_{t,k} = -p^a$
6: **end if**
7: Calculate $a^{dl}_{t,k}$ based on (7)
8: Calculate $C(s_{t,k}, a_{t,k})$ using (6)

derived from (8)–(11):

$$SOC_{t,k} - \frac{\psi^D \Delta t}{\phi^D B_{\text{cap}}} \leq SOC_{t+1,k} \leq SOC_{t,k} + \frac{\phi^C \psi^C \Delta t}{B_{\text{cap}}}. \tag{15}$$

The BESS power for each feasible BESS SOC $SOC_{t+1,k}$ can be calculated using (16):

$$\begin{cases} a^{B,c}_{t,k} = \frac{(SOC_{t+1,k} - SOC_{t,k}) B_{\text{cap}}}{\Delta t \phi^C}, a^{B,d}_{t,k} = 0, & \text{if } SOC_{t,k} \leq SOC_{t+1,k} \\ a^{B,c}_{t,k} = 0, a^{B,d}_{t,k} = \frac{(SOC_{t,k} - SOC_{t+1,k}) B_{\text{cap}} \phi^D}{\Delta t}, & \text{if } SOC_{t,k} \geq SOC_{t+1,k}. \end{cases} \tag{16}$$

2. Next, the power output of DG and grid's export and import powers are determined for each feasible BESS power solution, as detailed in Algorithm 1. If DG's status is OFF at any feasible state, the power output and generation cost are set to be zero. Otherwise, the power output of DG is determined based on the net load and DG's rated capacity and operating requirements, and the corresponding cost is calculated. Then, the export and import powers of the grid are determined based on the net load of the system. Finally, the dumped or unserved load is calculated, which is usually zero, unless the power cannot be exported/imported to or from the main grid.
3. Thus after determining the action variables of all the feasible actions, the learning agent takes action using the $\epsilon - greedy$ technique.

After each interaction between the local-agent and the scenario environment, the aggregator collects tuples with state $s_{t,k}$, action $a_{t,k}$, cost $C(s_{t,k}, a_{t,k})$, and next-state $s_{t',k}$ information from all agents. To update the action value functions, this paper uses a function to estimate the advantage of taking action $a_{t,k}$ in state $s_{t,k}$, which provides feedback on how much better or worse the action taken was compared to the overall expected return. The theoretical foundation of the advantage estimate is well-explained in [28,36,40] and therefore is omitted here to conserve space. We calculate the advantage estimate as

$$A(s_{t,k}, a_{t,k}) = \Big(C(s_{t,k}, a_{t,k}) + \gamma V(s_{t+1,k}) \Big) - V(s_{t,k}), \tag{17}$$

and

$$A(s_{t-1,k}, a_{t-1,k}) = \Big(C(s_{t-1,k}, a_{t-1,k}) + \gamma C(s_{t,k}, a_{t,k}) + \gamma^2 V(s_{t+1,k}) \Big) - V(s_{t-1,k}), \tag{18}$$

where $V(s_{t,k}) = \min_a Q_k(s_{t,k}, a)$. The action value $Q_k(s_{t,k}, a_{t,k})$ is updated using the advantage estimate as

$$Q_k(s_{t,k}, a_{t,k}) = Q_k(s_{t,k}, a_{t,k}) + \alpha A(s_{t,k}, a_{t,k}). \tag{19}$$

From the given distributed interactions, the aggregator collects individual microgrid state as in (2) and action as in (4) from the local agents and assembles them to determine the global agent's state and action vectors, which are basically the state in (1) and action in (3). The global agent's cost function is calculated with the combination of the probabilistic local agents' cost functions as

$$C(S_t, a_t) = \sum_{k=1}^{K} p_k C(s_{t,k}, a_{t,k}). \tag{20}$$

Traditionally, solving stochastic optimization problem using Q-learning requires combining scenario states to define the state information. In this case, the number of states per time step could increase intensively and traditional single- and multi-agent Q-learning approaches are computationally expensive to find the near optimal policy. Our proposed design employs local learning agents to interact with the individual microgrid scenarios. Therefore, each microgrid scenario has a dedicated local learning agent, and every agent has a Q-table which can be initialized with zeros or an approximated solution. Through distributed learning, the local learning agents interact with their individual microgrid environments and learn control policies updating the action values of their Q-tables. Due to the distributed design, the number of states per time step decreases considerably compared to traditional approaches, and the agents can effectively explore the solution space. Next, the knowledge obtained from the local agents are used to (1) build the state and action vectors for the global agent; (2) calculate the expected cost; and (3) learn the policy of the overall stochastic optimization problem. This procedure significantly improves the global agent's learning capacity and helps direct the global agent to approximate the optimization policy efficiently.

The backward pass of the proposed approach is dedicated to update the global agent's action values using the advantage estimate functions and learn the policy for the stochastic microgrid scheduling problem. At the end of forward-pass, the global agent's tuples are extracted from the data storage, and the advantage function is used to estimate the return of the steps in the backward sweeps as

$$A(S_t, a_t) = \Big(C(S_t, a_t) + \gamma V(S_{t+1}) \Big) - V(S_t), \tag{21}$$

and the global agent's action value $Q(S_t, a_t)$ is updated using the advantage estimate as

$$Q(S_t, a_t) = Q(S_t, a_t) + \alpha A(S_t, a_t). \tag{22}$$

This process helps to efficiently pass the future outcomes to the earlier time steps and improves the learning efficiency. After finishing the backward-pass process, the algorithm increments the iteration n and restarts the procedure again until the algorithm reaches to the maximum iteration number N. The detailed algorithm is presented in Algorithm 2.

4. Simulation results and analysis

In this section, we provide the simulation setup information and report different case studies to examine the performance of the proposed approach. We report the microgrid operating costs, computation time, and percentage of improvements for results analysis. Also, we present the comparisons with several existing approaches to justify the performance improvement.

The microgrid DER parameters are listed in Table 1. The microgrid exogenous information including a small residential community load-demand, RG output, and electricity price are plotted in Fig. 5. The system advisory model by National Renewable Energy Laboratory is used to obtain RG system parameters and outputs for the city of Phoenix, AZ [41]. A small residential community load-demand data is collected from [42] and used as the microgrid load. For the BESS, we assume charging and discharging efficiencies and maximum powers are the same. The optimization problem time horizon is set to be $T = 24$ hours with an one-hour interval.

According to [34], some natural disasters like hurricanes and blizzards are predictable $24 - 72$ hours before happening. In our case

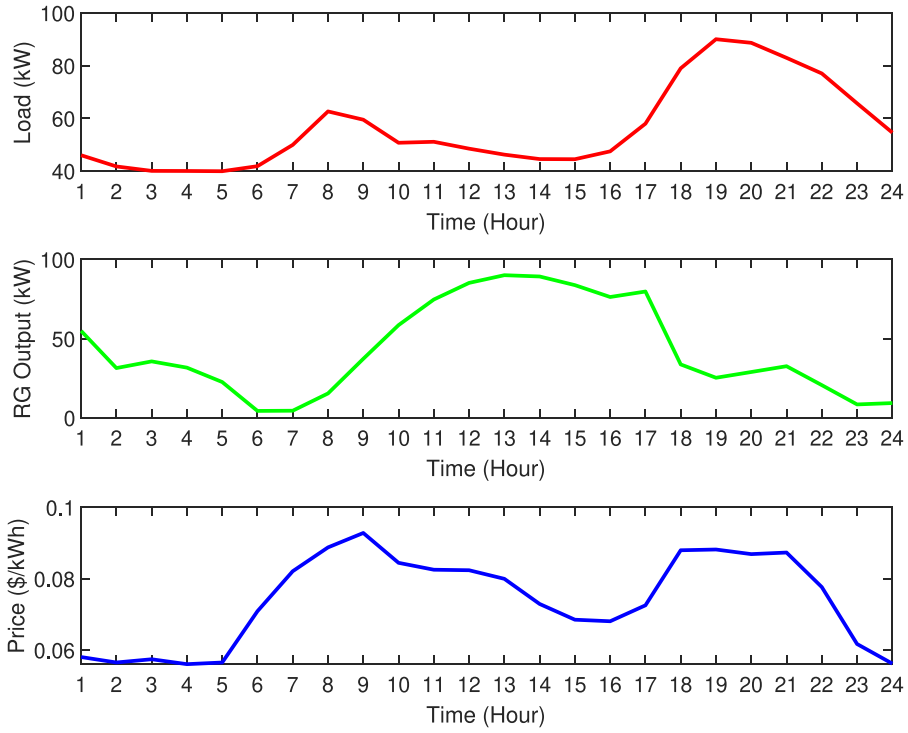


Fig. 5. Microgrid system load, RG output, and electricity price from the utility grid.

Algorithm 2 The proposed PRL algorithm.

Initialization:
Define global Q-table (Q) and local Q-tables (Q_k), set the iteration $n = 1$, and N, set the initial state S_1^1, set exploration probability rates, ϵ
for $t = 1 : T$ **do** ▷ Forward Pass
  for $k = 1 : K$ **do** ▷ Local Agents
    if $rand > \epsilon_1$ **then**
      Solve (14) for determining decision
    else
      Choose a decision randomly
    end if
    if $t < T$ **then**
      Find the next state, $s^n_{t+1,k}$
      Update action value $Q_k(s_{t,k}, a_{t,k})$ as (19)
    end if
  end for
  Determine S_t and S_{t+1} as (1), a_t as (3), and $C(S_t, a_t)$ as (20)
  Store the transitions in a buffer
end for
for $t = T : 1$ **do** ▷ Backward Pass
  Calculate the advantage estimate as (21)
  Update the global Q-value functions as (22)
end for
Increment n. If $n \leq N$ go to Step 2.
Return the global and local Q-value functions $(Q^N)_{t=1}^T$ and $(Q_k^N)_{t=1}^T$

studies, we assume the forecast of having extreme event at 12 PM with 2 hours of uncertainty. Therefore, we have a total of five scenarios, and the proposed PRL approach uses five local learning agents with a dedicated Q-table for each. U.S. EIA data shows that, on average, 4 hours of power interruption may occur due to extreme events [43]. Hence, we set the extreme event duration as 4 hours in our case study. Since the problem is formulated with discrete state and action variables,

Table 1
Microgrid DER parameters.

RG	Photovoltaic Capacity	50 kW
	Wind Turbine Capacity	100 kW
BESS	Capacity	150 kWh
	Char. and Dischar. Eff.	90%
	Maximum Power	30 kW
DG	Rated Power	100 kW
	Min. Dispatch Percentage	0.3
	Cost Coefficients	0.0009 ($/(kW)^2),
	x, y, and z	0.0213 ($/kW) and 1.1 ($)

the lookup table implementation is a suitable option to approximate the action values in a timely manner. With the binary DG ON/OFF status and 9 discretized BESS states, each scenario has 18 discretized states. Therefore, the given stochastic optimization problem has $1.9\times10^6 = 18^5$ states after combining all five scenarios, challenging existing learning approaches with computational complexity to find a near-optimal solution. The proposed PRL approach employs local RL agents to interact with the environments in a distributed manner. Hence at any time step, each local RL agent has 18 or fewer possible actions which significantly reduces the action space and lets the RL agents explore the solution space effectively and report the promising solutions to the global agent. Note that the training of the local and global agents happens offline using the forecast data. The proposed PRL approach is trained for 4000 iterations, where local RL agents are trained synchronously at each iteration. The replay buffer capacity is defined as 10 tuples for the Q-learning with ER approach. For the multi-agent Q-learning, $M = 4$ is used while defining the number of agents. We set iteration number as 4000 and exploration probability as 0.6. We assume equal probability for all the microgrid operating scenarios. Load and RG power output uncertainties are considered using the following equations as [44]

$$R_{t,k} = \min\{\max(\hat{R}_{t,k} + \varepsilon_r, R_{\min}), R_{\max}\}, \tag{23}$$

and

$$D_{t,k} = \min\{\max(\hat{D}_{t,k} + \varepsilon_d, D_{\min}), D_{\max}\}, \tag{24}$$

Table 2
Stochastic optimization results with expected cost and computation time.

Approach		Expected cost	Computation time
Offline	DP (reference)	$19.1	6.5 hours
Online	Proposed PRL	$20.2	2.2 min
	Q-learning with ER	$22.2	9.1 min
	Multi-agent Q-learning	$23	1.1 h
	Q-learning	$26.5	3.9 min

Table 3
Distribution functions for generating test scenarios.

Problem	RG	Load
NO.	Noise	Noise
1	$U(-5,5)$	$U(-3,3)$
2	$U(-5,5)$	$N(0,3)$
3	$N(0,1)$	$U(-3,3)$
4	$N(0,2)$	$N(0,1.5)$

Table 4
Online testing results with performance improvement.

Approach	Average cost ($)	Improvement (%)
Proposed PRL	21.9	20.94
Q-learning with ER	22.7	18.05
Multi-agent Q-learning	23.1	16.61

where $\hat{R}_{t,k}$ and $\hat{D}_{t,k}$ represent the day-ahead RG and load data, respectively. ε_r and ε_d are the RG and load demand noises, respectively. The Eqs. (23) and (24) are used to generate RG and load information for training and testing purposes. All the simulations are conducted in MATLAB *R*2019*b* on a PC with Intel Core i7 – 8650U 4.2GHz and 16GB RAM. A MATLAB script is used to define the DER parameters of the microgrid, and the parameters of the proposed PRL approach. We define a MATLAB function which provides microgrid state information to the local RL agents based on their scheduling scenarios. The decision-making process of the RL agents and the value function updates are conducted in the main MATLAB script file. All the approaches are implemented in the same environment during the performance comparison.

4.1. Stochastic optimization

Stochastic optimization case study is important to evaluate the performance of an approach under uncertainties. This case study assumes that a power outage may happen at 12 PM with two hours of uncertainty. Considering that, we have five different scenarios with the outage happening time frame 10 AM to 2 PM. The results and key observations in terms of expected operation cost and different net loads are discussed as follows.

4.1.1. Expected cost

The expected cost represents the expected form of total microgrid operational cost of five different scenarios. The results are summarized in Table 2. Note that we use dynamic programming (DP) as the reference approach. DP is used offline as it requires accurate forecast information to achieve the optimal solution, which may not be obtainable in practice. Also, the DP approach is computationally expensive in order to achieve the optimal solution in this benchmark. The table shows that our proposed PRL approach can reach very close to the optimal solution with the expected operational cost as $20.2 and computation time 2.2 minutes. The existing learning approaches output stochastic microgrid scheduling results with extra-operating costs and considerable computation costs. The Q-learning with ER and multi-agent Q-learning takes around 4 times and 30 times longer to output the scheduling decisions, indicating our proposed approach's computational efficiency. The traditional Q-learning approach provides the most expensive expected cost and requires intensive training.

The average expected cost curves and microgrid scheduling results are plotted in Fig. 6. The expected cost curves in Fig. 6 are obtained averaging after 30 runs. The fluctuations on the curves represent the agent's explorations, and the exploration rate degraded after every 50 iterations by 1.1. The results show that the expected cost curve of the proposed approach drops rapidly comparing to the other approaches and converges to the minimum expected daily cost. Q-learning with ER and multi-agent Q-learning approaches show competitive performance till 2000 iterations. After that, due to decay of exploration rate, multi-agent Q-learning approach struggles to find the proper scheduling decision and stuck at a local minima. Note, for the multi-agent Q-learning approach, we employ four agents, therefore obtain four expected costs. The minimum expected cost at every iteration is plotted in the figure. In contrast, Q-learning with ER approach explores replay buffer every iteration, and shows better performance in later iterations. However, it also presents a noticeable gap in comparison with our proposed approach. Overall, our proposed approach outperforms three existing approaches in terms of both the expected operational cost and computational time.

4.1.2. Different net loads

Extreme weather events may also affect RG output. Therefore, a case study is conducted with different net loads by varying the RG outputs during the extreme event to analyze the effect. The microgrid operations for an outage time frame at 12 PM - 3 PM (time step 13–16) with different net loads are obtained using the proposed PRL approach and plotted in Fig. 7. In the figure, grid exchange represents $(a^{G,m}_{t,k} - a^{m,G}_{t,k})$, which means the value is positive when the microgrid imports power from the grid. The battery output represents $(a^{B,d}_{t,k} - a^{B,c}_{t,k})$, which means the value is positive when the battery is discharging. The results show that the proposed approach effectively utilizes RG outputs, uses the battery for intraday energy shifting in a cost-effective manner, and dispatches DG if needed to prevent load shedding. Therefore, the proposed approach is useful for cost-efficient microgrid operations during the extreme weather events.

4.2. Online testing

In this case study, we evaluate the optimization performance of the learning approaches in the uncertain environments and test the adaptivity of the proposed approach. For introducing uncertainties in the microgrid scheduling operation, we use RG noise for representing intermittent nature of RG, and load noise for addressing uncertain load situations. We define RG and load noises using uniform and normal probability distribution functions. We use four test problems with uncertainties, and the problems are summarized in Table 3. In the table, U and N represent uniform and normal probability distribution functions, respectively. In the table, $U(a,b)$ represents uniform probability distribution function with the range $[a,b]$. And $N(l,m)$ represents normal probability distribution function with the mean l and standard deviation m. For the RG, we consider the noise range of $[-5kW, 5kW]$ and for the load demand, we use the noise range of $[-3kW, 3kW]$. The noises to generate RG output and load demand profiles for testing purposes are around 25% and 6% of deviations.

4.2.1. Expected cost

We assess the performance in terms of microgrid expected operational cost during the extreme weather events for all the test problems. For each test problem, we generate 500 test scenarios, and the statistical results are plotted in Fig. 8. From the results, we can observe that the proposed PRL approach achieves minimum microgrid expected operational cost for all cases.

We calculate the performance improvements for this experiment in a similar way as that in [44], and the results are presented in

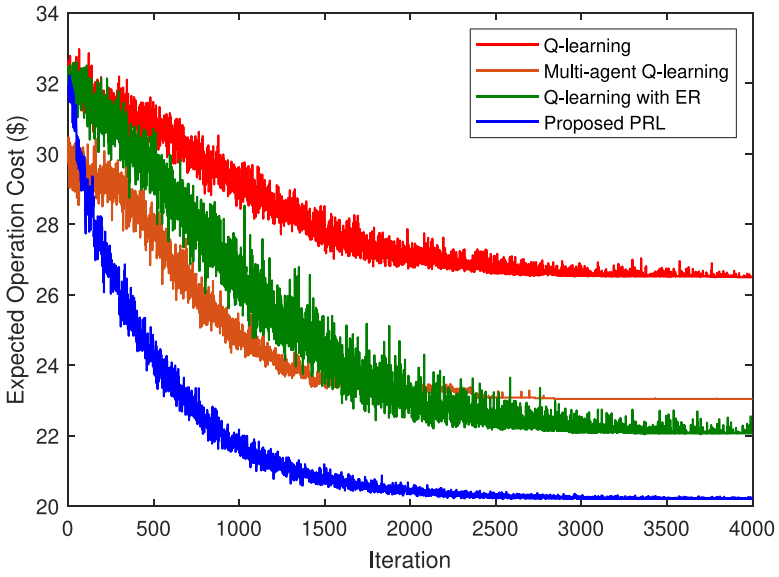


Fig. 6. Average expected cost convergence curve of the learning approaches after 30 runs.

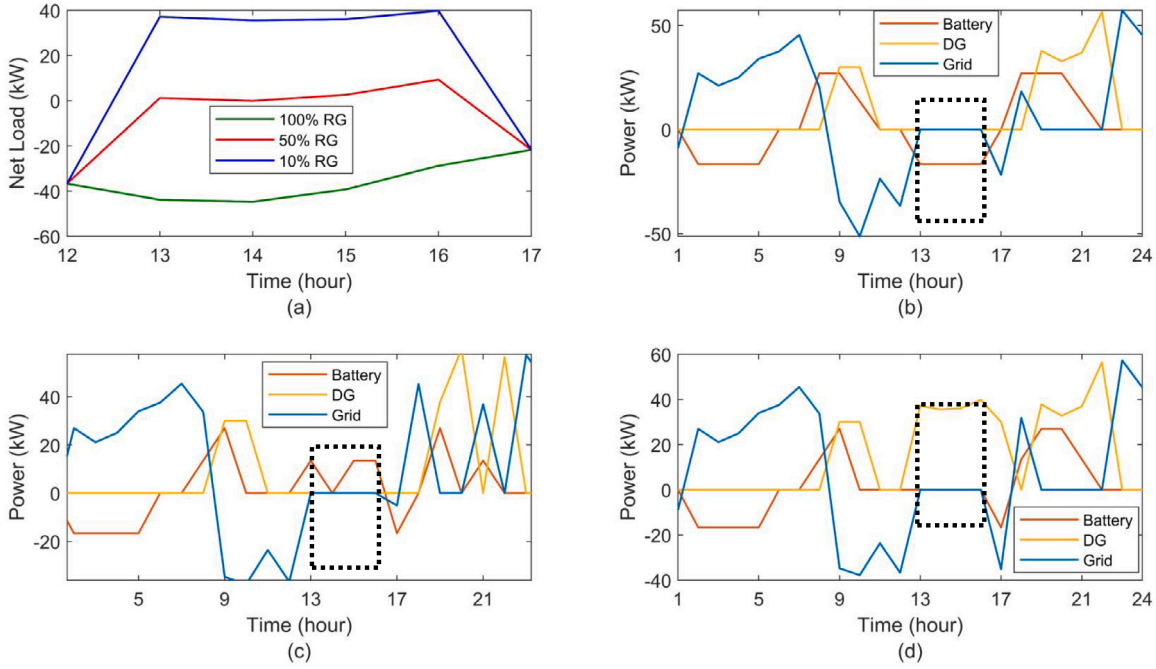


Fig. 7. Microgrid operations for an outage time frame at 12 PM - 3 PM (time step 13-16) with different RG outputs during the extreme weather event, (a) net load ($D_{t,k}$ - $R_{t,k}$), (b) microgrid operation with 100% RG, (c) microgrid operation with 50% RG, and (d) microgrid operation with 10% RG.

Table 4. The proposed PRL, Q-learning with ER, and multi-agent Q-learning methods output average expected daily operational costs as $21.9, $22.7, and $23.1, respectively. The performance improvements of these approaches are determined by comparing the results with the traditional Q-learning approach. In this case study, the proposed approach shows promising performance with a maximum of 20.94% of improvement in comparison.

4.2.2. Outage time uncertainties

In addition, we analyze the effect of uncertainties in different outage times and evaluate the decision-making skills of the learning approaches. For this case study, we use test problem 4 from Table 3, and the statistical results are obtained using 500 test scenarios. The results are reported in terms of microgrid operational cost for all possible outage hours and illustrated in Fig. 9. The results show the impact of having an outage in all different possible hours in terms of operation costs. The statistical box plots show how much microgrid operational cost we should expect of using the learning approaches at different outage times under uncertainties. The proposed approach achieves the minimum operation cost for all cases. The trends show that outage at hour 10 AM and 1 PM may cause maximum and minimum microgrid operating costs compared to other possible outage hours. For all cases, the proposed PRL approach shows promising adaptive performance and can be used for the economic assessment of extreme events. It also provides the microgrid scheduling decisions to minimize the operational loss of the events.

Moreover, we conduct a case study with outage time uncertainty and assess the adaptability performance of the learning approaches. In this case study, we use the test problem 3 from Table 3 to generate 500

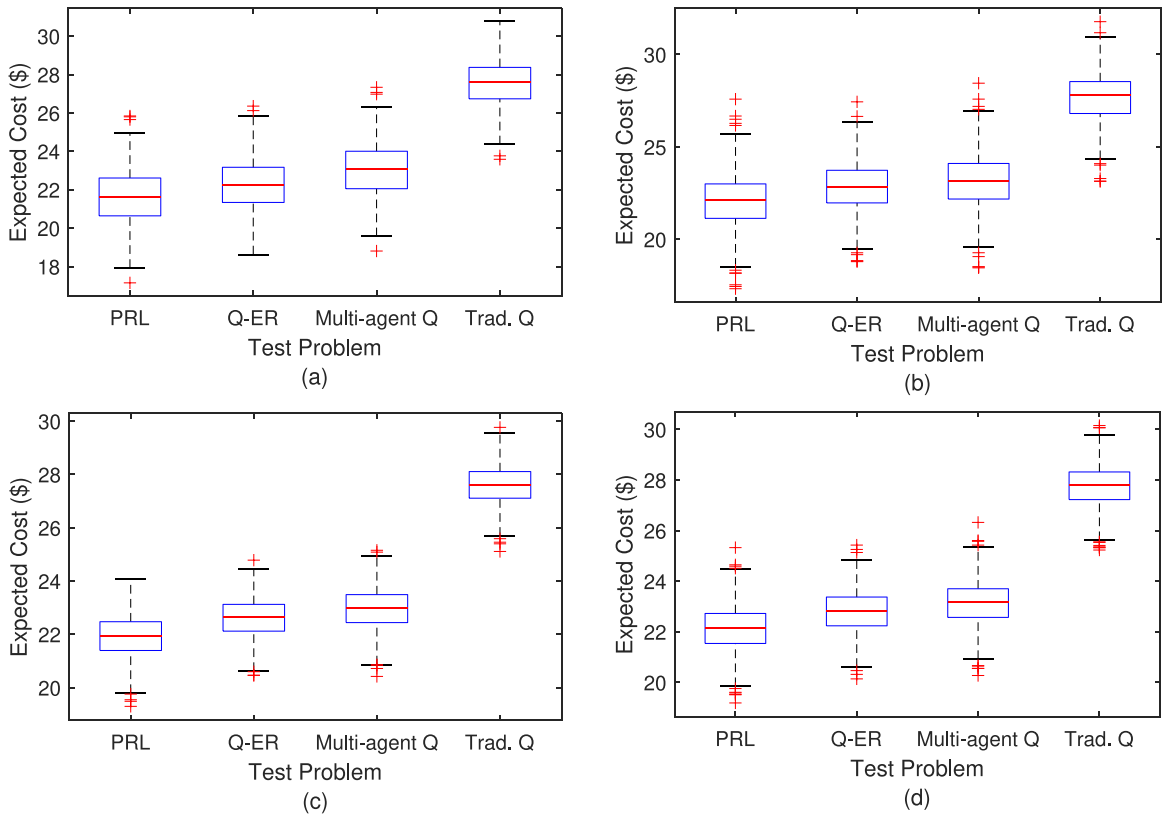


Fig. 8. Statistical results on microgrid expected cost with RG and load uncertainties during extreme weather events considering 500 test cases for each test problem.

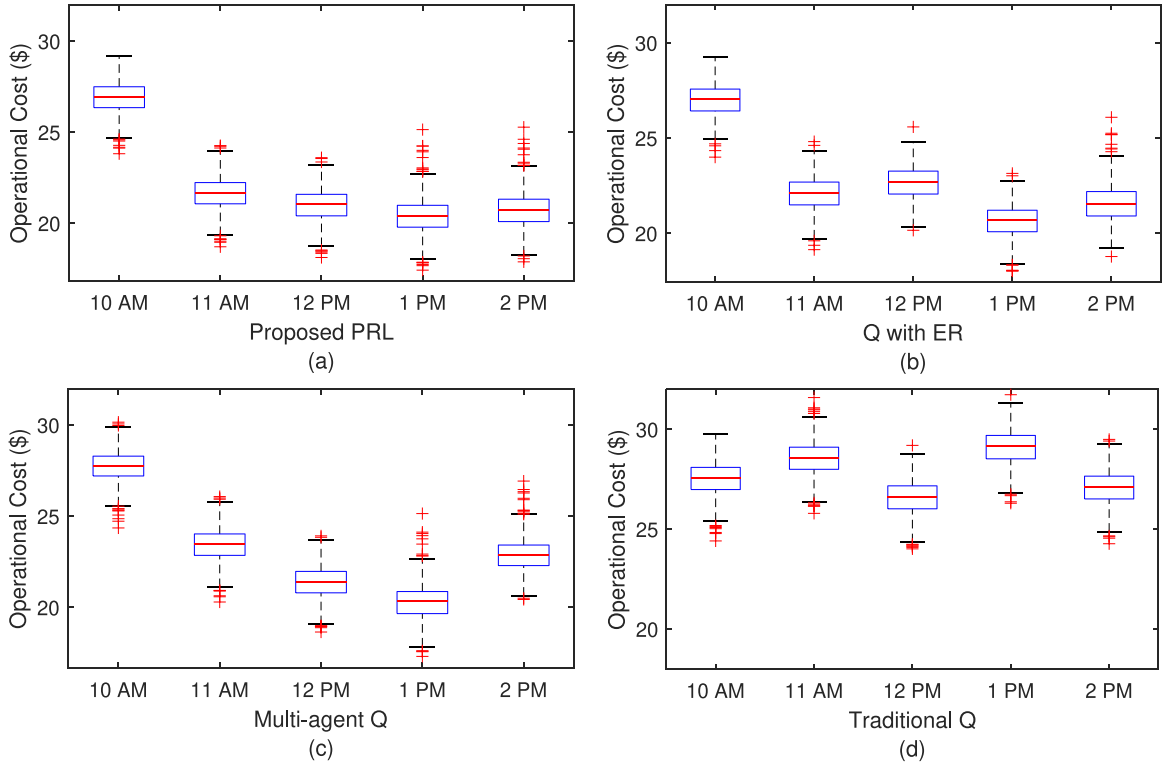


Fig. 9. The statistical results showing the impact of extreme weather event at different outage time with RG and load uncertainties.

test scenarios and randomly vary outage time within the outage time frame at each scenario. Specifically, in this cases study, we generate a random outage time for each test sample and evaluate the microgrid operation obtained from the learning approaches. When the agent senses an outage due to the extreme event, the agent follows the learned policy obtained for the corresponding scenario and determines

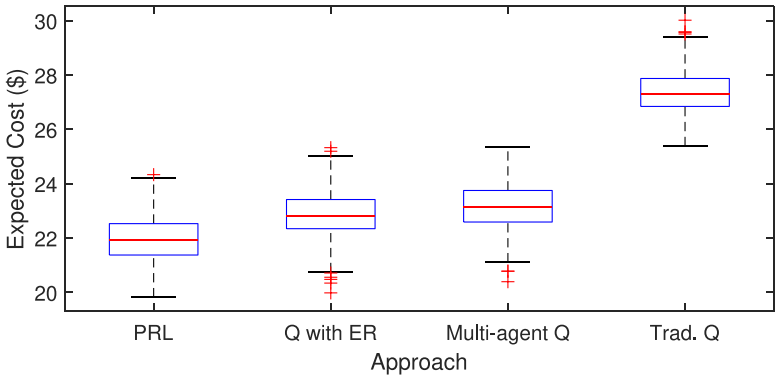


Fig. 10. Box plot for the online testing case study with random outage time.

Table 5
Online testing with random outage time results.

Approach	Average cost ($)	Improvement (%)
Proposed PRL	21.9	19.8
Q-learning with ER	22.8	16.5
Multi-agent Q-learning	23.2	15

microgrid operation accordingly. The Q-value function of the learning approaches is used to generate the scheduling decisions using a greedy technique. Statistical box plots are illustrated in Fig. 10 to summarize the results. The proposed PRL approach handles the outage time uncertainty well and outputs the minimum average expected cost comparing to other existing RL approaches.

Numerical results are also reported in Table 5. The proposed PRL approach outputs the minimum average cost with 19.8% of improvement comparing to the traditional Q-learning approach and outperforms other existing approaches. The existing Q-learning with ER and multi-agent Q-learning approaches also show 16.5% and 15% improvements and need an average of 4 and 28 times more computational time than the proposed approach, respectively. In conclusion, the proposed PRL approach shows promising performances in all case studies, indicating a potential advanced learning-based method for microgrid scheduling under extreme natural events.

5. Conclusion

This paper proposes a new RL approach with parallelized agents to efficiently solve the microgrid stochastic scheduling problem with resiliency considerations. Our proposed design employs local learning agents to interact with different microgrid operating environments under an extreme weather event in a distributed manner. It addresses the challenge of handling stochastic operation conditions in a timely manner. The information obtained from the local agents are used to build the state and action vectors for the global agent. Thus, we can compute the expected cost and efficiently generate the policy for the microgrid stochastic optimization problem. We formulate the proposed approach as a double-pass process, and the advantage estimate functions are used with a backward sweep to transfer the outcomes to the value function calculations efficiently. In the case study, stochastic optimization results show that the proposed PRL method is a computationally efficient approach that can achieve minimum expected microgrid operating costs compared to existing learning approaches. The proposed PRL approach also obtain around 20% of improvement in online testing case studies with 4 and 28 times less computation cost than Q-learning with ER and multi-agent Q-learning, respectively. Overall, the proposed PRL method performs microgrid scheduling efficiently considering the extreme event uncertainties.

CRediT authorship contribution statement

Avijit Das: Conception and design of study, Acquisition of data, Analysis and/or interpretation of data, Writing – original draft. **Zhen Ni:** Conception and design of study, Acquisition of data, Analysis and/or interpretation of data, Writing – original draft, Writing – review & editing. **Xiangnan Zhong:** Conception and design of study, Analysis and/or interpretation of data, Writing – original draft, Writing – review & editing.

Declaration of competing interest

There is no conflict of interest of this paper.

Data availability

Data will be made available on request

References

[1] Eskandarpour R, Lotfi H, Khodaei A. Optimal microgrid placement for enhancing power system resilience in response to weather events. In: 2016 North American power symposium. NAPS, IEEE; 2016, p. 1–6.
[2] Climate Central. Power OFF: Extreme weather and power outages. 2020, https://medialibrary.climatecentral.org/resources/power-outages.
[3] Sabouhi H, Doroudi A, Fotuhi-Firuzabad M, Bashiri M. Electrical power system resilience assessment: A comprehensive approach. IEEE Syst J 2019;14(2):2643–52.
[4] Jufri FH, Widiputra V, Jung J. State-of-the-art review on power grid resilience to extreme weather events: Definitions, frameworks, quantitative assessment methodologies, and enhancement strategies. Appl Energy 2019;239:1049–65.
[5] Panteli M, Mancarella P. Influence of extreme weather and climate change on the resilience of power systems: Impacts and possible mitigation strategies. Electr Power Syst Res 2015;127:259–70.
[6] Nie H, Chen Y, Xia Y, Huang S, Liu B. Optimizing the post-disaster control of islanded microgrid: A multi-agent deep reinforcement learning approach. IEEE Access 2020;8:153455–69.
[7] Frank T. U.S. suffers 147 big blackouts each year. That's rising. Oct., 2019, https://www.eenews.net/stories/1061245945.
[8] Ni Z, Paul S. A multistage game in smart grid security: A reinforcement learning solution. IEEE Trans Neural Netw Learn Syst 2019;30(9):2684–95.
[9] Venayagamoorthy GK, Sharma RK, Gautam PK, Ahmadi A. Dynamic energy management system for a smart microgrid. IEEE Trans Neural Netw Learn Syst 2016;27(8):1643–56.
[10] Wang MQ, Gooi H. Spinning reserve estimation in microgrids. IEEE Trans Power Syst 2011;26(3):1164–74.
[11] Khodaei A. Resiliency-oriented microgrid optimal scheduling. IEEE Trans Smart Grid 2014;5(4):1584–91.
[12] Xu Y, Liu C-C, Schneider KP, Tuffner FK, Ton DT. Microgrids for service restoration to critical load in a resilient distribution system. IEEE Trans Smart Grid 2016;9(1):426–37.
[13] Gao H, Chen Y, Xu Y, Liu C-C. Resilience-oriented critical load restoration using microgrids in distribution systems. IEEE Trans Smart Grid 2016;7(6):2837–48.

[14] Farzin H, Fotuhi-Firuzabad M, Moeini-Aghtaie M. Enhancing power system resilience through hierarchical outage management in multi-microgrids. IEEE Trans Smart Grid 2016;7(6):2869–79.
[15] Zhao Y, Lin Z, Ding Y, Liu Y, Sun L, Yan Y. A model predictive control based generator start-up optimization strategy for restoration with microgrids as black-start resources. IEEE Trans Power Syst 2018;33(6):7189–203.
[16] Sefidgar-Dezfouli A, Joorabian M, Mashhour E. A multiple chance-constrained model for optimal scheduling of microgrids considering normal and emergency operation. Int J Electr Power Energy Syst 2019;112:370–80.
[17] Nourollahi R, Salyani P, Zare K, Mohammadi-Ivatloo B. Resiliency-oriented optimal scheduling of microgrids in the presence of demand response programs using a hybrid stochastic-robust optimization approach. Int J Electr Power Energy Syst 2021;128:106723.
[18] Zhang J, Luo Y, Wang B, Lu C, Si J, Song J. Deep reinforcement learning for load shedding against short-term voltage instability in large power systems. IEEE Trans Neural Netw Learn Syst 2021;1–12. http://dx.doi.org/10.1109/TNNLS.2021.3121757.
[19] Gao X, Si J, Wen Y, Li M, Huang H. Reinforcement learning control of robotic knee with human-in-the-loop by flexible policy iteration. IEEE Trans Neural Netw Learn Syst 2021;1–15. http://dx.doi.org/10.1109/TNNLS.2021.3071727.
[20] Mu C, Peng J, Sun C. Hierarchical multiagent formation control scheme via actor-critic learning. IEEE Trans Neural Netw Learn Syst 2022;1–14. http://dx.doi.org/10.1109/TNNLS.2022.3153028.
[21] Srinivasan D, Venayagamoorthy GK. Guest editorial special issue on "neural networks and learning systems applications in smart grid". IEEE Trans Neural Netw Learn Syst 2016;27(8):1601–3.
[22] Kuznetsova E, Li Y-F, Ruiz C, Zio E, Ault G, Bell K. Reinforcement learning for microgrid energy management. Energy 2013;59:133–46.
[23] Kim B-G, Zhang Y, Van Der Schaar M, Lee J-W. Dynamic pricing and energy consumption scheduling with reinforcement learning. IEEE Trans Smart Grid 2015;7(5):2187–98.
[24] Fang X, Wang J, Song G, Han Y, Zhao Q, Cao Z. Multi-agent reinforcement learning approach for residential microgrid energy scheduling. Energies 2020;13(1):123.
[25] Luo F, Chen Y, Xu Z, Liang G, Zheng Y, Qiu J. Multiagent-based cooperative control framework for microgrids' energy imbalance. IEEE Trans Ind Inf 2016;13(3):1046–56.
[26] Ghorbani MJ, Choudhry MA, Feliachi A. A multiagent design for power distribution systems automation. IEEE Trans Smart Grid 2015;7(1):329–39.
[27] Ferreira LR, Aoki AR, Lambert-Torres G. A reinforcement learning approach to solve service restoration and load management simultaneously for distribution networks. IEEE Access 2019;7:145978–87.
[28] Mnih V, Badia AP, Mirza M, Graves A, Lillicrap T, Harley T, et al. Asynchronous methods for deep reinforcement learning. In: International conference on machine learning. 2016, p. 1928–37.
[29] Nair A, Srinivasan P, Blackwell S, Alcicek C, Fearon R, De Maria A, et al. Massively parallel methods for deep reinforcement learning. 2015, arXiv preprint arXiv:1507.04296.
[30] Silver D, Newnham L, Barker D, Weller S, McFall J. Concurrent reinforcement learning from customer interactions. In: International conference on machine learning. 2013, p. 924–32.
[31] Andreas J, Klein D, Levine S. Modular multitask reinforcement learning with policy sketches. In: International conference on machine learning. PMLR; 2017, p. 166–75.
[32] Teh Y, Bapst V, Czarnecki WM, Quan J, Kirkpatrick J, Hadsell R, et al. Distral: Robust multitask reinforcement learning. Adv Neural Inf Process Syst 2017;30.
[33] Das A, Ni Z, Zhong X. Aggregating learning agents for microgrid energy scheduling during extreme weather events. In: 2021 IEEE power & energy society general meeting. IEEE; 2021, p. 1–5.
[34] Mansour-lakouraj M, Shahabi M. Comprehensive analysis of risk-based energy management for dependent micro-grid under normal and emergency operations. Energy 2019;171:928–43.
[35] Das A, Ni Z. A computationally efficient optimization approach for battery systems in islanded microgrid. IEEE Trans Smart Grid 2017;9(6):6489–99.
[36] Sutton RS, Barto AG. Reinforcement learning: An introduction. MIT Press; 2018.
[37] Zhang S, Sutton RS. A deeper look at experience replay. 2017, arXiv preprint arXiv:1712.01275.
[38] Ni Z, Malla N, Zhong X. Prioritizing useful experience replay for heuristic dynamic programming-based learning systems. IEEE Trans Cybern 2019;49(11):3911–22. http://dx.doi.org/10.1109/TCYB.2018.2853582.
[39] Das A, Wu D, Ni Z. Approximate dynamic programming with policy-based exploration for microgrid dispatch under uncertainties. Int J Electr Power Energy Syst 2022;142:108359. http://dx.doi.org/10.1016/j.ijepes.2022.108359.
[40] Degris T, Pilarski P, Sutton R. Model-free reinforcement learning with continuous action in practice. In: American control conference. 2012, p. 2177–82.
[41] Blair N, Dobos AP, Freeman J, Neises T, Wagner M, Ferguson T, et al. System advisor model, sam 2014.1. 14: General description. NREL/TP-6A20-61019, National Renewable Energy Lab. (NREL), Golden, CO (United States); 2014.
[42] NREL. Openei. 2020, https://openei.org/community/blog/commercial-and-residential-hourly-load-data-now-available-openei.
[43] EIA US. Annual electric power industry report, June. 2020, URL https://www.eia.gov/electricity/data/eia861/.
[44] Das A, Ni Z. A novel fitted rolling horizon control approach for real-time policy making in microgrid. IEEE Trans Smart Grid 2020;11(4):3535–44.