SimBuild
2024

Eleventh National Conference of IBPSA-USA
Denver, Colorado
May 21 – 23, 2024

Reinforcement Learning to Enhance Optimal Operation of Resilient Community Energy Systems

Zhuorui Li[1], Xu Han[1], Jing Wang[2], and Wangda Zuo[2,3]
[1]Department of Civil, Environmental and Architectural Engineering, University of Kansas, Lawrence, KS 66045, USA
[2]National Renewable Energy Laboratory, Golden, CO
[3] Department of Architectural Engineering, Pennsylvania State University, University Park, 16802, PA, USA

Abstract

This paper presents a novel model-free multi-agent Reinforcement Learning (RL) control method to enhance the resilience of community energy systems in island mode, which coordinates multiple objectives without the necessity of identifying system models that require expert knowledge. Specifically, a community-level coordinator agent is designed to allocate renewable energy resources among different buildings, and multiple building-level agents are developed to optimize load schedules based on limited energy resources and requirements of building loads and occupants' comfort. In a two-day evaluation, our RL approach demonstrated a similar performance against MPC without requiring system models and formulation of optimization problems as required in MPC.

Introduction

In recent years, extreme weather conditions have increasingly affected the stability and reliability of the grid (Wang et al., 2019). Events such as wildfires, hurricanes, heatwaves, heavy snow, and extreme cold can cause significant disruptions to the power grid. Photovoltaic (PV) systems have become increasingly popular as a source of emergency power when grid electricity is unavailable, especially in areas frequently affected by power outages due to extreme weather or other reasons (Wang et al., 2019). Therefore, improving the resilience and reliability of power distribution from PV panels to satisfy building needs and occupants' thermal comfort requirements is crucial.

"Community resilience" refers to the ability of a community to actively participate in crisis planning, respond effectively to crises, and engage in recovery work (Milofsky and Carl, 2023). Unlike grid power, which provides electricity continuously, PV panels only generate power during daylight hours with uncertainties due to weather conditions. As a result, the distribution of this PV power to meet the varying and specific demands of each building becomes a critical challenge as it should coordinate multiple objectives simultaneously. This brings additional challenges to conventional Rule-Based Control (RBC), which is commonly used in practice (Bukar et al., 2020).

Therefore, Model Predictive Control (MPC) has been proposed to optimize power distribution and building load scheduling. Zhang et al. (2019) utilized MPC with stochastic programming and variable constraints to optimize the sizing and siting schemes of battery storage and PV generation systems, aimed at enhancing power system resilience. Rasheed et al. (2019) proposed a hierarchical control system based on multiple agents for managing residential loads in a real-time pricing context. Wang et al. (2020) utilized MPC to achieve optimal controls of a resilient community when experiencing an outage. Their approach optimized the distribution of PV power, HVAC load, and building load scheduling to ensure the occupants' thermal comfort and other needs. Cai et al. (2020) proposed a microgrid scheduling method based on MPC, aimed at enhancing the resilience of distribution networks during outages. Yan et al. (2019) introduced a resilient energy management scheme based on MPC for a prototype building community, equipped with community-scale PV panels and a Battery Energy Storage System (BESS). The aim was to augment the community's self-power supply capability during periods of grid outage.

While MPC can effectively enhance the resilience of microgrids and buildings, it still possesses several notable disadvantages. Firstly, the identification of system models and formulation of the optimization problem, including the objective function and constraints, necessitates expert knowledge. Secondly, the excessive effort required to develop and customize system models for different buildings limits its applications in scale, which has been reported to constitute approximately 70% of the required effort (Henze et al., 2013). Thirdly, it is even more challenging when large-scale buildings are involved, leading to an increased dimension and complexity of the optimization problem that is hard to solve.

To overcome the limitations of the MPC, model-free RL offers the possibility of learning optimal control policy without constructing the model of the environment. Of the studies of resilient communities that applied RL most focused on reducing recovery time and economic loss. Elsayed et al. (2020) used Q-learning to build an algorithm that efficiently allocated resources and scheduled users' loads. Kim and Lee (2020) applied Deep Q-network (DQN) to assess peer-to-peer automatic energy trading to achieve a higher profit for each prosumer. Fang et al. (2021) adopted a multi-agent DQN to attain distributed energy scheduling and strategy formulation for microgrids, and the RL algorithm can guarantee efficient and autonomous operations within the microgrid market, ensuring a balanced benefit for each component of the microgrid.

In addition to the economic aspect, satisfying occupants' needs is also critical. Zhou et al. (2021) applied three kinds of RL algorithms, including Q-learning, Policy Gradient (PG), and Proximal Policy Optimization (PPO), to enhance the resilience of the distribution system of power systems with the moving of a typhoon to maintain essential load that was related to occupants' needs and reduce load shedding. Mocanu et al. (2019) applied both the PG and DQN algorithms to a smart grid to manage multiple actions of devices concurrently to reduce peak load consumption and cost. Their approaches effectively manage the intrinsic uncertainty and variability associated with PV power generation and the occupancy's various load requirements. However, as one of the crucial factors of the occupants' needs, the thermal comfort of the occupants was not considered in these studies.

Based on our review, limitations exist in the abovementioned studies that applied RL to enhance the resilience of communities. First, these studies usually simplified building load types. For example, they may utilize simple aggregated building loads rather than detailed breakdown loads considering different patterns and priorities. Second, they often overlook occupants' thermal comfort when the community is operating in island mode. Furthermore, there is a noticeable absence of research dedicated to exploring both scheduling of diverse types of controllable loads and maintaining occupants' thermal comfort with limited renewable energy generation within a community-scale context in an island mode.

Therefore, in this paper, we propose a novel model-free multi-agent RL control method to enhance the resilience of community energy systems in an island mode, which considers meeting various types of building loads and occupants' thermal comfort needs with limited renewable energy resources when the community is disconnected from the power grid. Specifically, the RL-based controller optimizes the PV power allocation among different buildings, scheduling various loads, including critical, shiftable, sheddable, and modulatable loads, at each building, HVAC load, battery charge, and discharge. The control objective is to minimize the PV curtailment, satisfy essential loads, and meet occupants' thermal comfort requirements without violating the constraints of battery capacities. The RL-based controller was developed based on the Deep Deterministic Policy Gradient (DDPG) algorithm. A community-level coordinator agent is designed to allocate renewable energy resources among different buildings, and multiple building-level agents are developed to optimize load schedules based on limited energy resources and the requirement of building loads and occupants' comfort to solve the complex community-scale control problem in a distributed manner. Then, a case study is conducted that compares a conventional model-based MPC method and the proposed model-free RL-based approach from various perspectives, which was suggested in previous studies (Zhou et al., 2021). In conclusion, the main contributions of this paper are as follows:

- Proposed a novel model-free control method to enhance the resilience of community energy systems in an island mode, considering meeting various types of building loads and occupants' thermal comfort requirements with limited renewable energy resources, which achieves a similar performance as an MPC while not requiring identifying the system models.
- Designed a multi-agent RL control algorithm that coordinates community-level renewable energy allocation and optimizes the scheduling of various building loads and maintains occupants' thermal comfort for different buildings in a resource among different buildings in a distributed manner, which enhances the scalability of the control algorithm in large scale community.
- Compared the model-based MPC method and the model-free RL-based approach for optimal controls of resilient community energy systems, which provided the assessment of the effectiveness and capability of the RL-based approach against a well-designed model-based optimal controller.

Methodology

The methodology section comprehensively outlines the (1) Problem description, (2) Multi-Agent RL Control Method, a) State Definition, b) Action Definition, c)

Reward Definition, (3) Two-phase multi-agent RL training approach, and (4) Environment Setup.

Problem Description

The study was based on a net-zero energy community in Anna Maria Island, FL, consisting of residential and commercial buildings, and equipped with community-level solar panels (Bernard, 2016). This study focused on enhancing the resilience of the community considering three buildings in an island mode when the grid power is not available after a disaster: a residential house (93.8 m^2), an ice cream shop (160.5 m^2), and a bakery shop (410 m^2). The three buildings have different construction materials, heat pump nominal powers, and battery sizes. All these buildings employ heat pumps for heating and cooling needs. The weather data came from a standard yearly weather file measured at Tampa International Airport (Wang et al. (2020) except for the solar radiation, which was collected from a local weather station. Furthermore, the building load data was provided by the community (Wang et al., 2020). In this study, we aim to optimize distribution and building load scheduling for a community, ensuring resilience during a two-day outage of the main power grid. The available energy source is the community-scale PV panels. Each building has a battery energy storage system to supply energy when the distributed PV power is insufficient. Under these circumstances, the objective is to optimize PV power distributions for three buildings, satisfying each building's various types of loads, especially critical loads (for example, Lights, refrigerators, and so on), balancing charge and discharge of battery storage systems.

For the community-level coordinator, since there is no electricity storage system installed in the PV generation plant and batteries are equipped within each building, the sum of distribution ratios of the three buildings should equal one. Based on the different kinds of occupants' needs, building loads are classified into four categories: critical, shiftable, sheddable, and modulatable loads. Critical loads, such as lighting and refrigerators, are essential for daily life; Shiftable loads, such as dishwashers, must be satisfied once a day with continuous operating time but flexible starting times. Sheddable loads, such as computers or coffee makers, can only be turned on or off. Modulatable load, like HVAC, can choose a suitable amplitude of the original load.

The control objective for the community level is to determine the optimal PV distribution ratios among different buildings to minimize PV energy curtailment and load shedding while satisfying each building's loads. The control purpose for the building level is to minimize PV curtailment, satisfy building loads, and meet occupants' thermal comfort requirements while avoiding over-discharge from the battery.

The optimal controls of the energy systems of this resilient community have been studied by Wang et al. (2020) with a model-based MPC approach, which served as an upper-bound-performance benchmark for this study that applies a model-free RL approach.

Multi-Agent RL Control Method

In an RL algorithm, an agent learns the optimal control policy by interacting with the environment. The environment, in turn, gives states/rewards to the agent based on the actions taken. The agent's goal is to learn a policy that could result in the maximum expected cumulative reward over time.

In this study, a multi-agent RL-based approach has been proposed to enhance the optimal operation of resilient community energy systems as shown in Figure 1. The algorithm for the RL agent is DDPG, which employs an actor-critic architecture where the "actor" approximates the optimal policy, and the "critic" estimates the value of the action determined by the actor. The reason that we adopted DDPG is that it works for the environment with a continuous action space, e.g., PV distribution ratio in this study, and performs best among other algorithms in our initial tests. The actor's policy function, parameterized by $\emptyset$, maps states (s) to a specific action (a) as shown in Eq. (1).

$$a = \pi(s|\emptyset). \tag{1}$$

The critic action-value function, parameterized by θ, estimates the Q-value that describes the expected cumulated reward of the (s, a) , which can be calculated by Eq. (2), s is the current state and s' is next state:

$$Q_{target}(s, a) = r + \gamma\, Q(s, a|\theta)(s', \pi(s')). \tag{2}$$

The loss of the critic network (L) is the Mean Squared Error (MSE) between the estimated Q-value and the target Q-value as shown in Eq. (3), where E stands for expected value.

$$L\,(\theta, D) = E_{(s,a,r,s',d)\sim D}[(Q_{target}(s, a) - Q(s, a|\theta))^2]. \tag{3}$$

The loss of the actor-network is the negative expected Q-value as shown in Eq. (4), which is negative to maximize the expected Q-value.

$$L(\phi,D) = -E_{s\sim D}\,(\,Q\,(s,\pi(s|\emptyset)\,|\theta). \tag{4}$$

In this work, there are multiple agents on two layers. One RL agent is implemented at the first layer serving as the community-level PV allocation coordinator, which the PV power distribution ratio for them. At the second

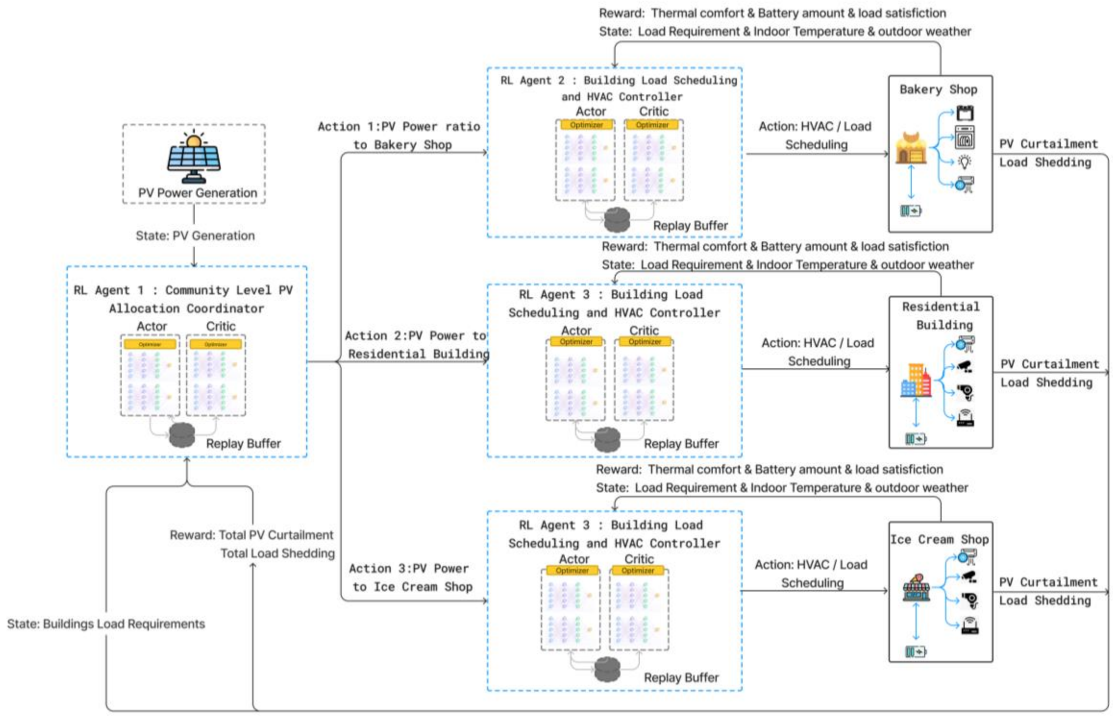

Figure 1 Structure of the multi-agent RL-based approach for controlling resilient community energy

layer, three building RL agents are developed to optimize building load scheduling, and HVAC operations for each building in the building level. The detailed design of the states, action, rewards are described in the following sub-sections.

State Definition:

For the community-level agent, the state features include:

- PV flexible power: the surplus PV energy production after satisfying building-level agents' essential load requirements.
- Building agents' flexible loads: flexible loads of the three buildings that can be calculated using Eq. (5):

$$P_{load,flex,i}^{t} = \sum_{j=1}^{N_{shed,i}} p_{shed,j,i}^{t} + \sum_{j=1}^{N_{modu,i}} p_{modu,j,i}^{t} + \sum_{j=1}^{N_{shif,i}} p_{shif,j,i}^{t}. \tag{5}$$

Where $p_{shed,j,i}^{t}$, $p_{modu,j,i}^{t}$, $p_{shif,j,i}^{t}$ represent the load requirement data for sheddable, modulatable, and shiftable loads in building i, respectively. The $N_{shif,i}$, $N_{modu,i}$, $N_{shed,i}$, denote the number of appliances for each type of load, respectively.

For building-level agents, the state features include:

- Distributed PV flexible power: PV flexible power allocated to each building.
- Each appliance's load requirements at time *t in building i*
- Indoor temperature at each building *i*
- Forecasted outdoor temperature, T_{out}^{t} that can be calculated using Eq. (6):

$$T_{out}^{t} = (T_{out}^{t+1} + T_{out}^{t+2})/2 \tag{6}$$

Action Definition:

The action vector for the community-level agent comprises three PV power distribution ratios as described in Eq. (7). Each ratio ranges from 0 to 1, and the sum of them is equal to one.

$$A_{community}^{t} = (A_{residential}^{t}, A_{Ice\ Cream}^{t}, A_{Bakery}^{t}). \tag{7}$$

The action vector for building-level agents includes load scheduling actions related to sheddable load, modulatable load, shiftable load, and HVAC load:

$$A_{sheddable,i}^{t} = \left(A_{shed,i,1}^{t}, A_{shed,i,2}^{t}, \ldots, A_{shed,i,N_{shif,i}}^{t}\right). \tag{8}$$

$$A_{shed} = \begin{cases} 0, & If\ \ A_{shed} < 0.5. \\ 1, & If\ \ A_{shed} \geq 0.5. \end{cases} \tag{9}$$

$$A_{modu,i}^{t} = \left(A_{modu,i,1}^{t}\ ,\ A_{modu,i,2}^{t}\ , \dots ,\ A_{modu,i,N_{modu,i}}^{t}\ \right). \tag{10}$$

$$A_{shif,i}^{t} = \left(A_{shif,i,1}^{t}\ ,\ A_{shif,i,2}^{t}\ , \dots ,\ A_{shif,i,N_{modu,i}}^{t}\ \right). \tag{11}$$

$$A_{HVAC,i}^{t}$$

The outputs of all actions of the building agents are continuous, ranging from 0 to 1, which would be converted to the actual control actions based on different strategies for different types of loads. For the sheddable load, the action is converted to binary values representing on/off of the load status based on Eq. (9) because the appliances can only be totally turned on or turned off. For modulatable load and HVAC load, the actions representing the ratio of nominal loads would be scaled by multiplying the original load requirements or HVAC nominal power. For each appliance with a shiftable load, the time step when the action firstly exceeds a value of 0.5 would be selected as the starting time. Once the starting time is decided, the appliances' power demand would persist until the end of its operational duration.

After applying these actions, the charging and discharging status of the battery can be calculated. If there is a surplus of distributed PV power, the surplus energy would be charged to the battery until the battery is fully charged. If the distributed PV power is insufficient for the load requirement, power would be discharged from the battery.

Reward Definition：

For the community-level agent, the reward is defined by Eq. (12):

$$Reward_{community}(t, \{x^t\}_H) = \sum_{i=1}^{I} \sum_{t=1}^{H} \left(P_{curt,i}^{t} + P_{ls,i}^{t} \right). \tag{12}$$

$P_{curt,i}^{t}$ and $P_{shed,i}^{t}$ can be calculate by Eq. (13):

$$\alpha_i^t \times P_{pv}^t - P_{curt,i}^t = P_{load,i}^t - P_{shed,i}^t. \tag{13}$$

In Eq. (13), $\alpha_i^t \times P_{pv}^t$ is the amount of PV generated power allocated to building *i* at time *t*. The variable $P_{curt,i}^{t}$ means the part of allocated PV power exceeds the current loads' requirement. Similarly, $P_{shed,i}^{t}$ is the variable of load shedding for situations when the available PV power cannot satisfy the current energy requirement.

For building-level agents, the reward design considers thermal comfort of occupants, battery, and load satisfaction:

$$Reward_{building}(t, \{x^t\}_H) = \sum_{t=1}^{H} \left(R_{Temp,i}^{t} + R_{Load,i}^{t} + R_{Battery,i}^{t} \right). \tag{14}$$

The reward design for the building-level agent comprises three components as defined in Table 1, Table 2, and Table 3: The first reward is to maintain the indoor temperature within a comfortable range; The second is to meet sheddable and modulatable load requirements, and the third reward is to prevent the battery from over-discharge. Refer to Tables 1-3 for detailed description of the reward design.

Table 1 Indoor Temperature Reward Function

Reward function - Thermal comfort Requirements
For each time step t in building i, **do**
$Curent_{dis,i} = abs(24 - Tindoor_{t,i})$
$Last_{dis,i} = abs(24 - Tindoor_{t-1,i})$
If $Tindoor_{t,i}$ in comfort range:
$R_{Temp,i}^{t} = 1 + (Last_{dis,i} - Current_{dis,i})0.2 - Current_{dis,i} * 0.1$
If $HVAC_{t-1,i} == 0$ or $HVAC_{t-2,i} == HVAC_{t-1,i}$:
Add 0.1to reward $R_{Temp,i}^{t}$ += 0.1
Elif $Current_{dis,i} < Last_{dis,i}$:
Reward equal to 0.1 $R_{Temp,i}^{t} = 0.1$
If $HVAC_{t-1,i}$ = = 0 or $HVAC_{t-2,i} == HVAC_{t-1,i}$:
$R_{Temp,i}^{t}$ += 0.1
Else:
Negative one as penalty $R_{Temp,i}^{t} = -1$

Table 2 Load reward function

Reward function - Load Requirements
For each time step t in building i, **do**
For actions related to sheddable load:
If $a_{sheddable,i}^{t} \geq 0.5$:
$R_{Load,i}^{t}$ += 0.1
For actions related to modulatable load:
If $a_{modu,i}^{t} \geq 0.5$:
$R_{Load,i}^{t}$ += 0.1

Table 3 Battery reward function

Reward function - Battery Requirements
For each time step t , **do**
If $Battery_{residential} \leq 20$ and $P_{discharge,residential} > 28$:
$R_{Battery,residential}^{t}$ -= 0.1
If $Battery_{ice} \leq 60$ and $P_{discharge,ice\ cream} > 108$:
$R_{Battery,ice\ cream}^{t}$ -= 0.1
If $Battery_{bakery} \leq 60$ and $P_{discharge,barkery} > 112$:
$R_{Battery,bakery}^{t}$ -= 0.1

Two-phase multi-agent RL training approach

The increasing dimension and complexity when multiple agents for different tasks are involved in a large-scale community energy system control problem brings a lot of challenges for the efficient training and convergence of the RL algorithms. To address this challenge, this study proposes a two-phase training approach to convert the complex multi-agent RL training problem to multiple simple single-agent RL training.

In the first phase, the three building-level agents are trained separately with a randomized community PV power distribution ratio. At this stage, the goal is to initially train the building-level agents so that the trial-and-error process can be significantly shortened when multiple agents are integrated together, leading to a better convergence ability. Upon achieving satisfactory outcomes by the building agents, we transition to the next phase.

In this secondary phase, we integrate the building-level agents with the community-level agent. The knowledge of the building-level agents learned from the first phase is transferred to this phase by loading the trained coefficients of the building-level agents in the initialization stage. By implementing this two-phase training process, we can enhance the stability and efficiency of the training process, ensuring a better convergence from both the building and community agents.

The RL algorithm is trained by updating every 24 hours, with the accumulated reward described in Eq. (2). The hyperparameters are determined through initial tests and summarized in Table 4.

Table 4 RL Model Parameter

Building Agents	Actor	Critic
Episode	1000	1000
Learning Rate	3e-3	2e-2
Optimizer	Adam	Adam
γ	0.99	0.99
Number of layers	4	4
Size of Hidden Layers	20	64
Community Agent	**Actor**	**Critic**
Episode	200	200
Learning Rate	3e-6	2e-6
Optimizer	Adam	Adam
γ	0.99	0.99
Number of layers	4	4
Size of Hidden Layers	20	64

Environment Setup

The environment is implemented for training and testing the RL algorithm. At the community level, the environment incorporates data on solar radiation, PV power production, and building load requirements. In each time step, the environment receives the solar radiation data and predicts the PV production accordingly, which was precomputed and stored in a CSV file in this study to be consistent with our benchmark study (Wang et al., 2020).

The building-level environment provides the load requirements for various buildings and predicts the indoor temperature and battery status based on control actions from the agents. The environment determines the load requirements for each building at each step by reading a CSV file, which was provided by the community based on real-world data (Wang et al., 2020).

The indoor temperature is estimated based on a thermal dynamics model developed by Eq. (15):

$$T_{room}^{t+1} = \beta_1 T_{amb}^{t} + \beta_2 T_{amb}^{t-1} + \beta_3 T_{room}^{t} + \beta_4 T_{room}^{t-1} - \beta_5 r_{hvac}^{t} + \beta_6 Q_{sol}^{t} + \beta_7 Q_{sol}^{t-1}. \tag{15}$$

The battery status is predicted by following equation Eq. (16):

$$E_{battery,i}^{t+1} = E_{battery,i}^{t} + P_{charge,i}^{t} * 0.9 - P_{discharge,i}^{t} * 1.1. \tag{16}$$

The environment is based on 48-hour real-world data during the hurricane season on August 4th and 5th 2014. One episode for the RL training is defined as 48 hours simulations of the environment, and 48 (s^t,a^t,r^t,s^{t+1}) transitions would be generated from each episode. The control interval of the RL agent is set to 1 hour.

Discussion and result analysis

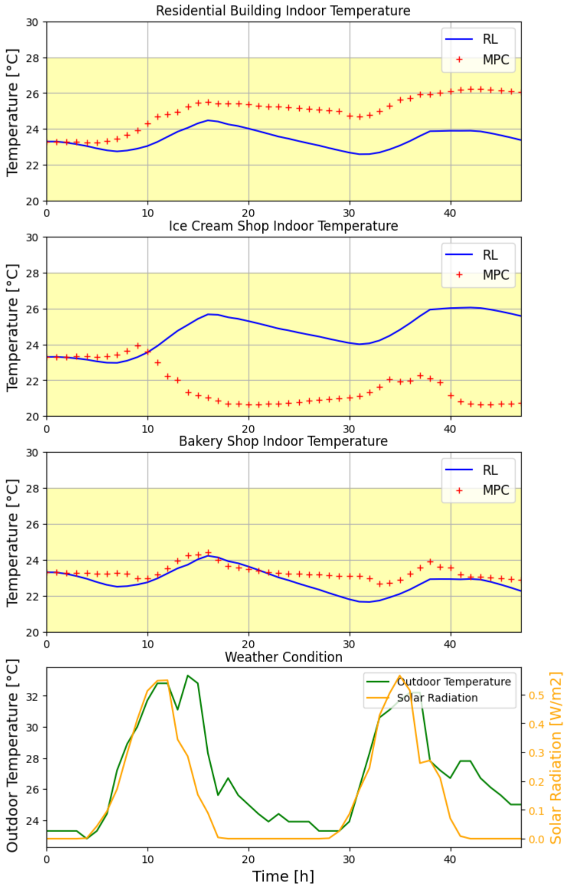

Figure 2 Indoor Temperature and Weather Condition (the comfort zone is defined from 20 to 28 °C as filled with yellow color in the figure to be consistent with (Wang et al., 2020))

Figure 2 illustrates the indoor temperature control performance. Both RL and MPC methods successfully maintain indoor temperatures within the comfort range, despite observable fluctuations over the two days. Notably, both methods effectively regulate indoor temperatures despite the disturbances of outdoor weather conditions. For the residential building, RL maintains a more stable temperature at about 24°C, whereas MPC sees a gradual temperature rise. For the ice cream shop, noticeable indoor temperature variations can be seen for both methods, which may be due to the thermal dynamics' characteristics of the building. The bakery shop's indoor environment remains remarkably stable for both methods, keeping between 22°C and 24°C. Overall, across diverse building types with varying construction materials and HVAC system configurations, RL and MPC demonstrate commendable and comparable indoor temperature control performance.

The changes in battery power are shown in Figure 3. *Both* the RL and MPC effectively protect the battery from overcharging and over-discharging in the two days. The RL-based approach controls the battery with similar daily patterns as MPC. The deviations between the battery power patterns by the two control methods are due to different PV power distributions to buildings and load patterns.

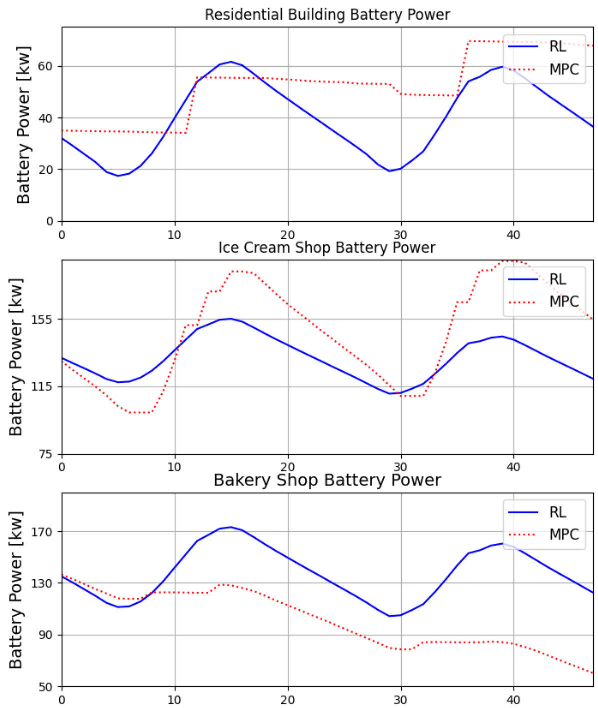

Figure 3 Battery Power

The three buildings' load status is shown in Figures 4,5 and 6. MPC, as a benchmark control method, satisfies all kinds of building loads requirements among three buildings. Compared with MPC control, RL can ensure that the critical loads are consistently met throughout the two days; shiftable loads are also satisfied with a similar shiftable load operation time in MPC; the sheddable load is only satisfied in the residential building, less than half of the original load requirement is met in the ice cream shop and no power is distributed for this load in the bakery shop; for the modulatable load in the bakery shop, RL only manages to fulfill approximately half of it.

While both well maintain the indoor temperature within the acceptable range, RL and MPC have shown different strategies for HVAC load management. RL maintains the HVAC system running with relatively low and stable power over the two days. MPC controls the HVAC load more dynamically. With RL control, HVAC systems in the residential building and the bakery shop consume more energy by 174.4% and 21.2%, respectively, than that with MPC, but the HVAC system in the ice cream shop controlled by RL consumes 56.4% less energy than that by MPC.

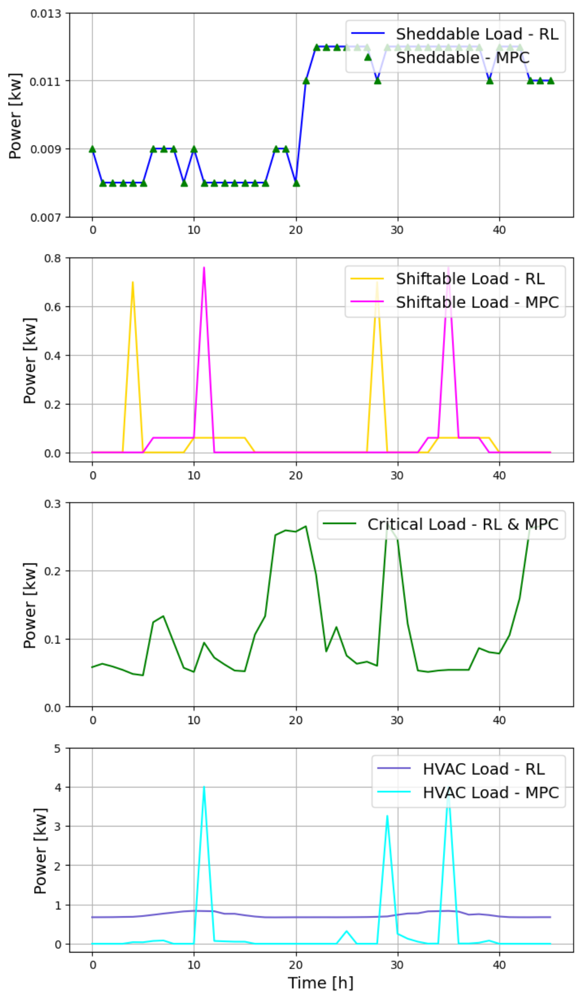

Figure 4 Residential building Loads

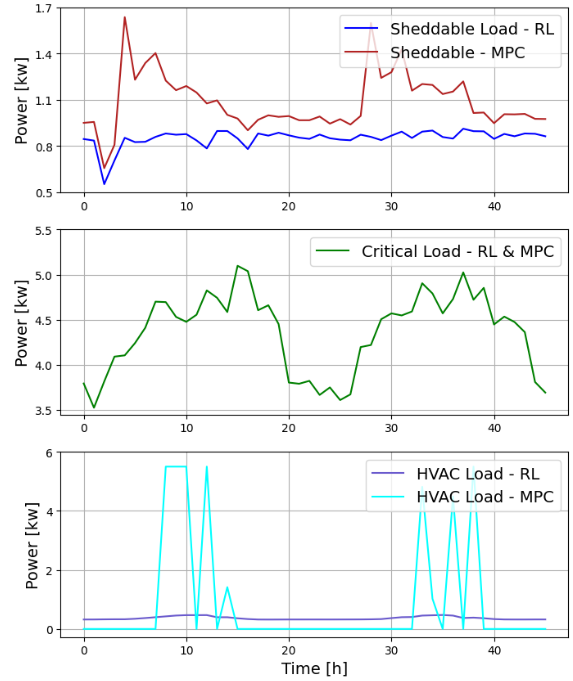

Figure 5 Ice Cream Shop Loads

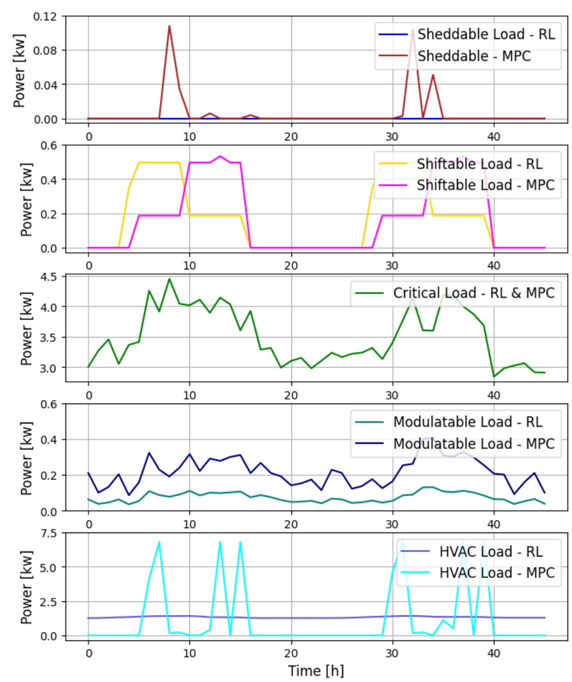

Figure 6 Bakery Loads

To further evaluate the performance of the two different controllers, we summarize the community's KPIs in Table 5. The KPIs include unmet hours for thermal comfort, PV curtailment, and required battery size. The unmet hours and PV curtailment for both MPC and RL are zero for all the buildings, which meets the control objectives perfectly. For battery size, RL outperforms MPC in terms of battery size as it requires a smaller battery size than MPC does. The battery size of the ice cream shop determined by RL is found to be much less than that by MPC. This is because the sheddable load and HVAC load controlled by RL are much smaller than that by MPC. While the indoor temperature is well maintained with the reduced HVAC load in RL, the sheddable load is not fully satisfied in RL, suggesting a further study to enhance the satisfaction of constraints in RL-based controls.

Conclusion

In this paper, we developed a multi-agent RL-based approach for coordinated control of PV power distribution and load scheduling across multifunctional buildings in a resilient community in an island mode. During a two-day assessment of community operations under two control methods (RL and MPC), the RL controller demonstrated comparable outcomes to the MPC. Across various building loads, RL showed similar load scheduling with MPC while satisfying occupancy’s essential load. RL can also effectively prevent batteries from overcharging and over-discharging without the formulation of an optimization problem with constraints used in MPC. While the operation patterns for HVAC loads differed between RL and MPC, both successfully maintained comfortable indoor temperatures across multifunctional buildings with varying construction materials and outdoor weather changes. Therefore, the proposed model-free RL-based approach has been demonstrated to achieve a similar control performance with the model-based MPC method for resilient communities. One of the advantages of the RL is that it does not need the development of system models and formulation of optimization problems as in MPC, which may require expert knowledge and customized design for each building, which suggests a better scalability and generalizability to be deployed in scale. This suggests great potential of using model-free RL to enhance the resilience of communities. The limitation of this paper is mainly on two aspects: short evaluation time and implementation challenges. Future research will extend the training and testing dataset to months to strengthen the efficacious model-free approach. With more data, split the dataset for training and testing to verify the generalization ability of the model-free approach; Implementing the RL model to a new community control system may need a long-term training process to learn the optimal control policy. We will consider applying simple rules for energy systems to shorten the warm-up period. Additionally, we have also been considering applying transfer learning to reduce the training time, i.e., training the RL agent for the target building by taking advantage of the knowledge learned by agents with a source building. In addition to the short training period, the model-free RL algorithms should be improved to better satisfy the constraints to make this approach more reliable and robust for practical uses.

Acknowledgments

This research is partially supported by National Science Foundation under Award No. CNS-2241361.

References

Bui, V.-H., Hussain, A., & Kim, H.-M. (2019). Double Deep Q-Learning-Based Distributed Operation of Battery Energy Storage System Considering Uncertainties. IEEE Transactions on Smart Grid, 1–https://doi.org/10.1109/tsg.2019.2924025

Table 5 Detailed KPIs for each building and community

	KPI	Residential Building	Ice Cream Shop	Bakery Shop	Community
MPC	Unmet hour for thermal comfort (h)	0	0	0	0
	PV Curtailment (kw)	0	0	0	0
	Battery Size	35.49	89.83	76.16	201.47
RL	Unmet hour for thermal comfort (h)	0	0	0	0
	PV Curtailment (kw)	0	0	0	0
	Battery Size	41.66	43.54	81.57	166.77

Bukar, Abba Lawan, et al. “A Rule-Based Energy Management Scheme for Long-Term Optimal Capacity Planning of Grid-Independent Microgrid Optimized by Multi-Objective Grasshopper Optimization Algorithm.” Energy Conversion and Management, vol. 221, Oct. 2020, p. 113161, https://doi.org/10.1016/j.enconman.2020.113161. Accessed 3 Feb. 2021.

Cai, S., Xie, Y., Wu, Q., & Xiang, Z. (2020). Robust MPC-based microgrid scheduling for resilience enhancement of distribution system. International Journal of Electrical Power & Energy Systems, 121, 106068–106068. https://doi.org/10.1016/j.ijepes.2020.106068

Elsayed, M., Melike Erol-Kantarci, Burak Kantarci, Wu, L., & Li, J. (2020). Low-Latency Communications for Community Resilience Microgrids: A Reinforcement Learning Approach. 11(2), 1091–1099. https://doi.org/10.1109/tsg.2019.2931753

Fang, X., Zhao, Q., Wang, J., Han, Y., & Li, Y. (2021). Multi-agent Deep Reinforcement Learning for Distributed Energy Management and Strategy Optimization of Microgrid Market. Sustainable Cities and Society, 74, 103163. https://doi.org/10.1016/j.scs.2021.103163

Henze, G. P. (2013). Model predictive control for buildings: a quantum leap? Journal of Building Performance Simulation, 6(3), 157–158. https://doi.org/10.1080/19401493.2013.778519

Huang, Q., Huang, R., Hao, W., Tan, J., Fan, R., & Huang, Z. (2019). Adaptive Power System Emergency Control using Deep Reinforcement Learning. IEEE Transactions on Smart Grid, 1–1. https://doi.org/10.1109/tsg.2019.2933191

Kim, J.-G., & Lee, B. (2020). Automatic P2P Energy Trading Model Based on Reinforcement Learning Using Long Short-Term Delayed Reward. Energies, 13(20), 5359. https://doi.org/10.3390/en13205359

Milofsky, Carl. "Resilient Communities in Disasters and Emergencies: Exploring their Characteristics." Societies 13.8 (2023): 188.

Misra, S., Line Roald, Vuffray, M., & Chertkov, M. (2017). Fast and Robust Determination of Power System Emergency Control Actions. ArXiv (Cornell University).

Mocanu, E., Mocanu, D. C., Nguyen, P. H., Liotta, A., Webber, M. E., Gibescu, M., & Slootweg, J. G. (2019). On-Line Building Energy Optimization Using Deep Reinforcement Learning. IEEE Transactions on Smart Grid, 10(4), 3698–3708. https://doi.org/10.1109/tsg.2018.2834219

Ojeda, J. M. (2023, March 7). Applied Reinforcement Learning VI: Deep Deterministic Policy Gradients (DDPG) for Continuous…. Medium. https://towardsdatascience.com/applied-reinforcement-learning-vi-deep-deterministic-policy-gradients-ddpg-for-continuous-dad372f6cb1d

Qiu, D., Ye, Y., Papadaskalopoulos, D., & Strbac, G. (2021). Scalable coordinated management of peer-to-peer energy trading: A multi-cluster deep reinforcement learning approach. Applied Energy, 292, 116940. https://doi.org/10.1016/j.apenergy.2021.116940

Rasheed, M. B., Javaid, N., Arshad Malik, M. S., Asif, M., Hanif, M. K., & Chaudary, M. H. (2019). Intelligent Multi-Agent Based Multilayered Control System for Opportunistic Load Scheduling in Smart Buildings. IEEE Access, 7, 23990–24006. https://doi.org/10.1109/access.2019.2900049

Sutton, R. S., & Barto, A. (2018). Reinforcement learning : an introduction. The Mit Press.

Yan, M., He, Y., Shahidehpour, M., Ai, X., Li, Z., & Wen, J. (2019). Coordinated Regional-District Operation of Integrated Energy Systems for Resilience Enhancement in Natural Disasters. IEEE Transactions on Smart Grid, 10(5), 4881–4892. https://doi.org/10.1109/tsg.2018.2870358

Yang, J. J., Yang, M., Wang, M. X., Du, P. J., & Yu, Y. X. (2020). A deep reinforcement learning method for managing wind farm uncertainties through energy storage system control and external reserve purchasing. International Journal of Electrical Power & Energy Systems, 119, 105928. https://doi.org/10.1016/j.ijepes.2020.105928

Yu, L., Xie, W., Xie, D., Zou, Y., Zhang, D., Sun, Z., Zhang, L., Zhang, Y., & Jiang, T. (2020). Deep Reinforcement Learning for Smart Home Energy Management. IEEE Internet of Things Journal, 7(4), 2751–2762. https://doi.org/10.1109/JIOT.2019.2957289

Zhang, B., Dehghanian, P., & Kezunovic, M. (2019). Optimal Allocation of PV Generation and Battery Storage for Enhanced Resilience. IEEE Transactions on Smart Grid, 10(1), 535–545. https://doi.org/10.1109/tsg.2017.2747136

Zhang, F., Zhang, D., Luo, F., Gianluca Ranzi, & Xu, Y. (2019). Resilient Energy Management for Residential Communities under Grid Outages. 2019 9th International Conference on Power and Energy Systems (ICPES). https://doi.org/10.1109/icpes47639.2019.9105455

Zhou, Z., Wu, Z., & Jin, T. (2021). Deep reinforcement learning framework for resilience enhancement of distribution systems under extreme weather events. International Journal of Electrical Power & Energy Systems, 128, 106676. https://doi.org/10.1016/j.ijepes.2020.106676

Wang, J., Garifi, K., Baker, K., Zuo, W., Zhang, Y., Huang, S., & Vrabie, D. (2020). Optimal Renewable Resource Allocation and Load Scheduling of Resilient Communities. Energies, 13(21), 5683. https://doi.org/10.3390/en13215683

Wang, J., Zuo, W., Rhode-Barbarigos, L., Lu, X., Wang, J., & Lin, Y. (2019). Literature review on modeling and simulation of energy infrastructures from a resilience perspective. Reliability Engineering & System Safety, 183, 360–373. https://doi.org/10.1016/j.ress.2018.11.029