Two-Stage Deep Reinforcement Learning for Distribution System Voltage Regulation and Peak Load Management

Preprint

Yansong Pei,[1] Yiyun Yao,[2] Junbo Zhao,[1] Fei Ding,[2] and Jiyu Wang[2]

1 University of Connecticut
2 National Renewable Energy Laboratory

Presented at the 2023 IEEE Power and Energy Society General Meeting
Orlando, Florida
July 16–20, 2023

NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Contract No. DE-AC36-08GO28308

Conference Paper
NREL/CP-5D00-84637
July 2023

Two-Stage Deep Reinforcement Learning for Distribution System Voltage Regulation and Peak Load Management

Preprint

Yansong Pei,[1] Yiyun Yao,[2] Junbo Zhao,[1] Fei Ding,[2] and Jiyu Wang[2]

1 University of Connecticut
2 National Renewable Energy Laboratory

Suggested Citation
Pei, Yansong, Yiyun Yao, Junbo Zhao, Fei Ding, and Jiyu Wang. 2023. *Two-Stage Deep Reinforcement Learning for Distribution System Voltage Regulation and Peak Load Management: Preprint*. Golden, CO: National Renewable Energy Laboratory. NREL/CP-5D00-84637. https://www.nrel.gov/docs/fy23osti/84637.pdf.

NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Contract No. DE-AC36-08GO28308

Conference Paper
NREL/CP-5D00-84637
July 2023

National Renewable Energy Laboratory
15013 Denver West Parkway
Golden, CO 80401
303-275-3000 • www.nrel.gov

NOTICE

This work was authored in part by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by U.S. Department of Energy Office of Energy Efficiency and Renewable Energy Solar Energy Technologies Office Agreement Number 37770. The views expressed herein do not necessarily represent the views of the DOE or the U.S. Government.

Cover Photos by Dennis Schroeder: (clockwise, left to right) NREL 51934, NREL 45897, NREL 42160, NREL 45891, NREL 48097, NREL 46526.

Two-Stage Deep Reinforcement Learning for Distribution System Voltage Regulation and Peak Demand Management

Yansong Pei *Student Member, IEEE*, Yiyun Yao, Junbo Zhao, *Senior Member, IEEE*, Fei Ding, Jiyu Wang

***Abstract*—The growing integration of distributed solar photovoltaic (PV) in distribution systems could result in adverse effects during grid operation. This paper develops a two-agent soft actor critic-based deep reinforcement learning (SAC-DRL) solution to simultaneously control PV inverters and battery energy storage systems for voltage regulation and peak demand reduction. The novel two-stage framework, featured with two different control agents, is applied for daytime and nighttime operations to enhance control performance. Comparison results with other control methods on a real feeder in Western Colorado demonstrate that the proposed method can provide advanced voltage regulation with modest active power curtailment and reduce peak load demand from feeder's head.**

***Index Terms*—Deep reinforcement learning, distribution system, voltage regulation, peak load management.**

I. Introduction

Increasing penetrations of solar photovoltaic (PV) systems in the active distribution network (ADN) raise concerns about system voltage quality [1]. With the development of smart inverter technology, PV inverters can simultaneously regulate active and reactive power, which can be involved as a non-wire option of distributed volt-var control(VVC) [2]. In addition, a battery energy storage system (BESS) can be deployed along with PV to shift excess PV power for supplying system load during peak hours [3]. Consequently, it is of significance to develop a control solution to simultaneously regulate PV and BESS for grid operation enhancement.

Various methods have been proposed to coordinate the control of PV and BESS. [4] proposes a battery energy management system to control the generator and BESS. Different modes of operation are listed for controlling the charge/discharge rate of the batteries. In [5], a demand-side management approach is proposed based on a fuzzy logic method to incorporate PV, wind turbines, and BESS. These approaches mainly focus on peak demand shaving without considering the voltage violation problem. A rule-based control strategy proposed in [6] achieves peak shaving control using grid-connected PV systems with BESS. It also demonstrates that the voltage violation and peak shaving problem can be simultaneously solved through proper coordination; however, when faced with a system with a massive number of PV and BESS, it faces high complexity in rules setting. Furthermore, accurate model information is required to design the control policy, which is hard to acquire in practice.

Recently, reinforcement learning as a model-free control method is being increasingly implemented, especially deep reinforcement learning (DRL)-based approaches, such as an attention-enabled multi-agent DRL (MADRL) [7], MADRL for a realistic distribution system [8], a multi-agent deep deterministic policy gradient algorithm-based algorithm [9]. Compared with local autonomous volt-var control [10] without communication between each inverter and optimal power flow methods [11], DRL-based approaches do not rely on voltage-reactive power piecewise curve or accurate model information. These algorithms coordinate the PV inverters and reduce voltage violations. A certain amount of PV active power is curtailed without considering the use of BESS, which could have been stored and used during the peak load time. There are studies using DRL for energy management, such as capacity scheduling of PV-BESS using proximal policy optimization (PPO) [12], MADRL for home energy management [13], and a two-level scheduling algorithm using DRL [14]. These methods have successfully used the battery to solve some economic problems by charging the batteries at low electricity price times and discharging them during peak load periods to reduce the users' cost.

This paper proposes a novel SAC-DRL solution for the coordinated control of PV inverters and BESS. The objective is to minimize the voltage violation while maintaining low PV active power curtailment and achieving an effective peak demand reduction. The contributions of this paper are:

- The proposed SAC-DRL is trained by a novel time variant reward, and the ADN can coordinate the control of a high number of PV inverters and BESS to simultaneously achieve voltage regulation and peak demand reduction.
- The proposed method requires little or no prior knowledge of ADN information. Comprehensive experiments on a real distribution feeder in western Colorado demon-

This work was authored by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by the U.S. Department of Energy Office of Energy Efficiency and Renewable Energy Solar Energy Technologies Office Agreement Number 37770. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

Y. Pei and J. Zhao are with the Department of Electrical and Computer Engineering, University of Connecticut, Storrs, CT 06269 USA. Y. Yao, F. Ding and J. Wang are with National Renewable Energy Laboratory, Golden, CO 80401, USA (e-mail: junbo@uconn.edu).

strate that the proposed solution has better performance than rule-based and other DRL-based control methods.

II. Problem Formulation

A. Distribution System Model

Consider a distribution system with n nodes denoted by the set $\mathcal{N}$:=1, .., n and branches by the set $\mathcal{M}$:=1, .., m. In the system, there are *E* nodes denoted by the set $\mathcal{E} \subseteq \mathcal{N}$ that have a voltage magnitude meter installed. For node $i \in \mathcal{E}$, define v_i^t as the voltage magnitude at *t* moment. There are *H* nodes denoted by the set $\mathcal{H} \subseteq \mathcal{N}$ that have PV with smart inverters deployed. For $i \in \mathcal{H}$, define the PV set points as $x_i^t := (P_i^t, Q_i^t)$, where *t* represents the time step. There are *B* nodes denoted by the set $\mathcal{B} \subseteq \mathcal{N}$ that have BESS deployed. For $i \in \mathcal{B}$, define BESS actions as $b_i^t := (P_{i,c}^t, P_{i,d}^t)$, where represents power charge/discharge rate of batteries. There are *L* nodes denoted by the set $\mathcal{L} \subseteq \mathcal{N}$ that have load demand PL_i^t, QL_i^t. Let PL_i^t, QL_i^t denote the real and reactive power load on node *i* at time *t*. There are *ev* nodes denoted by the set $\mathcal{EV} \subseteq \mathcal{N}$ that have electric vehicle (EV) deployed. For $i \in \mathcal{EV}$, define EV $ev_i^t := (PL_{ev}^t)$,where $PL_{ev}^t \in PL_i^t$, represents power demand from electric vehicles. In this paper, the EV is treated as a normal load, which only has an active power load. The power flow constraints are:

$$V_i^t (I_i^t)^* = (P_i^t - PL_i^t) + j(Q_i^t - QL_i^t), \forall i \in \mathcal{H}, \quad (1)$$

$$V_i^t (I_i^t)^* = -PL_i^t - jQL_i^t, \forall i \in \mathcal{N}/\mathcal{H}, \quad (2)$$

where V_i and I_i are the node's voltage magnitude and current injection. Throughout the operation of the distribution grid, the nodes equipped with voltage magnitude meters are required to remain within a predefined range. Any node voltage magnitude with greater than 1.05 p.u. or less than 0.95 p.u. will be counted as voltage violation nodes (VVNs). N_{vvn} stands for the total number of VVNs, yielding

$$0.95 \leqslant |V_i^t| \leqslant 1.05, \forall i \in \mathcal{N}/N_{vvn}, \quad (3)$$

$$V_n^t \leqslant 0.95 \ or \ V_n^t \geqslant 1.05, \forall n \in N_{vvn}, \quad (4)$$

B. PV Inverter Model

For each PV inverter, the power set point $x_i^t := (P_i^t, Q_i^t)$ is constrained, i.e, $x_i^t \in \mathcal{RE}_i^t$ for $\forall i \in \mathcal{H}$. The region $\mathcal{RE}_i^t$ is determined by the apparent power capacity and time-vary solar irradiance, λ_t. As described in California Rule 21 [15], the full reactive power, Q_i^t capability range is defined as 30% of the nameplate apparent power rating. Then, the region, $\mathcal{RE}_i^t$, can thus be defined as:

$$\mathcal{RE}_i^t = \{0 \leq P_i^t \leq P_{i,max}^t, \ -0.3S_i \leq Q_i^t \leq 0.3S_i \quad (5)$$

where $P_{i,max}^t = \lambda_t \times S_i$; $P_{i,max}^t$ is the maximum real power of the ith PV inverter, and S_i is the corresponding nameplate apparent power rating.

C. BESS Model

The dynamic model of the BESS can be modeled as the following discrete time equation:

$$\sigma_i^{t+1} = \sigma_i^t + (\eta_{i,c} P_{i,c}^t - \frac{1}{\eta_{i,d}} P_{i,d}^t), \quad (6)$$

where σ is the BESS state of charge (SOC); and $\eta_{i,c}$,$\eta_{i,d}$ are the charging and discharging efficiencies, which describe that the variation of the SOC change is proportional to the power of the charging and discharging, $P_{i,c}^t$ and $P_{i,d}^t$, for each BESS. There is a limit on the charging/discharging action subject to:

$$P_{i,c}^t \cdot P_{i,d}^t = 0, \quad (7)$$

which means that the charging and discharging action cannot happen on the same BESS at each time step. The SOC of the BESS is subject to the upper and lower boundaries:

$$\sigma_i^{min} \leq \sigma_i^t \leq \sigma_i^{max}. \quad (8)$$

For each BESS, there is the upper limit power of the charging and discharging constrained as:

$$0 \leq P_{i,c}^t \leq P_{i,c}^{max}, \forall i \in \mathcal{B}, t \quad (9)$$

$$0 \leq P_{i,d}^t \leq P_{i,d}^{max}, \forall i \in \mathcal{B}, t, \quad (10)$$

where $P_{i,c}^{max}$ and $P_{i,d}^{max}$ are the rated charging and discharging power of the BESS.

D. Power Demand at Feeder Head

The power balance of the system is constrained according to the following equation:

$$P_{head}^t = P_{loss}^t + \sum_{i=1}^{\mathcal{EV}} PL_i^t + \sum_{i=1}^{\mathcal{L}} PL_i^t - \sum_{i=1}^{\mathcal{H}} P_i^t - \sum_{i=1}^{\mathcal{B}} P_{i,d}^t + \sum_{i=1}^{\mathcal{B}} P_{i,c}^t, \quad (11)$$

where P_{loss}^t is the active power loss in the system; P_{head}^t is the power demand from the power feeder to compensate for the insufficient active power. Assume that the power demand from the feeder head is divided into two time periods: peak load time demand, $P_{head}^{t,peak}$ when $T \in (18, 24)$ and normal time demand, $P_{head}^{t,normal}$ when $T \in (1, 17)$. T represents the time in the real world (e.g., 17 indicates 17:00 p.m.).

In summary, the coordinated control of PV inverter and BESS can be formulated as the following optimal power flow (OPF) problem:

$$max : \sum_{i=1}^{|H|} f_i^t(P_i^t) - N_{vvn} - P_{head}^{t,peak}, \quad (12a)$$

$$subject \ to : P_i^t \in \mathcal{RE}_i^t, \forall i \in \mathcal{H}, \quad (12b)$$

$$(1) - (9). \quad (12c)$$

As (12) is a nonconvex problem that has multiple objectives, the proposed method should provide the PV inverters' set points and BESS actions to achieve a sub-optimal operation condition with a fast response. Further, the high PV penetration means a high number of control targets, which is hard to solve by traditional centralized or distributed approaches. This paper proposes the SAC-DRL approach to address them.

III. Proposed Two-Stage SAC-DRL Control Solution

A. Formulation of Markov Decision Process

The coordinated control of the PVs' active and reactive power set points and the BESS actions to regulate the voltage and reduce the peak load demand is formulated as a Markov decision process (MDP). The MDP comprises the environment, agents, observation, action, and reward, which are described as follows:

- **Environment**: An ADN, including the time-varying load profile. The PV real and reactive power set points and BESS charging/discharging power will be the input, and the output is the voltage of each node and power demand from the feeder head, which can be formulated as the following equation:

$$g(P_i^t, Q_i^t, PL_i^t, QL_i^t, P_{bess,d}^t, P_{bess,c}^t) \rightarrow V_m^t, P_{head}^t, m \in \mathcal{E} \tag{13}$$

  In this paper, the load shape, battery action, and PV set points will be fed into the simulation software as OpenDSS and the power demand and voltage are obtained after the simulation.
- **Agent**: The central controller of the system. The agent is responsible for controlling the PV inverter set points and battery actions. In the MDP, the agent makes the decision, A_t, based on the observation, S_t, at the t^{th} time step.
- **Observation**: The information observed by the agent. In this MDP, the agent will observe the time, T, PV maximum generation, $P_{i,max}^t$, the maximum reactive power capacity, $Q_{i,max}^t$, the load and EV information, PL_i^t, QL_i^t, and the SOC for the BESS. The set, S_t, including this information will be used for an agent to make decision A_t
- **Action**: The action set, A_t, includes all PV inverter set points and the BESS charging/discharging actions. For each PV inverter, $i \in \mathcal{H}$, the action is defined as $(\alpha_{PV,P}(i,t), \alpha_{PV,Q}(i,t))$, where $\alpha_{PV,P}(i,t) \in (0,1)$ and $\alpha_{PV,Q}(i,t) \in (-1,1)$. The PV set points in Eq.(5) can be calculated by the following equation: $P_i^t = \alpha_{PV,P}(i,t) \times P_{i,max}^t, Q_i^t = \alpha_{PV,Q}(i,t) \times 0.3S_i^t$. For each BESS, $i \in \mathcal{B}$, the action given by the agent is $\beta_i^t \in (-1,1)$. The BESS charging/discharging power is calculated by the following equation:

$$P_{i,c}^t = \beta_i^t \cdot P_{i,c}^{max}, P_{i,d}^t = 0, if \beta_i^t \geq 0 \tag{14a}$$

$$P_{i,d}^t = \beta_i^t \cdot P_{i,d}^{max}, P_{i,c}^t = 0, if \beta_i^t \leq 0 \tag{14b}$$

- **Reward**: R_t obtained after action A_t executed under the condition of S_t. Considering the different control strategies required for different time periods, two innovative reward functions were designed for training in this paper, as follows:

$$R_{day} = \gamma \sum_{i=1}^{n} v_{i,violation} + \varepsilon Pc^t + \rho \sum_{i=1}^{\mathcal{B}} P_{i,c}^t, \tag{15a}$$

$$R_{night} = \gamma \sum_{i=1}^{n} v_{i,violation} + \mu \sum_{i=1}^{\mathcal{B}} P_{i,d}^t, \tag{15b}$$

$$v_{i,violation} = (1 - min(\delta - |1 - v_i^t|, 0))^2 - 1, \tag{15c}$$

$$Pc^t = \frac{\sum_{i \in \mathcal{H}} P_i^t}{\sum_{i \in \mathcal{H}} P_{i,max}^t}, \tag{15d}$$

  where γ is the penalty coefficient of the voltage violation calculated by (15a); ε is the penalty coefficient of the PV active power curtailment according to the PV set points; ρ is the reward coefficient for the battery charging during the daytime, which encourages the BESS charging to reduce the curtailment and store the energy for peak time use; μ is the reward coefficient for the battery discharging during the nighttime, where μ has two different values depending on different T values to encourage a large amount of discharge during the peak period and a small amount of discharge during late nighttime; δ is the threshold used to optimize the voltage barrier function and Pc_t is the PV active power generation rate used to punish the curtailment.

B. SAC-Based DRL Agent

DRL is the process of trial and error to obtain a higher reward. During this process, the neural network of the agent constantly updates itself by iteratively adjusting the coefficient and weights along gradients of higher rewards. Actor-critic-based reinforcement learning, as an advanced algorithm, has one actor-network and one critic network. The actor takes the observation as the input and outputs the action accordingly, and the critic takes the environment observation with the actor's action as input and makes an assessment of the action along with the direction of how much to adjust. As the iteration progresses, the actions made by the actor will gain increasingly higher rewards, and the critic's state value estimation will be more accurate; hence, compared with other DRL methods, the actor-critic-based method has the characteristics of fast convergence and good performance. In this paper, the agent is trained and updated by using the off-policy SAC algorithm [16]. The actor-network in the SAC outputs the action by following the policy whose purpose is to maximize the sum of the reward, $R(S_t, A_t)$, and the entropy of the policy, $H(\pi(\cdot \mid s_t))$. There are three networks in the proposed SAC-based approach: two soft Q-function networks parameterized by θ, and a policy function network, π, parameterized by ϕ. The actor-network is a policy equation shown as follows:

$$\mathcal{J}(\pi) = argmax\ E[\sum_{t=0}^{\infty} \omega^t (R(s_t, a_t) + \alpha \times H(\pi(\cdot \mid s_t))], \tag{16}$$

where ω is the future discount coefficient; α is a temperature parameter that indicates the entropy's contribution to the reward. The α will initially be designed as a large value to obtain higher entropy rewards by increasing the exploration space of action. As the training proceeds, α will gradually decrease and shrink the exploration breadth, eventually approaching the

optimal policy. The critic network estimates the state value as the Q-function as:

$$y(S_t, R_t, S_{t+1}) = r + \omega(Q(S_{t+1}, A_{t+1} - \alpha log \pi_\theta(A_{t+1} \mid S_{t+1})), \tag{17}$$

The SAC algorithm relies on an experience replay buffer to update with enhancing sample efficiency. After the reward is obtained by the executed action, the replay buffer stores the observation, action, reward, and next step observation as a transition. A batch of transitions, $B = \{(S_t, A_t, S_{t+1}, R_t)\}$, will be randomly selected to update the neural network. The actor-network updates the coefficient using a gradient ascent by the following:

$$\nabla\theta_i \frac{1}{|B|} \sum_{((s_t))\in B} (\min_{i=1,2} Q_{\theta_i}(s_t, \pi(\cdot \mid s_t)) - \alpha \log \pi_\theta(\pi(\cdot \mid s_t) \mid s_t)), \tag{18}$$

the critic network updates the Q-function using gradient descent by the following:

$$\nabla\theta_i \frac{1}{|B|} \sum_{((s_t, a_t, s_{t+1}, r_t))\in B} (Q_{\theta_i}(s_t, a_t) - y(s_t, r_t, s_{t+1}))^2, \tag{19}$$

where the clipped double-Q method is used to obtain the smaller Q-value between the two Q approximators.

C. Two-Agent Control Solution

A single-agent SAC has enough capability to control a good number of PVs and BESSs during the daytime. During the training process, the agent constantly adjusts the PV active power set points in the neural network; however, setting the PV active power production at nighttime is meaningless since there is no solar. The update on the weight and bias during the nighttime will not reflect any reward change, which could mislead the update of the SAC agent. In addition, the agent seeks a balance among the active power curtailment, battery changing, and voltage violation during the daytime. When nighttime comes, the reward requires the agent to concentrate on solving the peak load demand reduction problem. The single-agent approach will be disturbed by the time-variance reward, resulting in performance degradation. Considering the generalization ability of the proposed solution to achieve 24-hour control, agents at two-stage are used to control different operating scenarios during the daytime and nighttime with different dimensions of actions. The agent applied in the daytime stage will be trained by following (15a) while that for the nighttime stage will be trained by following the (15b). The coordinated PV-BESS control based on the proposed two-agent layout is shown in Fig. 1.

IV. Results on A Realistic Distribution System

The proposed two-stage DRL-SAC solution is tested on a real feeder in Western Colorado. There are 759 nodes, 159 loads, 65 EVs, 95 BESS, and 112 PV units in the original model. The training process uses 28 days of historical data

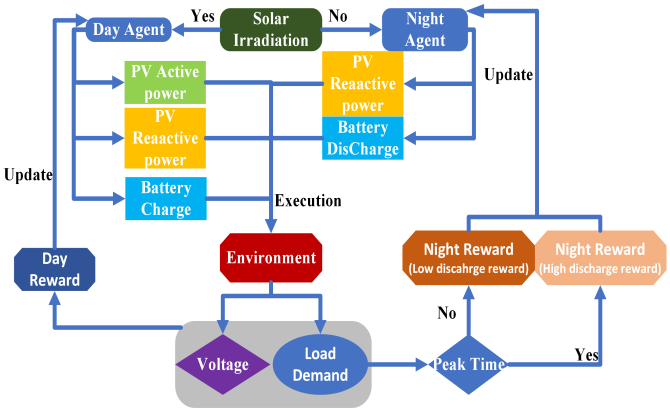

Fig. 1: Flowchart of the proposed method.

with a 1-hour time resolution. The DRL agents are trained using PyTorch 1.8.1. The test process using 7 days of hourly data is taken in OpenDSSDirect under Python 3.8.

Several existing methods are used to compare the proposed method to highlight the advantages, including 1) **No-control**: The PV inverter will produce the maximum active power according to solar irradiation. There are no reactive power set points of the PV or BESS actions. 2) **VVC-volt priority+ruled-based control** (VVCV+RBC): The VVCV controls the PV reactive power value requested by the volt-VAR curve to be limited until the desired voltage is achieved, which will curtail the active power. The BESS is controlled by following a rule of charging at 25% of maximum charging power during the daytime and discharging at 30% of maximum discharging power during the peak load period. The BESS discharges at 5% of maximum discharging power during the night. 3) **Single deep deterministic policy (DDPG)** [17]: The single DDPG agent will be trained by the same set of two reward functions and controls the PV inverter with the BESS. 4) **Single-SAC**: The single SAC agent trained by the same settings of the two reward functions is implemented to simultaneously control the PV inverter and BESS.

TABLE I: 7-Day simulation with different control scenarios

Agent	N_{vvn}	Peak(kW)	Max Volt	Min Volt	Curt
No-control	86	2.61×10^5	1.059	0.931	-
VVCW+RBC	26	2.26×10^5	1.053	0.953	4.6%
DDPG	42	2.61×10^5	1.052	0.928	2.3%
SAC	2	1.88×10^5	1.051	0.953	1.6%
Proposed	0	1.82×10^5	1.049	0.952	0.6%

The test results are summarized in Table I and Fig. 2. It suggests that both the SAC-based and proposed approaches have better performance than other control solutions. The No-control has the most voltage violations with 86 VVNs during the 7-day test without any curtailment. Because of using the BESS, the number of VVN controlled by VVCW+RBC is reduced to 26. Since there is no cooperation between the PV and BESS, the VVCW+RBC approaches will make the control decisions earlier than BESS starts causing the active power curtailment to be 4.6%. The result of the DDPG suggests that it cannot handle the coordinated control between PVs

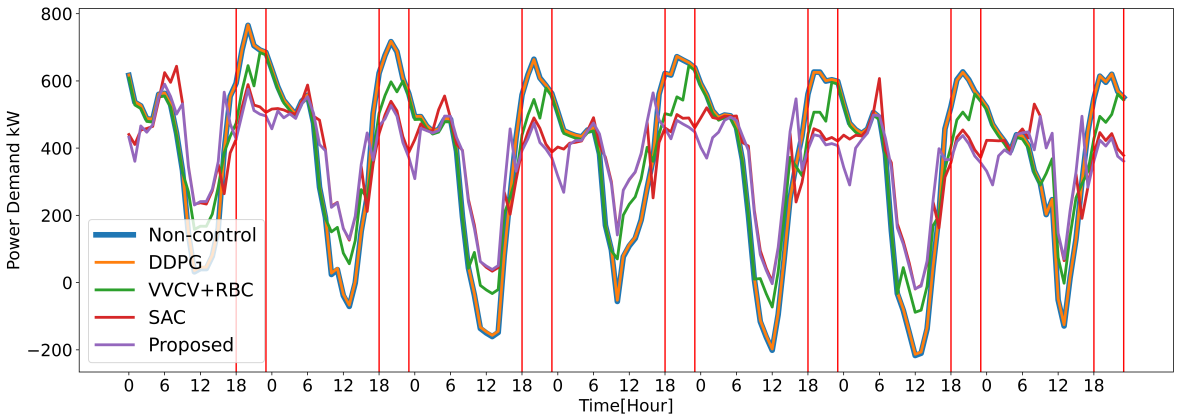

Fig. 2: Load demand from feeder's head for different approaches during 7-day testing.

and BESS well; the DDPG ignores the control of the active power and BESS, and focuses only on the reactive power set points causing the lowest minimum voltage in all approaches. The SAC as an advanced algorithm successfully achieves the coordinated control of PV and BESS, regulates voltage and reduces the number of VVN to 2 while maintaining active power curtailment at 1.6%. The proposed approach has the best performance with no voltage violation and the curtailment is only 0.6%.

Fig. 2 shows the 7-day load demand from the feeder head, where the regions between two adjacent red lines are the peak load period. The no-control method as the baseline shows that during the daytime, especially between 10 am and 2 pm when the solar irradiance is the strongest, the PV can generate a large amount of real power. At the noon on days 3, 5, and 6, it can even generate excess real power needed by the system. The DDPG agent has a similar load demand curve with No-control. Regular charging and discharging using the ruled-based control method relieves the pressure on the feeder head of the peak load period and reduces the load demand by 13%. By contrast, the SAC and proposed approaches reduce the load demand by 27.9% and 30.2%, respectively. Because a two-stage agent design for handling the daytime and nighttime is developed, the agents can perform stably than using a single agent, and it can be observed that there is another peak demand caused by the unstable BESS action on the 6^{th} day morning. These results confirm that the proposed two-stage SAC-DRL-based method has better performance on voltage regulation and peak load reduction.

V. Conclusion

This paper proposes a two-stage SAC-DRL approach to achieve the coordinated control of the PV inverter and BESS to regulate the system voltage and reduce the load during the peak time. Comparative tests on a real feeder with several existing approaches demonstrate that the proposed method can address the voltage violations by curtailing modest real power and charging the BESS to store the energy for load compensation during peak hours. The new two-stage framework, featured by two different agents, is applied to deal with two different reward functions for daytime and nighttime to improve performance without system knowledge.

References

[1] S. Eftekharnejad, V. Vittal, G. T. Heydt, B. Keel, and J. Loehr, "Impact of increased penetration of photovoltaic generation on power systems," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 893–901, 2013.

[2] P. Jahangiri and D. C. Aliprantis, "Distributed volt/var control by pv inverters," *IEEE Trans. Power Systems*, vol. 28, no. 3, pp. 3429–3439, 2013.

[3] H. Alharbi and K. Bhattacharya, "Stochastic optimal planning of battery energy storage systems for isolated microgrids," *IEEE IEEE Trans. Sustainable Energy*, vol. 9, no. 1, pp. 211–227, 2018.

[4] K. Thirugnanam, S. K. Kerk, C. Yuen, N. Liu, and M. Zhang, "Energy management for renewable microgrid in reducing diesel generators usage with multiple types of battery," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 8, pp. 6772–6786, 2018.

[5] R. Khezri, A. Mahmoudi, and M. H. Haque, "A demand side management approach for optimal sizing of standalone renewable-battery systems," *IEEE Trans. Sustainable Energy*, vol. 12, no. 4, pp. 2184–2194, 2021.

[6] R. Manojkumar, C. Kumar, S. Ganguly, and J. P. S. Catalão, "Optimal peak shaving control using dynamic demand and feed-in limits for grid-connected pv sources with batteries," *IEEE Systems Journal*, vol. 15, no. 4, pp. 5560–5570, 2021.

[7] D. Cao, J. Zhao, W. Hu, F. Ding, Q. Huang, and Z. Chen, "Attention enabled multi-agent drl for decentralized volt-var control of active distribution system using pv inverters and svcs," *IEEE Trans. Sustain. Energy.*, vol. 12, no. 3, pp. 1582–1592, 2021.

[8] Y. Pei, Y. Yao, J. Zhao, F. Ding, and J. Wang, "Multi-agent deep reinforcement learning for realistic distribution system voltage control using pv inverters," in *2022 IEEE Power Energy Society General Meeting (PESGM)*, 2022, pp. 1–5.

[9] D. Cao, W. Hu, J. Zhao, Q. Huang, Z. Chen, and F. Blaabjerg, "A multi-agent deep reinforcement learning based voltage regulation using coordinated pv inverters," *IEEE Trans. Power Systems*, vol. 35, no. 5, pp. 4120–4123, 2020.

[10] A. M. Howlader, S. Sadoyama, L. R. Roose, and S. Sepasi, "Distributed voltage control method using volt-var control curve of photovoltaic inverter for a smart power grid system," in *IEEE 12th International Conference on PEDS*, 2017, pp. 630–634.

[11] Y. Yao, F. Ding, K. Horowitz, and A. Jain, "Coordinated inverter control to increase dynamic pv hosting capacity: A real-time optimal power flow approach," *IEEE Syst. J*, pp. 1–12, 2021.

[12] B. Huang and J. Wang, "Deep-reinforcement-learning-based capacity scheduling for pv-battery storage system," *IEEE Transactions on Smart Grid*, vol. 12, no. 3, pp. 2272–2283, 2021.

[13] X. Xu, Y. Jia, Y. Xu, Z. Xu, S. Chai, and C. S. Lai, "A multi-agent reinforcement learning-based data-driven method for home energy management," *IEEE Trans. Smart Grid*, vol. 11, no. 4, pp. 3201–3211, 2020.