

Fair Reinforcement Learning Algorithm for PV Active Control in LV Distribution Networks

Maurizio Vassallo*, Amina Benzerga*, Alireza Bahmanyar*, Damien Ernst*[†]

* Department of Electrical Engineering and Computer Science, Liège, Belgium

[†] LTCI, Télécom Paris, Institut Polytechnique de Paris, France

{mvassallo, abenzerga, abahmanyar, dernst}@uliege.be

Abstract—The increasing adoption of distributed energy resources, particularly photovoltaic (PV) panels, has presented new and complex challenges for power network control. With the significant energy production from PV panels, voltage issues in the network have become a problem. Currently, PV smart inverters (SIs) are used to mitigate the voltage problems by controlling their active power generation and reactive power injection or absorption. However, reducing the active power output of PV panels can be perceived as unfair to some customers, discouraging future installations. To solve this issue, in this paper, a reinforcement learning technique is proposed to address voltage issues in a distribution network, while considering fairness in active power curtailment among customers. The feasibility of the proposed approach is explored through experiments, demonstrating its ability to effectively control voltage in a fair and efficient manner.

Index Terms—Distributed Generation, Fair Curtailment, Low-Voltage Distribution Network, Reinforcement Learning, Voltage Control

I. INTRODUCTION

The increase in the global energy consumption, the unstable cost of electricity and the goal of reducing CO₂ emissions are among the most important reasons for vast integration of distributed energy resources (DERs) throughout distribution networks.

However, due to the decentralised production of electricity, grid management has become more complex, demanding more sophisticated control methods to guarantee grid reliability. This presents a challenge for a distribution system operator (DSO), who needs to ensure that power networks operate within their operational limits. One of the outcomes of the widespread integration of DERs is the occurrence of reverse power flow that occurs when the energy generated exceeds the energy consumed. This reverse power flow can result in certain voltage-related problems. Voltage management in electricity distribution networks has been widely studied in the literature. In [1], the authors use a fuzzy logic algorithm to control the tap position of the transformers. Although this solution may be effective for traditional distribution networks, its implementation in DER-dominated networks is impractical due to the unpredictable and fluctuating nature of the DERs' output. This would require frequent tap adjustments, resulting in high maintenance cost. Another possible solution would be

to use the flexibility of DERs and controllable loads. For example, using an optimal power flow (OPF) technique, it is possible to optimise the operation of power systems by adjusting the generation and load levels to minimise the overall cost of operation while meeting various operational constraints. This involves solving a large-scale optimisation problem that considers multiple nonlinear operational constraints and objectives, making it computationally expensive [2].

In [3], the authors propose an active power curtailment technique with different types of droop-based methods and a dynamic programming method to minimise power curtailment. Authors in [4] discuss the potential of smart inverters (SIs) to mitigate voltage and frequency deviations using active power curtailment, volt-watt control, and frequency-watt control. The active power curtailment control can be one of the most effective controls for voltage issues. However, this technique reduces the energy output of the DERs to the grid, resulting in economic drawbacks for customers. For this reason, some works have preferred to control the reactive power instead. Paper [5] presents a reactive power-based control strategy for single-phase PV inverters to improve voltage unbalance and voltage regulation in low-voltage distribution networks. Similarly, in [6], authors propose a voltage control loop that can absorb or supply reactive power to maintain voltage within acceptable bounds. However, only reactive power control is not always enough to solve the voltage problems. Consequently, some studies have focused on addressing both active and reactive control to mitigate voltage-related issues. The method proposed in [7] provides an effective voltage regulation while decreasing power losses and maximising the revenue of the customers. The authors in [8] propose a new technique for mitigating over-voltage violations in distribution networks, using short-term PV power generation forecasts to adjust active and reactive power injection from the PV inverters.

All the solutions discussed so far have tried to solve the voltage problems without dealing with fairness during curtailment. Indeed, the curtailment of DERs is dependent on their location, and customers might be curtailed unequally and unfairly [9]. In [10], the authors propose to

share the required curtailment equally among the customers. Many definitions of fairness have been introduced. Authors in [11] introduce a definition of fairness and apply it in [12] where they propose a complex analytical approach for achieving fair active power curtailment and reactive power control in low-voltage distribution networks. Simpler solutions are proposed using reinforcement learning (RL) techniques. In [13], the authors apply a soft actor-critic (SAC) based RL algorithm to address the voltage issues using three definitions for fairness. The authors consider only active power curtailment, that could lead to suboptimal solutions.

The objective of this study is to explore the implementation of a deep RL technique for managing the active and reactive power outputs of SIs in an LV distribution network. The proposed approach aims to ensure fairness among the customers while complying with the voltage regulation and minimising active power curtailment. The main benefit of this RL-based approach lies in its capacity to deliver an optimal and fair solution with a limited amount of network data requirements.

The rest of the paper is organised as follows. Section II defines the problem. Section III presents some background information about fairness, SI and RL, respectively. Section IV presents the proposed method. The results are reported in Section V and Section VI concludes the work and discusses some possible future works.

II. PROBLEM STATEMENT

A power network can be represented as a directed graph, $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, with a set of nodes, \mathcal{N} , also called buses. Directed edges, \mathcal{E} , link two buses together, also called lines. A line $e_{i,j} \in \mathcal{E}$ connects node i to node j , with $i, j \in \mathcal{N}$. Several devices, which may inject or withdraw power from the grid, might be connected to each bus. The sets of all devices, either loads \mathcal{L} or generators \mathcal{DG} , are assumed to be known.

The DSO takes into account the behaviour of the network over a set of discrete time steps, t , over a studied time horizon, T .

We denote the active and reactive power of load i at time step t as $\mathcal{L}_{i,t}^p$ and $\mathcal{L}_{i,t}^q$, with $i \in \mathcal{L}$; the active and reactive power of the PVs as $\mathcal{DG}_{i,t}^p$ and $\mathcal{DG}_{i,t}^q$ with $i \in \mathcal{DG}$; the voltage of the node i at time step t as $V_{i,t}$, with $i \in \mathcal{N}$.

The network situation, in a given time step $t \in T$, is considered critical if the voltage of at least one of the buses is outwith the boundaries: $V_{i,t} < v^{\min}$ (under-voltage) or $V_{i,t} > v^{\max}$ (over-voltage) for $\forall i \in \mathcal{N}$, where v^{\max} and v^{\min} are the maximum and the minimum acceptable voltage, respectively.

III. CONTEXTUAL BACKGROUND

A. Fairness

Active power curtailment is not desirable from an economic and environmental point of view because it implies that some energy cannot be produced, and that almost the same amount of the curtailed DER production needs to be replaced by another energy source. However, sometimes it is needed to curtail the energy production to keep the distribution network functional.

The amount of energy curtailment needed depends on many factors, such as the location of the DER in the feeder, the amount of the energy consumption and the amount of energy generation. Therefore, customers may not be curtailed equally, which might feel unfair and could be discouraging for new customers.

In this paper, we define fairness in active power curtailment as the equality in the amount of energy that each customer is being curtailed in case of a voltage issue. Fairness is evaluated using the Gini index, which it is used to test the degree of inequality during curtailment. A low Gini index indicates a more equal distribution of curtailment, while a high Gini index indicates a more unequal distribution. In particular, a Gini index of 0 reflects perfect equality, where all customers are curtailed the same, while a Gini index of 1 reflects maximal inequality among customers' curtailment [14].

B. Smart inverters

Smart inverters for PV systems are electronic devices that convert the direct current (DC) produced by solar panels into alternating current (AC). Unlike traditional inverters, smart inverters incorporate advanced features and communication capabilities that allow them to actively manage and control the flow of electricity from the PV system into the grid. This enables them to regulate the voltage and frequency of the AC power, which helps maintain reliability and stability. The control is performed adjusting their active and reactive power outputs.

As shown in Fig. 1, the maximum output of a SI is

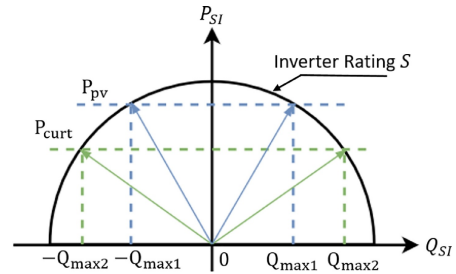


Fig. 1. SI capability curve [15]

constrained by its rated output S . The active power, P , is mainly dependent on the weather conditions. The active power, P , and reactive power, Q , can be controlled to satisfy

the formula $P^2 + Q^2 \leq S^2$. Therefore, in situations where critical voltage issues necessitate a certain amount of reactive power, $\pm Q_{max2}$, that exceeds the current available quantity, $\pm Q_{max1}$, and the active power output is already high, P_{pv} , curtailing the active power, P_{curt} , can create additional capacity for reactive power. The active power curtailment, P_{curt} , is constrained by $0 \leq P_{curt} \leq P_{pv}$.

In practice, however, the SIs have different capability curves and operational constraints, as documented in the literature [16].

C. Reinforcement learning

In a RL framework, an agent interacts with the environment over a series of discrete time steps represented by $t \in T$, where T denotes the set of all the time periods under consideration. As shown in Fig. 2, at each time step t , the agent receives some information from the environment, the state s_t , takes an action a_t , and receives a numerical feedback from the environment, the reward r_t . With the action chosen, the agent modifies the environment, ending up in a new state s_{t+1} .

The agent learns by repeatedly executing a loop of state-

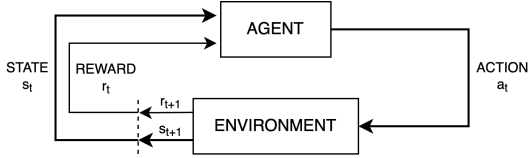


Fig. 2. Action-reward feedback loop of a generic RL framework [17]

action-reward-new state. Through this process, the agent becomes better at selecting actions that lead to positive rewards and avoiding actions that result in negative rewards.

The goal of the agent is to maximise the expected return:

$$R_t = \lim_{T \rightarrow \infty} \mathbb{E}_\pi \left[\sum_{i=t}^T \gamma^{i-t} r_i \right] \quad (1)$$

where π is the policy used to choose the action a_t given the state s_t and γ is the discount factor, $0 \leq \gamma \leq 1$, that it is used to express how the future rewards should be considered when making decisions in the present.

The action-value function, also known as Q-function, is defined as the expected return when the agent in state s_t , decides to take the action a_t , following the policy π :

$$Q_\pi(s_t, a_t) = \lim_{T \rightarrow \infty} \mathbb{E} \left[\sum_{i=t}^T \gamma^{i-t} r_i \mid s_0 = s_t, a_0 = a_t \right] \quad (2)$$

The general RL solution makes use of the recursive relationship known as the Bellman equation:

$$Q_\pi(s_t, a_t) = \lim_{T \rightarrow \infty} \mathbb{E} \left[r_t + \gamma \mathbb{E} [Q_\pi(s_{t+1}, a_{t+1})] \right] \quad (3)$$

This recursive relationship allows the agent to estimate the value of a state by considering the values of its future states

and their associated rewards.

In standard Q-Learning, the action a_{t+1} in Eq. 3 is chosen with a greedy policy $a_{t+1} = \mu(s_{t+1}) = \text{argmax}_a Q(s_{t+1}, a)$, which chooses the action that maximises the Q-value of the next state. This approach has been widely used in RL and has proven to be effective in a wide range of applications [18].

The greedy approach adopted by Q-Learning to optimise the policy π renders it unsuitable for continuous action spaces, as the *argmax* operation is a limiting factor.

The deterministic policy gradient (DPG) implementation tries to solve this problem. The idea behind DPG is to learn a deterministic function, $\mu(s_t)$, that can approximate the *argmax* operation. Generally, the $\mu(s_t)$ function is constructed as a deep neural network (Deep DPG) [19].

In particular, DDPG is an actor-critic algorithm that uses two neural networks: an actor (characterised by weight parameters ϕ) decides the action to perform in a given state, and a critic (characterised by weight parameters θ) evaluates the action chosen by the actor.

The critic is updated using the Bellman equation (Eq. 3).

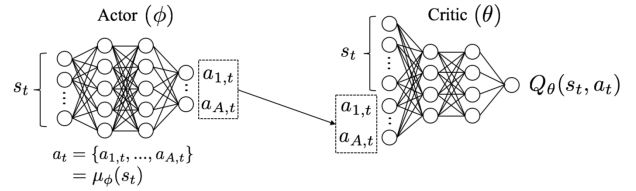


Fig. 3. Actor-critic architecture in DDPG [20]

The actor is updated using the gradient of the expected return from a start distribution, J , with respect to the actor parameters.

$$\nabla_\phi J \approx \mathbb{E} [\nabla_\phi Q_\theta(s_t, a) \mid a = \mu_\phi(s_t)]$$

Additionally, in order to enhance the performance of the model, various techniques are commonly applied in RL problems. These include memory buffers to store experiences, target networks to stabilise learning by using separate networks for estimation and actions selection, and batch normalization to improve the convergence of the model.

IV. PROPOSED METHOD

A. Distribution network

Results are obtained on a distribution network with a single feeder. The network is a low-voltage distribution network, and it is served by a 0.4 MVA MV/LV transformer station. The network has 20 customers and each customer has a PV generator. Hence, for this work, the PV penetration rate is set to 100%.

The time series are considered for one year with 15 minutes resolution, for a total of 35040 time steps. Different load profile time series are used for the customers, and similar

generators profiles are used for the PV generators. The time series are taken from the SimBench dataset [21].

B. RL elements

1) Environment:

The environment is the LV distribution grid as described in II.

2) State:

The state comprises the information accessible to the agent. In this case, the active and reactive power of loads, the active power of generators and the voltage magnitude of each bus at time step t .

$$s_t = [\mathcal{L}_t^p, \mathcal{L}_t^q, \mathcal{DG}_t^p, V_t] \quad (4)$$

The state space, \mathcal{S} , is given by all the possible combinations of $[\mathcal{L}_t^p, \mathcal{L}_t^q, \mathcal{DG}_t^p, V_t]$. Thus, at any given instant, s_t is a subset of \mathbb{R}^n , where n is given by: $n = |\mathcal{L}| + |\mathcal{L}| + |\mathcal{DG}| + |\mathcal{N}|$.

3) Action:

Similar to [15], the agent can choose the reactive power output of each SI, from $-S$ to $+S$. As mentioned in Section III-B, the action, $a_{i,t}$, can change the active and reactive power of each PV i , with $i \in \mathcal{DG}$. In particular, the agent chooses an action $a_{i,t}$ to control the amount of reactive power, $Q_{i,t}$, of the SI i at time step t . The reactive power chosen is given by $Q_{i,t} = S_i \cdot a_{i,t}$. If the value of $Q_{i,t}$ chosen by the agent satisfies the relationship $P_{i,t}^2 + Q_{i,t}^2 \leq S_i^2$, no curtailment is applied; otherwise some curtailment is applied and the new active power generation is given by: $\bar{P}_{i,t} = \sqrt{S_i^2 - Q_{i,t}^2}$.

The action space, \mathcal{A} , is given by all the possible combinations of $[-1, 1]^m$, where m is given by: $m = |\mathcal{DG}|$.

4) Reward:

The reward at time step t , r_t , consists of the sum of three terms:

- A reward, r_v , regarding voltage violation. The agent is punished if the voltage magnitude of the buses deviates from 1 pu. The idea is to punish the agent by a low value when the voltage bus is inside the range $[v^{min}, v^{max}]$ and to punish it more when the voltage bus is farther away. This is ensured with a bowl-shaped voltage violation function, f_v , [22]. For this paper, v^{min} and v^{max} are equal to 0.95 and 1.05, respectively.

$$r_v = - \sum_{i=0}^{\mathcal{N}} f_v(V_{i,t}) \quad (5)$$

- A reward, r_a , limiting the magnitude of each action.

$$r_a = - \sum_{i=0}^{\mathcal{DG}} |a_{i,t}| \quad (6)$$

This is used to train the agent to choose action that are as low as possible in magnitude to avoid too much curtailment and overuse of the reactive power.

- A reward for the fairness, r_f .

$$r_f = - \sum_{i=0}^{\mathcal{DG}} |\bar{c}_t - c_{i,t}| \quad (7)$$

where $c_{i,t}$ is the active power curtailment of the PV i decided by the agent when choosing the action $a_{i,t}$, and \bar{c}_t is the mean value of the active power curtailment among the customers at time step t .

The purpose of this reward is to encourage the agent to select actions that have relatively similar consequences on customers' curtailment, so that if the agent decides to curtail one customer more than others, it receives a larger penalty.

The total reward is given by the algebraic sum of all the previous terms multiplied by a scaling factor.

$$r_t = \alpha \cdot r_v + \beta \cdot r_a + \omega \cdot r_f \quad (8)$$

The weights α , β and, ω are selected with careful consideration to balance the magnitude of each parameter, ensuring that the agent is able to optimise its performance while also exhibiting the desired behaviour. In this case study, the values for these parameters, considering the different scale of the single rewards, are chosen to be 1000, 5, and 25 respectively. These terms are chosen to have the following impacts on the reward function: $\alpha \cdot r_v \gg \beta \cdot r_a \simeq \omega \cdot r_f$.

The agent's neural network architectures are composed by two hidden layers with 256 neurons each. The learning rate for the actor and critic network is set to 0.0001, the discount factor γ is equal to 0.99 and the soft update parameter τ is equal to 0.005. The batch size is 64 and the maximum number of experiences is 200000.

V. RESULTS

The results are presented for four different scenarios.

- 1) Scenario *a*. This is considered as the baseline: the situation where no control is applied and in the network are present some over and under-voltages that a DSO would like to solve.
- 2) Scenario *b*. The objective of the DSO is solely to address the voltage issues within the network. To reflect this goal in the reward function, only the r_v term is taken into account.
- 3) Scenario *c*. The DSO aims to mitigate voltage problems while minimising active power curtailment. To reflect this objective in the reward function, both the r_v and r_a terms are taken into consideration.
- 4) Scenario *d*. The DSO wants to address the voltage problems while minimising active power curtailment

and maintaining fairness in the distribution of the curtailment among customers. This is achieved by taking into account all the three factors: voltage deviation, actions' magnitude, and fairness.

In each scenario, the agent undergoes training across several episodes, with each episode covering 30% of the total time steps T , corresponding to approximately the first 4 months of the year. After training, the agent is evaluated on the remaining 70% of the time series, corresponding to approximately the remaining 8 months of the year.

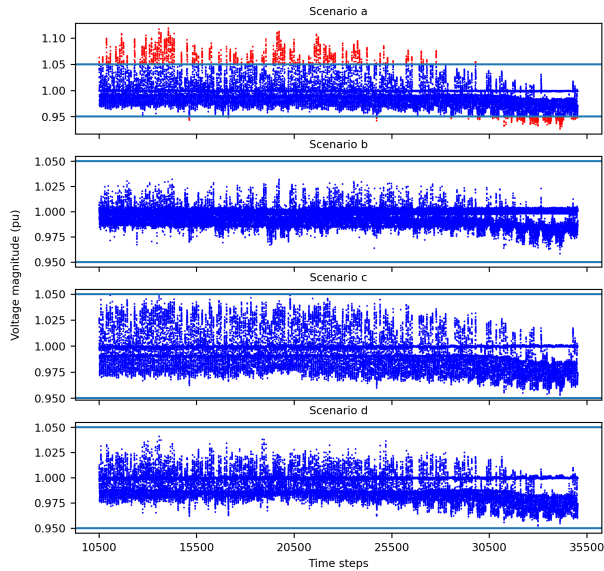


Fig. 4. The buses' voltage situation for each scenario

Fig. 4 provides a visual representation of the voltage levels across the buses in each of the different scenarios. For each time steps, two dots are drawn: representing the maximum voltage of all nodes recorded at that time step and a dot to represent the minimum. Moreover, the dots are blue or red whether they are inside or outside the safe voltage range, respectively.

Scenario *a* shows some voltage problems: some over-voltages on the left-hand side of the plot, corresponding to more sunny days, and some under-voltages on the right-hand side, corresponding to the winter period.

It is possible to see that in all the control scenarios the agents are able to solve the voltage problems. In particular, in Scenario *b* the voltage distribution of the nodes is close to 1. This is obtained only thanks to an excessive curtailment of the customers. In Scenarios *c* and *d* the agents solve the voltage problems, while keeping the curtailment not too high. In both scenarios, the maximum and minimum voltage are close to the boundaries but inside the safe range, making the agents choosing actions without overdoing it. It is worth noting that the low values of active power curtailment obtained in the different scenarios, can be attributed to the

efficient control of both active and reactive power by the SIs. It should be emphasised that if only active power curtailment had been utilised, the amount of curtailed power would have been substantially greater, and the resolution of under-voltage issues would not have been achieved, as evident in other studies ([5], [6], [13]).

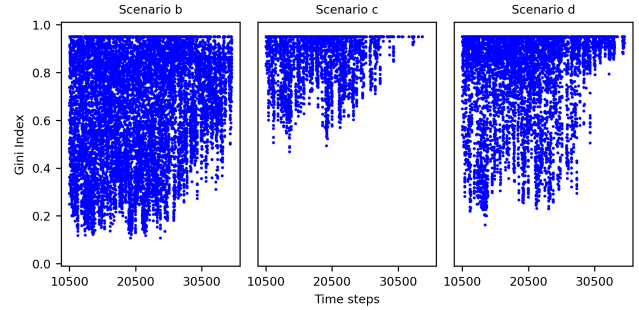


Fig. 5. Gini index for each scenario where some control is applied

The fairness of the curtailment in each scenario can be evaluated using the Gini index, as shown in Fig. 5. The figure shows the Gini index of the customers' curtailment for each time step where at least one customer is being curtailed. In Scenario *b*, where only the voltage deviation punishment is considered, the curtailment appears to be fair, likely due to the large amount of curtailment applied. In this scenario the total curtailment is 20.7 MWh equivalent to 44% of the total production. In Scenario *c*, the curtailment has a higher Gini index for most of the time steps, leading to a less fair result than Scenario *b*. In this scenario the energy curtailed is 2.65 MWh equivalent to 5.6% of the total production. Scenario *c* required less curtailment to solve the voltage problems, but this means that often some customers are curtailed more than others. In Scenario *d*, the curtailment is more fair than in Scenario *c*, since the amount of power needed to be curtailed is distributed among all customers, even if this requires more curtailment overall. In this scenario, the energy curtailed is 8.77 MWh, equivalent to 18% of the total energy produced.

It is important to note that, the agent in Scenario *d* presents some limitations. In particular, from Fig. 5, it is possible to see that the curtailment is never totally fair, since the Gini index is never 0 (the minimum observed Gini index value is 0.16). At the same time, for some time steps the curtailment is unfair since the Gini index value is high (the maximum observed Gini index value is 0.95, that corresponds to only one customer being curtailed) and the agent behaves similar to the other scenarios where fairness is not considered.

Table I shows a summary of the results obtained in each scenario. In particular, it shows the total PV energy produced by the customers in the testing period of approximately eight months, the energy curtailed by the RL agent, the number of time steps that present a voltage

TABLE I.
Summary of the results for each scenario

Scenario	Total PV generation (MWh)	Energy Curtailed (MWh)	# Under Voltages	# Over Voltages	Min V (pu)	Max V (pu)
<i>a</i>	47.5	0	493	1677	0.9263	1.1191
<i>b</i>		20.7	0	0	0.9582	1.0321
<i>c</i>		2.65	0	0	0.9530	1.0497
<i>d</i>		8.77	0	0	0.9520	1.0409

problem, either under or over-voltage, and the minimum and maximum voltage recorded at any bus in the eight months. It is possible to see that the agents in scenarios *c* and *d* solved all the voltage problems while keeping the energy curtailed low.

It is worth noting that the weights of the single terms have a great impact on the reward function and so on the results. In this experiment, the weights for the magnitude of the actions, β , and the fairness, ω , are chosen to have similar impacts on the reward function as discussed in Section IV. However, in different situations, either β or ω may have a greater impact depending on the specific goals of the DSO. Ultimately, it is a trade-off between how much active power curtailment the DSO is willing to accept without considering fairness versus increasing fairness at the cost of higher active power curtailment.

VI. CONCLUSION

This paper investigates the control of SIs using a reinforcement learning technique to solve the voltage problems in a power distribution network. The model used is a deep deterministic policy gradient algorithm. This algorithm is able, using only the network's information, to solve the critical situations in the network. In particular, the agent is able to solve the voltage problems, with a cost for the distribution system operator of a few MWh of energy not produced in the entire time span considered. Moreover, the focus of this paper is solving the voltage issues with a fair curtailment strategy. The results show that it is possible to obtain a fair active power curtailment for the customers, controlling the active and reactive power of PV smart inverters, but at the cost of a higher curtailment.

Some future works may examine more complex networks; studying the impact of the weights β and ω on the results; using a decentralised reinforcement learning technique to solve the voltage problems in a fair way, and considering the limitations of SIs.

REFERENCES

- [1] S. K. Salman and Z. G. Wan, "Fuzzy logic-based avc relay for voltage control of distribution network with and without distributed/embedded generation," in *2007 IEEE Lausanne Power Tech*, 2007, pp. 2128–2132. DOI: 10.1109/PCT.2007.4538647.
- [2] S. Magnússon, G. Qu, and N. Li, *Distributed optimal voltage control with asynchronous and delayed communication*, 2019. DOI: 10.48550/ARXIV.1903.01065. [Online]. Available: <https://arxiv.org/abs/1903.01065>.
- [3] M. Maharjan, U. Tamrakar, Z. Ni, B. Bhattarai, and R. Tonkoski, "Overvoltage prevention and curtailment reduction using adaptive droop-based supplementary control in smart inverters," *Applied Sciences*, vol. 11, no. 17, 2021, ISSN: 2076-3417. DOI: 10.3390/app11177900.
- [4] A. M. Howlader, S. Sadoyama, L. R. Roose, and Y. Chen, "Active power control to mitigate voltage and frequency deviations for the smart grid using smart PV inverters," *Applied Energy*, vol. 258, p. 114000, 2020, ISSN: 0306-2619. DOI: <https://doi.org/10.1016/j.apenergy.2019.114000>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0306261919316873>.
- [5] Y. Z. Gerdroodbari, R. Razzaghi, and F. Shahnia, "Improving voltage regulation and unbalance in distribution networks using peer-to-peer data sharing between single-phase PV inverters," *IEEE Transactions on Power Delivery*, vol. 37, no. 4, pp. 2629–2639, 2022. DOI: 10.1109/TPWRD.2021.3113011.
- [6] P. Jahangiri and D. C. Aliprantis, "Distributed Volt/VAr control by PV inverters," *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 3429–3439, 2013. DOI: 10.1109/TPWRS.2013.2256375.
- [7] F. Olivier, P. Aristidou, D. Ernst, and T. Van Cutsem, "Active management of low-voltage networks for mitigating overvoltages due to photovoltaic units," *IEEE Transactions on Smart Grid*, vol. 7, no. 2, pp. 926–936, 2016. DOI: 10.1109/TSG.2015.2410171.
- [8] S. Ghosh, S. Rahman, and M. Pipattanasomporn, "Distribution voltage regulation through active power curtailment with PV inverters and solar generation forecasts," *IEEE Transactions on Sustainable Energy*, vol. 8, no. 1, pp. 13–22, 2017. DOI: 10.1109/TSTE.2016.2577559.
- [9] R. Tonkoski and L. A. Lopes, "Impact of active power curtailment on overvoltage prevention and energy production of PV inverters connected to low voltage residential feeders," *Renewable Energy*, vol. 36, no. 12, pp. 3566–3574, 2011, ISSN: 0960-1481. DOI: <https://doi.org/10.1016/j.renene.2011.05.031>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S096014811100259X>.
- [10] R. Tonkoski, L. A. C. Lopes, and T. H. M. El-Fouly, "Coordinated active power curtailment of grid connected PV inverters for overvoltage prevention," *IEEE Transactions on Sustainable Energy*, vol. 2,

- no. 2, pp. 139–147, 2011. DOI: 10.1109/TSTE.2010.2098483.
- [11] A. R. Vadavathi, G. Hoogsteen, and J. L. Hurink, “Comparison of fairness based coordinated grid voltage control methods for PV inverters,” in *2021 IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe)*, 2021, pp. 1–5. DOI: 10.1109/ISGTEurope52324.2021.9639982.
- [12] A. Vadavathi, G. Hoogsteen, and J. Hurink, “PV inverter based fair power quality control,” *IEEE Transactions on Smart Grid*, pp. 1–1, 2023. DOI: 10.1109/TSG.2023.3244601.
- [13] Z. Wei, F. Nijs, J. Li, and H. Wang, “Model-free approach to fair solar PV curtailment using reinforcement learning,” Jun. 2023. DOI: 10.1145/3575813.3576871.
- [14] Wikipedia contributors, *Gini coefficient — Wikipedia, the free encyclopedia*, [Online; accessed 10-April-2023], 2023. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Gini_coefficient&oldid=1148472029.
- [15] C. Li, Y.-A. Chen, C. Jin, R. Sharma, and J. Kleissl, “Online PV smart inverter coordination using deep deterministic policy gradient,” *Electric Power Systems Research*, vol. 209, p. 107988, 2022, ISSN: 0378-7796. DOI: <https://doi.org/10.1016/j.epsr.2022.107988>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0378779622002164>.
- [16] A. Nouri, A. Soroudi, and A. Keane, “Resilient decentralised control of inverter-interfaced distributed energy sources in low-voltage distribution grids,” *IET Smart Grid*, vol. 3, no. 2, pp. 153–161, 2020. DOI: 10.1049/iet-stg.2019.0163. [Online]. Available: <https://doi.org/10.1049/2Fiet-stg.2019.0163>.
- [17] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.
- [18] Y. Li, *Deep reinforcement learning: An overview*, 2018. arXiv: 1701.07274 [cs.LG].
- [19] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, *Continuous control with deep reinforcement learning*, 2019. arXiv: 1509.02971 [cs.LG].
- [20] J.-F. Toubeau, B. Bakhshideh Zad, M. Hupez, Z. De Grève, and F. Vallée, “Deep reinforcement learning-based voltage control to deal with model uncertainties in distribution networks,” *Energies*, vol. 13, no. 15, 2020, ISSN: 1996-1073. DOI: 10.3390/en13153928. [Online]. Available: <https://www.mdpi.com/1996-1073/13/15/3928>.
- [21] S. Meinecke, D. Sarajlić, S. R. Drauz, A. Klettke, L.-P. Lauven, C. Rehtanz, A. Moser, and M. Braun, “Simbench—a benchmark dataset of electric power systems to compare innovative solutions based on power flow analysis,” *Energies*, vol. 13, no. 12, 2020, ISSN: 1996-1073. DOI: 10.3390/en13123290. [Online]. Available: <https://www.mdpi.com/1996-1073/13/12/3290>.
- [22] J. Wang, W. Xu, Y. Gu, W. Song, and T. C. Green, *Multi-agent reinforcement learning for active voltage control on power distribution networks*, 2021. DOI: 10.48550/ARXIV.2110.14300. [Online]. Available: <https://arxiv.org/abs/2110.14300>.