



Model-Free Approach to Fair Solar PV Curtailment Using Reinforcement Learning

Zhuo Wei, Frits de Nijs, Jinhao Li, Hao Wang*

Department of Data Science and AI

Monash University

Melbourne, Victoria, Australia

ABSTRACT

The rapid adoption of residential solar photovoltaics (PV) has resulted in regular overvoltage events, due to correlated reverse power flows. Currently, PV inverters prevent damage to electronics by curtailing energy production in response to overvoltage. However, this disproportionately affects households at the far end of the feeder, leading to an unfair allocation of the potential value of energy produced. Globally optimizing for fair curtailment requires accurate feeder parameters, which are often unknown. This paper investigates reinforcement learning, which gradually optimizes a fair PV curtailment strategy by interacting with the system. We evaluate six fairness metrics on how well they can be learned compared to an optimal solution oracle. We show that all definitions permit efficient learning, suggesting that reinforcement learning is a promising approach to achieving both safe and fair PV coordination.

CCS CONCEPTS

• **Theory of computation** → **Theory and algorithms for application domains**; • **Hardware** → **Power and energy**.

KEYWORDS

Renewable Energy, PV Curtailment, Voltage Control, Energy Fairness, Reinforcement Learning

ACM Reference Format:

Zhuo Wei, Frits de Nijs, Jinhao Li, Hao Wang. 2023. Model-Free Approach to Fair Solar PV Curtailment Using Reinforcement Learning. In *The 14th ACM International Conference on Future Energy Systems (e-Energy '23)*, June 20–23, 2023, Orlando, FL, USA. ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/3575813.3576871>

1 INTRODUCTION

Distributed energy resources owned by households are growing in popularity. Installed capacity of photovoltaic (PV) systems has grown at an average annual rate of 49% over the last decade [12], with the global PV capacity expected to more than double in the next decade [9]. However, PV systems can result in significant challenges to the local grid because of the midday PV power injection and

reverse power flow [16]. One of the most severe problems is that the massive increase in PV systems can lead to significant excess generation and overvoltage problems on distribution feeders [4].

Active power curtailment has been proposed as an important solution to the technical and economic challenges associated with PV power generation [12]. Inverter reactive power compensation as a method of PV curtailment can be an effective solution to solve overvoltage problems. However, this method often results in more PV curtailment for households at the end of the feeder [6]. Distribution network operators urgently need a fair PV curtailment solution. For example, households with PV in Australia currently lose a total value of 1.2-4.5 million Australian dollars per year due to unfair PV curtailment [13]. Also, according to Carley and Konisky [2], the transition to low-carbon energy sources, including PV generation, can perpetuate existing energy inequity if the fairness problem cannot be solved. Therefore, the risk of unfairness may become a barrier to entry for late adopters of PV panels, slowing down the overall transition to a renewable grid and net zero.

Although fair PV curtailment solutions exist, practical technical challenges remain unsolved. Gebbran et al. [6] used an optimization-based method to achieve fair PV curtailment, assuming known distribution network parameters and accurate load forecasts. However, distribution network visibility is a challenging problem, and network parameters are often not known [8, 14]. In particular, distribution network operators may obtain incorrect grid parameters due to insufficient data updates [3]. A recent study by Yeh et al. [15] investigated voltage control with an unknown grid topology to ensure stability, but without a fairness objective. Furthermore, globally fair curtailment of solar panels also requires accurate predictions of future generations and demand to be able to ensure fairness over the full day.

We propose to approach the problem of fair PV curtailment using reinforcement learning (RL), for RL can be used in real-time and does not require exact physical models of the distribution system [1, 5]. Therefore, RL particularly suits the PV curtailment problem, in which the distribution network visibility is low. We train an RL algorithm, which learns a fair PV curtailment policy from simulated control on historical data, and continuous interactions with the distribution network environment. The learned control policy aims to minimize overall electricity costs while fairly curtailing households' PV generation to keep their voltages in a safe range.

The contributions of this paper are as follows.

- We propose a new approach for fair PV curtailment using RL. Our approach does not require any knowledge of the distribution network parameters but interacts with the network and households to learn a fair PV curtailment strategy.

*Corresponding author: hao.wang2@monash.edu (Hao Wang).

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

e-Energy '23, June 20–23, 2023, Orlando, FL, USA

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-0032-3/23/06...\$15.00

<https://doi.org/10.1145/3575813.3576871>

- We specify three fairness definitions for PV curtailment over two different periods, instant and accumulative over a day, resulting in six cases. These six cases are demonstrated to be effective in learning fair PV curtailment strategies.
- We compare the RL solutions to optimization solutions which rely on an accurate network model by characterizing the trade-off between electricity cost reduction and fairness.

2 MODEL

To use RL methods to minimize households' electricity costs and ensure voltage safety and fair PV curtailment, we need to build an environment for a distribution network, including voltage constraints and fairness definitions for PV curtailment.

2.1 Network Model

We introduce a linear feeder network [10] as the environment to capture the interactions between households and voltage changes on the network caused by PV injection, a simplification which allows us to calculate and compare with the optimal curtailment. The RL algorithm presented in Section 2.2 will interact with the developed environment to learn a fair PV curtailment strategy.

Specifically, our feeder network, shown in Fig. 1, captures the dependency of voltage changes with power import and export on each node. We denote $p_{h,t}^{br}$ and $q_{h,t}^{br}$ as the active and reactive power at time t on branch h , and j is the imaginary unit. For each household h , we denote $p_{h,t}$ and $q_{h,t}$ as the active power and reactive power exchanged between the household and the grid at time t , respectively. When $p_{h,t} < 0$, household h imports power from the grid; otherwise when $p_{h,t} > 0$, the household exports power to the grid. Similarly for reactive power, $q_{h,t} < 0$ indicates absorption, and $q_{h,t} > 0$ indicates release of reactive power. According to Zhong et al. [17], the power flows in Fig. 1 are described by

$$p_{h+1,t}^{br} = p_{h,t}^{br} + p_{h,t}, \quad (1)$$

$$q_{h+1,t}^{br} = q_{h,t}^{br} + q_{h,t}, \quad (2)$$

$$v_{h+1,t} = v_{h,t} - (r_{h+1,t} p_{h+1,t}^{br} + x_{h+1,t} q_{h+1,t}^{br}) / v_{0,t}, \quad (3)$$

where $v_{h,t}$ is the voltage of node h at t , $r_{h,t}$ and $x_{h,t}$ are the resistance and reactance of branch h , and node 0 connects to the grid.

2.2 Electricity Cost and Feed-in Tariff

We aim to minimize households' electricity costs, which means maximizing households' profits from exporting PV. Therefore, the objective of the problem is described as

$$\min \sum_{h \in \mathcal{H}} \sum_{t \in \mathcal{T}} c(t) \times (p_{h,t}^L - p_{h,t}^G \times (1 - \alpha_{h,t}^C)), \quad (4)$$

where \mathcal{H} is the set of households, and \mathcal{T} is the horizon in which fair PV curtailment is implemented. The electricity price is denoted as $c(t)$, as it is often a time-of-use (ToU) price [6]. For household h , PV generation at t is $p_{h,t}^G$, and $p_{h,t}^L$ denotes household h 's load at t . For each household, a certain percentage of PV generation curtailed is a decision variable, which is denoted as $\alpha_{h,t}^C$. Thus, the net power exchange between household h and the grid $p_{h,t}$ is expressed as

$$p_{h,t} = p_{h,t}^G \times (1 - \alpha_{h,t}^C) - p_{h,t}^L. \quad (5)$$

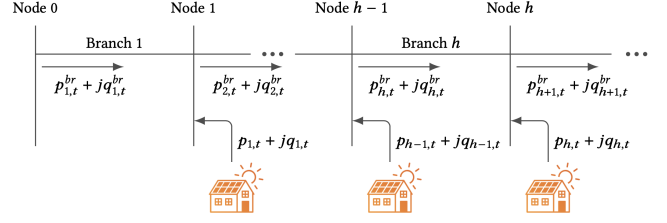


Figure 1: The structure of the network.

2.3 Voltage Control and Constraint

The safe range of the voltage is often set between 0.95 and 1.05 per unit (p.u.) [10]. Thus, the voltage constraints can be expressed as

$$0.95 \leq v_{h,t} \text{ (p.u.)} \leq 1.05, \forall h \in \mathcal{H}, t \in \mathcal{T}, \quad (6)$$

where $v_{h,t}$ is the voltage of node h at t .

PV inverters are used for voltage control to ensure that the voltages are within the safe range. Specifically, PV inverters support Volt-Var Control (VVC) [6], in which voltages are controlled by injecting or absorbing reactive power in response to voltage. The relationship between $q_{h,t}$ and $v_{h,t}$ can be expressed as

$$q_{h,t} = \bar{q} \times \mathbf{1}(v_{h,t} \leq V_1) + \bar{q} \frac{V_2 - v_{h,t}}{V_2 - V_1} \times \mathbf{1}(V_1 < v_{h,t} < V_2) - \bar{q} \frac{v_{h,t} - V_3}{V_4 - V_3} \times \mathbf{1}(V_3 < v_{h,t} < V_4) - \bar{q} \times \mathbf{1}(v_{h,t} \geq V_4), \quad (7)$$

where $\mathbf{1}(\cdot)$ is the indicator function, \bar{q} is the maximum reactive power output, which is normally defined as a percentage of the maximum capacity of the PV inverter in each household. The values of voltage thresholds V_1 to V_4 depend on different product standards.

Also, due to the use of PV inverters, it is necessary to consider the capacity of the inverters, whose relationship with the PV curtailment can be expressed as

$$\bar{s} \geq (p_{h,t}^G \times (1 - \alpha_{h,t}^C))^2 + q_{h,t}^2, \forall h \in \mathcal{H}, t \in \mathcal{T}, \quad (8)$$

where \bar{s} is the maximum capacity of PV inverters.

2.4 Definition of Fairness

Aiming to curtail PV fairly, we consider three definitions of fairness. According to [2], the core principle of energy fairness is to ensure that certain groups do not disproportionately share the burden or fail to receive the benefits. In our problem, fair PV curtailment means that no households suffer from more curtailment or low PV export compared with other households. We follow the approach of [6], who defines three fairness metrics:

- (1) Definition 1: egalitarian PV curtailment;
- (2) Definition 2: proportional PV curtailment;
- (3) Definition 3: egalitarian PV output.

In Definition 1, the PV curtailment of each household does not exceed a specific value. In Definition 2, the ratio of PV curtailment to the maximum available export power for each household does not exceed an upper limit. In Definition 3, the PV exports of households are not below the lower bound. The differences between the three definitions are illustrated in Appendix A.1.

The work of [6] proposed fairness definitions that should be optimized strictly, where fairness is maximized every timestep. We

consider a relaxed *accumulative* case, which allows the controller to gradually equalize fairness over the full length of a day. For Definition 1, fair PV curtailment means that the *maximum* amount of PV curtailment needs to be minimized. Therefore, in these two cases, the fairness penalty due to PV curtailment becomes

$$\alpha_t^C = \{\alpha_{h,t}^C \mid \forall h \in \mathcal{H}\}, \quad (9)$$

$$f_1^{ins}(\alpha_t^C) = \max_{h \in \mathcal{H}} (\alpha_{h,t}^C \cdot p_{h,t}^G), \quad (10)$$

$$f_1^{acc}(\alpha_T^C) = \max_{h \in \mathcal{H}} \left(\sum_{t=0}^T (\alpha_{h,t}^C \cdot p_{h,t}^G \cdot \Delta) \right), \quad (11)$$

where T is the control horizon of one day in thirty-minute steps. To show accumulative quantities in kWh, step size $\Delta = 30/60$, which means each half hour.

For the Definition 2, the ratio of PV curtailment to the maximum available export power needs to be minimized. Moreover, this definition only applies to the case where the PV generation of households is more than their power consumption. Thus, the fairness penalty in the two cases can be expressed as

$$f_2^{ins}(\alpha_t^C) = \max_{h \in \mathcal{H}} \left(\frac{\alpha_{h,t}^C \cdot p_{h,t}^G}{\max(\epsilon, p_{h,t}^G - p_{h,t}^L)} \right), \quad (12)$$

$$f_2^{acc}(\alpha_T^C) = \max_{h \in \mathcal{H}} \left(\frac{\sum_{t=0}^T (\alpha_{h,t}^C \cdot p_{h,t}^G \cdot \Delta)}{\max(\epsilon, e_h^E)} \right), \quad (13)$$

$$e_h^E = \sum_{t=0}^T ((p_{h,t}^G - p_{h,t}^L) \cdot \Delta), \quad (14)$$

where e_h^E is the total energy export in kWh, and $\epsilon > 0$ a small value.

For Definition 3, the PV exports of households need to be maximized. Same as Definition 2, this definition only applies to households where PV generation is larger than their daily consumption. Therefore, the fairness penalty in the two cases can be expressed as

$$f_3^{ins}(\alpha_t^C) = -\min_{h \in \mathcal{H}} \left(\mathbf{1}(\alpha_{h,t}^C > 0) \cdot ((1 - \alpha_{h,t}^C) \cdot p_{h,t}^G - p_{h,t}^L) \right. \\ \left. + \mathbf{1}(\alpha_{h,t}^C = 0) \cdot \max_{h \in \mathcal{H}} (p_{h,t}^G - p_{h,t}^L) \right), \quad (15)$$

$$f_3^{acc}(\alpha_T^C) = -\min_{h \in \mathcal{H}} \left(\mathbf{1}(\sum_{t=0}^T \alpha_{h,t}^C > 0) \cdot (\sum_{t=0}^T ((1 - \alpha_{h,t}^C) \cdot p_{h,t}^G - p_{h,t}^L) \cdot \Delta) \right. \\ \left. + \mathbf{1}(\sum_{t=0}^T \alpha_{h,t}^C = 0) \cdot \max_{h \in \mathcal{H}} (e_h^E) \right), \quad (16)$$

where the negative sign means that more PV exports will result in less penalty, thus maximizing PV exports.

3 METHODOLOGY

We use reinforcement learning (RL) to optimize a fair and safe control policy for PV curtailment. To do so, we first define a Markov Decision Process (MDP) on the single-feeder control problem, and train a controller with Soft Actor Critic [7].

MDP model. An MDP is a discrete-time model defined by its state space \mathcal{S} , action space \mathcal{A} , transition function $T(s, a) \rightarrow s'$ and reward function $R(s, a) \rightarrow \mathbb{R}$, with the objective to find a control policy μ that maximises the expected discounted cumulative reward, $\max_{\mu} \mathbb{E} \left[\sum_t \gamma^t R(s_t, a_t) \mid s_{t+1} \sim T(s_t, a_t), a_t \sim \mu(s_t) \right]$.

The state space of our agent depends on the definition of fairness we consider. The instantaneous case state s_t^{ins} contains the previous time's measured load, generation, and voltage, while the accumulative case s_t^{acc} also includes the total curtailment up to t

$$s_t^{ins} = (p_{h,t-1}^L, p_{h,t}^L, p_{h,t}^G, v_{h,t-1} \mid \forall h \in \mathcal{H}), \quad (17)$$

$$s_t^{acc} = (p_{h,t-1}^L, p_{h,t}^L, p_{h,t}^G, v_{h,t-1}, \sum_{\tau=0}^{t-1} \alpha_{h,\tau}^C \cdot p_{h,\tau}^G \mid \forall h \in \mathcal{H}). \quad (18)$$

Action space \mathcal{A} includes all possible actions \mathbf{a} that the agent can perform in the environment. In this paper, one central agent decides how much PV production to curtail for all households. Thus, the action \mathbf{a}_t is the proportion of curtailed power at moment t ,

$$\mathbf{a}_t = \{\alpha_{h,t}^C, \forall h \in \mathcal{H}\}. \quad (19)$$

Our objective is to maximize households' profits from selling PV, while ensuring fair PV curtailment to maintain voltage safety. To maintain the voltage within the safe range, we set how much the voltage exceeds the safe range $[0.95, 1.05]$ as the penalty term $g(v_{h,t})$, which can be expressed as

$$g(v_{h,t}) = \max(0.95 - v_{h,t}, 0) + \max(v_{h,t} - 1.05, 0). \quad (20)$$

For the fairness penalty caused by PV curtailment, we let $f(\alpha_t^C)$ denote the general form of fairness penalty, representing all six cases, such as $f_1^{ins}(\alpha_t^C)$, $f_1^{acc}(\alpha_T^C)$, $f_2^{ins}(\alpha_t^C)$, $f_2^{acc}(\alpha_T^C)$, $f_3^{ins}(\alpha_t^C)$, and $f_3^{acc}(\alpha_T^C)$.

Therefore, the reward r_t is expressed as

$$r_t = \sum_{h \in \mathcal{H}} \left[(1 - \alpha_{h,t}^C) - w_1 g(v_{h,t}) \right] - w_2 f(\alpha_t^C), \quad (21)$$

where w_1 and w_2 are weights. For accumulative cases, the fairness penalty only applies once per episode, at the last step in the episode.

Learning algorithm. We use the SAC method [7] to solve the fair PV curtailment problem. More specifically, we set a distribution network as the environment required by the RL model. Then we train an agent by the SAC algorithm to be responsible for the PV curtailment of all households. We use this agent to curtail PV generation for all households and simulate the impact of PV curtailment on the voltage and cost of electricity through the network model. Thus, the agent can interact with the environment and learn a reasonable PV curtailment policy. For our hyper-parameters, we set the discount factor to 0.99, learning rate to 0.0003, and batch size to 64. We set 80% of the data source as the training set and 20% as the test set. The sampling is from 0:00 midnight each day. The training episodes were set to 24 hours.

4 CASE STUDY

Preliminary experimental results are implemented in a network containing 10 households, using data from [11]. We use half-hourly historical data for one year (17, 520 data points per panel) as our training set. Also, we train the model for 3×10^4 simulated timesteps on this dataset. The learning curves of our model are shown in Appendix A.3. We use Australian dollars (\$) to show monetary values. According to [6], the time-of-use electricity prices are 0.12 \$/kWh for off-peak (23:00–7:00), 0.52 \$/kWh for peak (14:00–19:30),

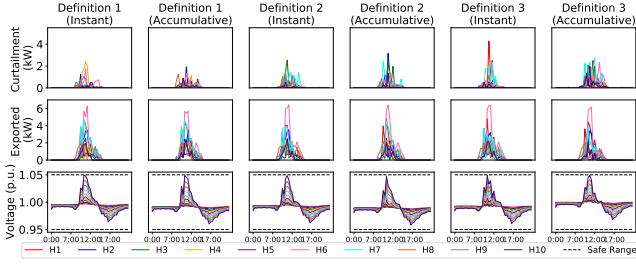


Figure 2: PV curtailment under different fairness definitions.

0.22 \$/kWh for other time periods. The feed-in tariff is 0.1 \$/kWh throughout the day.

Based on the suggestion of [6], the maximum capacity of PV inverters \bar{s} is set to the PV generation without PV curtailment. The maximum reactive power output \bar{q} is set to 10% of \bar{s} . The values of V_1 to V_4 are set to 0.94 p.u., 0.98 p.u., 1.06 p.u., and 1.1 p.u.

According to [17], we set the voltage at node 0 as 240 V and set it as 1 p.u. For the active power injected into node 1, it is the sum of the loads for all households at t moment. For the reactive power injected into node 1, which is $q_{1,t}^{br}$, we set it as 0. The reason is that $q_{h,t}^{br}$ means that the PV inverter at node $h - 1$ injects reactive power during PV curtailment, while there is no PV inverter at node 0.

Due to the reverse power flow caused by households selling excess power to the grid during PV generation peaks, if a household gets closer to the end of the feeder, its voltage will rise more. Thus, we trained the SAC model to achieve fair PV curtailment based on different fairness definitions, and the results are shown in Fig. 2, with more detailed heatmaps of the curtailment in Appendix A.4. We observe that the controller limits voltage to the safe range. Comparing across the instant and accumulative cases, we observe that instant curtailment is concentrated in the peak PV generation period, while the PV curtailment in accumulative cases is spread out. This was expected, because the instant case has more stringent per-timestep fairness requirements, while the other case is more concerned with total daily PV curtailment.

Comparing the three definitions, in both the Instant Case and Accumulative Case, the mean of total PV curtailments for each household in the Definition 1 is slightly smaller than others. The reason for this result is that in the Definition 1 scenario, the maximum PV curtailment can directly influence the reward, while other scenarios do not have this setting. In addition, because Definition 3 cares about exports instead of directly limiting PV curtailment value or ratio, the mean of total PV curtailments in the Definition 3 is the most in these definitions. More specifically, in the Instant Case, the means of total PV curtailments for Definition 1–3 are 3.25kW, 3.64kW, and 3.75kW. In the Accumulative Case, the means of total PV curtailments for Definition 1–3 are 3.02kW, 4.11kW, and 7.09kW. Due to different concerns of different fairness definitions, the agent chooses different policies to ensure that the voltage is within a safe range and no household is subject to unfair curtailments.

We plot the Pareto curve of the SAC model in Fig. 3 to explore how the model trade-offs between minimizing power cost and improving fairness. We compare our results with optimization assuming known parameters [6], which bounds the RL agent’s possible performance. The optimization problem is to minimize the cost

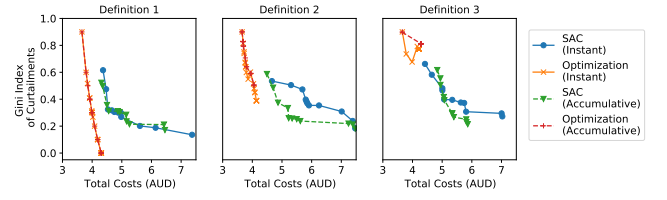


Figure 3: Pareto curves of SAC and optimization model.

of electricity while ensuring safety and improving fairness. The optimization model formulation is presented in Appendix A.2.

In Fig. 3, the x -axis represents the total cost of electricity for all households, and the y -axis represents the Gini Index of the total PV curtailment for households, which smaller values indicating better fairness. The decrease in Gini Index leads to a rise in electricity costs because a fair curtailment solution can result in some additional PV curtailment [6]. Since instant cases have fairness requirements for each moment, but accumulative cases are more concerned with the overall situation of the day, accumulative cases are usually more flexible. This means that accumulative cases can produce better results. Our model is slightly inferior to the optimization model in terms of saving electricity costs and improving the fairness of Definition 1. This result is because the optimization model uses all parameters of the entire distribution network to calculate the optimal solution. In contrast, the SAC model agent can observe only the parameters in the state. Our model performs better for reducing the Gini Index for Definition 2 and 3 because optimization models in these definitions do not target the amount of PV curtailment directly. In summary, the most crucial advantage of our model is that it does not need to observe all the distribution network information, and it still achieves a good effect in reducing the Gini Index, which means that it can effectively improve fairness in practice.

5 CONCLUSION

In conclusion, we developed an RL-based model to solve the fair curtailment problem. We first described this problem, which means minimizing electricity costs while ensuring that the voltage is in a safe range and improving fairness. We modeled a distribution network to explore the impact of PV curtailment on voltage and household interactions. We also specified several different definitions of fairness. Finally, we developed a SAC model and obtained preliminary experimental results.

Our vision is to build a model that can solve fairness problems in real-world scenarios. Therefore, we will consider reactive power control in the future and employ various fairness measures to obtain a more balanced view. We will also investigate how to select one of the three fairness definitions for a particular scenario. In addition, we will gradually extend our model to consider a longer control period, more households, and more complex network topologies. Moreover, phase imbalance and high-voltage grid fluctuations also need to be considered in our future work.

ACKNOWLEDGMENTS

We would like to thank anonymous reviewers for their constructive comments. This work has been supported in part by the Monash Energy Institute Professional development awards for ECR Women in

Energy and the Australian Research Council (ARC) Discovery Early Career Researcher Award (DECRA) under Grant DE230100046.

REFERENCES

- [1] Erick O Arwa and Komla A Folly. 2020. Reinforcement Learning Techniques for Optimal Power Control in Grid-Connected Microgrids: A Comprehensive Review. *IEEE Access* 8 (2020), 208992–209007.
- [2] Sanya Carley and David M Konisky. 2020. The justice and equity implications of the clean energy transition. *Nature Energy* 5, 8 (2020), 569–577.
- [3] Konstantina Christakou, Mario Paolone, and Ali Abur. 2017. Voltage control in active distribution networks under uncertainty in the system model: A robust optimization approach. *IEEE Transactions on Smart Grid* 9, 6 (2017), 5631–5642.
- [4] Paul Denholm, Matthew O’Connell, Gregory Brinkman, and Jennie Jorgenson. 2015. *Overgeneration from solar energy in california. a field guide to the duck chart*. Technical Report. National Renewable Energy Lab.(NREL), Golden, CO (United States).
- [5] Cong Feng, Yuanzhi Liu, and Jie Zhang. 2021. A taxonomical review on recent artificial intelligence applications to PV integration into power grids. *International Journal of Electrical Power & Energy Systems* 132 (2021), 107176.
- [6] Daniel Gebbran, Sleiman Mhanna, Yiju Ma, Archie C Chapman, and Gregor Verbič. 2021. Fair coordination of distributed energy resources with Volt-Var control and PV curtailment. *Applied Energy* 286 (2021), 116546.
- [7] Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel, and Sergey Levine. 2018. Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor. In *Proceedings of the 35th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 80)*. PMLR, 1861–1870.
- [8] Changfu Li, Yi-An Chen, Chenrui Jin, Ratnesh Sharma, and Jan Kleissl. 2022. Online PV Smart Inverter Coordination using Deep Deterministic Policy Gradient. *Electric Power Systems Research* 209 (2022), 107988.
- [9] Eric O’Shaughnessy, Jesse R Cruce, and Kaifeng Xu. 2020. Too much of a good thing? Global trends in the curtailment of solar PV. *Solar Energy* 208 (2020), 1068–1077.
- [10] Qiuyu Peng and Steven H Low. 2016. Distributed optimal power flow algorithm for radial networks, I: Balanced single phase case. *IEEE Transactions on Smart Grid* 9, 1 (2016), 111–121.
- [11] Elizabeth L Ratnam, Steven R Weller, Christopher M Kellett, and Alan T Murray. 2017. Residential load and rooftop PV generation: an Australian distribution network dataset. *International Journal of Sustainable Energy* 36, 8 (2017), 787–806.
- [12] Felix Rafael Segundo Sevilla, David Parra, Nicolas Wyrsh, Martin K Patel, Florian Kienzle, and Petr Korba. 2018. Techno-economic analysis of battery storage and curtailment in a distribution grid with high PV penetration. *Journal of Energy Storage* 17 (2018), 73–83.
- [13] Naomi Stringer, Navid Haghdadi, Anna Bruce, and Iain MacGill. 2021. Fair consumer outcomes in the balance: Data driven analysis of distributed PV curtailment. *Renewable Energy* 173 (2021), 972–986.
- [14] Sam Weckx, Carlos Gonzalez, Jeroen Tant, Tom De Rybel, and Johan Driesen. 2012. Parameter identification of unknown radial grids for theft detection. In *2012 3rd IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe)*. IEEE, 6 pages.
- [15] Christopher Yeh, Jing Yu, Yuanyuan Shi, and Adam Wierman. 2022. Robust online voltage control with an unknown grid topology. In *Proceedings of the Thirteenth ACM International Conference on Future Energy Systems*. 240–250.
- [16] Zhanqiang Zhang, Chunxia Dou, Dong Yue, Bo Zhang, and Peng Zhao. 2021. High-economic PV power compensation algorithm to mitigate voltage rise with minimal curtailment. *International Journal of Electrical Power & Energy Systems* 125 (2021), 106401.
- [17] Weifeng Zhong, Kan Xie, Yi Liu, Chao Yang, and Shengli Xie. 2018. Topology-aware vehicle-to-grid energy trading for active distribution systems. *IEEE Transactions on Smart Grid* 10, 2 (2018), 2137–2147.

A APPENDICES

A.1 Illustrated fairness definitions

Fig. 4 illustrates the differences between these definitions.

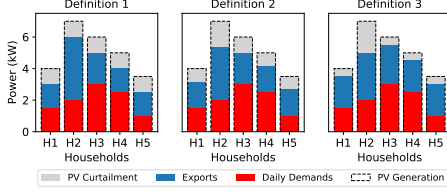


Figure 4: Differences between fairness definitions.

In Fig. 4, the dashed boxes represent the total amount of PV generation, which is split between curtailed (grey) and consumed (red). The remainder of the generation (blue) is sold to the grid.

A.2 Optimization model

For the optimization model, we aim to minimize the cost of electricity while ensuring safety and improving fairness. We modeled this optimization problem as a linear programming problem and used the MiniZinc solver. Our objective function includes the cost of electricity consumption and the penalty due to unfair curtailments. Thus, we can tradeoff the cost minimization and fairness maximization by the weight β in this objective function. The optimization problem is formulated as

$$\begin{aligned}
 \min \quad & \sum_{h \in \mathcal{H}} \sum_{t \in \mathcal{T}} c_t \times (p_{h,t}^L - p_{h,t}^G \times (1 - \alpha_{h,t}^C)) - \beta f(\alpha_t^C), \\
 \text{s.t.} \quad & 0 \leq \alpha_t^C \leq 1, \\
 & p_{h,t} = p_{h,t}^G \times (1 - \alpha_{h,t}^C) - p_{h,t}^L, \\
 & p_{h+1,t}^{br} = p_{h,t}^{br} + p_{h,t}, \\
 & q_{h+1,t}^{br} = q_{h,t}^{br} + q_{h,t}, \\
 & v_{h+1,t} = v_{h,t} - (r_{h+1,t} p_{h+1,t}^{br} + x_{h+1,t} q_{h+1,t}^{br}) / v_{0,t}, \\
 & 0.95 \leq v_{h,t} \text{ (p.u)} \leq 1.05, \\
 & q_{h,t} = \bar{q} \times \mathbf{1}(v_{h,t} \leq V_1) + \bar{q} \frac{V_2 - v_{h,t}}{V_2 - V_1} \times \mathbf{1}(V_1 < v_{h,t} < V_2), \\
 & -\bar{q} \frac{v_{h,t} - V_3}{V_4 - V_3} \times \mathbf{1}(V_3 < v_{h,t} < V_4) - \bar{q} \times \mathbf{1}(v_{h,t} \geq V_4),
 \end{aligned} \tag{22}$$

where β is the weight for fairness penalties.

A.3 Learning curves

We demonstrate the training process of the SAC model with learning curves, as shown in Fig. 5. Each learning curve shows how the per-step reward increases throughout a single run of the training process. The reward is defined in (21), taking different forms according to the type of fairness definitions and instant/accumulative cases. The reward value means the sum of the ratio of PV generation after curtailment to the generation before curtailment for all households, minus the penalties. The maximum reward per step of the policy is 10, which means no curtailment for each of the 10 households and no penalties.

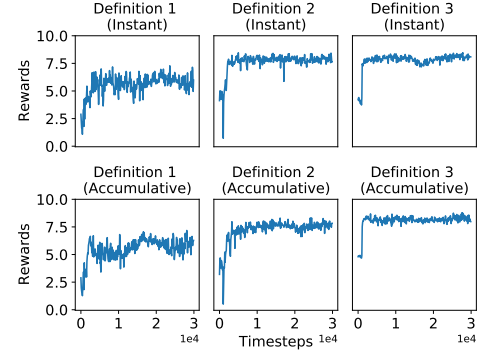


Figure 5: Learning curves for different fairness cases.

A.4 Learned curtailment policies

We complement Fig. 2 with heat maps to show exported and curtailed PV of households under different fairness cases. Since PV generation is not available at all times of the day, we only selected the time period with the highest PV production (8:00 - 15:00) to plot heat maps.

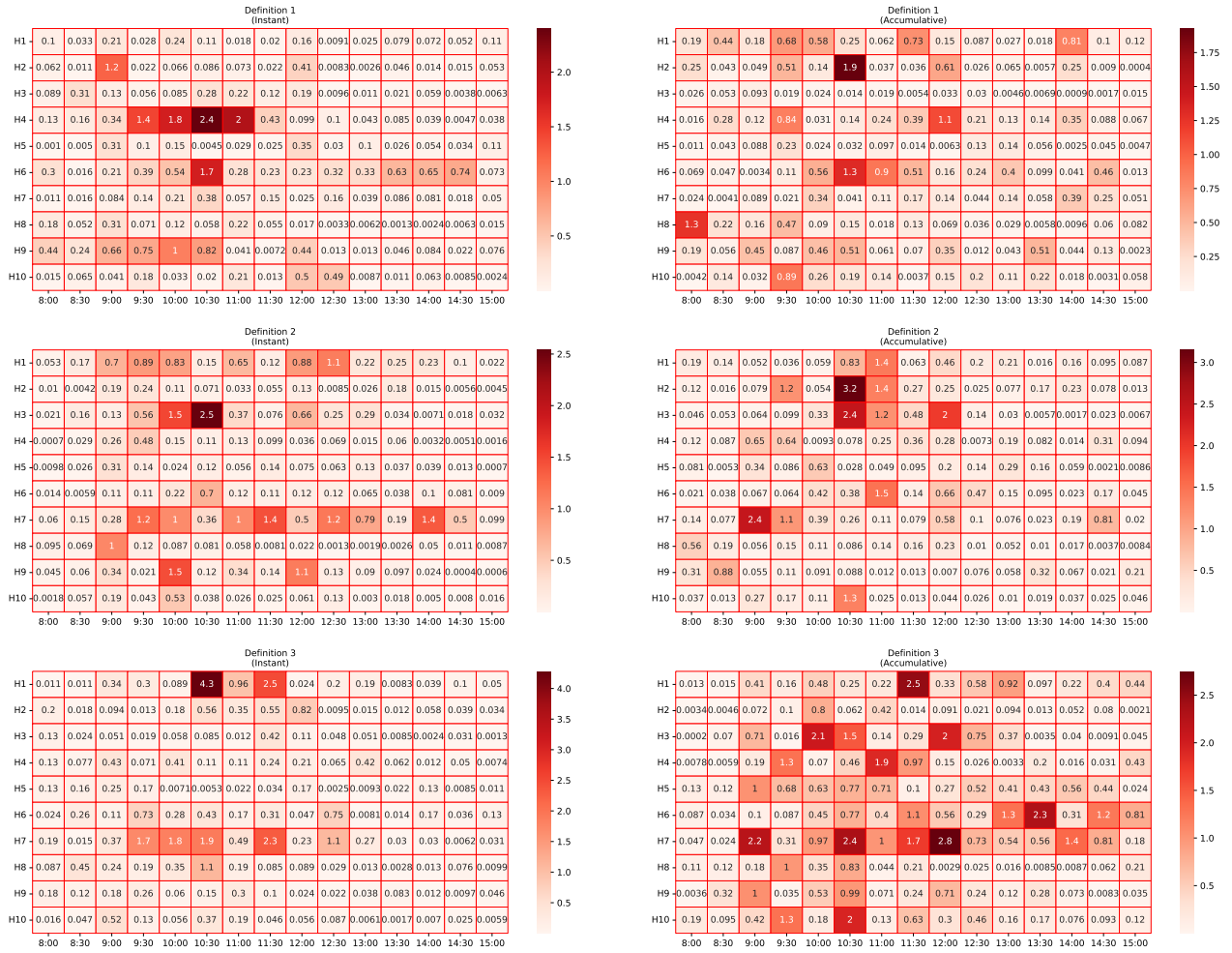


Figure 6: PV curtailments during 8:00 - 15:00.

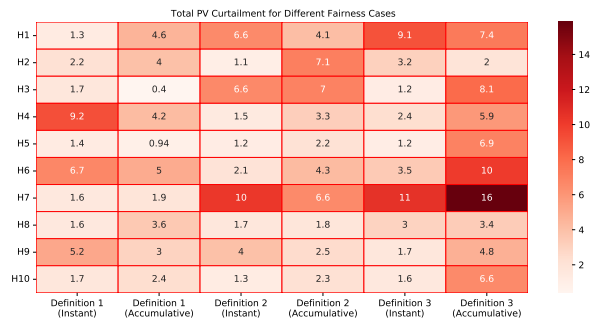


Figure 7: Total PV curtailments for the whole day.

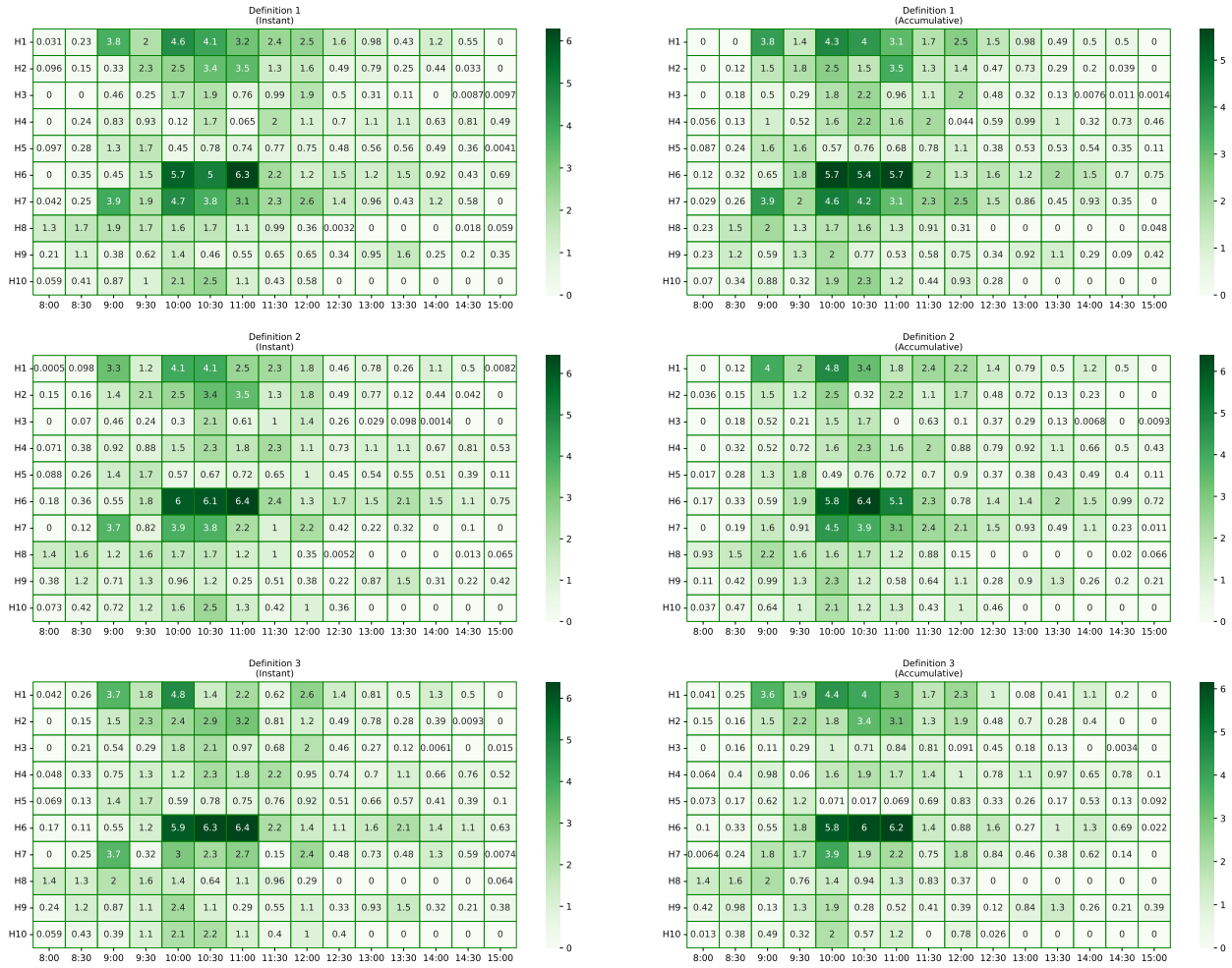


Figure 8: Exported PV during 8:00 - 15:00.

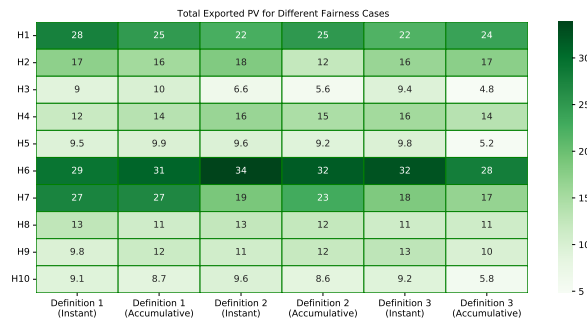


Figure 9: Total exported PV for the whole day.