SIEMENS

Master's Thesis

Autonomous Control of Low Voltage Grids Based on Data-Driven State Estimation and Reinforcement Learning

In partial fulfillment of the requirements for the degree Master of Science at the School of Computation, Information and Technology of the Technical University of Munich.

Advisors	Dr. Mathias Duckheim Dr. Edwin Camilo Mora Gil M.Sc. Jano Schubert Siemens AG Prof. Dr. rer. nat. Thomas Hamacher M.Sc. Anurag Mohapatra Chair of Renewable and Sustainable Energy Systems
Submitted by	Hakan Özlemiş
Submitted on	Munich, October, 16, 2023

Abstract

Increasing penetration of Distributed Energy Resources (DERs) in Low-Voltage (LV) grids necessitates novel, reliable control strategies combining grid monitoring and automated decision-making. However, this chained control task poses a significant challenge to Distribution System Operators (DSOs) due to the lack of accurate electrical grid models and automation infrastructure at the LV level. This thesis presents a model-free, data-driven control algorithm for voltage violation mitigation in LV grids, without requiring prior knowledge of physical models or expensive Real-Time (RT) measurements. The algorithm relies on RT measurements and historical Smart-Meter (SM) data to operate the grid within its limits. The method involves a model-free, data-driven State Estimation (SE) and an online-trained Reinforcement Learning (RL) agent. Based on estimated voltages, the RL agent decides on appropriate power factors and curtailment signals for dispatchable DERs to mitigate voltage violations. Simulation experiments show that the controller reduces voltage violations by 98% compared to two baseline algorithms and can handle uncertainties arising from SE deviations. The performance and stability of the controller is improved by extending the RL agent's observation space with multiple time steps using one-dimensional (1D) Convolutional Neural Networks (CNNs). Additionally, this work shows the effectiveness of the controller in low SM penetration scenarios and highlights the importance of SM locations. Overall, this thesis demonstrates that controlling LV grids reliably and cost-effectively with model-free, data-driven methods is feasible.

keywords - **autonomous power systems, data-driven, grid control, reinforcement learning, state estimation**

Statement of Academic Integrity

I,

Last name: Özlemiş
First name: Hakan
ID No.: 03702655

hereby confirm that the attached thesis,

Autonomous Control of Low Voltage Grids
Based on Data-Driven State Estimation and
Reinforcement Learning

was written independently by me without the use of any sources or aids beyond those cited, and all passages and ideas taken from other sources are indicated in the text and given the corresponding citation.
I confirm to respect the "Code of Conduct for Safeguarding Good Academic Practice and Procedures in Cases of Academic Misconduct at Technische Universität München, 2015", as can be read on the website of the Equal Opportunity Office of TUM.
Tools provided by the chair and its staff, such as models or programs, are also listed. These tools are property of the institute or of the individual staff member. I will not use them for any work beyond the attached thesis or make them available to third parties.

I agree to the further use of my work and its results (including programs produced and methods used) for research and instructional purposes.

I have not previously submitted this thesis for academic credit.

Munich, October, 16, 2023 _____

Declaration for the Transfer of the Thesis

I agree to the transfer of this thesis to:

- Students currently or in future writing their thesis at the chair:
  - ☐ Flat rate by employees
  - ☐ Only after particular prior consultation.
- Present or future employees at the chair
  - ☐ Flat rate by employees
  - ☐ Only after particular prior consultation.

My copyright and personal right of use remain unaffected.

Munich, October, 16, 2023 _____

Contents

List of Abbreviations and Acronyms

1D	one-dimensional
AC	Alternative Current
a.m.	ante meridiem (before midday)
ANN	Artificial Neural Network
AOVD	Average Overvoltage Violation Duration
API	Application Programming Interface
ASC	Active Sign Convention
AUVD	Average Undervoltage Violation Duration
CAMS	Copernicus Atmosphere Monitoring Service
CED	Change in EV Demand
CHD	Change in HP Demand
CI/CD	Continuous Integration and Continuous Delivery
CNN	Convolutional Neural Network
CPIL	Change in PV Inverter Loading
CPO	Change in PV Output
CVI	Critical Voltage Increase
DC	Direct Current
DER	Distributed Energy Resource
DSO	Distribution System Operator
EV	Electric Vehicle
FO	Feedback Optimization
HH	Household
HP	Heat Pump
IOV	Integral of Overvoltage Violation
IUV	Integral of Undervoltage Violation

KCL	Kirchhoff's Current Law
KPI	Key Performance Indicator
KVL	Kirchhoff's Voltage Law
LV	Low Voltage
MD	Maximum Deviation
MDP	Markov Decision Process
MLP	Multilayer Perceptron
MRD	Maximum Relative Deviation
NCVI	No Critical Voltage Increase
NR	Newton-Raphson
OPF	Optimal Power Flow
OV	Overvoltage
p.m.	post meridiem (after midday)
POA	Plane of Array
PSC	Passive Sign Convention
p.u.	per unit
PV	Photovoltaic
RGS	Random Grid Search
RL	Reinforcement Learning
RMSE	Root Mean Squared Error
RT	Real-Time
SAC	Soft Actor-Critic
SCADA	Supervisory Control and Data Acquisition
SE	State Estimation
SM	Smart-Meter
TDD	Test-Driven Development
TOVD	Total Overvoltage Violation Duration
TUVD	Total Undervoltage Violation Duration
UML	Unified Modeling Language
UTC	Coordinated Universal Time
UV	Undervoltage
WLS	Weighted Least Square

List of Symbols

Sets

$\mathcal{N} := \{0, \ldots, N\}$	Set of grid nodes
$\mathcal{N}_{\text{SM}}$	Set of nodes equipped with a Smart-Meter (SM)
$\mathcal{F} := \{1, \ldots, F\}$	Set of grid feeders
$\mathcal{D} := \{\text{PV}, \text{EV}, \text{HP}\}$	Set of Distributed Energy Resource (DER) types
$\mathcal{S}_D$	Set of DER indices of type $D \in \mathcal{D}$
$\mathcal{S}_D^{\text{u}}$	Set of controllable DER indices of type $D \in \mathcal{D}$

Scalars

t	Time step
$n \in \mathcal{N}$	Node index
$P_{n,t}$	Active power injection at node n at time step t in [kW]
$Q_{n,t}$	Reactive power injection at node n at time step t in [kVar]
$V_{n,t}$	Absolute voltage magnitude at node n at time step t in [V]
$v_{n,t}$	Relative voltage magnitude at node n at time step t in [p.u.]
$\varphi_{n,t}$	Phase angle at node n at time step t in [rad]
G_{nm}	Conductance of the line between buses n and m in [S]
B_{nm}	Susceptance of the line between buses n and m in [S]
$P_{i \to j,t}$	Active power flow from node i to node j at time step t in [kW]
$Q_{i \to j,t}$	Reactive power flow from node i to node j at time step t in [kVar]
$P_t^{\text{HH}_i}$	Active power demand of Household (HH) i at time step t in [kW]
$Q_t^{\text{HH}_i}$	Reactive power of HH i at time step t in [kVar]
P_t^d	Active power of a DER $d \in \mathcal{S}_D$ in type of $D \in \mathcal{D}$ at time step t in [kW]
Q_t^d	Reactive power of a DER $d \in \mathcal{S}_D$ in type of $D \in \mathcal{D}$ at time step t in [kVar]
P_{inst}^d	Installed power capacity of a DER $d \in \mathcal{S}_D$ in type of $D \in \mathcal{D}$ in [kW]
Q_{max}^d	Maximum reactive power potential of inverter of a DER $d \in \mathcal{S}_D$ in type of $D \in \mathcal{D}$ in [kVar]

$u_t^{\mathrm{PV,p}}$	Control signal for active power curtailment of Photovoltaics (PVs) at time step t
$u_t^{\mathrm{PV,q}}$	Control signal for reactive power control of PVs at time step t
$u_t^{\mathrm{EV,p}}$	Control signal for active power curtailment of Electric Vehicles (EVs) at time step t
$u_t^{\mathrm{HP,p}}$	Control signal for active power curtailment of Heat Pumps (HPs) at time step t
$P_t^{d,\mathrm{u}}$	Controlled active power of a controllable DER $d \in \mathcal{S}_D^{\mathrm{u}}$ in type of $D \in \mathcal{D}$ at time step t in [kW]
$Q_t^{d,\mathrm{u}}$	Controlled reactive power of a controllable PV $d \in \mathcal{S}_{\mathrm{PV}}^{\mathrm{u}}$ at time step t in [kVar]
r_t	Reinforcement Learning (RL) reward at time step t
w^{ov}	Penalty weight for Overvoltage (OV) instance
w^{uv}	Penalty weight for Undervoltage (UV) instance
$w^{\mathrm{PV,p}}$	Penalty weight for curtailing PV
$w^{\mathrm{PV,q}}$	Penalty weight for exploiting reactive power potential of PV inverters
$w_{\mathrm{NCVI}}^{\mathrm{PV,q}}$	Penalty weight for exploiting reactive power potential of PV inverters despite 'No Critical Voltage Increase (NCVI)'
$w^{\mathrm{EV,p}}$	Penalty weight for curtailing EV
$w^{\mathrm{HP,p}}$	Penalty weight for curtailing HP
v_{CVI}	Critical Voltage Increase (CVI) indication limit in [p.u.]
$v_{\max}$	Maximum nodal voltage magnitude limit in [p.u.]
$v_{\min}$	Minimum nodal voltage magnitude limit in [p.u.]

Vectors

$\mathbf{z}_t$	Vector of Real-Time (RT) measurements at time step t
$\mathbf{e}_t$	Vector of SM measurements at time step t
$\hat{\mathbf{e}}_t$	Vector of estimated SM measurements at time step t
$\mathbf{v}_t$	Vector of measured and estimated relative voltage magnitudes at time step t
$\mathbf{s}_t$	Vector of RL states at time step t
$\mathbf{a}_t$	Vector of RL actions at time step t

Functions

$\pi(\cdot)$	RL policy function
$\pi^*(\cdot)$	RL optimal policy function
$V(\cdot)$	RL state-value function
$r(\cdot)$	RL reward function
$R(\cdot)$	RL total discounted return function
$Q(\cdot)$	RL action-value function (Q-function)

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Distributed Energy Resources (DERs) are small-scale power generation and consumption units located near end-users and connected to the distribution grid. They encompass a variety of technologies, including but are not limited to Photovoltaic (PV) systems (or simply PVs), wind turbines, battery storage systems, Electric Vehicle (EV) chargers (or simply EVs), Heat Pumps (HPs) [1].

While DERs offer many advantages, such as decarbonization, reduced energy costs, and flexibility, their integration into distribution grids poses significant challenges [1, 2]. One of the key obstacles is the low predictability and lack of control over generation and consumption patterns associated with small-scale units [3]. Moreover, the rapid expansion of distributed generation and sector-coupled energy demands, characterized by their high simultaneity, can lead to voltage limit violations and overloading of power lines and substation transformers. Consequently, conventional grid management technologies must be adapted to ensure reliable and cost-effective distribution grids.

To achieve this adaptation, an active distribution grid management strategy is necessary, which includes grid monitoring and control of DERs. However, monitoring in Low Voltage (LV) grids is challenging due to limited Real-Time (RT) sensing and communication infrastructure [4]. RT monitoring is critical for effective control decisions, but incomplete knowledge of the grid model, including missing or inaccurate topology data, line parameter mismatches, and possibly unknown behind-the-meter DER installations, makes the grid control task more challenging [5, 6].

Developing a control strategy that can effectively manage the challenges associated with integrating DERs into LV grids is essential. This strategy should rely on a minimal amount of measurement and communication infrastructure and not depend on a power grid model. Its primary objective should be to ensure sustainable, reliable, and cost-effective grid operation.

1.2 Related Work

In recent years, there has been a growing interest in DER management systems. Research has relied on assumptions of (near-)perfect knowledge of the grid model and/or availability of RT measurements downstream to LV substation, such as Op-

timal Power Flow (OPF) [7]. The authors of [8, 9] propose a Feedback Optimization (FO) scheme that is based on model-based State Estimation (SE) to eliminate dependency on wide RT measurement infrastructure. However, limitations in real-world applications due to imperfect models and lack of downstream RT measurements have prompted researchers to explore the potential of data-driven control techniques like Reinforcement Learning (RL) in power system control [10].

To mitigate voltage issues in distribution grids, [11] deploys RL to learn Volt-Var control using transformer tap changers, voltage regulators, and capacitor banks. [12, 13] extend control flexibility by leveraging the reactive power potential of distributed smart inverters in an RL scheme. The studies in [14–16] investigate the use of RL for energy management of PVs, EVs, HPs, and energy storage systems. Although such RL schemes are model-free regarding knowledge of underlying system model, they often require RT observation of the system state. Therefore, RL cannot be used alone for active grid management without prior grid monitoring.

As conventional grid monitoring approaches require a grid model and costly downstream RT measurement infrastructure in LV grids [17], the authors of [18] propose a model-free, data-driven SE method that uses a Multilayer Perceptron (MLP) to map RT data (e.g., substation measurements, meteorological data) to historical grid data (e.g., downstream voltage magnitude, nodal power injections), which are only available after daily or weekly synchronization of Smart-Meter (SM) measurements with the database of the Distribution System Operator (DSO). However, these approaches do not consider the combined impact of grid monitoring and control limitations on both the model and RT observation. This thesis is motivated by addressing these limitations holistically.

The aforementioned state-of-the-art and the position of this thesis is illustrated in Figure 1.1. The x-axis of the figure represents the level of knowledge on the grid model and the availability of downstream RT measurements, while the y-axis differentiates between two perspectives of grid automation, i.e., grid monitoring and grid control. This thesis aims to fill the gap in autonomous LV grid control that does not depend on a grid model and costly downstream RT measurements. For a detailed comparison of the practical requirements for implementing the control algorithm introduced in this thesis, please refer to Table 3.1.

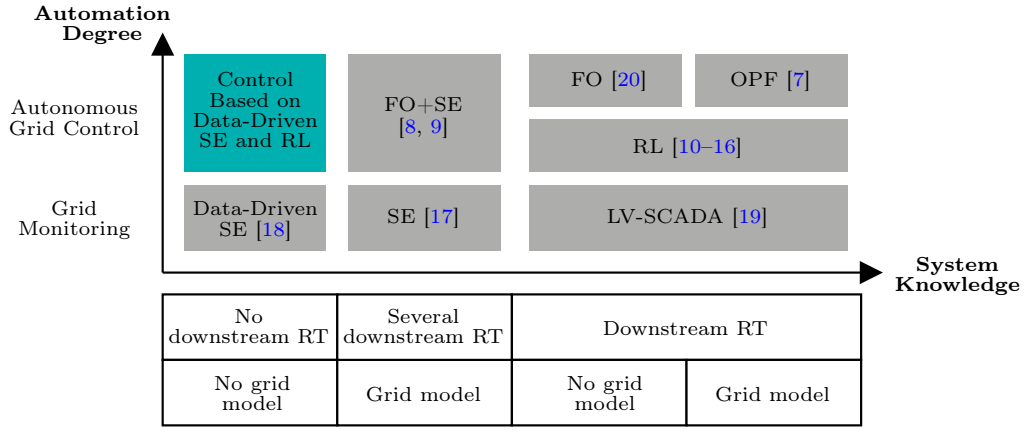


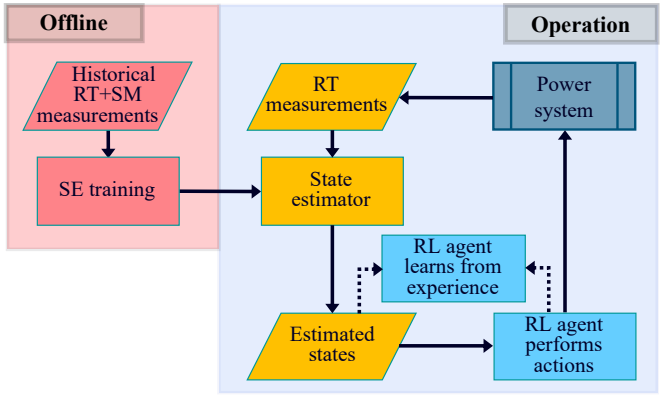
Figure 1.1 State-of-the-art in grid automation (gray) and this thesis' position (green)

1.3 Contribution

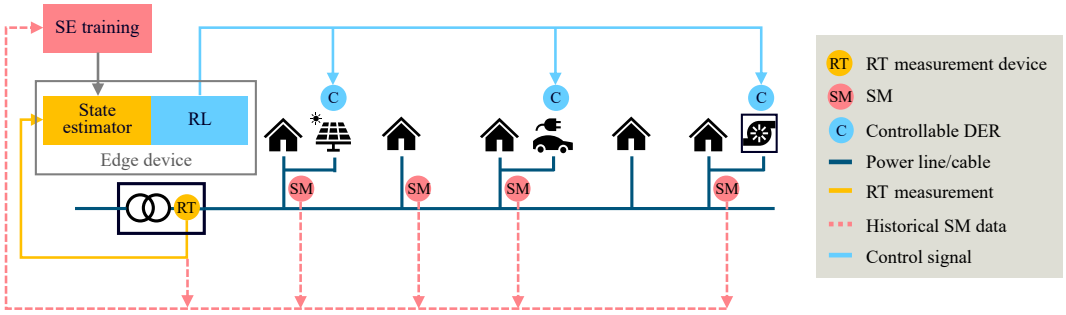
This thesis investigates a novel fully data-driven control algorithm for operating an LV grid that leverages RT substation and historical SM measurements to determine suitable power factors for PVs and active power curtailment factors for PVs, EVs, and HPs. As shown in Figure 1.2, the algorithm combines a data-driven state estimator with a model-free RL algorithm to maintain the voltage within the grid limits under varying conditions.

The state estimator is trained offline to predict downstream voltage magnitudes at nodes equipped with SMs based on the RT measurements at distribution substation. During RT grid operation, the RT measurements and the state estimator's output serve as inputs to the central grid controller that is responsible for minimizing voltage band violations with minimal DER interventions. This thesis focuses on using RL as the control algorithm that is trained during operation (online). The RL controller incurs penalties for estimated voltage violations and applies control signals to dispatchable DERs. The controller manages DERs through active power curtailment of PVs, EVs, and HPs, as well as reactive power control of PVs.

Furthermore, this thesis introduces a simulation framework to model grid control scenarios, enabling the development and evaluation of control algorithms. The framework combines time series power flow simulation, state estimation, and control algorithms in a single software tool. An RL-based control algorithm is introduced and extensively evaluated using this simulation framework. To conduct evaluations,



(a) Flow diagram



(b) Physical layer

Figure 1.2 Proposed control algorithm

a test setup is prepared that includes a future DER penetration scenario and DER profiles for a year with a resolution of 1 min.

The performance of the developed control algorithm in mitigating Overvoltage (OV) and Undervoltage (UV) problems is demonstrated by comparing it to baseline control algorithms. Moreover, the core concept is enriched by incorporating temporal learning through one-dimensional (1D) Convolutional Neural Networks (CNNs). The effectiveness of the control algorithm is also evaluated at lower SM penetration levels, and its sensitivity to the location of SMs is examined in this study.

To be more concise, the main contributions of this thesis can be summarized as follows:

1. A simulation framework to develop and evaluate grid control algorithms
2. A model-free control algorithm for LV grid operation that combines data-driven SE and RL without requiring a grid model or downstream RT measurements
3. Performance evaluation against baseline controllers and sensitivity analysis for two placement scenarios at a low SM penetration level
4. Introduction of 1D-CNNs for RL in grid control applications

1.4 Overview

Chapter 2 provides a technical background on electrical power system models, power system SE, and RL. In Chapter 3, the control problem and the proposed fully data-driven approach are described in detail. The software framework developed in this thesis is presented in Chapter 4. Chapter 5 contains the test and evaluation setup, while the results are presented in Chapter 6. Finally, Chapter 7 summarizes the main findings and provides an outlook on possible future work.

Chapter 2

Technical Background

This chapter provides a technical background on the main disciplines that are related to this thesis. The first subsection discusses electrical power system modeling, which is an essential aspect of power system analysis. The second subsection covers State Estimation (SE), which is a critical process used to estimate the underlying state of a power system based on limited measurements. The third and fourth subsections are dedicated to Markov Decision Process (MDP) and Reinforcement Learning (RL), which are widely used methodologies in control theory, including power system control.

2.1 Electric Power System Modeling

Power system models are used to analyze and predict the behavior of real power systems. These models consist of various components, including buses (also known as nodes), transformers, transmission lines, switches, shunts, electricity consumers and producers, and/or storage units. Considering that the power system dynamics and transient analysis are beyond the scope of this work, it is postulated that the power system is in steady-state. At steady-state, power systems can be modeled via a set of nonlinear algebraic equations. In light of this and disregarding time notation, the power flow equations for a network comprising $N+1$ buses can be written as in [21, p. 12]:

$$V_0 = V_{\text{slack}} \quad , \quad \varphi_0 = 0^\circ \tag{2.1a}$$

$$P_n = \sum_{m=1}^{N} V_n V_m (G_{nm} \cos\varphi_{nm} + B_{nm} \sin\varphi_{nm}) \tag{2.1b}$$

$$Q_n = \sum_{m=1}^{N} V_n V_m (G_{nm} \sin\varphi_{nm} - B_{nm} \cos\varphi_{nm}) \tag{2.1c}$$

$$S_n = P_n + \mathrm{j} Q_n \tag{2.1d}$$

where:

- the bus 0 is the slack node which is the reference node, having a fixed voltage V_{slack} and being the reference zero of phase angle.
- P_n and Q_n are the net active and reactive power injection at bus n.

- S_n is the net apparent power injection at bus n, and j denotes here the imaginary unit.
- V_n and φ_n are the nodal voltage magnitude and angle at bus n.
- G_{nm} and B_{nm} are the conductance and susceptance of the line between buses n and m.

LV grids are characterized by their radial topology, greater complexity in terms of number of delivery points, greater R/X ratios than transmission systems which makes the LV grids also more susceptible for voltage drop or rise. In addition, LV grids provide broad opportunities for the integration of DERs, such as PV, EV, HP. One other main difference of LV grids is that there are no synchronous generators which means there are no nodes with a fixed active power output and voltage magnitude (also known as *PV nodes*). All nodes except the slack node are assumed to have fixed active and reactive power outputs (also known as *PQ nodes*). The slack node has a fixed voltage magnitude and phase angle which builds up a reference of the grid, as stated in Equation (2.1a). Power injections at a bus n (at which an arbitrary collection of one or several Households (HHs), PVs, EVs, HPs might be located) are formulated as follows:

$$P_n = -\sum_{i:\,\mathrm{HH}_i \text{ at } n} P^{\mathrm{HH}_i} + \sum_{i:\,\mathrm{PV}_i \text{ at } n} P^{\mathrm{PV}_i} - \sum_{i:\,\mathrm{EV}_i \text{ at } n} P^{\mathrm{EV}_i} - \sum_{i:\,\mathrm{HP}_i \text{ at } n} P^{\mathrm{HP}_i} \tag{2.2a}$$

$$Q_n = -\sum_{i:\,\mathrm{HH}_i \text{ at } n} Q^{\mathrm{HH}_i} + \sum_{i:\,\mathrm{PV}_i \text{ at } n} Q^{\mathrm{PV}_i} - \sum_{i:\,\mathrm{EV}_i \text{ at } n} Q^{\mathrm{EV}_i} - \sum_{i:\,\mathrm{HP}_i \text{ at } n} Q^{\mathrm{HP}_i} \tag{2.2b}$$

In that way, the left-hand side of Equations (2.1b) and (2.1c) are modeled with Equations (2.2a) and (2.2b), respectively. It should be noted that load units, i.e., HHs, EVs, HPs, are in Passive Sign Convention (PSC), whereas the generators, i.e. PVs, in Active Sign Convention (ASC). As per the formulation of the bus "injections", the loads have a negative sign and the generators have a positive sign above.

The nonlinear power flow Equations (2.1) are typically solved by iterative algorithms such as the Newton-Raphson (NR) method. NR method exploits the Jacobian matrix which contains the partial derivatives of power flow equations. It is used to update the unknown variables of the power flow equations until convergence is reached. [22]

2.1.1 Voltage Fluctuation Along a Feeder

This section demonstrates the variation of voltage on a distribution line, encompassing both the rise and drop phenomena, by employing mathematical formulations of a line and power transfer to the end of this line.

To understand these phenomena, the single-phase π-equivalent circuit of power lines is considered, as shown in Figure 2.1 [23]. The electrical components in the circuit describe the parasitic effects with the external environment, such as coupling capacitance and inductance, as well as the finite conductivity of the lines.

In low and medium voltage networks, simplification of the circuit in Figure 2.1 is justified as the lines are short with respect to the transferred power. The parallel elements $\frac{G}{2} = \frac{G'l}{2}$ and $\frac{\mathrm{j}B}{2} = \frac{\mathrm{j}\omega C'l}{2}$ have relatively low values since the line length l is

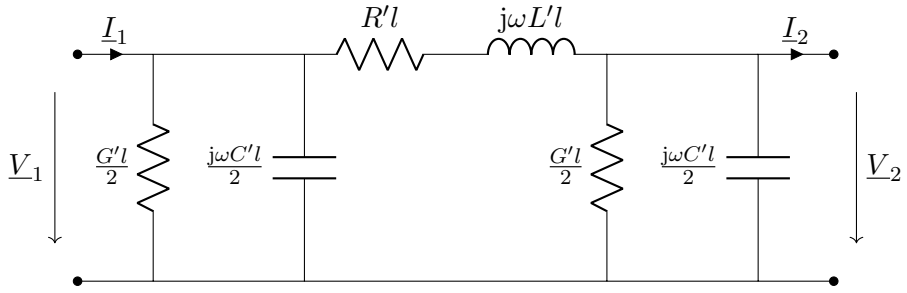


Figure 2.1 π-equivalent power line model

relatively small. As the parallel elements with vanishing conductance are connected in parallel with the terminals of the line, they can be disregarded. Thus, a power line can be modeled using a series resistance R and a series reactance X, which are dependent on the length of the line l as well as the primary impedance constants of the line: resistance per unit length R' and inductance per unit length L'. The resulting simplified equivalent circuit is shown in Figure 2.2. [24, p. 26]

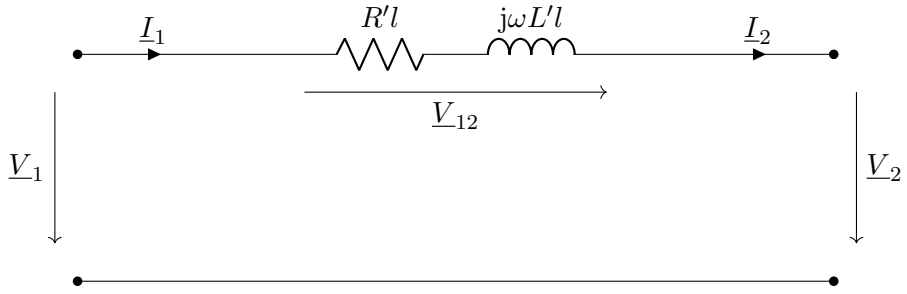


Figure 2.2 Simplified π-equivalent power line model

As a result of to Kirchhoff's Voltage Law (KVL), the voltage change over the power line is denoted by:

$$\begin{aligned} \underline{V}_{12} &= \underline{V}_1 - \underline{V}_2 \\ &= \Delta V + \mathrm{j}\delta V \end{aligned} \tag{2.3}$$

where the latter formulation is the rectangular form of the complex $\underline{V}_{12}$.

As the voltages $\underline{V}_1$ and $\underline{V}_2$ are compared relatively to each other, it can be assumed $\underline{V}_2$ has only real part, i.e., $\underline{V}_2 = V_2 \in \mathbb{R}$. This assumption allows for easier determination of the direction of active and reactive powers, P_2 and Q_2, at the end of the line by considering only the direction of real and imaginary parts of the current $\underline{I}_2$, denoted as $I_{2,r}$ and $I_{2,i}$, respectively. By virtue of the Kirchhoff's Current Law (KCL), it is evident that $\underline{I}_1 = \underline{I}_2$. Therefore, the subscript notation can be omitted, i.e., $\underline{I} = \underline{I}_1 = \underline{I}_2$.

The apparent power at the line end can be expressed in PSC as follows:

$$\begin{aligned} \underline{S}_2 &= P_2 + \mathrm{j}Q_2 \\ &= V_2 \underline{I}_2^* = V_2(I_r + \mathrm{j}I_i)^* = V_2(I_r - \mathrm{j}I_i) \end{aligned} \tag{2.4}$$

From Equation (2.4), P_2 and Q_2 can be formulated as:

$$\begin{aligned} P_2 &= V_2 I_r \\ Q_2 &= -V_2 I_i \end{aligned} \tag{2.5}$$

For a line with an impedance of $\underline{Z} = R + \mathrm{j}X$ transmitting complex current $\underline{I}$ to its end, voltage change Equation (2.3) can be rewritten as follows:

$$\begin{aligned} \Delta V + \mathrm{j}\delta V &= \underline{Z} \cdot \underline{I} \\ &= (R + \mathrm{j}X) \cdot (I_r + \mathrm{j}I_i) \\ &= (RI_r - XI_i) + \mathrm{j}(XI_r + RI_i) \end{aligned}$$

which implies:

$$\begin{aligned} \Delta V &= RI_r - XI_i \\ \delta V &= XI_r + RI_i \end{aligned} \tag{2.6}$$

Comparing the squared voltages $\left|\underline{V_1}\right|^2$ and V_2^2 allows for understanding whether the voltage at the end of the line increases or decreases compared to the voltage at the head of the line. The voltage $\underline{V_1}$ and its squared amplitude can be expressed using Equation (2.3):

$$\underline{V_1} = V_2 + (\Delta V + \mathrm{j}\delta V) \tag{2.7a}$$

$$\begin{aligned} \left|\underline{V_1}\right|^2 &= |V_2 + (\Delta V + \mathrm{j}\delta V)|^2 \\ &= |(V_2 + \Delta V) + \mathrm{j}\delta V|^2 \\ &= (V_2 + \Delta V)^2 + \delta V^2 \end{aligned} \tag{2.7b}$$

By substituting the equalities from (2.6) into Equation (2.7b), one obtains:

$$\begin{aligned} \left|\underline{V_1}\right|^2 &= (V_2 + RI_r - XI_i)^2 + (XI_r + RI_i)^2 \\ &= V_2^2 + 2V_2(RI_r - XI_i) + (R^2 + X^2)(I_r^2 + I_i^2) - 2RXI_rI_i + 2RXI_rI_i \\ &= V_2^2 + 2V_2(RI_r - XI_i) + (R^2 + X^2)(I_r^2 + I_i^2) \end{aligned} \tag{2.8}$$

The latter terms of Equation (2.8) that represent the quadratic voltage difference over the line can be expressed as a function f of V_2, I_r and I_i. A positive value of f indicates a voltage drop, while a negative value corresponds to a voltage rise.

$$\begin{aligned} f(\cdot) &= \left|\underline{V_1}\right|^2 - V_2^2 \\ &= 2V_2(RI_r - XI_i) + (R^2 + X^2)(I_r^2 + I_i^2) \end{aligned} \tag{2.9}$$

To derive the relationship between f and the real and imaginary parts of the current, the partial derivatives of f for a given V_2 are examined:

$$\begin{aligned} \frac{\partial f}{\partial I_r} &= 2V_2R + 2(R^2 + X^2)I_r \\ \frac{\partial f}{\partial I_i} &= -2V_2X + 2(R^2 + X^2)I_i \end{aligned} \tag{2.10}$$

To determine the sign of these derivatives, the constants can be replaced with their typical values. For example, the cable type “NAYY-J 4×150 SE” is commonly

used in German LV grids [25], which has a resistance per unit length of $R' = 0.206\,\frac{\Omega}{\mathrm{km}}$ and inductance per unit length of $L' = 0.08\,\frac{\Omega}{\mathrm{km}}$ at a grid frequency f = 50 Hz [26]. Additionally, according to [27, p. 177], the 95^{th} percentile of feeder lengths in Germany is equal to or less than 1.1 km. The voltage V_2 can be assumed to be the typical nominal line-to-ground voltage of European LV grids, i.e., $\frac{400}{\sqrt{3}}$V. Substituting these values into Equations (2.10) reveals that $\frac{\partial f}{\partial I_r}$ is positive if $I_r > -886\,\mathrm{A}$ and that $\frac{\partial f}{\partial I_i}$ is negative if $I_i < 344\,\mathrm{A}$. According to [26], the maximum thermal current for an underground "NAYY-J 4×150 SE" cable is 275 A. Therefore, both criteria are satisfied under normal conditions, leading to the relationship between f and the real and imaginary components of the current $\underline{I}$:

$$\begin{aligned} \frac{\partial f}{\partial I_r} > 0 \quad &\Rightarrow \quad f \propto I_r \\ \frac{\partial f}{\partial I_i} < 0 \quad &\Rightarrow \quad f \propto -I_i \end{aligned}$$

Further considering the relationship of P_2 and Q_2 with I_r and I_i in Equations (2.5), it can be verified that:

$$\begin{aligned} f &\propto P_2 \\ f &\propto Q_2 \end{aligned}$$

An increase in f corresponds to a greater voltage drop over the line, which can result from an increase in active and reactive power demands, P_2 and Q_2 (in PSC). In other words, increasing active power generation ($P_2 < 0$) increases the voltage at the feeder end, while increasing reactive power demand ($Q_2 > 0$) acts in the opposite direction, and decreases the voltage.

This phenomenon explains the undervoltage violations in highly loaded feeders as well as overvoltage violations in feeders with high power generation. However, reactive power feed-in or demand can support the grid voltage, correspondingly.

2.2 Power System State Estimation

SE is an essential tool for power system operation, monitoring, and control, as it provides important information for system operators to make informed decisions in RT. Accurate state estimation is crucial for maintaining the reliability of power systems and ensuring the safe and efficient operation of the grid. [28]

SE is a computational method that infers the state variables (e.g., voltage magnitude and phase angle) of a power system using a limited number of measurements within the power grid. It is fundamentally a numerical function $\mathbf{x} = g(\mathbf{z})$ that maps a vector of measured data $\mathbf{z}$ to a state vector $\mathbf{x}$.

The survey [17] exposes that the approaches based on *Weighted Least Square (WLS)* are commonly used for SE. The WLS approach expresses the available measurements $\mathbf{z}$ as a function of system states $\mathbf{x}$ and the measurement error vector $\boldsymbol{\epsilon}$.

$$\mathbf{z} = h(\mathbf{x}) + \boldsymbol{\epsilon}$$

where h represents the nonlinear power flow equations (see Section 2.1), and the $\boldsymbol{\epsilon}$ is the measurement error vector. The WLS suggests approximating the searched state

variables $\mathbf{x}$ by solving the following quadratic optimization problem:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} (\mathbf{z} - h(\mathbf{x}))^T \mathbf{W} (\mathbf{z} - h(\mathbf{x})) \tag{2.11}$$

where $\hat{\mathbf{x}}$ is the vector of estimated states, and $\mathbf{W}$ is the matrix *"that represents the user's confidence in the measured data"* [17].

Although SE is a well-established and frequently used concept in transmission systems, it remains an active research topic in distribution systems, especially for LV grids. Extending conventional SE approaches to active distribution grids poses significant challenges due to considerable differences between transmission and distribution systems, such as low RT observability, missing grid models or inaccuracies due to parameter mismatches [17, 29]. For instance, the optimization problem in Equation (2.11) requires power flow equations and a precise knowledge about the LV grid model including network topology, line lengths, line characteristics. However, this information is often unavailable to DSOs, and requires considerable effort for system identification [18]. Moreover, current LV grids lack sufficient RT capable measurement infrastructure at downstream nodes [4], which is crucial for SE. Consequently, model-based SE algorithms are not suitable for operation, monitoring, and control tasks related to today's LV grids.

2.3 Markov Decision Process

Markov Decision Process (MDP) is a mathematical framework that models decision-making problems where outcomes are partly stochastic and partly under the control of a decision-maker. MDP consists of a set of states, actions, and rewards. A *'state'* (s) is a specific condition or situation in which the decision-maker finds itself at a given point in time. An *'action'* (a) is a decision made by the decision-maker to make a transition from one state to another. Finally, a *'reward'* (r) is a numerical value associated with each state or action that reflects the desirability of that state or action. The changes between states are called transitions and the probability of moving from one state to another after taking an action is called *'transition probability'*. [30, p. 333–335]

Figure 2.3 shows a simple MDP diagram with three states (s_0, s_1, s_2) and two actions (a_0, a_1). Each arrow represents a state transition which is annotated with a set of transition probability and reward $\{p, r\}$. For example, from state s_0, there is a probability of 0.7 of transitioning to state s_1 with a reward of 10 by applying action a_0, and a probability of 0.3 of transitioning to state s_2 with a reward of -5 by applying action a_1. It is worth to mention that a negative reward can be interpreted as a penalty.

2.4 Reinforcement Learning

Reinforcement Learning (RL) is a type of machine learning that involves training a decision-maker called *'agent'* to make decisions based on feedback received from its *'environment'* which makes it different from supervised and unsupervised learning. RL is closely associated with MDP because MDP provides a framework for modeling state transition of the environment in which the agent operates.

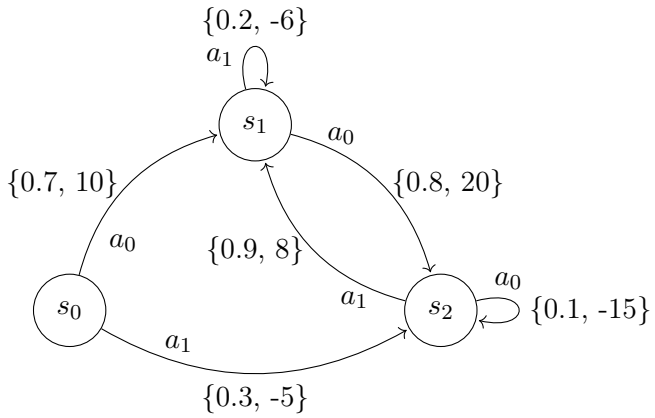


Figure 2.3 A simple MDP state diagram

RL offers several benefits over MDP, especially when the model is unknown. In MDP, it needs to be assumed that the model is known, which includes the transition probabilities and rewards. However, in many real-world scenarios, this assumption may not hold, and the consequences of the actions may not be known. RL manages this by learning from experience. In RL, the agent learns from experience by interacting with the environment. This means that RL can adapt to new situations and environments, and it can handle situations where the transition probabilities and rewards are unknown or uncertain [31, p. 124].

RL can also tackle non-Markovian decision processes, where the current state may not contain all the information necessary to make a good decision. MDP assumes that the current state contains all the information necessary to make a decision, but this is not always the case in real-world scenarios [31, p. 465]. Finally, numerous RL algorithms can deal with continuous state and action spaces, which is a significant advantage over conventional MDP. MDP assumes that the state and action spaces are discrete (or discretized) as in Figure 2.3, but in many real-world scenarios, these spaces may be continuous. RL can handle continuous state and action spaces by using function approximation techniques, such as neural networks [31, p. 322].

RL has two main types of functions which model the learning process and decision-making [32, p. 13-14][33]:

1. **Policy function:** Rule used by the agent to decide for an action to take in state s_t:

$$\pi(s_t) = a_t$$

2. **Value functions:**

   - **State-Value function:** Expected total discounted reward that an agent can expect to receive starting from a given state s_t:

$$V(s_t) = \mathbb{E}[R(t) \mid s_0 = s_t]$$

where the $R(t)$ is the discounted return function:

$$R(t) = r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \dots = \sum_{k=t}^{\infty} \gamma^{k-t} r_t$$

and γ is the discount factor which represents the relative importance of immediate and future rewards, and holds a value in range of $[0, 1]$.

- **Action-Value function (also known as *Q*-function):** Expected total discounted reward that an agent can expect to receive starting from a given state s_t and taking the action a_t:

$$Q(s_t, a_t) = \mathbb{E}[R(t) \mid s_0 = s_t,\, a_0 = a_t]$$

The relation between the state-value function and the action-value function is worth to note for better understanding. The action-value function is the expected reward from current state for a given specific action. However, the state-value function is the expected reward without imposing a specific action. It can be interpreted as the sum of all possible $Q(s_t, a')$ multiplied with the probability of taking the action a' at state s_t over all possible actions a' from the action space $\mathcal{A}$ [33]. The state-value function can be described in respect to the action-value function as follows:

$$V(s_t) = \mathop{\mathbb{E}}_{a' \in \mathcal{A}}[Q(s_t, a')]$$

The objective of the RL is to find an optimal policy π^* which maximizes the expected return $R(.)$:

$$\pi^* = \arg\max_{\pi} \mathbb{E}_{\pi}[R(.)] \tag{2.12}$$

For instance, the optimal policy to the MDP in Figure 2.3 where a discrete decision-maker (i.e., an agent for RL problems) could decide for deterministic actions with the starting state s_0 would be:

$$\pi^*(s_0) = a_0,\ \pi^*(s_1) = a_0,\ \pi^*(s_2) = a_1$$

which leads to an obvious reward maximization by choosing positive rewards over negative rewards (penalties). The resulting state transitions for the case of infinite horizon, without loss of generality, would be:

$$s_0 \xrightarrow{a_0} s_1 \xrightarrow{a_0} s_2 \xrightarrow{a_1} s_1 \xrightarrow{a_0} s_2 \xrightarrow{a_1} \dots$$

Learning such an optimal policy occurs through the updating of either one or both of the function types stated before, value and policy functions. A taxonomic comparison of RL algorithms plays an important role to decide which algorithm to use. Section 2.7 of book [34] offers a comprehensive analysis of three primary learning schemes, namely 'value-based', 'policy-based', 'actor-critic', which are succinctly summarized below.

Algorithms that use gradient descent to update a value function are referred to as value-based algorithms, such as *Deep Q-Learning (DQN)*. These algorithms learn the values of actions and select the action with the best value. They are particularly

effective for discrete action spaces as they directly estimate the optimal value for each action. The value-based algorithms are also known for their computational efficiency and sensitivity to function approximation errors.

Conversely, algorithms that update the policy function are called policy-based algorithms, such as *Policy Gradient*. Such algorithms determine the optimal policy that returns the best action based on the current state. Unlike value-based algorithms, policy-based algorithms are better suited for continuous action spaces since the policy function is a continuous function. Policy-based methods are mostly characterized by their better learning stability and less sample efficiency.

Concurrent learning of both value and policy functions is also possible with so-called actor-critic algorithms such as *Deep Deterministic Policy Gradient (DDPG)*, *Soft Actor-Critic (SAC)*. These provide a trade-off between both learning schemes by combining strengths of both. [34, 35]

In this work, the focus is primarily on the SAC algorithm, which sets itself apart from other algorithms in several key perspectives. SAC is recently often preferred over the other algorithms due to its several advantages. First, it is suited for continuous control problems. Second, it is a more sample-efficient algorithm, which means it can learn faster from fewer samples. Third, it is more stable, which means it is less likely to suffer from oscillations or divergences during agent training. Finally, SAC is able to encourage exploration by using an entropy regularization term, and prevents premature convergence to suboptimal policies. This is in contrast to other algorithms like DDPG, which are prone to getting stuck in local optima. However, SAC also has some disadvantages, such as being more computationally expensive and having more hyperparameters to tune than other algorithms as it has a more complex learning structure. [36]

Figure 2.4 illustrates an overview of SAC algorithm which outlines its main components and basic information flow. SAC comprises three main components: 'actor', 'critic', and 'target critic'. These components collectively contribute to the algorithm's learning process, guided by random samples from the experience replay buffer. This buffer stores past state transitions, enabling efficient learning by revisiting and utilizing previous experiences.

The actor is responsible for learning the policy function. Its policy learning is inherently stochastic, setting it apart from deterministic policy algorithms. Utilizing a neural network, the actor learns both the mean μ and standard deviation σ of the normal distribution of an action given a specific state. In training mode, an action is sampled from that normal distribution. This approach facilitates exploration and adaptability in continuous action spaces. On the other side, in deployment mode, the actor network outputs a deterministic prediction that corresponds to the mean value output of the network. It is worth to note that the action distribution is a multivariate distribution of n independent variables if the action is n-dimensional.

The critic, on the other hand, focuses on learning the Q-function. This function contributes to the optimization objectives in the policy learning process. The critic's action-value estimations guide the learning of the policy network.

A unique feature of SAC is the presence of the target critic. The target critic serves to enhance learning stability by introducing a lag compared to the main critic network. SAC employs a "soft" update mechanism, gradually blending the parameters of the target critic network with those of the main critic network. This gradual transition ensures a balanced learning process, preventing abrupt changes that could

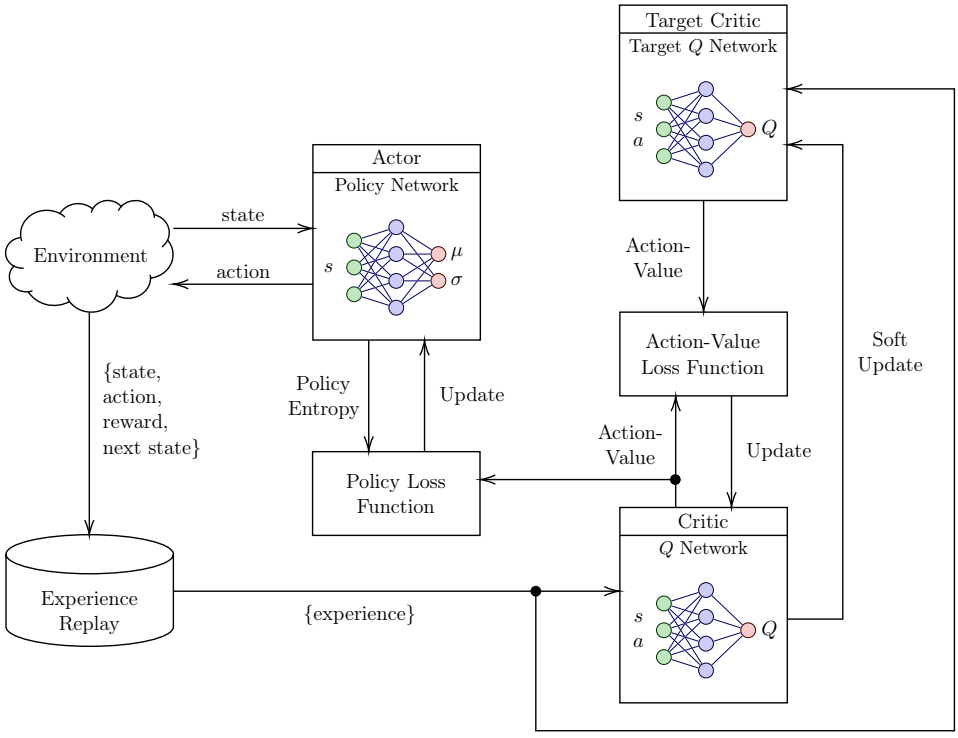


Figure 2.4 Overview of SAC algorithm (drawn based on [37, 38])

hinder stability. In this respect, an analogy can be made with the integral term of Proportional-Integral-Derivative (PID) controller in control theory.

In essence, SAC's actor-critic architecture, stochastic policy learning, and incorporation of a target critic all contribute to its effectiveness in solving complex problems with continuous action spaces [36].

Chapter 3

Methodology

3.1 Starting Point

As depicted in Figure 1.2b, the focus of this thesis is on a balanced, radial LV grid with $N+1$ nodes ($\mathcal{N} := \{0, \dots, N\}$) and F feeders ($\mathcal{F} := \{1, \dots, F\}$) connected to the substation. It is assumed that the DSO has incomplete knowledge of the grid's topology and its electrical parameters. Regardless, the DSO should operate the grid in a way that minimizes voltage band violations and DER interventions. To achieve this, the DSO aims to implement a central controller solution that makes use of the available automation infrastructure, which is described in the following.

3.1.1 Measurement Availability

Referring to [4, 17], it is assumed that the DSO measures a limited number of quantities in RT. Specifically, only the voltage magnitude at the substation's secondary side and the active and reactive power flow at each feeder-head are measured in RT. The RT measurements are assumed to have a time resolution of 1 min.

Without loss of generality, the substation is connected between nodes 0 and 1 where node 0 is the slack node. The 'relative' voltage magnitude measurement (in [p.u.]) at node 1 at time step t is represented as $v_{1,t}$. The active and reactive power transmitted from node 1 to the feeder root lines l_f of each feeder $f \in \mathcal{F}$ are denoted as $P_{1\to l_f,t}$ and $Q_{1\to l_f,t}$, respectively.

Furthermore, the DSO periodically collects historical measurements of nodal voltage magnitudes at nodes equipped with SMs ($n \in \mathcal{N}_{\text{SM}} \subseteq \mathcal{N}$) that are distributed throughout the grid [18]. The SM measurements have a resolution of 15 min and are typically acquired on a daily or weekly basis, thus not available to use during RT operation. RT substation and historical SM measurements are denoted with $\mathbf{z}_t$ and $\mathbf{e}_t$, respectively:

$$\mathbf{z}_t = \begin{bmatrix} v_{1,t} & P_{1\to l_1,t} & \dots & P_{1\to l_F,t} & Q_{1\to l_1,t} & \dots & Q_{1\to l_F,t} \end{bmatrix}^{\mathrm{T}} \tag{3.1a}$$

$$\mathbf{e}_t = (v_{n,t})_{n\in\mathcal{N}_{\text{SM}}} \tag{3.1b}$$

3.1.2 Control Degrees of Freedom

In accordance with [39, 40], it is assumed that the DSO has the ability to centrally issue power factor set points to controllable PV inverters, as well as active power curtailment signals to controllable DERs of type $D \in \mathcal{D}$, with $\mathcal{D} := \{\mathrm{PV}, \mathrm{EV}, \mathrm{HP}\}$.

The control problem is defined as non-discriminatory, which implies that the controller should not show preference for a specific group of end-users that may have a greater impact on voltage mitigation [40]. This is to prevent unfair curtailment or resource utilization of any particular end-user group. Subsequently, there are 4 control signals in total, 3 for P-curtailment, and 1 for Q-control.

The DER control can be mathematically summarized as:

$$P_t^{d,\mathrm{u}} = P_t^d \cdot (1 - u_t^{D,\mathrm{p}}) \ , \ \forall d \in \mathcal{S}_D^{\mathrm{u}} \ , \ u_t^{D,\mathrm{p}} \in [0,1], \ \forall D \in \mathcal{D} \tag{3.2a}$$

$$Q_t^{d,\mathrm{u}} = Q_{\max}^d \cdot u_t^{\mathrm{PV,q}} \ , \ \forall d \in \mathcal{S}_{\mathrm{PV}}^{\mathrm{u}} \ , \ u_t^{\mathrm{PV,q}} \in [-1,1] \tag{3.2b}$$

where the set $\mathcal{S}_D^{\mathrm{u}}$ denote the index set of controllable DERs of type $D \in \mathcal{D}$. Note that the P-curtailment signals $u_t^{D,\mathrm{p}}$ are applied to the uncontrolled active powers P_t^d. In contrast, the Q-control signal $u_t^{\mathrm{PV,q}}$ is applied to the maximum allowed reactive power $Q_{\max}^d = P_{\mathrm{inst}}^d \cdot \tan(|\cos(\varphi_{\max})|)$, where P_{inst}^d is the installed capacity of each PV d from $\mathcal{S}_{\mathrm{PV}}^{\mathrm{u}}$, and $\cos(\varphi_{\max})$ is the maximum allowed power factor of inverters.

As inverters can operate both inductive and capacitive, the Q-control signal of PVs may be negative or positive, respectively. It is important to note that PV power is denoted in ASC, while EV and HP power are in PSC.

To avoid interacting with DERs when flexibility requests are significantly low, control signals are suppressed if their absolute value is less than 10% of the control range, i.e., $u_t^{D,\mathrm{p}} < 0.1$ and $|u_t^{\mathrm{PV,q}}| < 0.1$.

In accordance with [39, §5.7.2.3], the Q exchange of PV inverters is suppressed also when a PV system generates less than 10% of its installed capacity, even if the controller decides for a non-zero Q:

$$Q_t^{d,\mathrm{u}} = \begin{cases} 0 & , \text{ if } P_t^{d,\mathrm{u}} < P_{\mathrm{inst}}^d \cdot 10\% \\ Q_t^{d,\mathrm{u}} & , \text{ else} \end{cases} \qquad \forall d \in \mathcal{S}_{\mathrm{PV}}^{\mathrm{u}} \tag{3.3}$$

3.2 Fully Data-Driven Grid Control

As shown in Figure 1.2 and introduced in Section 1.3, the investigated fully data-driven control algorithm comprises two stages: grid monitoring and DER control. Grid monitoring is realized with a model-free, data-driven SE approach that estimates voltage magnitudes at nodes equipped with SMs in RT. DERs are controlled in RT to mitigate voltage violations by deploying RL that is trained online during operation. As both stages are model-free, the investigated control algorithm does not require any knowledge about the topology and electrical parameters of the grid. Both stages of the algorithm are discussed in the following sections.

3.2.1 Data-Driven State Estimation

The state estimator, highlighted in yellow rectangle in Figure 1.2, is implemented using the MLP-based methodology described in [18], which is highlighted in pink.

An MLP with $L+1$ layers, comprising one input layer, $L-1$ hidden layers, and one output layer, can be expressed in recursive notation as follows:

$$\begin{aligned}\mathbf{y} = f_L(\mathbf{x}) &= a_L(\mathbf{W}_L \cdot f_{L-1}(\mathbf{x}) + \mathbf{b}_L) \\ &\text{with } f_0(\mathbf{x}) = \mathbf{x} \quad \text{(input layer)}\end{aligned}$$

where $\mathbf{x}$ is the input vector and $\mathbf{y}$ is the output vector of the Artificial Neural Network (ANN) [41]. For any given layer $l \in \{0, \ldots, L\}$, the function f_l represents the output of that layer. Specifically, f_l is a vector representation of the results of the activation function a_l, which is applied element-wise to the linear combination of the layer inputs (the output of the previous layer, i.e., f_{l-1}), along with the neuron weights $\mathbf{W}_l$ and bias $\mathbf{b}_l$.

In the context of the SE formulation of [18], the input vector comprises RT measurements taken at the substation, i.e., $\mathbf{x} = \mathbf{z}_t$, while the output vector represents the downstream voltage magnitudes where SMs are installed, i.e., $\mathbf{y} = \mathbf{e}_t$. In the first stage, the SE acting as a regressor maps the RT measurement $\mathbf{z}_t$ to voltage magnitude estimates at those nodes equipped with SMs. The SE thus approximates a mapping of the form $\mathbf{e}_t = f_L(\mathbf{z}_t)$.

During the offline training phase, the parameters $\mathbf{W}_l$ and $\mathbf{b}_l$ are trained through backpropagation using historical labeled data of $(\mathbf{z}_t, \mathbf{e}_t)$ using mini-batch gradient descent that updates the parameters with small random batches of training data. Root Mean Squared Error (RMSE) is used as loss function (see Section 5.2.1 for mathematical formulation) to minimize the estimation error between the predicted values $\hat{\mathbf{e}}_t$ and real values $\mathbf{e}_t$.

During grid operation, RT voltage estimates $\hat{\mathbf{e}}_t$ are readily obtained by evaluating the trained MLP for given RT measurements $\mathbf{z}_t$, i.e., $\hat{\mathbf{e}}_t = f_L(\mathbf{z}_t)$.

3.2.2 Grid Control with Reinforcement Learning

Referring to Section 2.4, an RL agent serves as the central controller of the DSO in this work, with the goal of mitigating voltage issues while keeping DER intervention minimal. The agent's observation is the composition of the RT measurements and estimated voltage magnitudes.

$$\mathbf{s}_t = \begin{bmatrix}\mathbf{z}_t{}^{\mathrm{T}} & \hat{\mathbf{e}}_t{}^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}} \tag{3.4}$$

Furthermore, the action of the agent at time t is represented by:

$$\mathbf{a}_t \coloneqq \begin{bmatrix}u_t^{\mathrm{PV,p}} & u_t^{\mathrm{PV,q}} & u_t^{\mathrm{EV,p}} & u_t^{\mathrm{HP,p}}\end{bmatrix}^{\mathrm{T}} \tag{3.5}$$

The DSO's operation goal is characterized with the agent's reward function $r(\mathbf{s}_t, \mathbf{a}_t)$ using weighting coefficients:

$$\begin{aligned}r_t = r(\mathbf{s}_t, \mathbf{a}_t) = &- w^{\mathrm{ov}} \cdot \max\left(\max(\mathbf{v}_t) - v_{\max}, 0\right) && \text{(3.6a)}\\ &- w^{\mathrm{uv}} \cdot \max\left(v_{\min} - \min(\mathbf{v}_t), 0\right) && \text{(3.6b)}\\ &- w^{\mathrm{PV,p}} \cdot u_t^{\mathrm{PV,p}} && \text{(3.6c)}\\ &- w^{\mathrm{EV,p}} \cdot u_t^{\mathrm{EV,p}} && \text{(3.6d)}\\ &- w^{\mathrm{HP,p}} \cdot u_t^{\mathrm{HP,p}} && \text{(3.6e)}\end{aligned}$$

$$- w^{\mathrm{PV,q}} \cdot \mid u_t^{\mathrm{PV,q}} \mid \tag{3.6f}$$

$$- w_{\mathrm{NCVI}}^{\mathrm{PV,q}} \cdot \max\left(\frac{\max(\mathbf{v}_t) - v_{\mathrm{CVI}}}{|\max(\mathbf{v}_t) - v_{\mathrm{CVI}}|}, 0\right) \cdot \mid u_t^{\mathrm{PV,q}} \mid \tag{3.6g}$$

where $\mathbf{v}_t \in \mathbb{R}^{1+N_{\mathrm{SM}}}$ is the vector of the lastly observed relative nodal voltage magnitudes, i.e., comprising $\hat{\mathbf{e}}_t$ and the LV-side voltage of the substation at time t, and the scalars $w^{(\cdot)} \in \mathbb{R}_{>0}$ denote the penalty factors.

The approach aims to control DERs to mitigate OV and UV issues. Therefore, negative rewards (penalty) are assigned using equation terms (3.6a) and (3.6b) if the maximum or minimum voltage observations exceed the allowed voltage range $[v_{\min}, v_{\max}]$. To ensure the sustainable operation of the system, DERs should only be curtailed in the presence of voltage violations. Thus, P-curtailments are penalized with terms (3.6c), (3.6d), and (3.6e). Additionally, unnecessary exploitation of the Q potential of PV inverters is avoided to prevent overloading of end-user inverters and to reduce thermal losses associated with increasing grid loading. This is achieved through the use of terms (3.6f) and (3.6g). The former term penalizes Q-control interventions in general, similar to the previous control penalties. The latter term penalizes Q interventions substantially intensified if there is 'No Critical Voltage Increase (NCVI)'. The big max operator in (3.6g) models this if-case, which is triggered when the observed maximum voltage magnitude exceeds a voltage level v_{CVI} indicating a voltage increase, also denoted as 'Critical Voltage Increase (CVI)'.

The SAC algorithm (see Figure 2.4) that has shown promising results in recent power system studies [11, 12] is chosen to solve the RL problem introduced in Section 2.4 and later characterized by Equations (3.1)–(3.6).

3.3 Real-World Application Requirements

Table 3.1 provides a comparative overview of the necessary data, system knowledge, measurement, and communication infrastructure that the DSO would require to apply the methodology outlined in Section 3.2 to real LV grids.

Table 3.1 Comparison of Requirements for Real-World Applications

Required	**Not Required**
1. RT voltage magnitude measurement at the substation's secondary side	1. Grid topology or parameters
2. RT active and reactive power measurement at the feeder root lines	2. Downstream RT measurements
3. SMs at critical downstream nodes and periodical data register synchronization	3. High-resolution (<1 min) measurements
4. Communication from the substation to energy system communities or DERs, e.g., TCP/IP connection of SMs	4. Voltage phase measurements
5. Edge device at the substation that is capable of performing computations such as periodic training of SE and RL	5. Installed DER capacities
	6. Total number of installed DERs
	7. Actual DER power
	8. Synchronous communication with downstream nodes

Chapter 4

Simulation Framework Development

This chapter offers a comprehensive overview of simulation environment designed and developed within the scope of this thesis. First section introduces the software architecture that guided the implementation of the methodology described in Chapter 3. The actual implementation will be addressed in the subsequent section.

4.1 Software Architecture

Software architecture is the process of designing and defining the fundamental structure of a software system. It involves making decisions regarding the organization of software components, the interface between them, and the interactions among them. Furthermore, it serves as the blueprint for a software system and provides the basis for all future development and maintenance efforts. [42]

In consideration of the generalizability and scalability of the framework, the architecture is designed such that its modular structure enables integration of various state estimation algorithms, control algorithms, and grid operation scenarios. This provides a plug-in platform for developing and validating control algorithms.

The developed architecture is demonstrated with a *Unified Modeling Language (UML)* diagram in Figure 4.1 by using conventional notation, see [43]. The contribution of this thesis is the so-called *Grid Operation Framework* highlighted in a blue envelope. This framework leverages three external classes:

1. *PowerSystem*: Class of power grid models that consists of grid topology, grid parameters such as line and transformer characteristics, electrical loads (such as HHs, EVs, HPs), generation units (such as PVs), time series profiles assigned to these units, measurement devices in the grid, and a set that defines which of the DERs are controllable.

2. *PowerFlowSolverInterface*: Interface to run power flow calculations of a given power system with compatible solvers, e.g., MATPOWER [44], OpenDSS [45].

3. *StateEstimator*: Model-free, data-driven state estimation algorithm as described in Section 3.2.1.

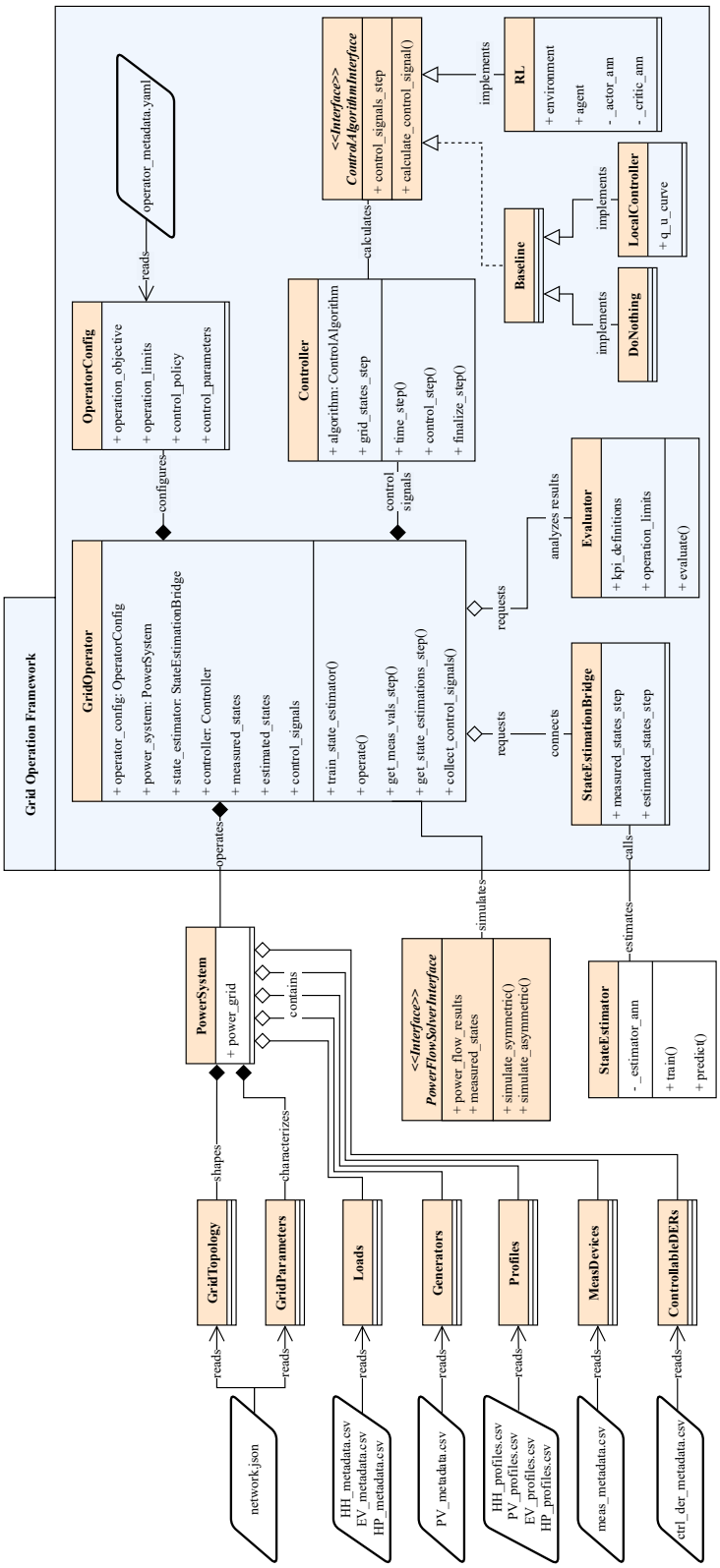


Figure 4.1 UML diagram of *Grid Operation Framework*

The central unit of the framework is *GridOperator*, aptly named after power system operators. It is tasked with overseeing all aspects related to simulating a grid operation cycle, i.e., preparation and initialization of time series simulation, accessing Supervisory Control and Data Acquisition (SCADA) measurements, requesting state estimation, running control algorithm and forwarding control signals to DERs. Prior to delving into the specific information flow and control process, it is essential to introduce the primary peripheral classes of the framework:

1. *OperatorConfig*: Configuration describing operation goal and grid limits for *GridOperator* as well as parameters for control algorithm. A configuration file of base simulation setup is provided in Appendix A, which specifies to the *GridOperator* that voltages below 0.95 p.u. and above 1.045 p.u. should be mitigated by using an RL-based controller. The entire parameter set is introduced in detail in Section 5.4.

2. *StateEstimationBridge*: Connector to external state estimation packages, responsible for delivering estimated states at runtime.

3. *ControlAlgorithmInterface*: Abstraction layer for grid control algorithms. Actual control algorithms inherit from that class and implements its abstract methods according to their specific methodology. These could include Volt-Var curve, RL or other control schemes such as OPF and FO.

4. *Controller*: *GridOperator*'s control unit that overtakes the responsibility of executing its control algorithm specified in *OperatorConfig*. Its plug-in control algorithm takes the observed states, consisting of RT measurements and estimated states, as input. Based on the algorithm's decision mechanism, the controller delivers control signals that are described in Section 3.2.2.

5. *Evaluator*: Grid analyzer that evaluates Key Performance Indicators (KPIs) that are a measure how well the grid was operated in its limits and how effective the control has been. To provide a clearer picture, the output file containing comprehensive results for a test case of the investigated approach is provided in Appendix B. These results are discussed in depth in Section 6.2.

Along with the class diagram in Figure 4.1, the behavioral diagram of the framework is given in a flow diagram, Figure 4.2. Without loss of generality, RL is used as an exemplary control algorithm in this diagram as it falls within the scope of this work. However, the respective control block is replaceable with an alternative control algorithm.

An end-to-end simulation of the proposed control algorithm starts with preparing profiles for a pre-defined simulation scenario, as defined by the *PowerSystem* instance. To prevent data leakage between the learning processes of SE and RL, the simulation data is separated into two parts based on shuffled days. Prior to starting the control simulation, SE training and validation take place offline. A time series simulation is performed to generate labeled data for supervised learning of SE. These data represent historical data that a DSO typically possesses. Consequently, a pre-trained SE model is prepared and made available for use during grid operation that follows immediately.

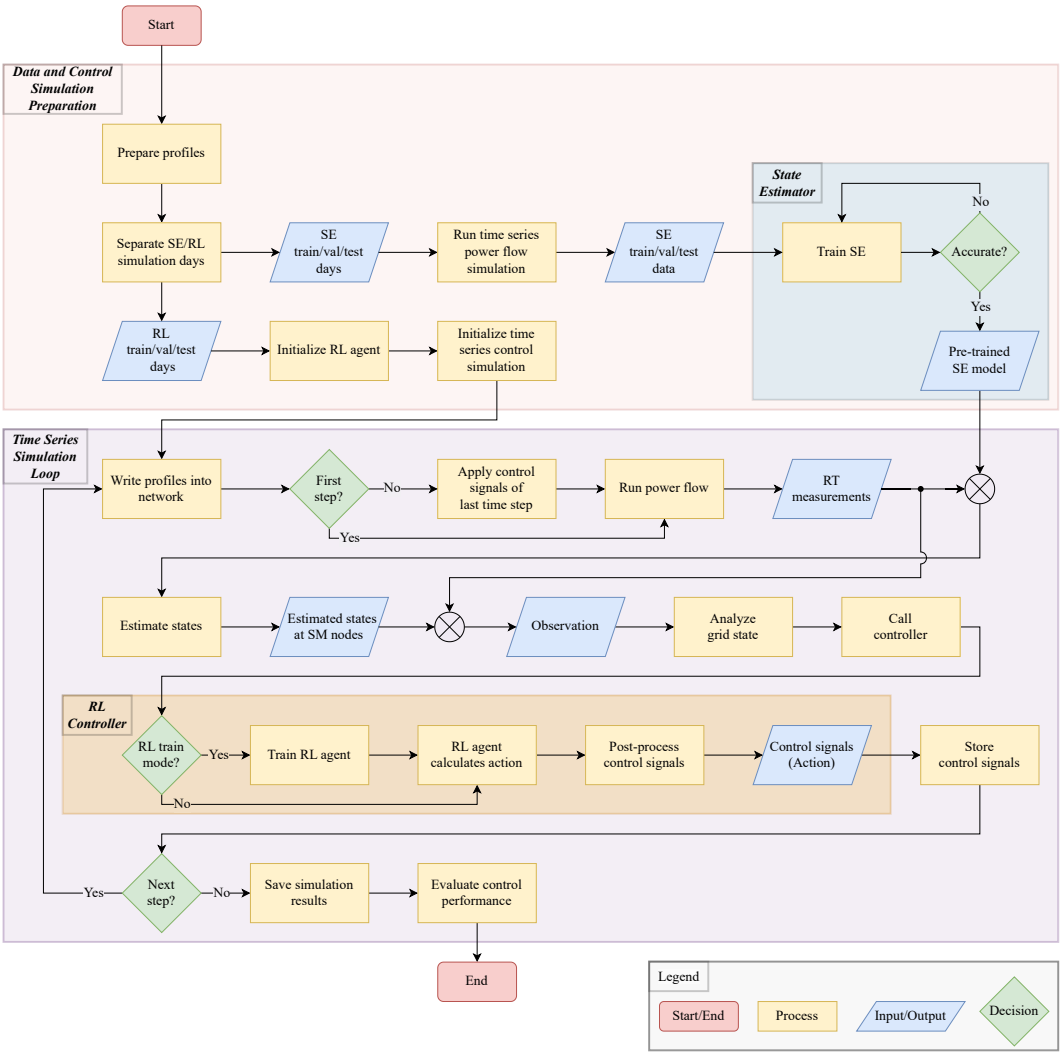


Figure 4.2 Flow diagram of *Grid Operation Framework* with RL as control algorithm

Each step of grid operation loop begins with updating the P and Q of consumption and generation units in the grid, which are read from the previously prepared profiles. Except the initial simulation step, all controllable DERs are overwritten in accordance with the corresponding control signal calculated at the end of the previous simulation step. The lag of one simulation step accounts for the controller's perception and reaction time, which may be caused by communication delays, computational expenses of the algorithm, and the realization of control signals.

A power flow simulation is performed to capture the behavior of a power system. The measurements assumed to be sensed in RT are extracted and passed to the pre-trained SE via the *StateEstimationBridge*. The estimated states are the downstream voltage measurements along the grid, where SMs are assumed to be installed, as described in Section 3.2.1. The collection of measurements available in RT and estimated in runtime comprise the set of observations for the system. The DSO analyzes the system based on its online observations, such as calculating the reward function for RL or evaluating other KPIs.

The *Controller* intervenes in DERs to take measures for the grid state, as described in Section 3.2.2. If the controller is an RL agent, it is trained from a batch of experiences during operation as long as it is in training mode. The SAC algorithm shapes a normal distribution based on the actor network's outputs (μ and σ, see Figure 2.4). During training mode, the agent samples the action from the resulting normal distribution. If the agent is not in training mode, it returns the calculated mean μ, which is the most probable action leading to the highest return.

Calculated actions are post-processed by the DSO before they are sent out to DERs, as described in Section 3.2.2. These are stored to apply first in the next simulation step, as mentioned earlier. If end of the simulation horizon is reached, the simulation results are saved, including power flow results, estimated states, calculated control signals. Based on these, the *Evaluator* analyzes the results taking the grid operation goal and limits specified in the *OperatorConfig* into account.

4.2 Implementation

4.2.1 Test-Driven Development

Due to the complexity and the planned long-term use of the designed architecture, the framework is implemented using *Test-Driven Development (TDD)* practices [46]. Before implementing a feature, determining its requirements helps to break the problem into pieces. Writing failing tests for these atomic units is the first step of TDD. This ensures that the "minimal working example" of the requested feature is continuously checked via automated tests. Development starts from the simplest milestones of the feature, and the cycle of development and failing tests continues until all tests are passed. The same procedure applies to each new feature from the beginning.

One of the key advantages of TDD is that it helps to maintain code quality and consistency through automated checks. Furthermore, testing each code piece using its simplest independent example ensures the collaborative functionality of code pieces. As tests are automatically performed periodically for all code pieces, error localization is simplified, and less debugging effort is required. As a result, a test coverage of 90% of the developed *Grid Operation Framework* is achieved with 79 unit tests. To provide a numerical understanding of this coverage level, it is worth noting that Atlassian, a leading software company in Continuous Integration and Continuous Delivery (CI/CD) solutions, considers a test coverage of 80% to be a good value [47].

4.2.2 Third-Party Software Components

The framework is implemented using the *Python* programming language, which offers a wide range of powerful open-source software tools. To implement the various components of the architecture described in Section 4.1, a variety of software tools is utilized. The following tools are deployed within the framework, along with their corresponding applications:

1. *Pandapower* [48]: Power system modeling and simulation tool. The *PowerSystem* class is primarily based on the modeling capabilities of Pandapower.

The *PowerFlowSolverInterface* uses Pandapower as the primary connector to a power flow solver. Although Pandapower is not a standalone power flow solver, it has an extension for the PYPOWER solver, which is a Python port of MATPOWER. Another advantage of Pandapower is its time series simulation interface, which enables the importation and automatic iteration of power profiles of end-users throughout the course of a time series simulation.

Pandapower provides also a control module for time series simulation. The internal *Controller* of the developed framework is inherited from Pandapower's controller interface. As such, the prework of automated data transfer in context of a time series simulation is adopted. However, the *Controller* class extends the scope with accepting estimated states to realize the methodology described in Section 3.2.

2. *Gym* [49]: Toolkit for developing system behavior in context of RL environments. It provides a generic environment structure that is compatible with many other RL agent algorithm toolkits, e.g., *StableBaselines3*.

   In the developed framework, power flow simulation results from Pandapower are integrated into a Gym environment for the RL controller.

3. *StableBaselines3* [50]: Toolkit with various RL agent algorithm implementations that are compatible with Gym-like environments. The implementation of the SAC algorithm [36] is adopted from this tool. The agent of the RL-based grid controller is an instance of this type.

4. *TensorFlow* [51]: Machine learning framework to train and evaluate artificial intelligence models.

   The *StateEstimator* is implemented using TensorFlow, which is not done within the scope of this thesis. The *StateEstimationBridge*, on the other hand, uses pre-trained TensorFlow models provided by the *StateEstimator*.

Besides implementing functional blocks, the code quality is ensured by deploying a Git CI/CD pipeline. This pipeline automatically runs code checks after each code commit. The checks include code functionality verification with the *Pytest* package, as described in Section 4.2.1, code beautifying with the *Black* package, syntactical and stylistic analysis with the *Pylint* package.

Chapter 5

Simulation Design

The control algorithm is validated through simulations. To conduct experiments, a realistic test scenario is defined and detailed in Section 5.1. The evaluation of the control performance is based on pre-defined KPIs, which are explained in Section 5.2. The proposed method is compared with two reference control techniques, which are discussed in Section 5.3. The foundational parameters of the control algorithm are presented in Section 5.4 and serve as the basis for the results presented in Chapter 6.

5.1 Test Scenario

The test scenario comprises three main components. First, the test grid will be introduced, followed by the penetration levels of DERs. This section will conclude with how the DER profiles are collected, generated, and deployed.

5.1.1 Test Grid

The experiments are designed to simulate a future scenario on the 'rural2' benchmark grid from the Simbench LV networks [52] that consists of 4 feeders and 93 supply points. In anticipation of increased installed DER capacities in the LV grid, the original substation is substituted with a larger 0.63 MVA 10/0.4 kV transformer. The voltage of the slack node, on the primary side of the transformer, is set to 1.0 p.u. The topology of the test grid can be found in Appendix E.

As this thesis focuses on residential LV grids, commercial loads are substituted with HHs based on the ratio between average commercial and household capacities and the corresponding diversity factor. For example, an average commercial building in the rural2 grid had an installed demand capacity of 7.9 kW, while an average single HH had an installed capacity of 1.6 kW, resulting in a ratio of approximately 5. A diversity factor of 0.78, drawn from the French standard diversity factor table given in [53, p. 19], is assumed for 5 HHs. As such, the disaggregated installed capacity of each HH in an apartment building is calculated by multiplying with $\frac{1}{5 \cdot 0.78}$. For instance, a commercial building with 10 kW is replaced with 5 HHs, each with 2.56 kW. These multiple HHs at the same node can be interpreted as an 'apartment building'. After this modification, the number of HHs in the grid increased to 127, with a total installed capacity of 217 kW. Of these, 92 are single-family houses, and 35 are in 7 apartment buildings.

5.1.2 DER Penetration

The DER penetration in the original rural2 grid is not high enough to cause voltage issues. Therefore, a progressive future scenario is defined based on future expectations from literature. Table 5.1 summarizes the defined scenario, which will be explained in detail below.

Table 5.1 DER Penetration Scenario

Type	Penetration	Number	Unit capacity levels in [kW]	Total installed capacity in [kW]
PV [54–56]	70% of buildings	69	15 (for single HHs) [55, 57] 45 (for apartments)	1185
EV [58–60]	70% of fleet	103	3.7 (20%) 11.0 (70%) 22.0 (10%)	1097
HP [61, 62]	70% of buildings	69	9 (for single HHs) [63] 27 (for apartments)	693

As per [54], there were 19.3 million residential buildings in Germany by 2020, out of which 11.7 million were eligible for PV installation, according to [55]. This represents a penetration of 60% in both rural and urban areas. Additionally, [56] suggests that buildings in rural areas have larger rooftops, leading to higher PV potential than the average. Therefore, a PV penetration of 70% is assumed for the rural test grid, resulting in 69 PV installations.

According to [57, p. 333–334], PV panel efficiency is expected to double by 2050. Additionally, [55] reports that the average installed capacity of small-scale PV systems was 7.5 kW_p as of 2020. Therefore, for the futuristic scenario, a capacity of 15 kW_p for single-family households is assumed. For apartment buildings each with 5 HHs, it is assumed that each building has three times the rooftop area, resulting in 45 kW_p PV installations.

The EV penetration in the test grid is determined based on the vehicle fleet. According to [58], there are 0.58 vehicles per resident in Germany, and each household has 1.99 residents [59]. Thus, there are 1.15 vehicles per household. Based on [60], a realistic EV penetration rate is 70% of all vehicles. This results in 103 EVs in the test grid, assumed each with its own residential EV charger. The residential EV chargers are dimensioned in three power classes: 3.7, 11, 22 kW. Relying on the rapidly growing EV charging infrastructure [64], it is assumed that the majority of the chargers will be 11 kW, representing 70% of all chargers. To realize innovative households, 10% of the chargers are 22 kW. The remaining 20% of the chargers are assumed to be old-generation 3.7 kW chargers. It is worth noting that for brevity, EV chargers are simply referred as EVs.

According to [61, 62], it is predicted that the number of HPs in Germany will reach 14–16 million by 2050. Considering that there are 19.3 million residential buildings in Germany, a realistic penetration level of 70% for HPs is expected. The installed capacity of HPs in single-family HHs is assumed to be equal to the total capacity of heating devices installed in the HHs [63]. These HHs have both HPs

with an installed capacity of either 1.9 or 3 kW, and heating rods with a capacity of 6 kW. As this thesis focuses on a progressive scenario, only HPs are assumed to supply the heating demand and are therefore dimensioned with a capacity of 9 kW. Similar to PVs, the capacity of HPs in apartment buildings is assumed to be three times that of an HP in a single-family HH.

Finally, for the given numbers of DERs in Table 5.1, the DERs are randomly distributed to the single-family HHs or apartment buildings in the test grid.

5.1.3 Power Profiles

To conduct time series simulations, a pool of residential profiles is compiled. In order to satisfy the conditions necessary for testing control algorithms and their dynamics, the simulations must possess a high temporal resolution and model various realistic conditions, including both low-demand-high-generation and high-demand-low-generation seasons. With these requirements in mind, open-source profiles with a temporal resolution of 1 min for a period of 1 year have been researched. The results of the research are compiled in Appendix C [65–67].

It is reasonable to assert that there exist sufficient HH profiles with high temporal resolution and long timespan. However, the available profiles for EV charging primarily focus on public charging due to its better accessibility rather than residential charging. Except for the *WPuQ* dataset [63], no significant profiles for HPs were identified. Additionally, there were no notable profiles available for PVs since PV output is specific to its location and installation. Consequently, PV profiles are mostly generated using simulation tools. A list of identified interesting simulation tools for generating HH and DER profiles is provided in Appendix D.

The real-measured *WPuQ* profiles from the German town of Hamelin are used for HH and HP profiles. The profiles from the year 2019 are selected for further deployment since they had the most complete profiles over a year. Out of the 38 houses, 26 had no data gaps and were ready to use without further effort. The profiles are normalized to use them for varying installed capacities of HHs and HPs. The P and Q measurements of these houses formed the profile pool for HHs. For HPs, the P measurements were supplemented with Q profiles using a constant inductive power factor of $\cos(\varphi^{\text{ind}}) = 0.97$, instead of using the given Q measurements for consistency with constant power factor for PV and EV profiles that will be described next.

The PV profiles are generated using the simulation tool *PVLIB* [68] for the German town Hamelin. The process for generating the profiles is depicted in Figure 5.1. The irradiance time series are obtained from the Copernicus Atmosphere Monitoring Service (CAMS) [69] using the Application Programming Interface (API) provided by *PVLIB*. The highest time resolution available for irradiance data is 15 min. The position of the Sun is calculated for year 2019 given the coordinates of Hamelin (N 52°6′15″, E 9°21′42″) using PVLIB. To generate a pool of 50 PV profiles, 50 different PV system installation settings (tilt and azimuth of the panels) are randomly sampled from their corresponding normal distributions based on the statistics in Germany from [70]. The authors of this work have determined that PV installations in Germany have a normal distribution for tilt with $\mathcal{N}(20.6°, 10.8°)$ and for azimuth with $\mathcal{N}(180°, 19.3°)$. For reference, a tilt of 0° represents a flat panel, and an azimuth of 180° is directed towards the south. Furthermore, it is assumed that all PV panel

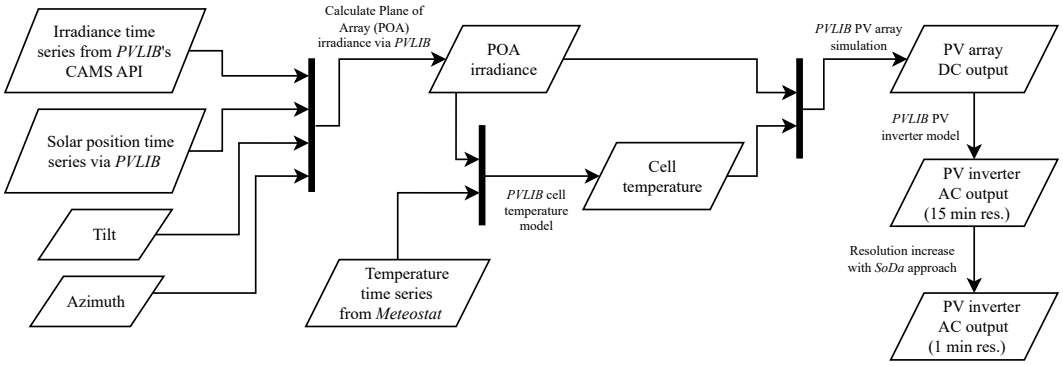


Figure 5.1 Flow diagram of PV profiles generation

arrays have a base power of 1 kW to provide normalized profiles for each installation setting.

Using PVLIB, the Plane of Array (POA) irradiance for each profile is calculated. The POA irradiance represents the actual irradiance that arrives at a tilted and azimuthed PV array, including direct and diffuse irradiance. The output of PV panels depends on the efficiency of the cells, which is dependent on the cell temperature. The cell temperature, in turn, depends on the irradiance and the air temperature. The air temperature data for Hamelin is obtained using the Python API of the weather data tool *Meteostat* [71]. A cell temperature coefficient of $-0.4\,\%/\mathrm{C}^\circ$ is assumed. The time series of the cell temperature is also calculated using PVLIB. Given the irradiance and cell temperature time series as inputs to the PV model of PVLIB, the Direct Current (DC) output of the PV system is calculated, which is then converted to Alternative Current (AC) output using the default inverter model of PVLIB.

The input irradiance data has a resolution of 15 min, and as a result, the outputs of the PVLIB models also have a 15 min resolution. However, the nature of PV generation is highly time-varying, making it crucial to have a high resolution of PV profiles to test control algorithms effectively. Therefore, a stochastic approach proposed in [72] (called *SoDa*) is applied to increase the resolution to 1 min. The SoDa approach uses stochastic noise to generate high-resolution irradiance time series based on parameters that are trained in a supervised manner beforehand. This approach provides a more realistic representation of the variability of PV generation and enables testing of control algorithms with a higher temporal resolution.

The EV profiles are generated using a closed-source simulation tool called *Agent-based eMobility Model (ABM eMob)*, which is developed as part of the PhD dissertation [73]. The *ABM eMob* simulates the charging behavior of EVs at home, public, and work charging stations with 15 min resolution by modeling EVs bottom-up as individual agents. The agent-based model is applied to a region that shares similar characteristics with Hamelin. In order to carry out the simulation, certain key parameters are adopted from [73] and presented in Table 5.2. It is worth noting that only home charging profiles are included in the profile pool since the test grid only models residential loads (see Section 5.1.1). As the demand for EV charging is not as dynamic as HH or PV, the charging profiles with 15 min resolution are leveraged by using forward-filling to increase the resolution from 15 min to 1 min. Forward-filling

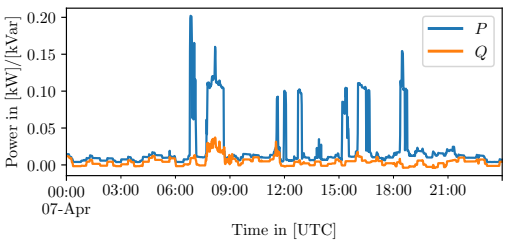
Table 5.2 EV Profile Generation Parameters [73]

EV types	small range, medium range, long range
Average EV consumption in [kWh/100 km][1,2]	15, 18, 16
EV battery capacities in [kWh][2]	40, 80, 100
Commuters ratio	55%
Daily average mileage in [km]	39

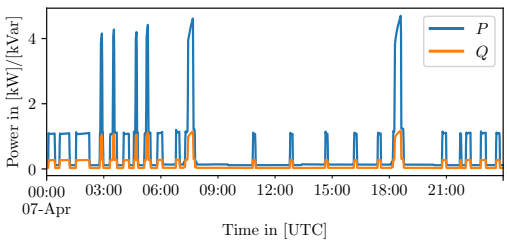
involves propagating the last observed 15 min value forward in time to fill in the missing values at 1 min intervals.

The resulting normalized HH, HP, and PV profiles are randomly assigned to the DERs defined in Table 5.1, and scaled up according to the installed capacity of each DER. However, as the installed charging capacity of EVs affects their profile shape, the EV profiles are assigned based on their absolute profile shape. It is worth noting that, as of the writing of this thesis, the tool *ABM eMob* takes into account residential chargers with a maximum capacity of 11 kW. The innovative EVs with an installed home charger capacity of 22 kW are assumed to have larger battery capacities, which would lead to longer charging sessions and profiles that are similar to 11 kW chargers. Therefore, the 22 kW charging points are assigned up-scaled profiles with the shape of 11 kW profiles.

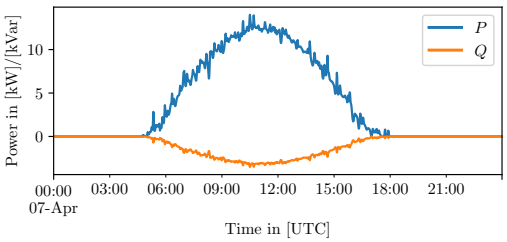
To facilitate visualization, one sample daily profile for each asset type is displayed in Figure 5.2.



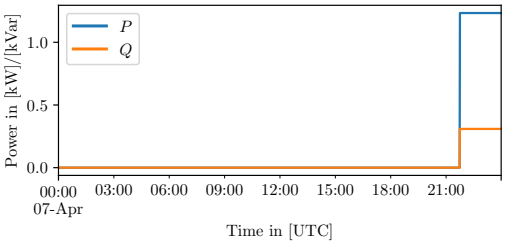
(a) A sample HH profile in PSC



(b) A sample HP profile in PSC



(c) A sample PV profile in ASC



(d) A sample EV profile in PSC

Figure 5.2 Sample daily DER profiles on day April 7, 2019

[1]Time-varying dependent on actual velocity and monthly average temperature
[2]For each EV type, in the order of [small range, medium range, long range]

5.2 Key Performance Indicators

Both stages of the algorithm, SE and control, are evaluated separately. Thus, the KPIs for both are discussed separately in this section.

5.2.1 State Estimation

To evaluate the performance of the SE, three KPIs are defined for evaluating time series results. Evaluating time series of multivariable systems requires a two-dimensional matrix evaluation, with one dimension representing the quantities under investigation at each time step and the other representing the time steps of test data:

- **RMSE**:

$$\mathrm{RMSE}(\mathbf{E}, \hat{\mathbf{E}}) = \sqrt{\frac{1}{N \cdot T} \sum_{i=1}^{T} \sum_{j=1}^{N} |\hat{\mathrm{E}}_{ij} - \mathrm{E}_{ij}|^2}$$

- **Maximum Deviation (MD)**:

$$\mathrm{MD}(\mathbf{E}, \hat{\mathbf{E}}) = \max_{i \in \{1,\dots,T\}} \left\{ \max_{j \in \{1,\dots,N\}} |\hat{\mathrm{E}}_{ij} - \mathrm{E}_{ij}| \right\}$$

- **Maximum Relative Deviation (MRD)**:

$$\mathrm{MRD}(\mathbf{E}, \hat{\mathbf{E}}) = \max_{i \in \{1,\dots,T\}} \left\{ \max_{j \in \{1,\dots,N\}} \left| \frac{\hat{\mathrm{E}}_{ij} - \mathrm{E}_{ij}}{\mathrm{E}_{ij}} \right| \right\}$$

where $\mathbf{E}$ is the matrix representation of real SM measurement vectors $\mathbf{e}_t$ for each evaluation time step $t \in \{1, \dots, T\}$, with each vector having a length of N. On the other hand, $\hat{\mathbf{E}}$ represents the estimated measurement matrix.

5.2.2 Grid Control

The primary control objective is to mitigate OV and UV. A voltage violation is said to occur when the maximum or minimum voltage magnitude in the grid falls outside the voltage band $[v_{\min}, v_{\max}]$. Three KPIs are defined for each:

- **Total Overvoltage/Undervoltage Violation Duration (TOVD/TUVD)**: Total duration of voltage violation in all test data after control has been applied. It is expressed in [min].

- **Average Overvoltage/Undervoltage Violation Duration (AOVD/AUVD)**: Average duration of voltage violation in all test data after control has been applied. It is a measure of how long an occuring incident lasts in average. It is expressed in [min].

- **Integral of Overvoltage/Undervoltage Violation (IOV/IUV)**: Cumulative sum of the deviation between the maximum/minimum voltage recorded

during the discrete test simulation period and the voltage limits. This can be mathematically formulated as:

$$\mathrm{IOV} = \sum_{t=1}^{T} \max(\max(\mathbf{v}_t) - v_{\max}, 0)$$
$$\mathrm{IUV} = \sum_{t=1}^{T} \max(v_{\min} - \min(\mathbf{v}_t), 0)$$

This can be interpreted as the area under the curve of the deviation function over the time period of interest, which gives the integral of voltage violation. It is expressed in [p.u. · min].

While voltage mitigation is the primary control objective, it is important to avoid unnecessary intervention of DERs by the controller to ensure sustainable grid operation. Therefore, the degree of DER intervention required for voltage mitigation is also used as a control KPI. These KPIs also serve as a measure of potential inconvenience for DER owners.

- **Change in PV Output (CPO)**:

$$\mathrm{CPO} = \frac{\sum_{t=1}^{T} \sum_{d \in \mathcal{S}_{\mathrm{PV}}} P_t^{d,\mathrm{u}}}{\sum_{t=1}^{T} \sum_{d \in \mathcal{S}_{\mathrm{PV}}} P_t^{d}} - 1$$

- **Change in PV Inverter Loading (CPIL)**:

$$\mathrm{CPIL} = \frac{\sum_{t=1}^{T} \sum_{d \in \mathcal{S}_{\mathrm{PV}}} \sqrt{{P_t^{d,\mathrm{u}}}^2 + {Q_t^{d,\mathrm{u}}}^2}}{\sum_{t=1}^{T} \sum_{d \in \mathcal{S}_{\mathrm{PV}}} \sqrt{{P_t^{d}}^2 + {Q_t^{d}}^2}} - 1$$

- **Change in EV Demand (CED)**:

$$\mathrm{CED} = \frac{\sum_{t=1}^{T} \sum_{d \in \mathcal{S}_{\mathrm{EV}}} P_t^{d,\mathrm{u}}}{\sum_{t=1}^{T} \sum_{d \in \mathcal{S}_{\mathrm{EV}}} P_t^{d}} - 1$$

- **Change in HP Demand (CHD)**:

$$\mathrm{CHD} = \frac{\sum_{t=1}^{T} \sum_{d \in \mathcal{S}_{\mathrm{HP}}} P_t^{d,\mathrm{u}}}{\sum_{t=1}^{T} \sum_{d \in \mathcal{S}_{\mathrm{HP}}} P_t^{d}} - 1$$

Note that the sets of DERs are denoted without 'u' in superscript as the change in total DERs output is considered, not only of the controllable DERs.

5.3 Reference Control Algorithms

The investigated controller is compared against two reference control algorithms:

1. **Constant cos(φ)**: All DERs operate with a pre-defined constant inductive power factor $\cos(\varphi^{\text{ind}}_{\text{const}})$. This is a very primitive static method that is provided by various inverter producers as a measure for overvoltage violations, e.g., [74, p. 11].

   $$Q^d_t = P^d_t \cdot \tan(\varphi^{\text{ind}}_{\text{const}}), \quad \forall d \in \mathcal{S}_D,\ \forall D \in \mathcal{D}$$

   Note that designating a power factor as 'inductive' offers a general description applicable to any sign convention, regardless of whether a DER is a producer or a consumer. For instance, an inductive power factor results in a positive Q in PSC for consumers, and a negative Q in ASC for generators. However, this case distinction is avoidable by denoting the inductive power factor with an 'ind' in superscript.

2. **Volt-Var Curve**: It is a decentralized active grid support method and is widely employed by state-of-the-art PV inverters available in the market, e.g., [74, p. 12]. In this scheme, PV inverters locally adhere to the so-called Volt-Var curve, which essentially serves as a reactive power lookup in relation to the local voltage magnitude at the inverter's connection node. The Volt-Var curve is parametrized as suggested by the standard settings of [39, §5.7.2.4], i.e., the interpolation points (0.93, 0.97, 1.03, 1.07) p.u. with a power factor limit of $\cos(\varphi_{\max}) = 0.9$ as depicted in Figure 5.3. For instance, a PV system with an array capacity of $10\,\text{kW}_\text{p}$ would have a reserve capacity for a corresponding $Q_{\max} = 10\,\text{kW}_\text{p} \cdot \tan(\arccos(0.9)) = 4.84\,\text{kVar}$. For the sake of completeness, the inverter of this PV system would be dimensioned as approximately $11.1\,\text{kVA}$.

   The inverters respond rapidly to changes in local grid voltage by adjusting their Q according to the Volt-Var curve, which in turn affects the voltage again. This situation presents a convergence problem where each of the local PV inverter controllers must decide at every simulation step for such a Q that aligns with the actual local voltage according to the Volt-Var curve. This convergence iteration is implemented within the framework outlined in Chapter 4. The Volt-Var curve strategy employed by the PV inverters is termed *LocalController*, as shown in Figure 4.1.

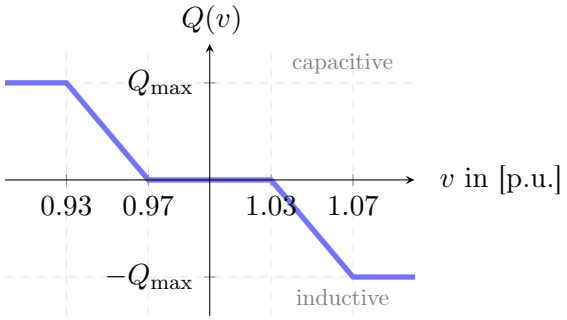


Figure 5.3 Adopted Volt-Var curve from [39, §5.7.2.4] in ASC for PVs

5.4 Base Parameter Set

5.4.1 State Estimation

The SE algorithm outlined in Section 3.2.1 is based on MLP. As such, the SE possesses hyperparameters that influence the learning process. To identify an appropriate parameter set for the state estimator, hyperparameter tuning is performed using Random Grid Search (RGS) [75]. Random combinations of the following hyperparameter sets are evaluated:

Table 5.3 Parameter Grid for Hyperparameter Tuning

Hyperparameter	**Set**
Hidden layers[3]	{1, 2, 3}
Hidden neurons	{8, 16, 32, 64}
Output layer	Only linear
Activation function	{ReLU, ELU, tanh} (see [76])
Batch size	{16, 32, 64}

In addition to the 40 randomly chosen parameter sets from Table 5.3, the simple linear regression [77, p. 19] (i.e., a single-layer perceptron with a linear activation function) is also included in the hyperparameter tuning, based on the experiences and feedback from the authors of [18]. The hyperparameter tuning results showed that the linear regression performs comparably to more complex neural networks, such as an MLP with 3 hidden layers each containing 64 neurons and using $\tanh(\cdot)$ activations followed by a linear output layer.

It is speculated that the obtained linear regressor may be associated with well-known linear branch-flow physical models, such as *LinDistFlow* [78]. In the subsequent controller tests, linear regression also exhibited better extrapolation characteristics compared to a nonlinear MLP, making it favorable to combine with, for example, a Q-controller that discovers grid operation points not present during the training phase of the state estimator. Therefore, linear regression is selected for use in the state estimator, see parameter *'state_estimation_architecture'* of Appendix A.

As outlined in Section 3.1.1, the grid measurement infrastructure is presumed to incorporate RT measurements at a 1 min resolution, specifically, the voltage at the secondary side of the substation, and P and Q measurements for each feeder. For primary investigation in this study, full SM penetration is assumed, with a total of 93 SMs providing historical voltage data at a 15 min temporal resolution.

5.4.2 Grid Control

As detailed in Chapter 3, RL is implemented as the fundamental algorithm for the control task. The configuration of the base controller parameters is provided in Appendix A, which is essentially part of the configuration file of the *Grid Operation Framework*, as seen in Chapter 4. In this subsection, the parameters related to the

[3] Note that the output layer is not included in the given numbers.

general control task and RL are elaborated upon in detail. Unless otherwise specified, these parameters are primarily used.

The control objective is set to minimize UV and OV under *'issues_to_mitigate'*. Voltage limits are defined as 0.95 p.u. and 1.045 p.u., respectively. The upper limit aligns with recommended LV voltage band usage [79]. Simulations executed under the defined scenario in Section 5.1 revealed that voltages falling below the UV limit of 0.935 p.u., as recommended by [79], occur rarely. Based on this observation, the lower voltage limit is selected to be slightly higher than recommended. This adjustment allows for testing of the control algorithm's performance against UV instances as well.

The *'mode'* parameter of the controller determines whether an RL agent should be trained during its current interactions with the environment or if it is being validated or deployed. The SAC algorithm is chosen as the RL learning algorithm. The ANN structure of SAC (refer to Section 2.4) is favored as the MLP for primary investigations. Another contribution of this thesis, 1D-CNNs for the control task, provides an alternative approach which will be introduced later along with the modifications for its implementation. Each of MLP-based ANNs of SAC have 3 hidden layers each containing 256 neurons and $ReLU$ as activation function. The output layers of ANNs are always linear.

Each episode of the RL environment comprises 1440 time steps, corresponding to a day at a 1 min resolution. As such, the RL agent starts a new episode at midnight each day, and the time horizon consists of 1440 transitions. In training mode, the RL agent collects data through random exploitation for the first four episodes in order to accumulate transitions to fill its replay buffer before it begins learning. As suggested by [50], this warm-up phase aids in collecting the underlying transition distribution to improve the entropy coefficient of SAC for faster learning. Based on [80], the agent is trained after each simulation step (*'train_freq'*) with a learning rate of $5 \cdot 10^{-4}$ and a batch size of 512 transitions. Additionally, the size of the replay buffer is capped at 10^5 transitions. The RL agent's ANNs are trained using a discount factor of $\gamma = 0$ to prioritize immediate rewards.

The DER control problem introduced in Section 3.2.2 is parametrized with the following parameters. All DERs in the grid are assumed to be dispatchable, that is, $\mathcal{S}_D = \mathcal{S}_D^{\mathrm{u}}$ for all DER types $D \in \mathcal{D} := \{\mathrm{PV}, \mathrm{EV}, \mathrm{HP}\}$. The parameter *'no_discriminative_ctrl'* is used to indicate that the DER control is applied uniformly to all controllable DERs, as described in Section 3.2.2. The inductive or capacitive power factor limit of PV inverters, represented by $\cos(\varphi_{\max})$, is set to 0.9. As explained in Section 5.3, the constant inductive power factor of EVs and HPs is defined as 0.97.

The penalty factors utilized in the reward function (3.6) are empirically determined and shown in Table 5.4. For OV violations, a heuristic is used which equates the penalty of sacrificing 10% of PV generation to an additional OV violation of

Table 5.4 Weighting Coefficients of Reward Function (3.6)

w^{ov}	w^{uv}	$w^{\mathrm{PV,p}}$	$w^{\mathrm{EV,p}}$	$w^{\mathrm{HP,p}}$	$w^{\mathrm{PV,q}}$	$w_{\mathrm{NCVI}}^{\mathrm{PV,q}}$
100	300	1	0.5	0.5	0.01	20

0.001 p.u. The Q-control is set to be 100 times less expensive than PV curtailment, as it does not require renewable energy curtailment. Nevertheless, unnecessary Q-control is penalized by an additional factor of 20 if the voltage at any grid node does not exceed $v_{\mathrm{CVI}} = 1.03$ p.u. To address UV mitigation with limited UV issues in data, the penalty for UV violations is increased, while the penalties for EV and HP curtailment are decreased compared to PV curtailment.

Chapter 6

Results

Simulation data are randomly selected on a daily basis from the 1-year data prepared in Section 5.1.3. Since the training, validation and test days are randomly selected, the data are not sequential, effectively eliminating biases due to seasonal trends.

To prevent data leakage, training and validation data of the state estimator and the RL agent are kept isolated from each other. The state estimator is trained with 64 days, validated with 7 days, and evaluated on 21 days. The labeled SE data are at a resolution of 15 min, but predictions from the state estimator use 1 min RT measurements, resulting in voltage estimates at a 1 min resolution. The RL agent is primarily trained with 32 days, with each day forming an episode. 8 days are set aside for periodic validation of the learning process, and 10 randomly selected days from the remaining data are used to assess the RL agent's performance.

6.1 State Estimation

The test performance of the linear regression-based SE, as per the KPIs outlined in Section 5.2.1, is presented in Table 6.1. Utilizing the derived linear regressor, the SE accomplishes an RMSE of 0.00902 p.u. when predicting $\mathbf{e}_t$ from $\mathbf{z}_t$. Given the heightened interest in critical voltages and the increasing inaccuracies associated with such due to their infrequent occurrence, the MD and MRD are also displayed. They have values of 0.01914 p.u. and 0.02003 p.u./p.u., respectively.

A comprehensive analysis of the SE performance is carried out on node 55 of the test grid (see Appendix E), which experiences the highest MD as well as MRD. The time series results, corresponding to the night of the highest MD and MRD, are depicted in Figure 6.1 over a span of three hours, which reveals that the SE effectively tracks the voltage trend. However, it does not flawlessly capture the voltage drop at 3:12 a.m., which is when the greatest MD and MRD occur.

Table 6.1 Test Accuracy of State Estimation

RMSE in [p.u.]	**MD in [p.u.]**	**MRD in [p.u./p.u.]**
0.00902	0.01914	0.02003

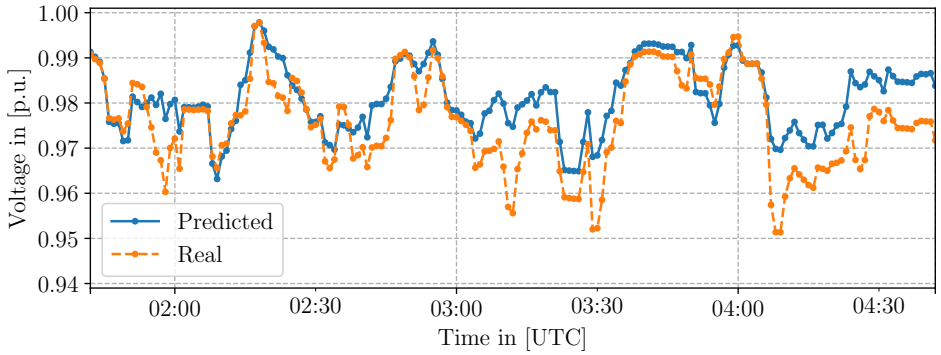


Figure 6.1 Predicted (——) and real (----) voltages at node 55 on 10.02.2019 of the highest estimation deviation 0.01914 p.u. at 3:12 a.m.

For a more thorough analysis, the estimation for this node across all test data is depicted as a scatter plot in Figure 6.2. The actual voltage measurement is plotted on the x-axis, while the estimated values are presented on the y-axis. A noteworthy observation from this figure is that the accuracy of the state estimation diminishes for lower voltages. Conversely, the predictions for higher voltages closely align with the optimal linear line. Given that the future scenario detailed in Section 5.1 results in fewer UV issues compared to OV, the SE demonstrates more accurate estimations for higher voltages than for lower ones, thereby explaining this observation.

It can be observed from Figure 6.1 that the majority of the predictions are placed above the actual values, which signals an overestimation trend. This inference is further confirmed by the histogram and box plot of the estimation deviations for the node being analyzed. The histogram, displayed in Figure 6.3, represents the frequency of estimation deviations' occurrences across 50 equally sized bins. It is evident that the estimation deviation is primarily centered around zero. The box plot, shown in Figure 6.3, encapsulates the 99% confidence interval of voltage estimations with its whiskers, with the box limits signifying the 50% confidence interval. The outliers, which fall outside the 99% confidence interval, are represented by the green dots. The box plot, in essence, is a single-axis representation of the histogram plot statistics. The median of the box plot, represented by the orange line, falls slightly below zero, $-8 \cdot 10^{-5}$ p.u., suggesting that a very small majority of voltages are

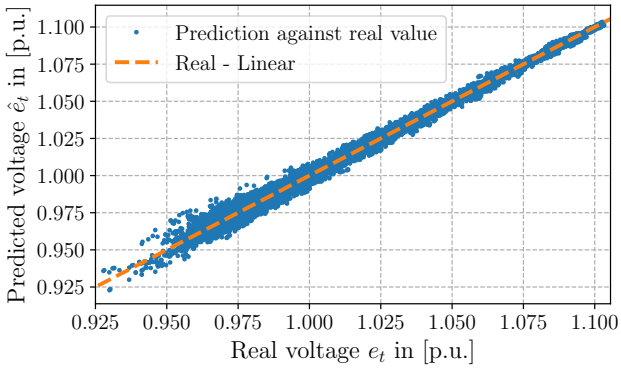


Figure 6.2 Voltage estimation for node 55 over all test data

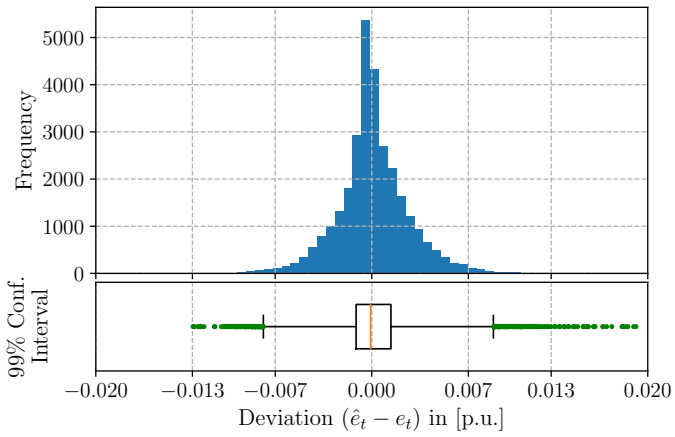


Figure 6.3 Voltage estimation deviation statistics (top: histogram, bottom: box plot) for node 55 over all test data

underestimated. However, the magnitude of these underestimations is not as high as that of the overestimations. It is important to highlight that this work does not delve into the underlying cause of such overestimations, as it exceeds the scope of this research. Nevertheless, the overestimations could be attributed to a physical characteristic of linear approximations of distribution grids that might tend to overestimate or could be a result of the prevalence of higher voltages in the defined test scenario.

6.2 Grid Control

This section analyzes the performance of the fully data-driven control algorithm introduced in Section 3.2 that exploits the state estimator obtained in Section 6.1 and whose controller is primarily parametrized as such in Section 5.4.2. The controller is tested on 10 days with a total of 14,400 time steps. It is compared against two reference controllers introduced in Section 5.3 and one hypothetical RL agent, called *State Expert RL*, that has access to RT voltage measurements at all nodes and therefore does not require SE. The KPIs described in Section 5.2.2 are used here for performance evaluation. Appendix B summarizes the detailed output of the *Evaluator* introduced in Section 4.1 for the primary results.

Table 6.2 demonstrates that the proposed controller significantly mitigates OV violations, reducing the Total Overvoltage Violation Duration (TOVD) from over

Table 6.2 Overvoltage Mitigation Results

	Constant cos(φ)	**Volt–Var Curve**	**State Expert RL**	**SE+RL**
TOVD in [min]	2708	2932	33	6
AOVD in [min]	90.3	127.7	1.7	2.0
IOV in [p.u. · min]	83.83	67.78	0.10	0.14

Table 6.3 Undervoltage Mitigation Results

	Constant $\cos(\varphi)$	Volt–Var Curve	State Expert RL	SE+RL
TUVD in [min]	263	257	59	171
AUVD in [min]	4.5	4.8	1.4	2.5
IUV in [p.u. · min]	2.25	2.24	0.29	1.18

45 hours to 6 min. Likewise, the proposed controller reduces the Average Overvoltage Violation Duration (AOVD) from over 90 min of the baseline controllers to 2 min. The cumulative OV impact, Integral of Overvoltage Violation (IOV), is reduced significantly with the investigated control algorithm, compared to baseline algorithms. Compared to State Expert RL agent, the IOV is slightly higher around by 0.04 p.u.·min due to SE deviations.

Notably, the Volt-Var curve cannot improve the occurrence or duration of voltage violations compared to the constant $\cos(\varphi)$, but it softens greater OV as the obtained integral indicator shows. This is due to the deadband of the Volt-Var curve, which limits the potential of Q. Another reason is that the inverters are controlled in a decentralized manner, meaning not all PV inverters detect a voltage incident and respond simultaneously, as would be the case with a constant power factor. Nonetheless, the inverters observing a significant OV operate with an inductive power factor up to 0.9, resulting in an improved voltage drop at these more critical nodes.

Table 6.3 reports the performance of the studied controllers in UV situations. Although the future scenario described in Section 5.1 results in fewer UV issues compared to OV, the proposed controller still reduces the integral UV impact indicator by 48% compared to the baseline controllers. Note that the absence of UV training data also affects the overall performance of the proposed data-driven controller.

Table 6.4 sums up the DER interventions leading to the results in Table 6.2 and Table 6.3. For voltage mitigation support, the Volt-Var controller results in an additional PV inverter loading of 0.62%. The proposed controller curtails 8.29% of PV generation to mitigate OV issues, slightly more than the 6.11% of the State Expert RL agent. This higher curtailment leads to a correspondingly lower PV inverter loading. It is observed that the proposed controller utilizes the available reactive power potential more effectively compared to the local Volt-Var controller, leading to a higher PV inverter loading.

Table 6.4 DER Intervention Results

Change	Constant $\cos(\varphi)$	Volt–Var Curve	State Expert RL	SE+RL
CPO in [%]	Reference	0.0	-6.11	-8.29
CPIL in [%]		+0.62	+6.08	+3.92
CED in [%]		0.0	-0.46	-0.45
CHD in [%]		0.0	-5.24	-1.49

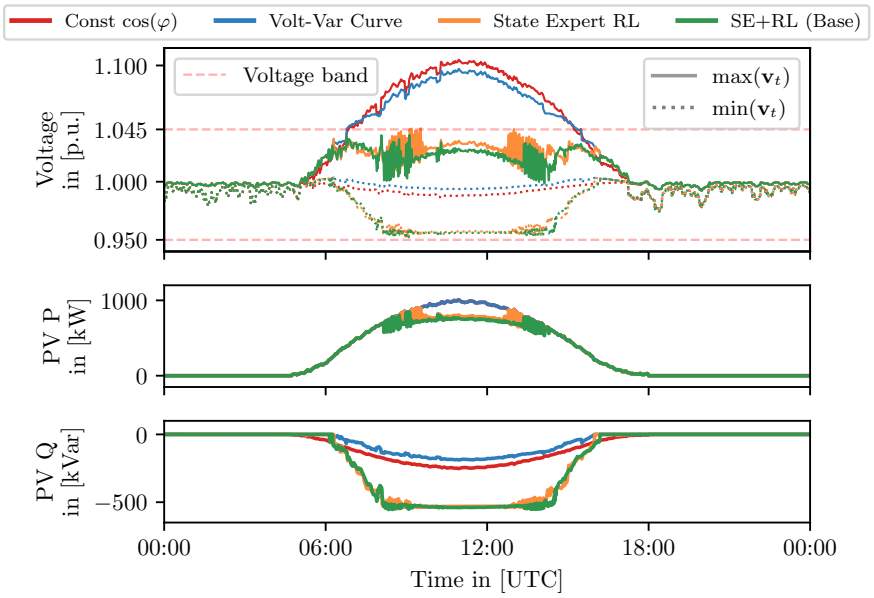


Figure 6.4 Maximum (——) and minimum (- - - -) voltages along all nodes (top), total P (middle) and Q (bottom) of all PVs for an exemplary sunny test day (April 7, 2019)

To mitigate UV issues, as Change in EV Demand (CED) and Change in HP Demand (CHD) show, the RL agent mainly curtails HPs instead of EVs due to the higher utilization factor and energy demand of HPs.

Across all test days, in summary, the investigated SE-based RL controller successfully reduces the cumulative impact of both OV and UV by 98% compared to two baseline methods by curtailing distributed generation only by 8.29% and consumption by less than 2%.

The performance of the studied controllers for an exemplary sunny test day (April 7, 2019) is depicted in Figure 6.4. The resulting voltage profiles indicate that the SE-based RL controller effectively keeps voltage within the admissible voltage band. The total PV power output for the sunny day reveals how the controller successfully prioritizes inductive Q over P-curtailment. Specifically, it begins curtailing PV only after the inverters reach their power factor limit at around 8 a.m. in UTC, and it ceases curtailing PV around 2:30 p.m when stressed times are over. As such, the RL controller achieves peak shaving.

To facilitate comparative analysis, the system states at times of peak voltages, using both the Volt-Var curve and the SE-based RL controller, are graphically represented on the test grid in Appendix E.1 and Appendix E.2, respectively. The nodal voltage magnitudes are represented in colors, corresponding to the colormap exhibited on the right. Compared to the Volt-Var curve, the investigated controller reduces the maximum voltage instance from 1.097 p.u. at nodes 55 and 66 to 1.04 p.u. Note that these peak voltages occur at different times, specifically at 10:57 a.m. and 9:07 a.m. On this particular sunny day, the RL agent only curtails the PVs by a total of 13.4% as it achieves a peak shaving. Significantly, it decides on a systematic 243% increase of inductive Q compared to the Volt-Var curve.

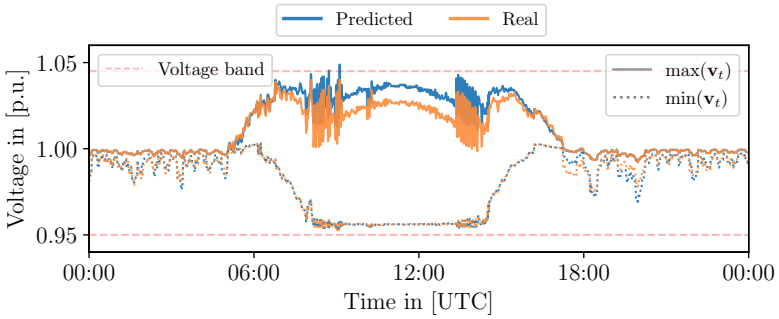


Figure 6.5 Maximum (——) and minimum (- - - -) voltages for the sunny day: Observed extremes (——) by SE-based RL controller and real extremes (——)

Note that the proposed controller behaves more conservative than the State Expert RL agent as seen in Figure 6.4. This is because the state estimator tends to slight voltage overestimation, see Section 6.1. The RL agent thus observes voltages that are higher than the actual voltages in the grid, as shown in Figure 6.5. This leads to a higher PV curtailment compared to State Expert RL.

Another observation from Figure 6.4 is that the SE-based RL controller shows slight fluctuations in PV output. This results in a slightly noisy voltage profile at the times when the controller starts and stops curtailing PV generation. This happens because the control characteristics are modelled to lag by one time step, as explained in Section 4.1. The controller decides to curtail when it observes OV. After the control is applied in the next simulation step, the voltage drops back to the desired range. As the controller no longer detects OV instances, it ceases curtailment. However, this cessation induces OV in the succeeding time step, thereby leading to an alternating pattern in the voltage profile. This issue is addressed in Section 6.2.1.

The controller's performance on an exemplary cloudy test day (February 9, 2019) is also depicted in Figure 6.6. The controller resists the UV issue during the evening peak, primarily by curtailing HPs, as previously mentioned. However, minor UVs persist. The fluctuations in control decisions of the SE-based RL controller are evident compared to State Expert RL. These fluctuations, which were explained earlier, are the primary cause of the performance degradation observed in the proposed controller. As the performance on UV incidents has been demonstrated once, and they appear to be less critical, the subsequent investigations will mainly focus on analyzing OV mitigation that is of greater interest.

6.2.1 Learning from Temporal Data

The primary results, particularly in Figure 6.4, demonstrate that the control characteristic that lags by one time step causes fluctuations in voltage mitigation. To address this issue, the SE-based RL controller in Section 3.2.2 has been slightly modified. Since the main cause of the fluctuations is identified as the alternating voltage observations, the temporal correlation and voltage changes are taken into account by modeling the RL observation $\mathbf{s}_t$ as multi-step:

$$\mathbf{s}_t = \begin{bmatrix} \mathbf{z}_t^{\mathrm{T}} & \hat{\mathbf{e}}_t^{\mathrm{T}} & \mathbf{z}_{t-1}^{\mathrm{T}} & \hat{\mathbf{e}}_{t-1}^{\mathrm{T}} & \mathbf{a}_{t-1}^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}} \tag{6.1}$$

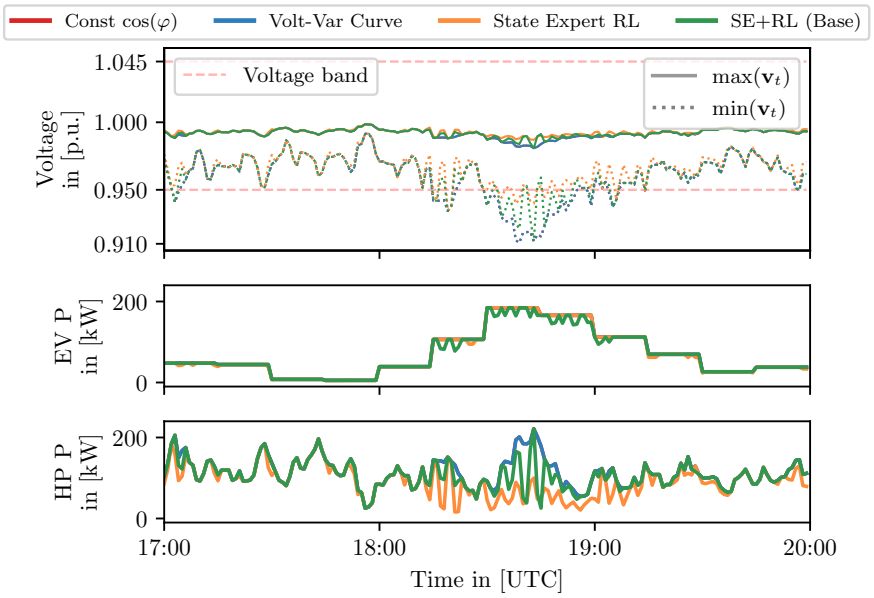


Figure 6.6 Maximum (——) and minimum (- - - -) voltages along all nodes (top), total P of all EVs (middle) and all HPs (bottom) for an exemplary cloudy test day (February 9, 2019)

i.e., the RL agent decides on the P-curtailment and Q-control factors introduced in Section 3.2.2 based on both the previously taken action and the prior and current RT measurements and voltage estimates. Taking previous state and action into consideration allows inferring from time series and providing a short-term (two-step) memory to the RL agent. In terms of control theory, two-step observation implicitly provides the differentials of the system state.

There are various ANN architectures possible for realizing such a temporal feature extraction problem. One straightforward solution is to keep temporal sequential as shown in Equation (6.1) and feed straightforward into an MLP-based RL as introduced in Section 5.4.2. This *Multi-Step MLP RL* has the disadvantage of that the input dimension of the MLPs explode with increasing number of considered time steps.

One alternative to this problem is utilizing 1D-CNNs. Unlike the two-dimensional CNNs, the kernels in 1D-CNNs are one-dimensional, hence the name. In this work, CNN layers are placed prior to actor and critic networks of SAC for feature extraction. This enables temporal learning to occur in the CNN layers, providing only the most relevant information to the following MLP-based actor and critic networks. 1D-CNNs are chosen due to their efficient learning from time series data [81].

The state inputs for 1D-CNNs are two-dimensional, with one dimension representing the features and the other representing time:

$$\mathbf{s}_t \coloneqq \left[\begin{bmatrix} \mathbf{z}_{t-1}^{\mathrm{T}} & \hat{\mathbf{e}}_{t-1}^{\mathrm{T}} & \mathbf{a}_{t-1}^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}} \quad \begin{bmatrix} \mathbf{z}_{t}^{\mathrm{T}} & \hat{\mathbf{e}}_{t}^{\mathrm{T}} & \mathbf{a}^{0\mathrm{T}} \end{bmatrix}^{\mathrm{T}} \right] \tag{6.2}$$

Note that a padding is applied in Equation (6.2) with the ineffective action vector $\mathbf{a}^0$ to keep the length of two vectors consistent, as the action $\mathbf{a}_t$ to be decided based on $\mathbf{s}_t$ is not known yet.

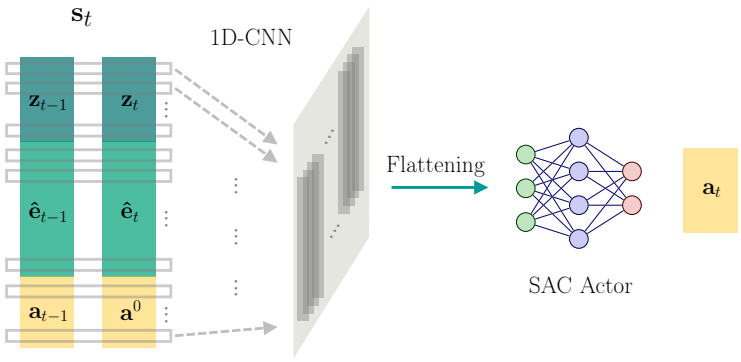


Figure 6.7 Temporal learning from two-step observation with 1D-CNN, illustrated for SAC actor

Figure 6.7 illustrates the described application of 1D-CNN, exemplary for SAC's actor, see Figure 2.4.

1D-CNNs of actor and critic networks have two layers and 64 output channels, with kernel sizes of 2 and 1 for the former and latter layers, respectively. It is important to note that the CNN layers are followed by actor and critic networks that are further modeled as introduced in Section 5.4.2. This CNN-based version of the algorithm is called *Multi-Step CNN RL*, which is compared against the single-step and MLP-based versions in the following analysis.

Since the complexity of the ANN increases with an increase in input parameters and/or network depth, the training size is extended from 32 days to 48 days. Similarly, the number of validation days is also increased from 8 to 12 days.

Before taking a closer look at improvements in fluctuations on a sunny day, Table 6.5 and Table 6.6 serve to numerical comparison of voltage mitigation performance between the primary results and the RL agent enhanced with temporal learning. In comparison to the single-step results in Table 6.2 and Table 6.3 (rightmost columns), both SE-based multi-step agents outperform the single-step version in regard to two important KPIs. Although the number of voltage instances increases slightly, the integral impact of OVs, the IOV, is reduced by 78% and 50% with the MLP and CNN versions, respectively. The AOVD is improved by 15% with both. The decrease in IOV and AOVD despite the increase in TOVD indicates that the multi-step version softens the OV violation amplitudes but allows more frequent smaller violations. It is worth noting that the investigated SE-based RL controller closes the gap to the hypothetical State Expert RL agent after introducing temporal learning both with MLP and CNN. On the other hand, the performance in

Table 6.5 Overvoltage Mitigation Results with Temporal Learning

	Multi-Step MLP RL	**Multi-Step CNN RL**
TOVD in [min]	11	24
AOVD in [min]	1.7	1.7
IOV in [p.u. · min]	0.03	0.07

Table 6.6 Undervoltage Mitigation Results with Temporal Learning

	Multi-Step MLP RL	**Multi-Step CNN RL**
TUVD in [min]	173	179
AUVD in [min]	2.6	2.6
IUV in [p.u. · min]	1.03	1.05

UV mitigation does not significantly change. The Average Undervoltage Violation Duration (AUVD) increases by 0.1 min, i.e., an increase of 4%, and the Integral of Undervoltage Violation (IUV) drops by $11 - 13\%$.

Table 6.7 summarizes the DER interventions by both multi-step agents. It is important to note that the values are again referenced against the constant $\cos(\varphi)$ from Table 6.4. A highly significant outcome of these results is the reduction of PV curtailment by 35% and 37% with the MLP and CNN versions, respectively, compared to the single-step version. This indicates that the grid is operated more sustainably, resulting in approximately 3% higher DER power.

Regarding the numerical comparison between MLP-based and CNN-based temporal learning, the CNN-based agent operates the grid in a slightly more sustainable manner, whereas it allows for slightly higher voltage violations. However, the difference between the two is not significant compared to baseline controllers. This could be attributed to the fact that the number of time steps incorporated for temporal learning is limited to 2. The learning efficiency of the CNN-based agent may become clearer with the inclusion of more time steps, as the MLP-based agent may face issues arising from the so-called "curse of dimensionality" phenomenon [82]. However, this thesis does not provide further comparison with respect to dimensional complexity. Nonetheless, this thesis showcases the potential of 1D-CNNs in RL applications in the power system domain. Drawing from the survey [10], this study appears to be a pioneering application of 1D-CNNs in grid control.

Figure 6.8 demonstrates the performance of the multi-step agents on the same sunny day (April 7, 2019), alongside the primary results. It is important to note that the green curve in Figure 6.8 is the same curve as in Figure 6.4, and is included here for easier visual comparisons. It is evident that both multi-step agents perform similarly, as previously discussed, which is reflected in the significant overlap of their curves. The main takeaway from this comparison is that involving temporal learning significantly reduces voltage fluctuations, resulting in a more stable and reliable grid.

Table 6.7 DER Intervention Results with Temporal Learning

	Multi-Step MLP RL	**Multi-Step CNN RL**
CPO in [%]	-5.36	-5.20
CPIL in [%]	+5.23	+5.20
CED in [%]	-1.16	-0.37
CHD in [%]	-1.34	-1.36

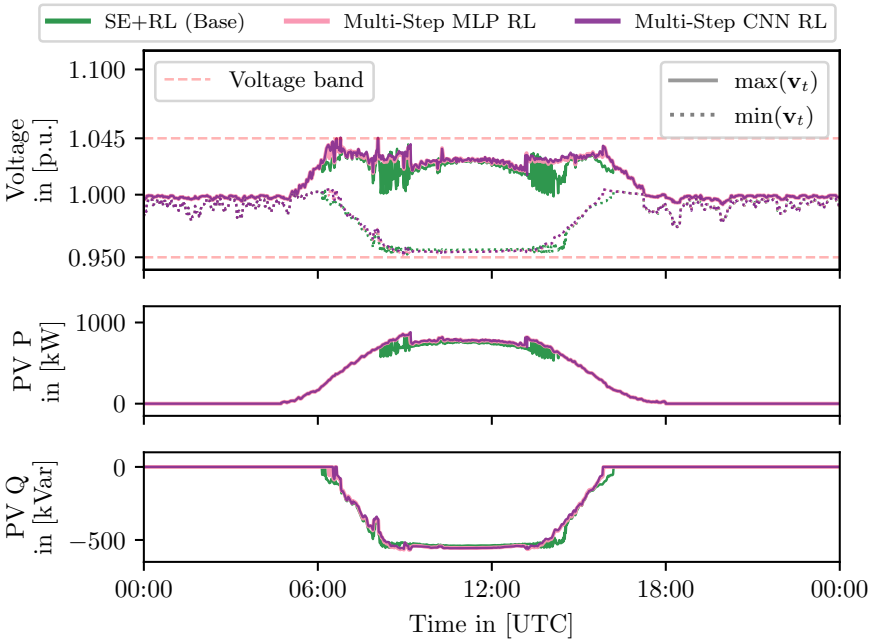


Figure 6.8 Maximum (——) and minimum (- - - -) voltages along all nodes (top), total P (middle) and Q (bottom) of all PVs for the sunny day with SE-based multi-step RL controllers

Additionally, the leveraged algorithm demonstrates a more efficient use of the voltage band, approaching its limits more closely and therefore curtailing less PV.

6.2.2 Sensitivity on Smart-Meter Penetration and Location

This section investigates the validity of the SE-based RL controller with fewer SM installations, after the concept has been proven with previous results using SMs installed at all downstream nodes. To this end, two test cases are designed. The first test case, called *SMs Only Ends*, assumes that only the 6 nodes at the end of the feeders have an SM installed, as most critical voltage incidents emerge at those nodes, as shown in previous results. As a counterexample, the second test case, called *SMs Only Mids*, assumes that the 6 nodes are not located at the end of the feeders but instead at intermediate nodes located away from the end of the feeders. The SM locations for both cases are shown in figures in Appendix E, with numbers in blue for the *SMs Only Ends* test case and in grey for the *SMs Only Mids* test case. Note that the colormaps do not represent to the results of this section. They are only referred to demonstrate the SM locations in the existing topology.

It is worth noting that the analysis further assumes that all DERs are controllable, to ensure comparability of sensitivity analysis with the same flexibility capacities. While not all nodes have an SM installed, it is practically feasible to assume that DERs at nodes without an SM are also controllable, as discussed in [83].

For each of both of the test cases, a separate state estimator is trained using the corresponding SM penetration and locations. The RL agents are also trained separately for each case, with the respective state estimator. The agents are modeled

Table 6.8 Overvoltage Mitigation Results with Low SM Penetration

	SMs Only Ends	**SMs Only Mids**
TOVD in [min]	35	3387
AOVD in [min]	2.3	211.7
IOV in [p.u. · min]	0.10	111.75

with a CNN architecture for two-step observations, as described in Section 6.2.1. Therefore, the analysis conducted here is compared against the results of *Multi-Step CNN RL* from Section 6.2.1, ensuring comparability.

Table 6.8 and Table 6.9 provide a summary of the results for both scenarios concerning OV and UV mitigation, respectively. Compared to the results of *Multi-Step CNN RL* in Table 6.5 and Table 6.6 (rightmost columns), the *SMs Only Ends* version shows similar results to the full SM penetration case. The AOVD only increases by 0.6 min and the IOV by 0.03 p.u.·min although the number of SMs is reduced from 93 to 6. Furthermore, the RL agent curtails PVs 3.06%, as demonstrated in Table 6.10, compared to 5.2% of the full SM penetration case. Note that the DER intervention results are again referenced against the constant $\cos(\varphi)$ from Table 6.4. In addition, the UV mitigation performance remains almost the same.

These results reveal that the proposed SE-based RL control algorithm is feasible even with a low number of SMs. However, it is important to note that the SMs in the first test case, *SMs Only Ends*, are located at the most favorable nodes.

On the other hand, the second test case, *SMs Only Mids*, indicates that the location of the SMs plays a more critical role than the SM penetration level. As the RL agent cannot observe significant voltage increases with the estimations for nodes located away from the feeder ends, the agent does not find DER intervention necessary, as shown in Table 6.10. This leads to more problematic OV issues, as given in the right columns of Table 6.8 and Table 6.9. It underperforms even when compared to baseline algorithms, cf. Table 6.2, since the agent suppresses Q-control.

Another remarkable finding is that UV mitigation results of *SMs Only Mids* are identical to the Volt-Var curve's results in Table 6.3. This is because the Volt-Var feature is only available for PV inverters, which are typically inactive during times of UV. Consequently, capacitive Q support to increase voltage is not possible for both control algorithms. Furthermore, the curtailment of EVs and HPs are ruled out for *SMs Only Mids* as well, as the nodes with SMs do not encounter significantly low voltages. Therefore, both methods yield the same KPI values for UV cases.

Table 6.9 Undervoltage Mitigation Results with Low SM Penetration

	SMs Only Ends	**SMs Only Mids**
TUVD in [min]	174	257
AUVD in [min]	2.8	4.8
IUV in [p.u. · min]	1.02	2.24

Table 6.10 DER Intervention Results with Low SM Penetration

	SMs Only Ends	**SMs Only Mids**
CPO in [%]	-3.06	0.0
CPIL in [%]	+6.54	-1.66
CED in [%]	-0.68	0.0
CHD in [%]	-1.32	0.0

Figure 6.9 depicts the OV mitigation performance in both test cases. As shown, the performance of *SMs Only Ends* is comparable to that of the case with complete SM penetration. However, *SMs Only Mids* leads to a prolonged OV that reaches its peak at 1.108 p.u.

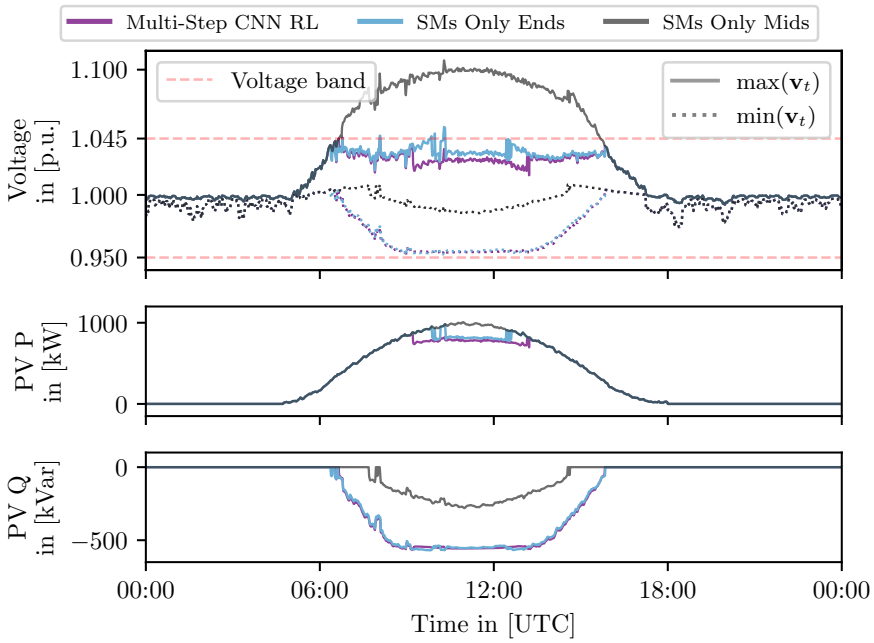


Figure 6.9 Maximum (——) and minimum (- - - -) voltages along all nodes (top), total P (middle) and Q (bottom) of all PVs for the sunny day with only 6 SMs

The examined SE-based RL controller has demonstrated effective performance even with a limited number of SMs, provided they are located at critical nodes. Although higher penetration levels increase the likelihood of covering critical nodes, identifying these nodes in advance can reduce costs. Therefore, this study proposes that the SM rollout should primarily focus on covering critical nodes and promoting SM installations at these locations, rather than relying on unconscious or uncontrolled SM installations in regard to economic benefits and effectiveness of the proposed control algorithm.

Chapter 7

Conclusion and Outlook

This thesis presents a fully data-driven control algorithm that addresses voltage violation mitigation in LV grids without prior physical model knowledge or costly downstream RT measurements. The algorithm relies on RT measurements at the substation and historical SM data to operate the grid within its limits. The approach includes a data-driven state estimator that feeds an online-trained RL agent. The RL agent determines suitable power factors and curtailment signals for dispatchable DERs to mitigate voltage violations.

To validate the proposed control algorithm, an experimental setup is established. Two lists of open-source datasets containing DER profiles and profile generation tools are compiled after a literature review. Based on this review, a test dataset with a temporal resolution of 1 min for a period of 1 year is created to serve as a benchmark for evaluating the approach. Furthermore, a simulation framework is designed and developed to enable the development and testing of various grid control algorithms under different grid operation scenarios. This internal framework provides a valuable platform for future research on new control algorithms for LV grid operation.

The model-free, data-driven state estimator shows its best performance using a simple linear regression. The DER control task and the RL environment are modeled based on a realistic use case, taking into account the lack of accurate grid models and automation infrastructure at the LV level. The reward function of the RL agent is empirically parametrized to minimize overvoltage and undervoltage issues as well as DER interventions. The SAC algorithm is utilized as the learning scheme of the RL agent.

Primary experiments on the test grid show that the fully data-driven control algorithm reduces voltage violations by 98% compared to two baseline methods. This reduction in voltage violations is achieved by curtailing distributed PV generation by only 8.29% and consumption of EVs and HPs by less than 2%.

In addition, the control algorithm is enhanced by extending the observation space of the RL agent with multiple time steps. Learning from time series data is realized using 1D-CNNs. This leads to twofold improvements: first, temporal learning addresses fluctuations in DER control, ensuring a more stable grid. Second, the controller intervenes in DERs less, curtailing PVs by only 5.2%. This study demonstrates the potential of learning from time series data and 1D-CNNs for grid control applications.

Examining the effect of SM penetration level and their location on the perfor-

mance of the proposed control algorithm is another important contribution of this thesis. The results show that the algorithm delivers promising results even with a lower penetration level of SMs, as long as they are installed at critical grid nodes. This information can be used by grid operators to optimize the placement of SMs for better performance and lower SM installation costs.

To summarize, this thesis demonstrates the feasibility of reliable and cost-effective LV grid control without prior physical model knowledge or costly downstream RT measurements. The investigated controller outperforms baseline methods, showing significant reduction in voltage violations and overall improvement in service quality. Simulation experiments also indicate that the controller can handle uncertainties arising from SE deviations, suggesting its robustness for practical applications.

Despite the thorough analysis in this thesis, there are still avenues for future work. Moving forward, one potential area is to examine the impact of measurement device errors on the proposed approach's performance. Another open research question is the transferability of pre-trained RL agents to unknown grids with different characteristics where online training might not be possible. Future research can focus on developing "transfer learning" techniques for such use cases or exploring the use of an already trained RL agent on a grid with changing DER penetration or changing topology to enhance the approach's adaptability. Additionally, augmenting the RL controller with primitive physical models, such as sensitivity models that may be available to a grid operator, is another interesting use case and research direction.

Appendix A

Base Configuration File for Grid Operation Framework

```yaml
issues_to_mitigate:
  - undervoltage
  - overvoltage
operation_limits:
  undervoltage: 0.95  # in [p.u.]
  overvoltage: 1.045  # in [p.u.]
state_estimation:
  architecture: "linear_regression"
  rt_available: "secondary_v_and_root_line_flows"
  sm_available: "all_downstream_nodes"
control_policy: "rl"  # "RL"/"const_cos_phi"/"Volt-Var"/...
controller_parameters:
  mode: "train"  # RL mode: "train"/"validate"/"test"
  agent_type: "SAC"  # RL algorithm
  ann_architecture:
    arch_type: "MLP"  # "MLP"/"CNN"
    params:
      hidden_layers: 3  # no. of MLP layers in RL ANN's
      hidden_neurons: 256  # no. of hidden neurons in each layer
  episode_length: 1440  # length of each episode, 1 day w/ 1min resolution
  learning_starts_after_episode: 4  # 4 days to collect data
  train_freq: 1  # after how many sim. steps to train RL
  learning_rate: 0.0005  # default is 0.0003 in SAC paper
  batch_size: 512  # gradient descent batch size
  buffer_size: 100000  # replay buffer max. transition capacity
  seed: 0  # SAC random sampling seed during exploitation
  gamma: 0.0  # discount factor
  controllable_ders: "all"  # all DERs are controllable
  no_discriminative_ctrl: True  # whether discriminative control
  max_cos_phi: 0.9  # maximum power factor, both inductive or capacitive
  ev_hp_const_cos_phi: 0.97  # EVs' and HPs' constant power factor
  w_ov: 100  # penalty factor for overvoltage
  w_uv: 300  # penalty factor for undervoltage
  w_pv_p: 1  # penalty factor for PV curtailment
  w_ev_p: 0.5  # penalty factor for EV curtailment
  w_hp_p: 0.5  # penalty factor for HP curtailment
  w_pv_q: 0.01  # penalty factor for PV Q control
  w_pv_q_ncvi: 20  # multiplier PV Q penalty if NCVI
  v_cvi: 1.03  # critical voltage increase indication limit, in [p.u.]
```

Appendix B

An Evaluation Log File

```
-------------- Installed DER Flexibility Capacities --------------
    Installed PV capacity: 1185.0 kW
    Installed EV capacity: 1096.9 kW
    Installed HP capacity: 693.0 kW

---------------- DER Flexibility Control Analysis ----------------
  Uncontrolled:
    Total PV P injection: 45842.96 kWh
    Total PV Q injection: 0.0 kVarh
    Total PV Q demand: 11489.71 kVarh
    Total EV demand: 3398.58 kWh
    Total HP demand: 13949.36 kWh

  After control:
    Total PV P injection: 42043.45 kWh
    Total PV Q injection: 87.21 kVarh
    Total PV Q demand: 23053.25 kVarh
    Total EV demand: 3383.41 kWh
    Total HP demand: 13740.9 kWh

    PV curtailment: 8.29 %
    PV apparent power change: +3.92 %
    EV curtailment: 0.45 %
    HP curtailment: 1.49 %

------------------------ Voltage Analysis ------------------------
  Overvoltage:
    Steps with overvoltage: 6
    Duration in [h]: 0.1
    Ratio of all test time steps: 0.04%
    Integral of overvoltage in [pu*h]: 0.00235
    Sequential overvoltage violations more than:
        (2 steps): 3
        (5 steps): 0
        (10 steps): 0
    Average overvoltage duration: 2.0 min

  Undervoltage:
    Steps with undervoltage: 171
    Duration in [h]: 2.85
    Ratio of all test time steps: 1.19%
    Integral of undervoltage in [pu*h]: 0.0196
    Sequential undervoltage violations more than:
        (2 steps): 48
        (5 steps): 8
        (10 steps): 1
    Average undervoltage duration: 2.49 min
```

Appendix C

Open-Source Profile Datasets

Table C.1 Open-Source Profile Datasets

	No. Profiles	Profile Type	Resolution	Timespan	Region	Link[1,2]
Low Carbon London	5567	HH	30 min	2.5 years (2011-2014)	London, UK	Hyperlink
WPuQ	38	HH (38), HP (38), PV (3)	10 s	2.5 years (2018-2020)	Hamelin, Germany	Hyperlink
HTW	74	HH	1 s, 1 min	1 year (2010)	Berlin, Germany	Hyperlink
Simbench	52	HH (5), Comm (15), PV (8), EV (12), HP (12)	15 min	1 year (2010)	post-processed HTW	Hyperlink
NREL	> 900,000 (synthetic)	HH, Comm	15 min	1 year (2016)	USA	Hyperlink
Open Power System Data – Household Data	6 HH, 3 Comm, 2 Public	HH, Comm	1 min	4 years	Konstanz, Germany	Hyperlink
Adaptive Charging Network (ACN)	112 (over 30,000 charging sessions)	EV (public)	No time series data, only session info	2018 - present	California, USA	Hyperlink
FfE Synthetic HH Load Profiles	940 (synthetic)	HH	1 min	1 year	Germany	Hyperlink
FfE Real BEV Session Profiles	2054 sessions	EV (public)	5 min	3 years (2019-2021)	Munich, Germany	Hyperlink
FfE Synthetic BEV Profiles	5 (synthetic)	EV (residen-tial+public)	1 h	1 year (2030)	Germany	Hyperlink
London Solar Generation Data	6	HH, PV, Solar Irradiance	10 min	1.5 years (2013-2014)	London, UK	Hyperlink
NextGen (NG)	5000	HH	5 min	2016 - present	Canberra, Australia	Hyperlink

[1]Links are only provided in digital form as hyperlinks, in terms of usefulness.
[2]Links are accessible as of the submission date of this thesis.

SHED	300	HH, PV, hot water tank	30 min	∼3 years (2010-2013)	Sydney, Australia	Hyperlink
GREEND	9	HH	1 s	8 months	Austria and Italy	Hyperlink
RBSA Metering Data	101	HH	15 min	5 years (2006 - 2010)	Pacific Northwest, USA	Hyperlink
ECO (Electricity Consumption & Occupancy)	6	HH	1 s	8 months (2012-2013)	Switzerland	Hyperlink
REFIT: Electrical Load Measurements	20	HH	8 s	2 years (2013 - 2015)	UK	Hyperlink
SustDataED	50	HH	1 min	2010 - present	Portugal	Hyperlink
IDEAL	39-255	HH	1 s	20 months (39 HH)	UK	Hyperlink
UCSD Uni Microgrid	> 10 Public, 26 PV, 210 EV, 1 CHP, 2 BESS	Microgrid w/ Public, PV, EV, HP, BESS	15 min	5 years (2015-2020)	San Diego, USA	Hyperlink
BLOND-50	1	Comm	0.02 ms	213 days	Germany	Hyperlink
Borealis	30	HH	6 s	1 year	Waterloo, Canada	Hyperlink
CoSSMic	11	Residential, Comm (+ PV, EV)	1 min, 15 min, 1 h	1-3 years	Germany	Hyperlink
Energy Informatics Group Pakistan	42	HH	1 min	1 year	Pakistan	Hyperlink
HIPE	10	Industrial	5 s	3 month	Karlsruhe, Germany	Hyperlink
Home Electricity Survey (HES)	255 (1 month), 26 (1 year)	HH	2 min	1 month-1 year	UK	Hyperlink
SMART* Apartment	114	HH	1 min	2 years	USA	Hyperlink
SMART* Home 2017	7	HH	1 s	> 2 years	USA	Hyperlink
SMART* Microgrid	443	HH	1 min	1 day	USA	Hyperlink
Sustainable Building Energy Systems 2017	23	HH	1 min	1 year	Ottawa, Canada	Hyperlink
OPEN EI Data Lake				Hyperlink		

Appendix D

Open-Source Profile Generation Tools

Table D.1 Open-Source Profile Generation Tools

	Profile Type	**Possible Resolution**	**Link**[1,2]
SIMONA	HH, PV, EV	1 h	Hyperlink
WRMC-BSRN/PANGAEA	Solar Irradiance	1 min, 3 min	Hyperlink
PVLIB	PV	15 min via CAMS	Hyperlink
PROCSIM	HH, PV, Wind	HH: 1 s, PV: 5 min	Hyperlink
LoadProfileGenerator	HH	1 min - 1 h	Hyperlink
Helioclim-3 Archives	Solar Irradiance	1 min	Hyperlink
Meteostat	Weather Data (Solar Irradiance, Wind, Temperature)	1 h	Hyperlink
PySoDa	PV Resolution Increase	5 s	Hyperlink
EHDO	HP, PV, BESS	1 h	Hyperlink
Emobpy	EV	1 min - 1 h	Hyperlink
FfE PublicEVChargingLoadProfileTool	EV	5 min	Hyperlink
EVLPG	EV	Several minutes	Hyperlink
NILMTK	HH appliances	1 min	Hyperlink

[1]Links are only provided in digital form as hyperlinks, in terms of usefulness.
[2]Links are accessible as of the submission date of this thesis.

Appendix E

Voltage Colormaps

E.1 Using Volt-Var Curve

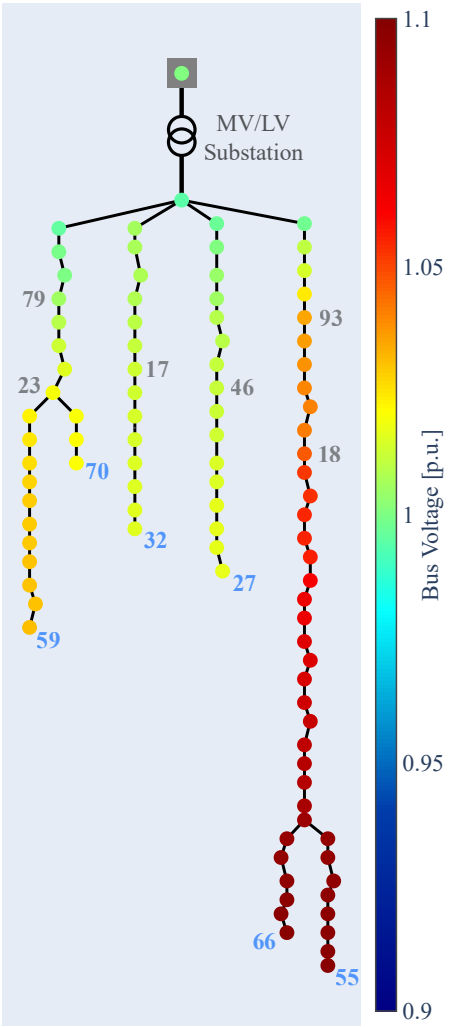


Figure E.1 Voltage colormap at time of the highest voltage (10:57 a.m. (UTC)) on a sunny test day (April 7, 2019) with the Volt-Var curve controller

E.2 Using Proposed SE-Based RL

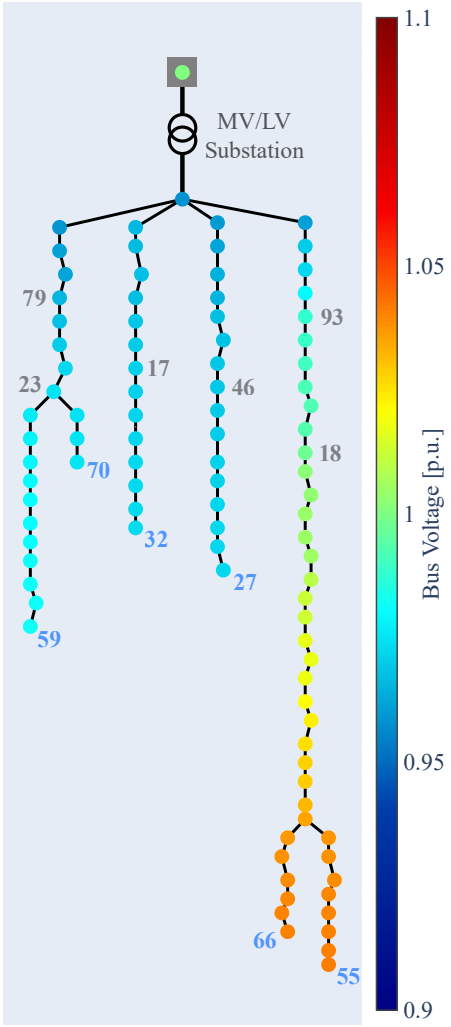


Figure E.2 Voltage colormap at time of the highest voltage (9:07 a.m. (UTC)) on a sunny test day (April 7, 2019) with the SE-based RL controller using base parameter set in Appendix A

Bibliography

[1] *Unlocking the Potential of Distributed Energy Resources*. Tech. rep. International Energy Agency (IEA), 2022. URL: https://www.iea.org/reports/unlocking-the-potential-of-distributed-energy-resources (visited on 10/15/2023).

[2] Vasileios A Evangelopoulos, Pavlos S Georgilakis, and Nikos D Hatziargyriou. "Optimal operation of smart distribution networks: A review of models, methods and future research". In: *Electric Power Systems Research* 140 (2016), pp. 95–106. DOI: 10.1016/j.epsr.2016.06.035.

[3] Priti Paudyal et al. "The Impact of Behind-the-Meter Heterogeneous Distributed Energy Resources on Distribution Grids". In: *2020 47th IEEE Photovoltaic Specialists Conference (PVSC)*. IEEE. 2020, pp. 857–862. DOI: 10.1109/PVSC45281.2020.9300626.

[4] M. C. Alvarez-Herault et al. "A survey based on the state of the art and perspectives in the monitoring and the control of LV networks". In: *CIRED 2021 - The 26th International Conference and Exhibition on Electricity Distribution*. 2021, pp. 1495–1499. DOI: 10.1049/icp.2021.1629.

[5] Haoran Li et al. "Distribution grid impedance & topology estimation with limited or no micro-PMUs". In: *International Journal of Electrical Power & Energy Systems* 129 (2021). DOI: 10.1016/j.ijepes.2021.106794.

[6] Xinlei Chen and Omid Ardakanian. "Data efficient energy disaggregation with behind-the-meter energy resources". In: *Sustainable Energy, Grids and Networks* 32 (2022). DOI: 10.1016/j.segan.2022.100813.

[7] Benedetto-Giuseppe Risi, Francesco Riganti-Fulginei, and Antonino Laudani. "Modern Techniques for the Optimal Power Flow Problem: State of the Art". In: *Energies* 15.17 (2022). DOI: 10.3390/en15176387.

[8] Miguel Picallo, Saverio Bolognani, and Florian Dörfler. "Closing the loop: Dynamic state estimation and feedback optimization of power grids". In: *Electric Power Systems Research* 189 (2020). DOI: 10.1016/j.epsr.2020.106753.

[9] Miguel Picallo et al. "Cross-layer design for real-time grid operation: Estimation, optimization and power flow". In: *Electric Power Systems Research* 212 (2022). DOI: 10.1016/j.epsr.2022.108378.

[10] Xin Chen et al. "Reinforcement Learning for Selective Key Applications in Power Systems: Recent Advances and Future Challenges". In: *IEEE Transactions on Smart Grid* 13.4 (2022), pp. 2935–2958. DOI: 10.1109/TSG.2022.3154718.

[11] Yuanqi Gao, Wei Wang, and Nanpeng Yu. "Consensus Multi-Agent Reinforcement Learning for Volt-VAR Control in Power Distribution Networks". In: *IEEE Transactions on Smart Grid* 12.4 (2021), pp. 3594–3604. DOI: 10.1109/TSG.2021.3058996.

[12] Xianzhuo Sun and Jing Qiu. "Two-Stage Volt/Var Control in Active Distribution Networks With Multi-Agent Deep Reinforcement Learning Method". In: *IEEE Transactions on Smart Grid* 12.4 (2021), pp. 2903–2912. DOI: 10.1109/TSG.2021.3052998.

[13] Ying Zhang et al. "Deep Reinforcement Learning Based Volt-VAR Optimization in Smart Distribution Systems". In: *IEEE Transactions on Smart Grid* 12.1 (2021), pp. 361–371. DOI: 10.1109/TSG.2020.3010130.

[14] Eleni Stai et al. "Reinforcement Learning Models for Adaptive Low Voltage Power System Operation". In: *2023 IEEE Belgrade PowerTech*. 2023. DOI: 10.1109/PowerTech55446.2023.10202750.

[15] Liang Yu et al. "Multi-Agent Deep Reinforcement Learning for HVAC Control in Commercial Buildings". In: *IEEE Transactions on Smart Grid* 12.1 (2020), pp. 407–419. DOI: 10.1109/TSG.2020.3011739.

[16] Jun Cao et al. "Deep Reinforcement Learning-Based Energy Storage Arbitrage With Accurate Lithium-Ion Battery Degradation Model". In: *IEEE Transactions on Smart Grid* 11.5 (2020), pp. 4513–4521. DOI: 10.1109/TSG.2020.2986333.

[17] Kaveh Dehghanpour et al. "A Survey on State Estimation Techniques and Challenges in Smart Distribution Systems". In: *IEEE Transactions on Smart Grid* 10.2 (2019), pp. 2312–2322. DOI: 10.1109/TSG.2018.2870600.

[18] Thomas Werner et al. "Distributed State Estimation in Digitized Low-Voltage Networks". In: *NEIS 2018; Conference on Sustainable Energy Supply and Energy Storage Systems*. 2018. ISBN: 978-3-8007-4821-1.

[19] Shengye Lu et al. "Real-Time Low Voltage Network Monitoring—ICT Architecture and Field Test Experience". In: *IEEE Transactions on Smart Grid* 6.4 (2014), pp. 2002–2012. DOI: 10.1109/TSG.2014.2371853.

[20] Miguel Picallo et al. "Adaptive real-time grid operation via Online Feedback Optimization with sensitivity estimation". In: *Electric Power Systems Research* 212 (2022). DOI: 10.1016/j.epsr.2022.108405.

[21] Daniel K. Molzahn and Ian A. Hiskens. *A Survey of Relaxations and Approximations of the Power Flow Equations*. Vol. 4. 1-2. Foundations and Trends® in Electric Energy Systems, 2019, pp. 1–221. DOI: 10.1561/3100000012.

[22] Baljinnyam Sereeter, Cornelis Vuik, and Cees Witteveen. "On a comparison of Newton–Raphson solvers for power flow problems". In: *Journal of Computational and Applied Mathematics* 360 (2019), pp. 157–169. DOI: 10.1016/j.cam.2019.04.007.

[23] Haitham Safar. "Power Transmission Line Analysis Using Exact, Nominal π, and Modified π Models". In: *2010 The 2nd International Conference on Computer and Automation Engineering (ICCAE)*. Vol. 5. 2010, pp. 128–134. DOI: 10.1109/ICCAE.2010.5451500.

[24] Georg Kerber. "Aufnahmefähigkeit von Niederspannungsverteilnetzen für die Einspeisung aus Photovoltaikkleinanlagen". PhD thesis. Technische Universität München, 2011.

[25] Johanna Geis-Schroer et al. "Modeling of German Low Voltage Cables with Ground Return Path". In: *Energies* 14.5 (2021). DOI: 10.3390/en14051265.

[26] *Starkstromkabel NAYYJ/O nach VDE 0276-603*. URL: https://www.hesselmann.de/data-download/produktinfo/he-nayy.pdf (visited on 10/15/2023).

[27] Thomas Aundrup et al. "VDE-Studie: Batteriespeicher in der Nieder- und Mittelspannungsebene, Anwendungen und Wirtchaftlichkeit sowie Auswirkungen auf die elektrischen Netze". In: (2015).

[28] Mesut E Baran and Arthur W Kelley. "State estimation for real-time monitoring of distribution systems". In: *IEEE Transactions on Power systems* 9.3 (1994), pp. 1601–1609. DOI: 10.1109/59.336098.

[29] Lei Wang, Qun Zhou, and Shuangshuang Jin. "Physics-guided Deep Learning for Power System State Estimation". In: *Journal of Modern Power Systems and Clean Energy* 8.4 (2020), pp. 607–615. DOI: 10.35833/MPCE.2019.000565.

[30] Martin L Puterman. "Markov decision processes". In: *Handbooks in Operations Research and Management Science* 2 (1990), pp. 331–434. DOI: 10.1016/S0927-0507(05)80172-0.

[31] Richard S Sutton and Andrew G Barto. *Reinforcement Learning: An Introduction*. Second Edition. MIT Press, 2018. ISBN: 978-0-262-03924-6.

[32] Csaba Szepesvári. *Algorithms for Reinforcement Learning*. Springer, 2009. DOI: 10.1007/978-3-031-01551-9.

[33] *Part 1: Key Concepts in RL*. URL: https://spinningup.openai.com/en/latest/spinningup/rl_intro.html (visited on 10/15/2023).

[34] Hao Dong et al. *Deep Reinforcement Learning*. Springer, 2020. DOI: 10.1007/978-981-15-4095-0.

[35] *Part 2: Kinds of RL Algorithms*. URL: https://spinningup.openai.com/en/latest/spinningup/rl_intro2.html (visited on 10/15/2023).

[36] Tuomas Haarnoja et al. "Soft Actor-Critic Algorithms and Applications". In: *CoRR* (2018). DOI: 10.48550/arXiv.1812.05905.

[37] Yuting Xu et al. "Deep Reinforcement Learning Based Decision Making for Complex Jamming Waveforms". In: *Entropy* 24.10 (2022). DOI: 10.3390/e24101441.

[38] Ning Rao et al. "Soft Actor-Critic Deep Reinforcement Learning Based Interference Resource Allocation". In: *Proceedings of the 2021 7th International Conference on Computing and Artificial Intelligence*. Association for Computing Machinery, 2021, pp. 392–398. DOI: 10.1145/3467707.3467766.

[39] VDE. *VDE-AR-N 4105: Technical requirements for the connection to and parallel operation with low-voltage distribution networks*. 2018.

[40] German Federal Network Agency. *Beschlusskammer 6, BK6-22-300, Festlegung zur Durchführung der netzorientierten Steuerung von steuerbaren Verbrauchseinrichtungen und steuerbaren Netzanschlüssen nach § 14a EnWG*. 2023.

[41] Hassan Ramchoun et al. “Multilayer Perceptron: Architecture Optimization and Training”. In: *International Journal of Interactive Multimedia and Artificial Intelligence* (2016). DOI: 10.9781/ijimai.2016.415.

[42] David Garlan. “Software architecture: a roadmap”. In: *Proceedings of the Conference on the Future of Software Engineering*. 2000, pp. 91–101. DOI: 10.1145/336512.336537.

[43] Grady Booch, James Rumbaugh, and Ivar Jacobson. *Unified Modeling Language User Guide, The, Second Edition*. Addison-Wesley Professional, 2005. ISBN: 0321267974.

[44] Ray Daniel Zimmerman, Carlos Edmundo Murillo-Sánchez, and Robert John Thomas. “MATPOWER: Steady-State Operations, Planning and Analysis Tools for Power Systems Research and Education”. In: *IEEE Transactions on Power Systems* 26.1 (2011), pp. 12–19. DOI: 10.1109/TPWRS.2010.2051168.

[45] Roger C. Dugan. *Reference Guide, The Open Distribution System Simulator (OpenDSS)*. 2016. URL: https://spinengenharia.com.br/wp-content/uploads/2019/01/OpenDSSManual.pdf (visited on 10/15/2023).

[46] David Janzen and Hossein Saiedian. “Test-driven development concepts, taxonomy, and future direction”. In: *Computer* 38.9 (2005), pp. 43–50. DOI: 10.1109/MC.2005.314.

[47] Atlassian Corporation. *What is code coverage?* URL: https://www.atlassian.com/continuous-delivery/software-testing/code-coverage (visited on 10/15/2023).

[48] Leon Thurner et al. “Pandapower—An Open-Source Python Tool for Convenient Modeling, Analysis, and Optimization of Electric Power Systems”. In: *IEEE Transactions on Power Systems* 33.6 (2018), pp. 6510–6521. DOI: 10.1109/TPWRS.2018.2829021.

[49] Greg Brockman et al. “OpenAI Gym”. In: (2016). DOI: 10.48550/arXiv.1606.01540.

[50] Antonin Raffin et al. “Stable-Baselines3: Reliable Reinforcement Learning Implementations”. In: *Journal of Machine Learning Research* 22.1 (2021), pp. 12348–12355.

[51] Martín Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems*. 2015. DOI: 10.48550/arXiv.1603.04467.

[52] Steffen Meinecke et al. “SimBench—A Benchmark Dataset of Electric Power Systems to Compare Innovative Solutions Based on Power Flow Analysis”. In: *Energies* 13.12 (2020). DOI: 10.3390/en13123290.

[53] Schneider Electric. *Electrical installation guide, According to IEC International Standard*. 2016. ISBN: 978.2.9531643.3.6.

[54] Simon Becker et al. “dena-Gebäudereport 2022. Zahlen, Daten, Fakten”. In: (2021). URL: https://www.dena.de/fileadmin/dena/Publikationen/PDFs/2021/dena-Gebaeudereport_2022.pdf (visited on 10/15/2023).

[55] EUPD Research. *89 Prozent des Solarpotenzials noch ungenutzt*. URL: https://www.eupd-research.com/89-prozent-des-solarpotenzials-noch-ungenutzt/ (visited on 10/15/2023).

[56] Martin Lödl et al. "Abschätzung des Photovoltaik-Potentials auf Dachflächen in Deutschland". In: *11. Symposium Energieinnovation „Alte Ziele–Neue Wege "*. 2010.

[57] Gregory Kiss et al. "The 2050 City". In: *Procedia Engineering* 118 (2015). Defining the future of sustainability and resilience in design, engineering and construction, pp. 326–355. DOI: 10.1016/j.proeng.2015.08.434.

[58] German Federal Statistical Office. *Pkw-Dichte im Jahr 2022 erneut auf Rekordhoch*. URL: https://www.destatis.de/DE/Presse/Pressemitteilungen/2023/09/PD23_N048_46.html (visited on 10/15/2023).

[59] German Federal Agency for Civic Education (bpb). *Bevölkerung und Haushalte*. URL: https://www.bpb.de/kurz-knapp/zahlen-und-fakten/soziale-situation-in-deutschland/61584/bevoelkerung-und-haushalte/ (visited on 10/15/2023).

[60] Julian Reul et al. "Electric Vehicles in Germany: A Consumer Perspective". In: *31st Aachen Colloquium*. 2022.

[61] *Klimaneutrales Deutschland. Zusammenfassung im Auftrag von Agora Energiewende, Agora Verkehrswende und Stiftung Klimaneutralität*. Tech. rep. 2020.

[62] H Hecking et al. *Szenarien für eine marktwirtschaftliche Klima- und Ressourcenschutzpolitik 2050 im Gebäudesektor*. Tech. rep. 2017.

[63] Marlon Schlemminger et al. "Dataset on electrical single-family house and heat pump load profiles in Germany". In: *Scientific data* 9.1 (2022). DOI: 10.1038/s41597-022-01156-1.

[64] Regina Lamedica et al. "Integrating electric vehicles in microgrids: Overview on hosting capacity and new controls". In: *IEEE Transactions on Industry Applications* 55.6 (2019), pp. 7338–7346. DOI: 10.1109/TIA.2019.2933800.

[65] Shubhankar Kapoor, Björn Sturmberg, and Marnie Shaw. "A review of publicly available energy data sets". In: *Wattwatchers' My Energy Marketplace* (2020).

[66] Elisavet Proedrou. "A comprehensive review of residential electricity load profile models". In: *IEEE Access* 9 (2021), pp. 12114–12133. DOI: 10.1109/ACCESS.2021.3050074.

[67] Hussain Kazmi et al. "Towards data-driven energy communities: A review of open-source datasets, models and tools". In: *Renewable and Sustainable Energy Reviews* 148 (2021). DOI: 10.1016/j.rser.2021.111290.

[68] William F Holmgren, Clifford W Hansen, and Mark A Mikofski. "pvlib python: A python package for modeling solar energy systems". In: *Journal of Open Source Software* 3.29 (2018). DOI: 10.21105/joss.00884.

[69] Marion Schroedter-Homscheidt et al. "The Copernicus Atmosphere Monitoring Service (CAMS) Radiation Service in a nutshell". In: *22nd SolarPACES Conference 2016*. 2016.

[70] Yves-Marie Saint-Drenan et al. "An approach for the estimation of the aggregated photovoltaic power generated in several European countries from meteorological data". In: *Advances in Science and Research* 15 (2018), pp. 51–62. DOI: 10.5194/asr-15-51-2018.

[71] Christian Lamprecht. *Meteostat*. URL: https://meteostat.net/de/ (visited on 10/15/2023).

[72] Ignacio Losada Carreno et al. "SoDa: An Irradiance-Based Synthetic Solar DataGeneration Tool". In: *2020 IEEE International Conference on Smart Grid Communications (SmartGridComm)*. 2020. DOI: 10.1109/SmartGridComm47815.2020.9302941.

[73] Dominik Husarek. "Analysis of sector-coupling effects between the mobility sector and the energy system under consideration of energy transport and charging infrastructure". PhD dissertation. Technische Universität Darmstadt, 2023. DOI: 10.26083/tuprints-00023140.

[74] *Grid Support Utility Interactive Inverters for SMA Central Inverters*. SMA Solar Technology America LLC. URL: https://files.sma.de/downloads/SC_SCS-US-GridServices-TI-en-11.pdf (visited on 10/15/2023).

[75] James Bergstra and Yoshua Bengio. "Random Search for Hyper-Parameter Optimization". In: *Journal of Machine Learning Research* 13 (2012). DOI: 10.5555/2188385.2188395.

[76] Anh Nguyen et al. "An Analysis of State-of-the-art Activation Functions For Supervised Deep Neural Network". In: *2021 International Conference on System Science and Engineering (ICSSE)*. IEEE. 2021, pp. 215–220. DOI: 10.1109/ICSSE52999.2021.9538437.

[77] Sanford Weisberg. *Applied Linear Regression*. Vol. 528. John Wiley & Sons, 2005. DOI: 10.1002/0471704091.

[78] Mesut E. Baran and Felix F. Wu. "Network reconfiguration in distribution systems for loss reduction and load balancing". In: *IEEE Transactions on Power Delivery* 4.2 (1989), pp. 1401–1407. DOI: 10.1109/61.25627.

[79] Jens Buechner et al. *Moderne Verteilernetze für Deutschland (Verteilernetzstudie)*. Tech. rep. 2014.

[80] *Training with Soft-Actor Critic*. URL: https://github.com/yosider/ml-agents-1/blob/master/docs/Training-SAC.md (visited on 10/15/2023).

[81] Serkan Kiranyaz et al. "1D convolutional neural networks and applications: A survey". In: *Mechanical Systems and Signal Processing* 151 (2021). DOI: 10.1016/j.ymssp.2020.107398.

[82] Michel Verleysen and Damien François. "The Curse of Dimensionality in Data Mining and Time Series Prediction". In: *International Work-Conference on Artificial Neural Networks*. Springer. 2005, pp. 758–770. DOI: 10.1007/11494669_93.

[83] Erlanger Stadtwerke. *Elektromobilität, FAQs - Fragen und Antworten, "Benötige ich einen zusätzlichen Zähler?"* URL: https://netze.estw.de/de/Installateure/Elektromobilitaet/ (visited on 10/15/2023).