Reinforcement Learning Portfolio Optimization of Electric Vehicle Virtual Power Plants

Master Thesis

Author: Tobias Richter (Student ID: 558305)
Supervisor: Univ.-Prof. Dr. Wolfgang Ketter
Co-Supervisor: Karsten Schroer

Department of Information Systems for Sustainable Society
Faculty of Management, Economics and Social Sciences
University of Cologne

April 2, 2019

Eidesstattliche Versicherung

Hiermit versichere ich an Eides statt, dass ich die vorliegende Arbeit selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten Schriften entnommen wurden, sind als solche kenntlich gemacht. Die Arbeit ist in gleicher oder ähnlicher Form oder auszugsweise im Rahmen einer anderen Prüfung noch nicht vorgelegt worden. Ich versichere, dass die eingereichte elektronische Fassung der eingereichten Druckfassung vollständig entspricht.

Die Strafbarkeit einer falschen eidesstattlichen Versicherung ist mir bekannt, namentlich die Strafandrohung gemäß § 156 StGB bis zu drei Jahren Freiheitsstrafe oder Geldstrafe bei vorsätzlicher Begehung der Tat bzw. gemäß § 161 Abs. 1 StGB bis zu einem Jahr Freiheitsstrafe oder Geldstrafe bei fahrlässiger Begehung.

Tobias Richter

Köln, den 01.05.2019

Contents

List of Figures

List of Tables

List of Abbreviations

ANN Artificial Neural Network

DP Dynamic Programming

DSO Distribution System Operator

DDQN Double Deep Q-Networks

EPEX European Power Exchange

EV Electric Vehicle

GCRM German Control Reserve Market

GP Genetic Programming

MAW Mean Asymmetric Weighted Objective Function

MC Monte Carlo

ML Machine Learning

MDP Markov Decision Process

PDF Probability Density Function

RES Renewable Energy Sources

RL Reinforcement Learning

TD Temporal-Difference

TSO Transmission System Operator

V2G Vehicle-to-Grid

VPP Virtual Power Plant

Summary of Notation

Capital letters are used for random variables, whereas lower case letters are used for the values of random variables and for scalar functions. Quantities that are required to be real-valued vectors are written in bold and in lower case (even if random variables).

$\doteq$	equality relationship that is true by definition
$\approx$	approximately equal
$\mathbb{E}[X]$	expectation of a random variable X, i.e., $\mathbb{E}[X] \doteq \sum_x p(x)x$
$\mathbb{R}$	set of real numbers
$\leftarrow$	assignment
ε	probability of taking a random action in an ε-greedy policy
α	step-size parameter
γ	discount-rate parameter
λ	decay-rate parameter for eligibility traces
s, s'	states
a	an action
r	a reward
$\mathcal{S}$	set of all nonterminal states
$\mathcal{A}$	set of all available actions
$\mathcal{R}$	set of all possible rewards, a finite subset of $\mathbb{R}$
$\subset$	subset of; e.g., $\mathcal{R} \subset \mathbb{R}$
$\in$	is an element of; e.g., $s \in \mathcal{S}$, $r \in \mathcal{R}$
t	discrete time step
$T, T(t)$	final time step of an episode, or of the episode including time step t
A_t	action at time t
S_t	state at time t, typically due, stochastically, to S_{t-1} and A_{t-1}
R_t	reward at time t, typically due, stochastically, to S_{t-1} and A_{t-1}
π	policy (decision-making rule)
$\pi(s)$	action taken in state s under *deterministic* policy π
$\pi(a\|s)$	probability of taking action a in state s under *stochastic* policy π
G_t	return following time t
$p(s', r \mid s, a)$	probability of transition to state s' with reward r, from state s and action a
$p(s' \mid s, a)$	probability of transition to state s', from state s taking action a
$v_\pi(s)$	value of state s under policy π (expected return)

$v_*(s)$	value of state s under the optimal policy
$q_\pi(s,a)$	value of taking action a in state s under policy π
$q_*(s,a)$	value of taking action a in state s under the optimal policy
V, V_t	array estimates of state-value function v_π or v_*
Q, Q_t	array estimates of action-value function q_π or q_*
d	dimensionality—the number of components of $\mathbf{w}$
$\mathbf{w}$	d-vector of weights underlying an approximate value function
$\hat{v}(s,\mathbf{w})$	approximate value of state s given weight vector $\mathbf{w}$
$\mu(s)$	on-policy distribution over states
$\overline{\text{VE}}$	mean square value error

1 Empirical Setting

This research is embedded in the German carsharing and electricity markets. Germany is a suitable testbed, since it has a comparably high share of renewables in its energy mix and is pushing for an energy turnaround (German: *Energiewende*) since 2010 (BMU, 2010) The high renewable energy content in the energy mix causes electricity prices to be volatile, which makes Germany an attractive location for the use of VPPs.

Germany is home to the carsharing providers Car2Go[1] and DriveNow[2], which operate large EV fleets across the globe. It has been argued that electric carsharing can simultaneously solve several traditional mobility and environmental problems and are an important element of future smart cities (Firnkorn & Müller, 2015). Further, it is widely regarded that the future of mobility will be electric, shared, smart and eventually autonomous (Burns, 2013; Sterling, 2018). Carsharing providers are already contributing to the first two points by operating large fleets of electric vehicles. This research addresses the third point: Using electric carsharing fleets to smartly participate in electricity markets. Carsharing providers, like Car2Go and DriveNow, operate their carsharing fleets in a free-float model, which allows customers to pick up and drop vehicles at any place within the operating zone of the provider. Customers pay by the minute and are offered incentives to park the EVs at charging stations at the end of their trip.

We obtained real-world trip data from Daimler's carsharing service Car2Go. Additionally, we collected freely available balancing market data from the GCRM platform website `https://regelleistung.net`. The data of the EPEX Spot market have kindly been provided by ProCom GmbH[3] for research purposes. In the next chapters the different datasets are described, as well the most important processing steps outlined.

1.1 Electronic Vehicle Fleet Data

The Car2Go dataset consists of GPS data of around 500 Smart ED3 Fortwo vehicles in Stuttgart. These subcompact cars are equipped with a 17.6kWh battery and a standard 3.3kW on-board charger. They fully charge in about six to seven hours and can reach a maximum driving distance of 145km according to the manufacturer. When equipped with an additional 22kW fast charger the charging time reduces to about an hour.

In Table 1 the raw data is displayed, as we have obtained it by Car2Go. The

[1] `https://www.car2go.com`
[2] `https://www.drive-now.com`
[3] `https://procom-energy.de`

dataset contains spatio-temporal attributes, such as timestamp, coordinates, and the address of the EVs in 5 minute intervals. Additionally, status attributes of the interior and exterior are given (not displayed). Especially relevant for our research is the state of charge (SoC, in %) and information whether the EV is plugged into one of the 380 charging stations in Stuttgart. Note that the data only contain EVs that are *available for rent*, i.e., they are not currently rented out by a customer. EVs which are parked at a charging station are also not available until they have charged up approximately 70 SoC. Individual trips have to be reconstructed using the GPS data of the cars. The following preprocessing steps have been taken to prepare the data for further analysis. Table 2 depicts the dataset after all processing steps.

1. Drop unused data columns
   - *ID*: Number plate is already a unique identifier for every EV.
   - *Address*: Different addresses were given from same coordinates. *Latitude, Longitude* was used for locational data instead.
   - *Interior, Exterior*: Status attributes were not used in the analysis of this research. Although they could form interesting features for rental predictions.
   - *Engine Type*: All EVs in Stuttgart are electric vehicles.

2. Decrease GPS resolution to 10 meters

   The GPS accuracy of private industry sensors is approximately 5 meters under open sky, and worse near buildings, bridges and trees[4]. Rental trips are identified by changing GPS locations of the EV (See next point). To reduce the number of false identified trips, due to GPS measurement errors, the resolution is decreased.

3. Determine rental trips

   We infer that a customer rented an EV, if the position coordinates change between two data points of the same EV (see Table 1, 4^{th} to 5^{th} row). Note that we assume that customer do not undertake trips, which begin and end at the exact same location.

4. Infer charging stations

   The GPS location of the EVs is matched with the GPS locations, where an EV has been charged at least once in the dataset. We observed that the raw data do not show EVs that are parked at charging stations, but

[4]See `https://www.gps.gov/systems/gps/performance/accuracy`, accessed 23^{th} February 2019.

are not plugged in. This research assumes that all EVs, which are parked at charging stations are also plugged in. That is a valid assumption, since in Germany cars are only allowed to park at charging station if they are connected to it.

5. Clean data
   - *Service trips*: 999 rental trips were removed that had a trip duration longer than the maximum allowed rental time of two days. We assume that these trips were *service trips* undertaken by Car2Go. When the EVs returned with a higher SoC (e.g., they have been charged at the car repair shop), the previous trip had to be altered to end at a charging station to ensure charging consistency.
   - *Incorrectly charged EVs*: 999 EVs were removed that show incorrect charging behavior. The data of these EVs showed an increase of more than 20% SoC between trips or on trips, while not being located at a charging station.

1.2 Balancing Market Data

In this research, we use market balancing data from the German secondary reserve market. The following chapter will give an overview of the dataset and preprocessing steps that were taken. The data encompasses weekly lists of anonymized bids between 01.06.2016 and 01.01.2018 and a dataset of activated control reserve in Germany during the same period. For a detailed description about the market design of balancing markets refer to Chapter **??**.

The bidding data consists of the traded electricity product, the offered capacity P^{bal} (MW), the capacity price p^c ($\frac{€}{\text{MW}}$), and the energy price p^e ($\frac{€}{\text{MWh}}$) of each bid. Four different products are traded, which are a combination of positive control reserve (feed electricity into the grid) or negative control reserve (take electricity from the grid) and the provided time segment (peak or non-peak hours). Since negative prices are allowed on the secondary operating reserve market, the payment direction is included as well. Moreover, information about the amount of capacity that was accepted, i.e., either partially or fully, is listed. Bids, which were not accepted by the TSOs are not listed. An exemplary excerpt of the dataset is displayed in Table 3.

Table 1: Sample Raw Car2Go Data in Stuttgart

Number Plate	Timestamp	Latitude	Longitude	Street	Zip Code	Charging	SoC (%)
S-GO2471	24.12.2017 20:00	9.19121	48.68895	Parkplatz Flughafen	70692	no	94
S-GO2471	...	...	...	...	...	....	...
S-GO2471	24.12.2017 20:05	9.19121	48.68895	Parkplatz Flughafen	70692	no	94
S-GO2471	24.12.2017 20:10	9.19121	48.68895	Parkplatz Flughafen	70692	no	94
S-GO2471	24.12.2017 23:05	9.15922	48.78848	Salzmannweg 3	70192	no	71
S-GO2471	24.12.2017 23:10	9.15922	48.78848	Salzmannweg 3	70192	no	71
S-GO2471	25.12.2017 00:40	9.17496	48.74928	Felix-Dahn-Str. 45	70597	yes	62
S-GO2471	25.12.2017 00:45	9.17496	48.74928	Felix-Dahn-Str. 45	70597	yes	64
S-GO2471	...	...	...	...	...	....	...
S-GO2471	25.12.2017 06:50	9.17496	48.74928	Felix-Dahn-Str. 45	70597	no	100
S-GO2471	25.12.2017 08:25	9.2167	48.78742	Friedenaustraße 25	70188	no	42

Table 2: Sample Processed Car2Go Trip Data in Stuttgart

Number Plate	Trip	Start Time	Start Latitude	Start Longitude	Start SoC (%)
S-GO2471	1	24.12.2017 20:10	9.19121	48.6890	94
S-GO2471	2	24.12.2017 23:10	9.15922	48.7885	71
S-GO2471	3	25.12.2017 06:50	9.17496	48.7493	66
Number Plate	Trip	End Time	End Latitude	End Longitude	End SoC (%)
S-GO2471	1	24.12.2017 23:05	9.15922	48.7885	71
S-GO2471	2	25.12.2017 00:40	9.17496	48.7493	62
S-GO2471	3	25.12.2017 08:25	9.2167	48.7875	42
Number Plate	Trip	Trip Duration (min)	Trip Distance (km)	Trip Charge (%)	End Charging
S-GO2471	1	175	33.35	23	no
S-GO2471	2	90	13.05	9	yes
S-GO2471	3	155	29	20	no

Table 3: List of Bids of the German Secondary Reserve Market for the tender period 04.12.2017 - 11.12.2017.

Product	Capacity Price	Energy Price	Payment	Offered	Accepted
NEG-HT	0	1.1	TSO to bidder	5	5
NEG-HT	10.73	251	TSO to bidder	15	15
NEG-HT	200.3	564	TSO to bidder	22	22
...	...	...	...	...	...
NEG-NT	0	21.9	Bidder to TSO	5	5
NEG-NT	0	22.4	Bidder to TSO	5	5
...	...	...	...	...	...
POS-NT	696.6	1200	TSO to bidder	5	5
POS-NT	717.12	1210	TSO to bidder	10	7

In this study, we assume that bidding on 15-minute intervals in secondary operating reserve auctions will be possible in future energy markets. As mentioned in Chapter **??**, the market design of the GCRM secondary operating reserve tender was adjusted in 2017. Daily tenders with 4-hour bidding intervals were introduced in favor of weekly tenders with only two time segments. This change represents the trend by the TSOs to change the market design in order to better include RES into the operating reserve markets (Agricola et al., 2014). Due to the volatility of renewable electricity generation, providers are naturally dependent on accurate short-term forecasts, which are only possible with short tender periods and fine-grained bidding intervals.

In order to estimate the upper bound of profits that the EV fleet can earn by participating in the secondary operating reserve market, the *critical prices* $\overline{p}^c$ and $\overline{p}^e$ were determined for each auctioned interval. Following Brandt et al. (2017), we define $\overline{p}^c$ ($\frac{€}{\text{MW}}$) as the capacity price of the bid that was just barely accepted, whereas $\overline{p}^e$ ($\frac{€}{\text{MWh}}$) is the highest energy price that was payed for activated control reserve during that interval. For every 15-minute interval within the given tender period of one week, the activated control reserve in that interval was matched with the accepted bids in that tender period. At the point where supply, i.e., offered capacity of bids, met demand, i.e., activated control reserve, the critical price $\overline{p}^e$ was determined.

Example: The assumed critical prices for the secondary operating reserve tender interval of the 6$^{\text{th}}$ December 2017 between 08:00 and 08:15 are obtained as follows: Three suppliers submitted a reserve capacity of 5MW, 15MW and 22MW respectively (see Table 3). The critical capacity price $\overline{p}^c = 200.3\frac{€}{\text{MW}}$ is determined by the capacity price of the last (third) accepted bid in that time

segment. The TSO reported that 18MW of control reserve were activated between 08:00 and 08:15. Hence, the second bid determines the critical energy price $\overline{p}^e = 251\frac{€}{\text{MWh}}$, as control reserve capacity gets activated according to ascending order of the submitted energy prices. In this example the second bidder would get compensated with: $R = R^c + R^e = (10.73\frac{€}{\text{MW}} \times 15\text{MWh}) + (251\frac{€}{\text{MWh}} \times 13\text{MWh} \times 0.25\text{h}) = 976.7€$. Note that the second bidder get compensated for providing 13MW for 15 minutes (0.25h), instead of the submitted 15MW, since in total only 18MW of control capacity was activated, which was partly fulfilled by the 5MW of the first bidder.

1.3 Spot Market Data

The data from the EPEX Spot Intraday Continuous encompass order books and executed trades from 01.06.2016 until 01.01.2018. The list of trades contain information on the unit price p^u ($\frac{€}{\text{MWh}}$), the quantity (kW) and the traded product (hourly, quarterly or block). In this research, we focus on quarterly product times (15-minute intervals), as they provide the highest flexibility. Fleet controllers can promptly react to fluctuant electricity demand of the EV fleet by accurately adjusting the bid quantities. Future research could also consider other electricity products if lower prices justify decreased flexibility at that point in time. Additionally, the TSOs of the buyer and seller are listed in the dataset. They are only relevant if special conditions between TSO apply, e.g., when delivering electricity to other countries.

On the spot market, electricity trades can have a very short lead time of up to 5 minutes before delivery (see Table 4). This market characteristic is beneficial for our proposed trading strategy, since it allows the EV to procure electricity in almost real time. The controller can submit bids to the market, with accurate estimations of available charging capacity up to five minutes ahead. Similarly to the balancing market, the critical price $\overline{p}^u$ has to be determined for all intraday trading intervals. The critical unit price $\overline{p}^u$ is defined as the lowest price of all executed trades.

$$\overline{p}^u \doteq \min_{t \in \mathcal{T}} p_t^u \ ,$$

where $\mathcal{T}$ is the set of all trades in a bidding interval. *Example:* The critical unit price of the trades in Table 4 is $\overline{p}^u = 51.00\frac{€}{\text{MWh}}$ (trades 8031392, 8031387 and 8031375). All buyers that submitted bids with a price higher than the critical unit price, successfully procured electricity. Hence, accurate forecasts of the critical price allow to optimize the bidding behavior. For a detailed description of the intraday continuous market see Chapter **??**.

Table 4: List of Trades of the EPEX Spot Intraday Continuous Market

Execution time	ID	Unit Price	Quantity	Buyer Area	Seller Area	Product	Product Time	Delivery Date
2017-12-04 06:54:55	8031392	51.00	5500	Amprion	Amprion	Quarter	07:15 - 07:30	2017-12-04
2017-12-04 06:53:26	8031391	59.00	10000	TenneT	TenneT	Quarter	07:15 - 07:30	2017-12-04
2017-12-04 06:53:26	8031390	58.90	10000	TenneT	TenneT	Quarter	07:15 - 07:30	2017-12-04
2017-12-04 06:53:15	8031389	52.30	7000	50Hertz	50Hertz	Quarter	07:15 - 07:30	2017-12-04
2017-12-04 06:53:13	8031386	59.00	500	TenneT	TenneT	Quarter	07:15 - 07:30	2017-12-04
2017-12-04 06:53:13	8031387	51.00	3600	Amprion	Amprion	Quarter	07:15 - 07:30	2017-12-04
2017-12-04 06:53:13	8031388	52.00	1400	Amprion	Amprion	Quarter	07:15 - 07:30	2017-12-04
2017-12-04 06:53:02	8031385	58.90	11000	TenneT	TenneT	Quarter	07:15 - 07:30	2017-12-04
2017-12-04 06:52:38	8031380	60.00	10000	Amprion	Amprion	Quarter	07:15 - 07:30	2017-12-04
2017-12-04 06:52:38	8031381	57.50	8000	Amprion	Amprion	Quarter	07:15 - 07:30	2017-12-04
2017-12-04 06:52:38	8031382	58.00	2000	Amprion	Amprion	Quarter	07:15 - 07:30	2017-12-04
2017-12-04 06:52:38	8031383	58.90	4000	TenneT	TenneT	Quarter	07:15 - 07:30	2017-12-04
2017-12-04 06:52:38	8031384	60.00	4000	Amprion	Amprion	Quarter	07:15 - 07:30	2017-12-04
2017-12-04 06:52:27	8031379	52.30	8000	50Hertz	50Hertz	Quarter	07:15 - 07:30	2017-12-04
2017-12-04 06:51:33	8031378	66.00	5000	TransnetBW	TransnetBW	Quarter	07:15 - 07:30	2017-12-04
2017-12-04 06:51:28	8031377	54.00	8000	Amprion	Amprion	Quarter	07:15 - 07:30	2017-12-04
2017-12-04 06:51:24	8031376	54.00	7000	TenneT	TenneT	Quarter	07:15 - 07:30	2017-12-04
2017-12-04 06:49:34	8031375	51.00	4000	TenneT	TenneT	Quarter	07:15 - 07:30	2017-12-04
2017-12-04 06:49:26	8031374	54.00	5000	50Hertz	50Hertz	Quarter	07:15 - 07:30	2017-12-04
2017-12-04 06:49:23	8031373	55.10	8000	50Hertz	50Hertz	Quarter	07:15 - 07:30	2017-12-04

2 Model

The following chapter will introduce the model of this research. In its essence, we propose a solution for EV fleet providers to utilize a VPP portfolio to profitably provide balancing services to the grid on multiple markets. A control mechanism procures energy from electricity markets, allocates available EVs to VPPs, and intelligently dispatches EVs to charge the acquired amount of energy. The model employs a RL agent that learns an optimal bidding strategy by interacting with the electricity markets and reacts to changing rental demand of the EV fleet. This chapter is structured as follows: The information assumptions are listed first, the control mechanism is explained next, and finally, the RL approach is described in detail. The used notation in this chapter can be found in Table 5.

We formulate the problem as a *controlled EV charging* problem. The EV fleet operator represents the *controller*, which aims to charge the fleet at minimal costs. First, the controller predicts the amount of energy it can charge in a given *market period* h. The length of the market period Δh and the market closing time depend on the considered electricity market. Second, the controller places bids on one or multiple markets to procure the predicted amount of energy. Lastly, at electricity delivery time, the controller communicates with the EV fleet to control the charging in real-time. Online EV *control periods* t are typically shorter than market periods. In the empirical case that we consider, the market periods are 15 minutes long, while the EV control periods last 5 minutes. Nonetheless, the presented approach generalizes to other period lengths. During each control period, the controller has to take decisions which individual EVs it should dispatch to charge the procured amount of electricity. In times of unforeseen rental demand, this decision implies trading off commitments to the markets with compromising customer mobility by refusing customer rentals.

Table 5: Table of Notation

Symbol	Description	Unit
t	Control period.	-
h	Market period.	-
T	Number of control periods in a market period.	-
H	Number of market periods in day.	-
N_h	Total number of market periods.	-
Δt	Length of control period.	hours
Δh	Length of market period.	hours
P_h^{bal}	Amount of balancing power offered on the balancing market.	kW
$\overline{p}_h^c$	Critical capacity price in market period h.	$\frac{€}{\text{MW}}$

Continued on next page

Continued from previous page

Symbol	Description	Unit
$\overline{p}_h^e$	Critical energy price in market period h.	$\frac{€}{\text{MWh}}$
P_h^{intr}	Amount of power offered for the unit on the intraday market.	kW
$\overline{p}_h^u$	Critical unit price in market period h.	$\frac{€}{\text{MWh}}$
E_h^{bal}	Amount of energy charged from balancing market in market period h.	MWh
E_h^{intr}	Amount of energy charged from the intraday market in market period h.	MWh
P_t^{fleet}	Amount of available fleet charging power in control period t.	kW
$\widehat{P}_t^{fleet}$	Predicted amount of available fleet charging power in control period t.	kW
$C_h^{bal}(P)$	Cost function for procuring electricity from the balancing market.	€
$C_h^{intr}(P)$	Cost function for procuring electricity from the intraday market.	€
$\rho_{t,i}$	Opportunity costs of lost rental of EV i in control period t.	€
β_h	Imbalance costs in market period h.	€
λ_h^{bal}	Balancing market risk factor.	$[0,1]$
λ_h^{intr}	Intraday market risk factor.	$[0,1]$
θ_λ	Set of risk factors for all market periods $h \in \{1,...,N_h\}$.	-
$C^{fleet}(\theta_\lambda)$	Cost function for the fleets total costs over all market periods h.	€
C_h^{fleet}	Total accumulated fleet costs until market period h.	€
i	Electric Vehicle.	-
$\mathcal{F}$	Set of all EVs in the fleet	-
$\mathbf{c}_i$	Dummy variable if EV is connected to a charging station.	0/1
ω_i	Amount of electricity stored in EV.	kWh
Ω	Maximum battery capacity of EV.	kWh
δ	Charging power of EV at the charging station.	kW
p^{ind}	Industry tariff	$\frac{€}{\text{kWh}}$

2.1 Assumptions

In order to evaluate and operationalize our model, the following assumptions about the available information and the electricity market mechanism are taken:

2.1.1 Information Assumptions

1. Mobility demand

The controller is able to forecast the mobility demand of the EV fleet with different time-horizons based on historical data. More specifically, it can predict the amount of plugged-in EVs and consequently the available charging power P_t^{fleet} of the fleet at control period t. The prediction accuracy is increasing with shorter time horizons, from uncertain predictions one week ahead to very accurate predictions 30 minutes ahead. Past research presented successful mobility demand forecast algorithms in the context of free-float carsharing (Kahlen et al., 2018, 2017; Wagner et al., 2016).

2. Critical electricity prices

   The controller is able to forecast electricity prices of spot and balancing markets based on historical data. More specifically, it can estimate the critical prices $\overline{p}_h^c$, $\overline{p}_h^e$, and $\overline{p}_h^u$ for each market period with perfect accuracy (see Chapter 1.2 and Chapter 1.3 for the critical price definitions). Electricity price forecasting is an extensively studied research area with well-advanced prediction algorithms (Weron, 2014; Avci et al., 2018).

We are confident that taking the above assumptions is viable, assuming available forecasting information is common practice in the VPP and EV fleet charging literature, for example Vandael et al. (2015); Mashhour and Moghaddas-Tafreshi (2011); Tomić and Kempton (2007); Pandžić et al. (2013).

2.1.2 Market Assumptions

1. Balancing market

   The controller is able to submit bids of any quantity for single 15-minute market periods 7 days ahead. Since the critical capacity and energy prices are available (previous paragraph), the controller submits bids in the form $bid = (P^{bal}, \overline{p}^c, \overline{p}^e)$. Submitting a bid with the critical capacity price ensures that the bid will always get accepted by the TSOs. Submitting the bid with the critical energy price ensures that the balancing power will always get fully activated, which allows the fleet to charge at the submitted price for the market periods length.

2. Intraday market

   The controller submits bids to the intraday market 30 minutes ahead. The bids are submitted in the form $bid = (P^{intr}, \overline{p}^u)$. We assume that the order to buy will always get matched until the minimal lead time of the trade (e.g., 5 minutes on the EPEX Spot Intraday Continuous). In reality, this is not always the case since trades are executed immediately and it is not

guaranteed that a matching order to sell is submitted between the bidding time and the minimal lead time.

In essence, we are assuming that the controller always submits the optimal bid at the right time. In other words, every bid leads to the successful procurement of the desired amount of electricity. This assumption provides an upper bound for the fleet profits from trading EV battery storage on the electricity markets. However, the upper bound is only influenced by the accuracy of the electricity price forecasting algorithm, a research area that well exceeds the scope of this work. Furthermore, we assume that the controller is a price-taker. Due to the limited size of its bids, it is lacking the market share to influence prices on the markets. Similar assumptions have been made by Brandt et al. (2017) and Vandael et al. (2015).

2.2 Control Mechanism

The control mechanism constitutes the core of this research. It can be seen as a decision support system that can be deployed at an EV fleet operator to centrally control the charging of its fleet. Figure 1 depicts the control mechanism, which is divided into three distinct phases:

The first phase, *Bidding Phase I*, takes place just before the closing time of the balancing market, once every week (e.g., Wednesdays at 3pm at the GCRM). In this phase, the controller can place bids for every market period h of the following week on the balancing market. The second phase, *Bidding Phase II*, takes places in every market period of $\Delta h = 15$ minutes. At this point, the controller has the opportunity to place bids to the intraday market for the market period 30 minutes ahead. The third phase, *Dispatch Phase*, takes places in every control period of $\Delta t = 5$ minutes. In this phase the controller has to dispatch available EVs to charge the procured electricity from the markets. The phase involves allocating individual EVs to the VPP and potentially refusing customer rentals to assure that all market commitments can be fulfilled.

The following chapters will highlight the important parts of the three phases and provide detailed explanation and mathematical formulations.

2.2.1 Fleet Charging Power Prediction

In a first step, the controller has to predict the available fleet charging power for the market period of interest (see (A) in Figure 1). The actual available fleet charging power P_t^{fleet} in a control period t is given by the number of EVs that are connected to a charging station, with enough free battery capacity to charge the next control period $t+1$.

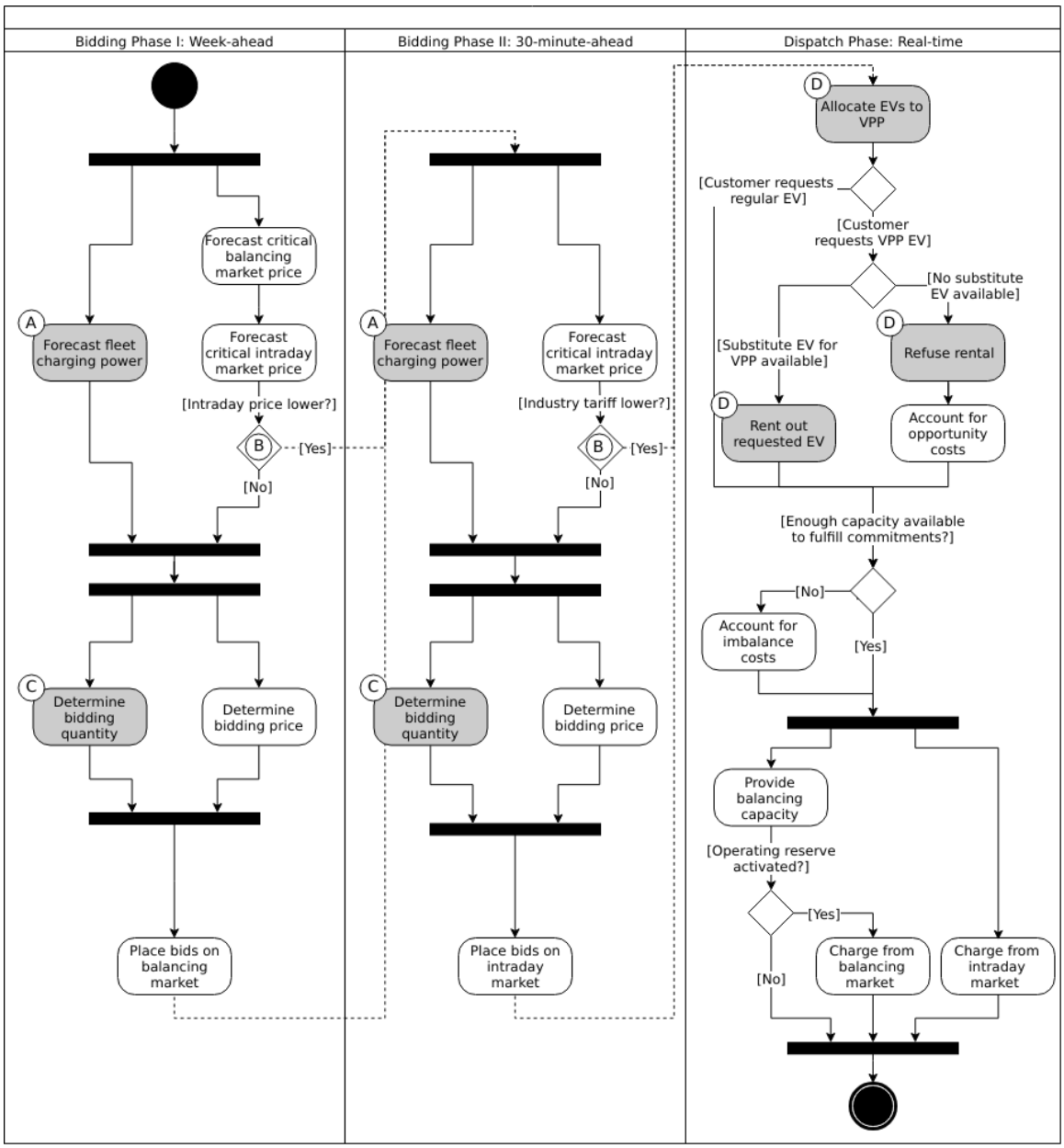

Figure 1: Control Mechanism

When the controller procures electricity from the markets, the fleet has to charge with the committed charging power during all control periods of the market period h, otherwise imbalance costs occur. To minimize the risk of not being able to charge the committed amount of energy during the whole market period, the predicted fleet charging power in a market period is defined as the minimal predicted fleet charging power of all control periods in that market period:

$$\widehat{P}_h^{fleet} \doteq \min_{n\in\{1,..,T\}} \widehat{P}_{t+n}^{fleet} \ , \tag{1}$$

where h is the market period of interest, t its first control period and T the number of control periods in a market period.

2.2.2 Market Decision

In a second step, the controller has to decide from which market it should procure the desired amount of energy (see (B) in Figure 1). Therefore, it compares the costs for charging electricity from the balancing market with the costs for charging from the intraday market. The cost function for procuring electricity from the balancing market is defined as follows:

$$\begin{aligned} C_h^{bal}(P) &\doteq -(P\times 10^{-3} \times \overline{p}_h^c) + (E_h^{bal} \times \overline{p}_h^e) \\ &= -(P\times 10^{-3} \times \overline{p}_h^c) + (P\frac{\Delta h}{10^3} \times \overline{p}_h^e) \ , \end{aligned} \tag{2}$$

where P (kW) is the amount of offered balancing power. The first term of the equation corresponds to the compensation the controller retrieves for keeping the balancing capacity available, while the second term corresponds to the costs for charging the activated balancing energy E_h^{bal} (MWh). Energy is power over time, hence E_h^{bal} can be substituted with P times the market periods length Δh, divided by the unit conversion term from kW to MW. Note that the critical energy price $\overline{p}^e \in \mathbb{R}$, can also take negative values, resulting in profits for the fleet, while the critical capacity price $\overline{p}^c \in \mathbb{R}_0^+$ is never negative and therefore never results in costs for the fleet.

The cost function for charging from the intraday market is defined similarly to (2):

$$\begin{aligned} C_h^{intr}(P) &\doteq E_h^{intr} \times \overline{p}_h^u \\ &= P\frac{\Delta h}{10^3} \times \overline{p}_h^u \end{aligned} \tag{3}$$

Again, depending on the market situation, $\overline{p}^u \in \mathbb{R}$ can either be negative or positive, resulting in costs or profits for the fleet. Contrarily to the balancing

market, on the intraday market the fleet does not get compensated for keeping the charging power available; only the charged energy affects the costs. If the costs for charging from the balancing market 7 days ahead $C^{bal}_{h+(7\times H)}(\widehat{P}^{fleet}_{h+(7\times H)})$ are higher than the costs of charging from the intraday market at the same market period $C^{intr}_{h+(7\times H)}(\widehat{P}^{fleet}_{h+(7\times H)})$, the controller does not procure electricity from the balancing market.

2.2.3 Determining the Bidding Quantity

In a third step, the controller has to take a decision on the amount of energy it should procure from the markets (see (C) in Figure 1). Determining the bidding quantity is the core challenge of the controlled charging problem. The bidding quantity determines the profits that can be made, by charging at a cheaper market price than the flat industry tariff. On one hand, the controller aims to maximize its profits by procuring as much electricity as possible from the markets. On the other hand, it needs to balance the risk of (a) procuring more energy that it can maximally charge and (b) not procuring enough energy from the market to sufficiently charge the fleet.

In case (a), the fleet is facing costs of compromising customer mobility, or worse, high imbalance penalties from the markets. Renting out EVs is considerably more profitable than using their batteries as a VPP. Refusing customer rentals, in order to fulfill market commitments, induces opportunity costs of lost rentals ρ on the fleet. Imbalance costs β occur, when the fleet can not charge the committed amount energy at all, even with refusing rentals. In case (b), the fleet also faces opportunity costs of lost rentals when individual EVs do not have enough SoC for planned trips of arriving customers.

The controller faces additional risks by bidding one week ahead on the balancing market, in contrast to only 30 minutes ahead on the intraday market: Predictions of available charging power are more uncertain with the larger time horizon. To account for all mentioned risks, we introduce a *risk factor* $\lambda \in \mathbb{R}_{0\leq\lambda\leq 1}$, where $\lambda=0$ indicates no risk, and $\lambda=1$ indicates a high risk. The controller determines the bidding quantity P^{bal}_h by discounting the predicted available fleet charging power $\widehat{P}^{fleet}_h$ with the possible risk λ_h of imbalance or opportunity costs:

$$P^{bal}_h \doteq \begin{cases} 0, & \text{if } C^{bal}_h(\widehat{P}^{fleet}_h) \geq E^{bal}_h 10^3 \times p^{ind} \\ 0, & \text{if } C^{bal}_h(\widehat{P}^{fleet}_h) \geq C^{intr}_h(\widehat{P}^{fleet}_h) \\ \widehat{P}^{fleet}_h \times (1-\lambda^{bal}_h), & \text{otherwise} \end{cases} \tag{4}$$

where h is the market period of interest one week ahead. If the controller can buy electricity at the intraday market at a lower price, it does not place a bid at

the balancing market. If the controller can charge cheaper at the regular industry tariff p^{ind}, it does not place a bid either. In all other cases, the controller submits P_h^{bal} to the market.

The bidding quantity for the intraday market P_h^{intr} depends on the previously committed charging power P_h^{bal} and the newly predicted charging power $\widehat{P}_h^{fleet}$:

$$P_h^{intr} \doteq \begin{cases} 0, & \text{if } C_h^{intr}(\widehat{P}_h^{fleet} - P_h^{bal}) \geq E_h^{intr} 10^3 \times p^{ind} \\ (\widehat{P}_h^{fleet} - P_h^{bal}) \times (1 - \lambda_h^{intr}), & \text{otherwise} \end{cases} \tag{5}$$

where h is the market period of interest 30 minutes ahead. Note that any amount of electricity that the controller procured from the balancing market P_h^{bal}, does not need to be bought from intraday market for the same market period. Since the predicted charging power $\widehat{P}_h^{fleet}$ is expected to be more accurate 30 minutes ahead than one week ahead, the controller is able to correct bidding errors it made in the first decision phase, and optimally charge the whole EV fleet.

2.2.4 Dispatching Electronic Vehicle Charging

In the last step, at electricity delivery time, the EVs have to be assigned to the VPP and be *dispatched* to charge (see (D) in Figure 1). Therefore the controller needs to detect how many EVs are eligible to be used as VPP in the control period t. An EV i is eligible if (a) it is connected to a charging station ($\mathbf{c}_i = 1$), and (b) it has enough free battery storage available ($\Omega - \omega_i$) to charge the next control period. Hence, the VPP is defined as:

$$VPP \doteq \{i \in \mathcal{F} \mid \mathbf{c}_i = 1 \vee \Omega - \omega_i \geq \gamma \Delta t\} \ , \tag{6}$$

where $\gamma \Delta t$ (kWh) denotes the amount of energy that can be charged with the charging speed of γ (kW) in control period t. γ is limited by either the EVs build-in charger, or the charging power of the connected charging station. In this model we assume γ is equal for all considered EVs and charging stations. *Example:* Assuming a charging power of $\gamma = 3.3$kW, an EV battery capacity of $\Omega = 17.6$kWh, and control periods of 5 minutes, the amount of energy charged in one control period is $3.3\text{kW} \times \frac{5}{60}\text{h} = 0.275\text{kWh}$. Hence, the maximal battery capacity to be eligible for VPP use is $17.6 - 0.275 = 17.325$kWh.

Remember that the fleet has to provide the total committed charging power $P_h^{bal} + P_h^{intr}$ across all control periods t of the market period h, independent of which individual EVs are actually charging the electricity. This fact allows the controller to dynamically dispatch EVs every control period and react to unforeseen rental demand. If a customer wants to rent out an EV that is assigned to the VPP,

the controller only has to refuse the rental, if no other EV is available to charge instead. When no replacement EV is available, the controller has to account for lost rental profits $\rho_{t,i}$. If the VPPs total amount of available charging power $|VPP|_t \times \gamma$ is not sufficient to provide the total market commitments $P_h^{bal} + P_h^{intr}$, the fleet gets charged imbalance costs β_h. Otherwise all the committed energy can be charged by the VPP.

2.2.5 Evaluating the Bidding Risk

The controllers main goal is to choose the risk factors λ_h^{bal}, λ_h^{intr} for every market period h, that minimize the cost of charging, while avoiding the risks of lost rental profits $\rho_{t,i}$ or imbalance costs β_h. The total fleet costs are defined as follows:

$$C^{fleet}(\theta_\lambda) \doteq \sum_{h}^{N_h} \left[C_h^{bal}(P_h^{bal}) + C_h^{intr}(P_h^{intr}) + \beta_h + \sum_{t}^{T} \sum_{i}^{|\mathcal{F}|} \rho_{t,i} \right], \tag{7}$$

where $\theta_\lambda \in \mathbb{R}^{2 \times N_h}_{0 \leq \lambda \leq 1}$ is the matrix of the risk factors λ_h^{bal}, λ_h^{intr} for all considered market periods N_h. $\mathcal{F}$ denotes the set of all EVs i in the fleet and $|\mathcal{F}|$ the fleet size. The costs for charging $C_h^{bal}(P_h^{bal})$, $C_h^{intr}(P_h^{intr})$ are clearly dependent on the chosen risk factors λ_h^{bal}, λ_h^{intr} (see (4) and (5)). In summary, the problem can be formulated as minimizing the total costs of the fleet, by choosing the optimal set of risk factors θ_λ:

$$\begin{aligned} \underset{\theta_\lambda}{\text{minimize}} \quad & C^{fleet}(\theta_\lambda) \\ \text{subject to} \quad & 0 \leq \lambda_h^{bal} \leq 1, \ \forall \lambda_h^{bal} \in \theta_\lambda \\ & 0 \leq \lambda_h^{intr} \leq 1, \ \forall \lambda_h^{intr} \in \theta_\lambda \end{aligned} \tag{8}$$

Solving this optimization problem with common methods like stochastic programming is a difficult task, assuming that complete information of available charging power and future electricity market prices is not always available. Since one goal of this research is to develop a model that can be applied to previously unknown settings and learn from uncertain environments, as mobility and electricity markets, we chose to solve the problem with a RL approach that is explained in detail in Chapter 2.3.

2.2.6 Example

At 3pm on the 9$^{\text{th}}$ of August 2017, the controller enters the first bidding phase for the market period $h =$ *16.08.2017 15:00-15:15*. It predicts that at that point in time 250 EVs are connected to a charging station, resulting in 900kW available fleet charging power ($\widehat{P}_h^{fleet} = 900\text{kW}$), given the charging power of 3.6kW per

EV. Assuming the available critical prices are $\overline{p}_h^c = 5\frac{€}{\text{MW}}$, $\overline{p}_h^e = -10\frac{€}{\text{MWh}}$, and $\overline{p}_h^u = 10\frac{€}{\text{MWh}}$ in that market period, the controller now evaluates the cheapest charging option. The flat industry electricity tariff is assumed to be $p^{ind} = 0.15\frac{€}{\text{kWh}}$. The costs for charging with the maximal predicted amount of available power $\widehat{P}_h^{fleet}$ from the balancing market ($C_h^{bal}(900\text{kW}) = -6.25€$) are less than charging from the intraday market ($C_h^{intr}(900\text{kW}) = 2.25€$) or charging at the industry tariff ($900\text{kW} \times 0.25\text{h} \times 0.15\frac{€}{\text{kWh}} = 33.75€$). In this example, the fleet operator will even get compensated for charging its fleet, by choosing the balancing market.

In the next step, the controller has to submit bids to the balancing market. The RL agent determined that the risk of bidding on the balancing market is $\lambda_h^{bal} = 0.3$. Consequently, the controller sets the bidding quantity to $P_h^{bal} = \widehat{P}_h^{fleet} \times (1 - \lambda_h^{bal}) = 900\text{kW} \times 0.7 = 630\text{kW}$ and submits a bid to the market and updates its account with $C_h^{bal}(630\text{kW}) = -4.725€$.

One week later, 30 minutes before electricity delivery time, the controller enters the second bidding phase. Due to the short time horizon, it predicts with high accuracy that only $\widehat{P}_h^{fleet} = 810\text{kW}$ is available for the same market period *16.08.2017-15:00*. By trading at the intraday market, the controller can now charge the remaining available EVs with a low risk of procuring more energy than it can maximally charge. At this point in time, the RL agent determines a remaining risk of $\lambda_h^{intr} = 0.05$, and sets the bidding quantity to $P_h^{intr} = (810\text{kW} - 630\text{kW}) \times (1 - 0.05) = 171\text{kW}$. The controller procures 171kW from the intraday market and updates its account with $C_h^{intr}(171\text{kW}) = 0.4275€$.

At electricity delivery time, the 16$^{\text{th}}$ of August 2017 at 3:00pm, the controller detects 255 available EVs; EVs which are connected to a charging station and have enough battery capacity left to be charged in the next control period. It assigns 223 EVs to provide the total committed 801kW charging power for the market period time Δh of 15 minutes. During that time, three customers want to rent out EVs that are allocated to the VPP. The first two rentals are accepted because two other EVs are available to charge instead. The third rental has be to refused, since no EV is remaining as substitution. The controller has to account for the opportunity costs of the lost rental $\rho_{t,i}$.

2.3 Reinforcement Learning Approach

In the following chapter the developed RL approach is outlined. First, we define the charging problem as a MDP, and second, the learning algorithm is explained. Remember that the goal of the controlled charging problem is to choose a set of risk factors θ_λ that minimize the fleets total costs across all market periods. The controller is able to influence the costs, by setting the risk factors λ^{bal}, λ^{intr} each

market period h. The risk factors influence the bidding quantities P_h^{bal}, P_h^{intr} that the controller submits to the balancing and intraday market, which in the end determine the fleet costs. The RL agent decides on the risk factors (i.e., takes an action) based on the observed state $\mathcal{S}$ every time step h (usually denoted as t in the RL literature). The optimal set of risk factors is learned by the RL agent through estimating a policy $\pi(a|s)$ that maps every state $s \in \mathcal{S}$ to an action $a \in \mathcal{A}$.

2.3.1 Markov Decision Process Definition

MDPs are defined by the state space $\mathcal{S}$, the action space $\mathcal{A}$, a set of reward signals $\mathcal{R}$ and the state-transition probabilities $p(s'|a, s)$. When $p(s'|a, s)$ is unknown, as it is in our case, it is possible to use a model-free approach (see Chapter **??**). The state space compromises the observed information the agent uses to decide on the action it is going to take. We observed the following factors that are associated with the bidding risk:

1. The bidding period's time of the day

   In times of volatile customer rental demand (e.g., during rush hour), the uncertainty on the guaranteed amount of available EVs increases. Bidding for these periods involves a higher risk of not being able to fulfill market commitments.

2. The current and estimated future size of the VPP

   Large VPPs benefit from the *risk-pooling* effect (Kahlen et al., 2017). Intuitively that means, larger VPPs are exposed to smaller risks: They have an increased probability that "lost" charging power, due to unforeseen rentals, can be substituted by the EVs of the VPP.

Since forecasts of available charging power are already available, we define the predicted VPP size $|\widehat{VPP}|_h$ as the as the necessary amount of EVs to provide the predicted charging power $\widehat{P}^{fleet}$ in time period h:

$$|\widehat{VPP}|_h \doteq \left\lceil \frac{\widehat{P}_h^{fleet}}{\gamma} \right\rceil , \tag{9}$$

where γ is the charging power per EV. *Example:* When the controller predicted 910kW available charging power, the estimated future size of the VPP to charge with the predicted power is `ceil(`$910\text{kW}/3.6\text{kW}$`)` $= 253$.

The state space is then defined as the set of all valid values of the elements of

the following tuple:

$$\mathcal{S} \doteq \left\langle t(h), |VPP|_h, |\widehat{VPP}|_{h+2}, |\widehat{VPP}|_{h+(7\times H)} \right\rangle \ , \tag{10}$$

where:

- $t(h)$ is the current daytime in hours, with discrete values in the range $\left[0,\ 23\right] \in \mathbb{N}$.
- $|VPP|_t$ is the current VPP size, with discrete values in the range $\left[0,\ |\mathcal{F}|\right] \in \mathbb{N}$.
- $|\widehat{VPP}|_{h+2}$ is the predicted VPP size 30 minutes ahead, with discrete values in the range $\left[0,\ |\mathcal{F}|\right] \in \mathbb{N}$.
- $|\widehat{VPP}|_{h+(7\times H)}$ is the predicted VPP size 7 days ahead, with discrete values in the range $\left[0,\ |\mathcal{F}|\right] \in \mathbb{N}$.

The state space encompasses $|\mathcal{F}|^3 \times 24$ states. Assuming a fleet size $|\mathcal{F}|$ of 500 EVs, the state space consists of 3×10^9 different states.

The agent takes actions by determining the risk that is associated with bidding on the electricity markets at each market period h. Hence, the action space is constituted by all combinations of valid values of the risk factors $\lambda^{bal}, \lambda^{intr}$:

$$\mathcal{A} \doteq \left\{\lambda^{bal}, \lambda^{intr} \in \mathbb{R}_{0\leq\lambda\leq1}\right\} \ , \tag{11}$$

where:

- λ^{bal} is the risk factor for bidding on the balancing market 7 days ahead, with discrete values in the range $\left[0, 1\right]$ in 0.05 increments.
- λ^{intr} is the risk factor for bidding on the intraday market 30 minutes ahead, with discrete values in the range $\left[0, 1\right]$ in 0.05 increments.

The action space encompasses $20^2 = 400$ actions. The state space and action space were consciously discretized to achieve faster learning rates. Convergence in continuous spaces is theoretically achievable, but computationally more complex (Sutton & Barto, 2018). In order to facilitate faster learning in real-world settings, where long training periods are not desirable, we chose to not pursue this direction.

The reward signal is naturally defined as the fleet costs that occurred in the last time step. When accumulating the rewards for all time steps, we arrive at the total fleet costs, which we aim to minimize:

$$R_{h+1} = C_h^{fleet} - C_{h-1}^{fleet} \ , \tag{12}$$

where C_h^{fleet} are the total accumulated fleet costs until the market period h. For a complete formulation the cost function see (7). The agent's actions directly determine the occurred costs or profits, and are presented to the agent in form

of a positive or negative reward signal. The particular challenge in the proposed RL problem is the significantly *delayed reward*. Choosing a risk factor in time step h determines the reward up to 672 time steps later (7 days, with 15-minute time steps), when the electricity from the balancing market has to be charged.

2.3.2 Learning Algorithm

This research proposes to solve the presented RL problem, with the Double Deep Q-Network algorithm (DDQN), developed by van Hasselt et al. (2016). DDQN is a state-of-the-art, model-free RL approach that uses a deep neural network as function approximator to estimate optimal Q-values (see Chapter **??** for a explanation of function approximation methods). It combines the revolutionary Deep Q-Network (DQN), originally proposed by Mnih et al. (2015) with Double Q-Learning (van Hasselt, 2010). In Double Q-Learning, experiences are randomly selected to update two different value functions to select and evaluate actions (in contrast to just one function for both tasks). DDQN has shown to reduce overoptimistic action-value estimates of the DQN algorithm, resulting in more stable and reliable learning results (van Hasselt et al., 2016). Combined with the *dueling network* architecture, proposed by Wang et al. (2015), this approach outperforms existing deep RL methods. Dueling networks lead to faster convergence rates in control problems with large action spaces than traditional single stream approaches. This property is especially beneficial for our proposed RL problem, as the defined action space (400 possible actions) is comparably large in comparison to classical control problems. In Figure 2, the conventional single stream approach (top) versus the dueling architecture (bottom) is depicted. The dueling architecture consists of a neural network of any shape with two streams that separately estimate the state-value and the action advantages. These estimates are later combined into Q-values (see Figure 2, green layer):

$$Q(s,a) = V(s) + \left(A(s,a) - \frac{1}{|\mathcal{A}|} \sum_{a'} A(s,a') \right) , \tag{13}$$

where V and A are estimates of the value function and action advantages respectively, represented by the two different streams in the network. By subtracting the mean action advantages (last term), identifiability (V and A can be recovered, given Q) and stability of the optimization is ensured. The separated streams allow to learn which states are valuable without having to learn each state-action interaction individually. Like this, a general state-value is learned that can be shared across many different actions, leading to faster convergence (Wang et al., 2015).

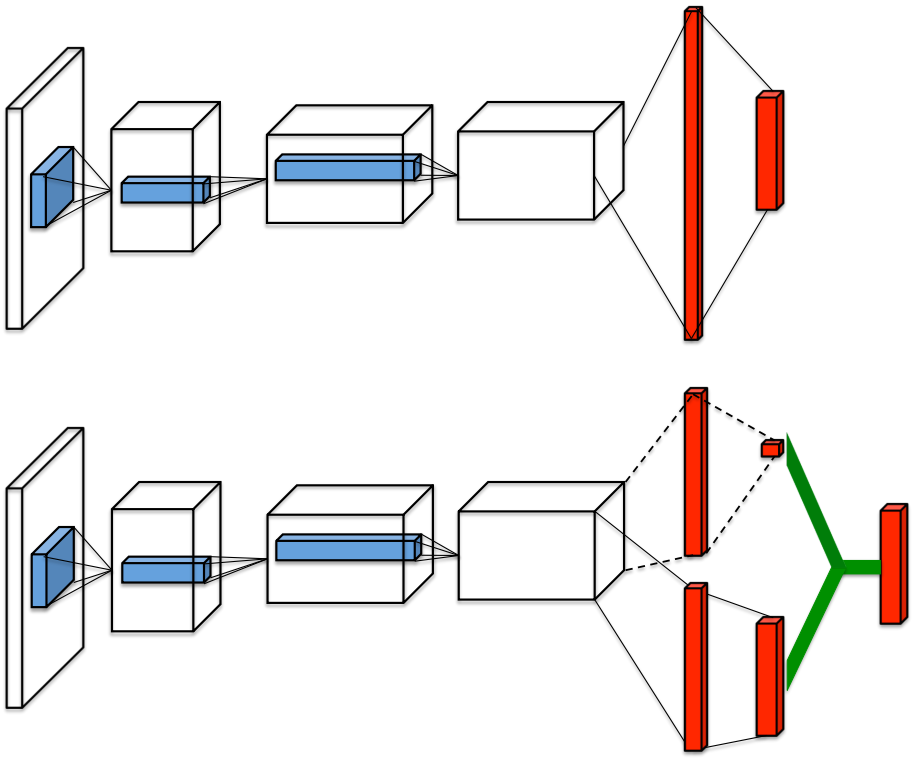

Figure 2: The dueling network architecture (Wang et al., 2015)

Our agent uses the dueling DDQN algorithm with a standard neural network architecture, similar to the one depicted in Figure **??**. It consist of four input nodes (number of states), three fully-connected hidden layers with ReLU activation functions, and a linear output layer with two nodes (number of actions). Further, an ϵ-greedy policy with a linear decreasing exploration rate was used. The RL agent was implemented with the neural networks API Keras[5, 6], which is a high-level abstraction layer of TensorFlow. TensorFlow is the de-facto standard for robust and scalable machine learning in industry and research (Abadi et al., 2016). Further, we used the shared research environment Google Colaboratory[7] to train and evaluate the agent. It offers free access to computing resources that are optimized for training machine learning models. More specifically, it provides a NVIDIA Tesla K80 GPU, with 2880 $\times$ 2 CUDA cores and 12GB GDDR5 VRAM . Additionally, the environment is equipped with a Intel(R) Xeon(R) CPU @ 2.30GHz (1 core, 2 threads), and over 12GB available memory. Google Colaboratory can be used up to 12 hours of consecutive training.

[5] `https://www.keras.io`
[6] `https://github.com/keras-rl/keras-rl`
[7] `https://colab.research.google.com`

3 Simulation Platform: FleetSim

- Simulation Platform to allow evaluating the performance of intelligent agents in smart charging/balancing the grid of EV fleets. Allows to test out bidding strategies and control mechanisms in a realistic EV fleet setting.
- Real life comparison graph: 10%.

3.1 Event-based Simulation

- Not only t+1
- Event e.g. denying rental, charge/little charge/no-charge has effects for the whole simulation
- Simpy
- Python

3.2 Architecture / Components

3.3 Modular Expandability

- Plug-in different Market designs
- Use different real-world data
- Change Fleet parameters
- Develop new strategy
-

4 Results

4.1 Simulation Settings

- Results heavily dependent on industry charging price, since on average the balancing prices are 50% cheaper, and intraday 30% cheaper.
- BMWi. (n.d.). Prices of electricity for the industry in Germany from 2008 to 2017 (in euro cents per kilowatt hour). In Statista - The Statistics Portal. Retrieved March 18, 2019, from `https://www.statista.com/statistics/595803/electricity-industry-price-germany/`.

4.2 FleetRL

- long-delayed rewards make RL hard (!?)
- Compare RL Algos eg. Q-learning vs DDQN –> deep learning makes difference in practice for complex system

- How much could have been used as VPP optimally/perfect information? Similar matrix to Kahlen, or even plot?

4.3 Sensitivity Analysis

- Prediction Accuracy
- Charging infrastructure

5 Conclusion

5.1 Contribution

- Compare to most simular studies:

(Kahlen et al., 2018; Vandael et al., 2015) etc..

- Business model for EV fleet owners with better results than previous studies
- Environmental impact by providing balancing power
- Decision Support System for controlled EV charging from multiple markets
- RL Algorithm that is designed to work in previously unknown environments and thus suited to deploy in real life settings of all kinds of EV fleets in all kinds of cities. E.g. scooters also?
- Event-based Simulation Platform to evaluate bidding strategies and RL agents, facilitate research

5.2 Limitations

- Model:
  - Bidding Mechanism: one week ahead, always accepted
  - Policy & Regulation: EVs not allowed to provide balancing power, minimum bidding quantities 1MW.
  - Markets: Fleet is a price-taker, what about larger fleets? Simulate market influence
- RL: See (Vázquez-Canteli & Nagy, 2019) conclusion for limitations.

5.3 Future Research

- Model: Current market design, i.e. daily w/ 4h slots. German "Mischpreisverfahren"
- RL: Long-delayed rewards, different reward structure, memory based

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... Zheng, X. (2016). TensorFlow: A System for Large-Scale Machine Learning. In *12th USENIX symposium on operating systems design and implementation (OSDI 16)*.

Agricola, A., Seidl, H., & Mischinger, S. (2014). DENA Ancillary Services Study 2030. Security and Reliability of a Power Supply With a High Percentage of Renewable Energy [Technical Report].

Avci, E., Ketter, W., & van Heck, E. (2018). Managing Electricity Price Modeling Risk Via Ensemble Forecasting: the Case of Turkey. *Energy Policy*, 390-403. Retrieved from `https://doi.org/10.1016/j.enpol.2018.08.053` doi: 10.1016/j.enpol.2018.08.053

BMU. (2010). *Energy Concept for an Environmentally Sound, Reliable and Affordable energy* [Technical Report]. Federal Ministry for the Environment, Nature Conservation and Nuclear Safety.

Brandt, T., Wagner, S., & Neumann, D. (2017). Evaluating a Business Model for Vehicle-Grid Integration: Evidence From Germany. *Transportation Research Part D: Transport and Environment*, 488-504. Retrieved from `https://doi.org/10.1016/j.trd.2016.11.017` doi: 10.1016/j.trd.2016.11.017

Burns, L. D. (2013). Sustainable Mobility: a Vision of Our Transport Future. *Nature*. Retrieved from `https://doi.org/10.1038/497181a` doi: 10.1038/497181a

Firnkorn, J., & Müller, M. (2015). Free-Floating Electric Carsharing-Fleets in Smart Cities: the Dawning of a Post-Private Car Era in Urban Environments? *Environmental Science & Policy*, 30-40. Retrieved from `https://doi.org/10.1016/j.envsci.2014.09.005` doi: 10.1016/j.envsci.2014.09.005

Kahlen, M., Ketter, W., & Gupta, A. (2017). Fleetpower: Creating Virtual Power Plants in Sustainable Smart Electricity Markets. *SSRN Electronic Journal*. Retrieved from `https://doi.org/10.2139/ssrn.3062433` doi: 10.2139/ssrn.3062433

Kahlen, M., Ketter, W., & van Dalen, J. (2018). Electric Vehicle Virtual Power Plant Dilemma: Grid Balancing Versus Customer Mobility. *Production and Operations Management*. Retrieved from `https://doi.org/10.1111/poms.12876` doi: 10.1111/poms.12876

Mashhour, E., & Moghaddas-Tafreshi, S. M. (2011). Bidding Strategy of Virtual Power Plant for Participating in Energy and Spinning Reserve Markets-Part I: Problem Formulation. *IEEE Transactions on Power Systems*, 949-956. Retrieved from https://doi.org/10.1109/tpwrs.2010.2070884 doi: 10.1109/tpwrs.2010.2070884

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... Hassabis, D. (2015). Human-Level Control Through Deep Reinforcement Learning. *Nature*. Retrieved from https://doi.org/10.1038/nature14236 doi: 10.1038/nature14236

Pandžić, H., Morales, J. M., Conejo, A. J., & Kuzle, I. (2013). Offering Model for a Virtual Power Plant Based on Stochastic Programming. *Applied Energy*. Retrieved from https://doi.org/10.1016/j.apenergy.2012.12.077 doi: 10.1016/j.apenergy.2012.12.077

Sterling, D. (2018). *Three Revolutions: Steering Automated, Shared, and Electric Vehicles to a Better Future*. Island Press/Center for Resource Economics. Retrieved from https://doi.org/10.5822/978-1-61091-906-7 doi: 10.5822/978-1-61091-906-7

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.

Tomić, J., & Kempton, W. (2007). Using Fleets of Electric-Drive Vehicles for Grid Support. *Journal of Power Sources*, 459-468. Retrieved from https://doi.org/10.1016/j.jpowsour.2007.03.010 doi: 10.1016/j.jpowsour.2007.03.010

Vandael, S., Claessens, B., Ernst, D., Holvoet, T., & Deconinck, G. (2015). Reinforcement Learning of Heuristic EV Fleet Charging in a Day-Ahead Electricity Market. *IEEE Transactions on Smart Grid*, 1795-1805. Retrieved from https://doi.org/10.1109/tsg.2015.2393059 doi: 10.1109/tsg.2015.2393059

van Hasselt, H. (2010). Double Q-learning. In *Advances in neural information processing systems*.

van Hasselt, H., Guez, A., & Silver, D. (2016). Deep Reinforcement Learning with Double Q-Learning. In *Aaai*.

Vázquez-Canteli, J. R., & Nagy, Z. (2019). Reinforcement Learning for Demand Response: a Review of Algorithms and Modeling Techniques. *Applied Energy*,

1072-1089. Retrieved from https://doi.org/10.1016/j.apenergy.2018.11.002 doi: 10.1016/j.apenergy.2018.11.002

Wagner, S., Brandt, T., & Neumann, D. (2016). In Free Float: Developing Business Analytics Support for Carsharing Providers. *Omega*, 4-14. Retrieved from https://doi.org/10.1016/j.omega.2015.02.011 doi: 10.1016/j.omega.2015.02.011

Wang, Z., Schaul, T., Hessel, M., Van Hasselt, H., Lanctot, M., & De Freitas, N. (2015). Dueling Network Architectures for Deep Reinforcement Learning. *arXiv preprint arXiv:1511.06581*.

Weron, R. (2014). Electricity Price Forecasting: a Review of the State-Of-The-Art With a Look Into the Future. *International Journal of Forecasting*, 1030-1081. Retrieved from https://doi.org/10.1016/j.ijforecast.2014.08.008 doi: 10.1016/j.ijforecast.2014.08.008