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 23, 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.

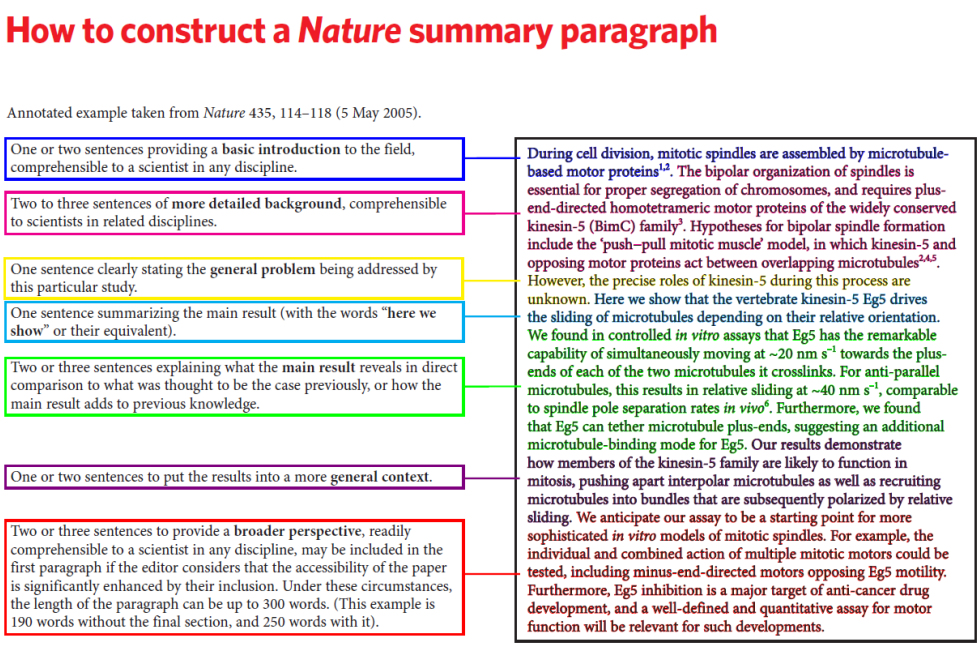
Tobias Richter

Köln, den 01.05.2019

Abstract

This is an abstract

- One or two sentences providing a basic introduction to the field, comprehensible to a scientist in any discipline.
- Two to three sentences of more detailed background, comprehensible to scientists in related disciplines.
- One sentence clearly stating the general problem being addressed by this particular study.
- One sentence summarising the main result (with the words "here we show" or their equivalent).
- Two or three sentences explaining what the main result reveals in direct comparison to what was thought to be the case previously, or how the main result adds to previous knowledge.
- One or two sentences to put the results into a more general context.
- Two or three sentences to provide a broader perspective, readily comprehensible to a scientist in any discipline, may be included in the first paragraph



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

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 Results

The following chapter will cover the main results of this research. First, the simulation environment is presented. Second, we examine the economic sustainability of an integrated bidding strategy. Third, the RL approach that aims to optimize the VPP portfolio is evaluated. In the last section, we will perform sensitivity analyses on the limiting algorithmic factor, the prediction accuracy, and a limiting physical factor, the charging infrastructure.

1.1 Simulation Environment

As part of this research, we developed an event-based simulation platform called *FleetSim*. The platform allows researchers to develop and test out different smart charging and bidding strategies in realistic environment based on real-world data. In FleetSim, intelligent agents (called controllers) centrally control the charging of an EV fleet, they are responsible to sufficiently charge their vehicles to satisfy real mobility demand. At the same time, the agents can create VPP of EVs, provide balancing services to the grid, and take part in electricity trading. Trips are simulated on an individual level, for example, not charging an individual EV at a particular point in time, can cause a whole series of lost rentals, due to an insufficient amount of battery for the next arriving customers. The agents are evaluated based on the profits of charging the fleet cheaper than the industry tariff, the costs of losing rentals and the imbalance they cause if they can not provide market commitments. Additionally, FleetSim facilitates easy sensitivity analyses, adaption to future market designs, and integration of novel data sets through its modular architecture and expandable design (see Figure 1). We consider FleetSim as a research platform for sustainable and smart mobility similar to PowerTac (?, ?). It builds on SimPy[1], a process-based discrete-event simulation framework. FleetSim is available open source[2] and can be readily installed as a Python package.

In order to simplify comparability and focus to real-world applicability of the analysis, we set the same parameters for all conducted experiments (see Table 1). They are corresponding to the real Car2Go specifications described in Chapter **??**. Further, we fixed the unknown prediction accuracy of the fleets available charging power $\widehat{P}^{fleet}$ to an estimate of modern forecast algorithms performance. The impact of the predictions uncertainty on the results will later be determined in a sensitivity analysis.

[1] `https://pypi.org/project/simpy/`
[2] `https://github.com/indyfree/fleetsim`

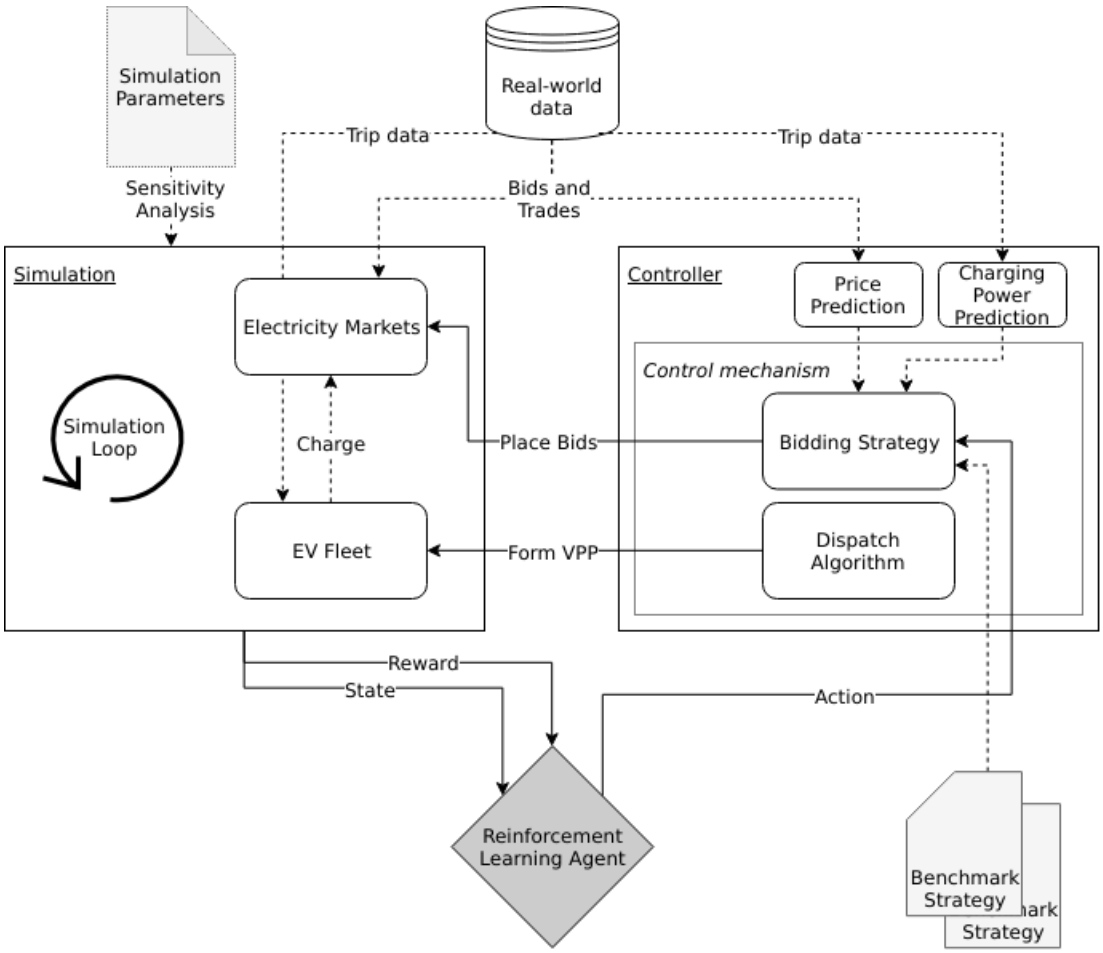


Figure 1: Architecture of FleetSim

Table 1: Simulation Parameters

Parameter	Value
EV battery capacity (Ω)	17.6 kWh
EV charging power (γ)	3.6 kW
EV range	145 km
Industry electricity price (p^{ind})	$0.15^3 \frac{€}{\text{kWh}}$
EV rental tariff	$0.24^4 \frac{€}{\text{min}}$
EV long distance fee (> 200 km)	$0.29^4 \frac{€}{\text{km}}$
Prediction accuracy $\widehat{P}^{fleet}$ week ahead	70%
Prediction accuracy $\widehat{P}^{fleet}$ 30 min ahead	90%

1.2 Integrated Bidding Strategy

Research Question 1 examines whether a fleet operator can use a VPP portfolio of EVs to profitably bid on multiple electricity markets. In Chapter **??**, we proposed a central control mechanism that charges the fleet with an integrated bidding strategy. The following section evaluates the results of the control mechanism in the simulation environment.

Table 2 shows the descriptive statistics of the fleet utilization during a simulation run, with data from June 1, 2016 to January 1, 2018. It can be observed that (a) the volatility of EVs parked at a charging station is remarkably high (large standard deviation), and (b) the fraction of EVs that can be utilized for VPP activities is diminishing low (3.55%). It is apparent that a high uncertainty and the low share of EVs that can possibly generate profits are challenging the economic sustainability of our proposed model. Figure 2 shows that despite a changing rental behavior throughout the day (e.g., rush hour peaks between 7:00-9:00 and 17:00-19:00), the amount of EVs that can utilized for VPP activities is comparably stable throughout the day.

Table 2: Fleet Statistics.

Statistic	Value
Fleet size	508
EVs available (min, max, **std**)	389.64 (165, 496, **49.18**)
EVs connected (min, max, **std**)	61.23 (34, 290, **61.11**)
VPP EVs (min, max, **std**)	13.84 (0, 94, **9.01**)

We defined several "naive" bidding strategies to evaluate and benchmark the performance of our developed model. The strategies are naive in that sense that they are assuming a fixed risk associated with bidding at a specific electricity market. As opposed to the developed RL agent, they do not take information of their environment into account and adjust the bidding quantities dynamically. Instead, the controller discounts the predicted amount of available charging power with a fixed risk factor λ (see (**??**) and (**??**)). Naturally, the controller estimates a higher risk for bidding on the balancing market week ahead than on the intraday market 30 minutes ahead. We defined following types of strategies:

1. Risk-averse ($\lambda^{bal} = 0.5$, $\lambda^{intr} = 0.3$)

[3] Average prices of electricity for the industry with an annual consumption of 500 MWh - 2000 MWh in Germany 2017 (?, ?).

[4] Rental fees according to the Car2Go pricing scheme. See `https://www.car2go.com/media/data/germany/legal-documents/de-de-pricing-information.pdf`, accessed 15th March 2019.

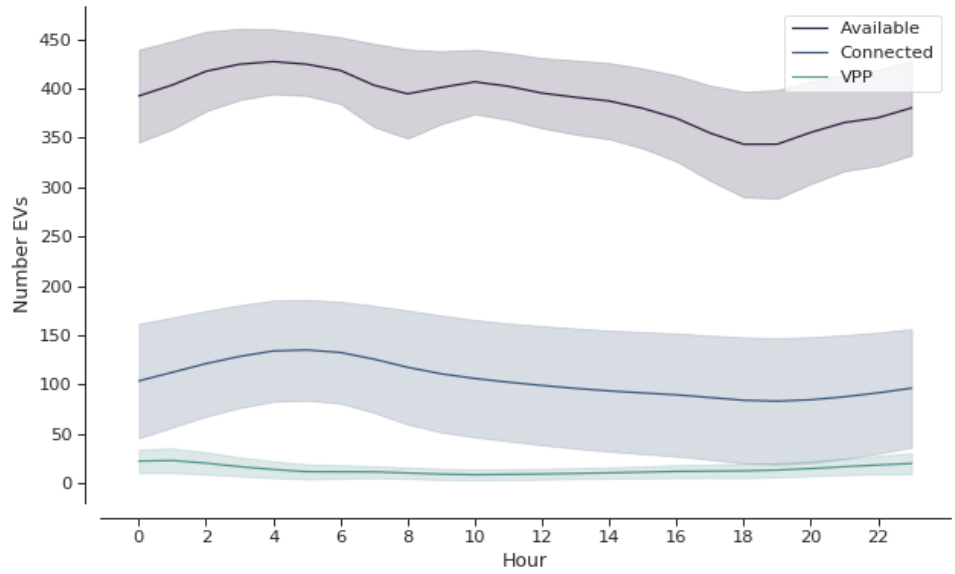


Figure 2: Daily fleet utilization (average, standard deviation) from June 1, 2016 to January 1, 2018. The blue error band is illustrating the large volatility in the amount of EVs that get parked at a charging station. The share of EVs that can be used as VPP is on average only 3.55% of the fleet's size. Most of the EVs are either not connected to a charging station or are already fully charged.

The controller avoids denying rentals and causing imbalances at all costs. In order to not commit more charging power that it can provide, it places only bids for conservative amounts of electricity on the markets. The risk-averse strategies *Balancing* and *Intraday* are comparable to similar strategies developed by ? (?, ?).

2. Risk-seeking ($\lambda^{bal} = 0.2$, $\lambda^{intr} = 0.0$)

   The controller aims to maximize its profits by trading as much electricity on the markets as possible. It strives to fully utilize the VPP and allocate a high percentage of available EVS to charge from the markets. Due to the rental uncertainty and a low estimated risk, the controller is prone to offering more charging power to the markets that it can provide. This may lead to lost rental costs or even imbalances.

3. Full information

   The optimal strategy to solve the controlled charging problem. The controller knows the bidding risks in advance and places the perfect bids on the markets. In other words, it charges the maximal amount of electricity from the markets without having to deny rentals or causing imbalances due to prediction uncertainties.

In Table 3, the simulation results of all tested strategies are listed. As expected, the developed integrated bidding strategies outperform their single market counterparts. The controller is able to capitalize on the most favorable market conditions and better utilizes the VPP by buying more electricity from the markets than charging the EVs regularly. The integrated strategies are resulting in 49%-54% more profits for the fleet than the single market strategies.

A controller bidding according to the *Integrated (risk-averse)* strategy, pays on average [0.10$\frac{€}{kWh}$ How much?] less for charging the fleet than other risk-averse strategies. A controller with an *Integrated (risk-seeking)* strategy, is even more profitable, despite having to account for lost rental profits. On the other side, the controller caused imbalances (highlighted red) which lead to high (unknown) market penalties or even exclusion from bidding activities. For this reason, imbalances need to be avoided, regardless of potential profits from a higher VPP utilization. We expect that the proposed RL agent learns a bidding strategy, which avoids imbalances while increasing profits at the same time. The upper bound of the optimal strategy *Integrated (full information)*.

[Balancing power, stability]

1.3 Reinforcement Learning Portfolio Optimization

Research Question 2 investigates whether a RL agent can optimize the integrated bidding strategy by dynamically adjusting the bidding quantities. The bidding quantities P^{bal}, P^{intr} are based on the evaluated risk associated with bidding on the individual electricity markets. In Chapter **??**, we introduced a RL approach that learns the risk factors λ^{bal}, λ^{intr} based on its observed environment and received reward signals. In Appendix **??**, the hyperparameters are presented which we used to train the dueling DDQN algorithm and solving the controlled charging problem under uncertainty. The values were determined manually through experimentation for the best results. The speed of convergence was also used as a criterion, since the training environment Google Colaboratory only allows up to 12 hours of computing time.

Further, the imbalance costs β were set to an artificially high value to incentivize the agent to learn to always avoid imbalances. Whenever the agents takes an action that causes imbalances (i.e., bid too much electricity), it will receive a highly negative reward signal, leading to a low estimated Q-value of that chosen action in a specific state.

In Figure 3, the performance of the optimized integrated bidding strategy is presented. The proposed RL algorithm increases the gross profits of the fleet on average (n=5) by approximately 72-75% when compared with the naive single

Table 3: Outcomes of naive bidding strategies over a 1.5 year period. Integrated bidding strategies outperform single market strategies.

	Balancing (risk-averse)	**Intraday (risk-averse)**	**Integrated (risk-averse)**	**Integrated (risk-seeking)**	**Integrated (full information)**
VPP utilization (%)	39	47	62	81	71
Energy bought (MWh)	803	985	1292	1681	1473
Energy charged regularly (MWh)	1278	1096	789	400	608
Lost rental profits (1000 €)	0	0	0	15.47	0
No. Lost rentals	0	0	0	1237	0
Imbalances (MWh)	0	0	0	1.01	0
Average electricity price ($\frac{€}{kWh}$)	-	-	-	-	-
Gross profit increase (1000 €)	43.62	45.08	**67.04**	**72.51**	77.36

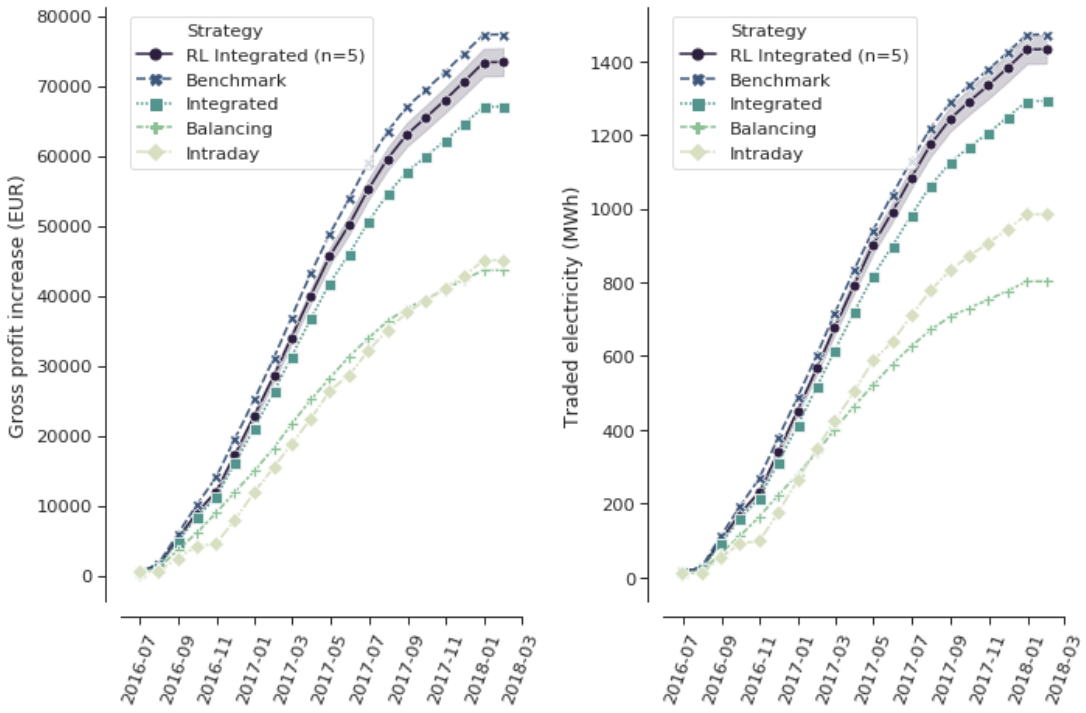


Figure 3: Comparison of gross profits and traded electricity between the proposed optimized integrated strategy and the other three naive charging strategies. The RL algorithm improves the achieved gross profit increase of the integrated bidding strategy on average by 12% and accomplishes nearly optimal results when compared to the benchmark strategy.

market strategies and by approximately 12% when compared with the naive integrated strategy. In none of the tested strategies the controller procured more electricity from the market that it can charge. To reach the optimal solution the gross profits would need to be increased by 3% further. Similarly, the RL algorithm increases the amount of electricity that the fleet charges from the electricity markets while avoiding imbalance at the same time. [Balancing power, renewables]

In another experiment, we evaluated the performance of the proposed RL algorithm in comparison to other RL algorithms with a simpler architecture. In particular, we were interested what impact modern advances in deep RL have on the ability to quickly learn to improve the agents policy, while still achieving good results after the whole training period. This question is especially relevant for the case, when no prior training for the fleet controller is possible and the agent has to quickly learn to avoid procuring more energy from the markets that it can charge. Therefore, we removed the notion of imbalance costs and changed the simulation setup to instantly stop the training episode when imbalances occur. In this way, the agents learns to maximize its reward while circumventing imbalances at all costs. The agent achieves a higher reward the longer it trades electricity

on the markets without committing to charge more electricity than it can. We compared the DQN algorithm (?, ?) with the Double DQN algorithm (?, ?) with and without the dueling architecture (?, ?). In Figure 4, the average (n=5) learning performances of the different RL approaches are displayed.

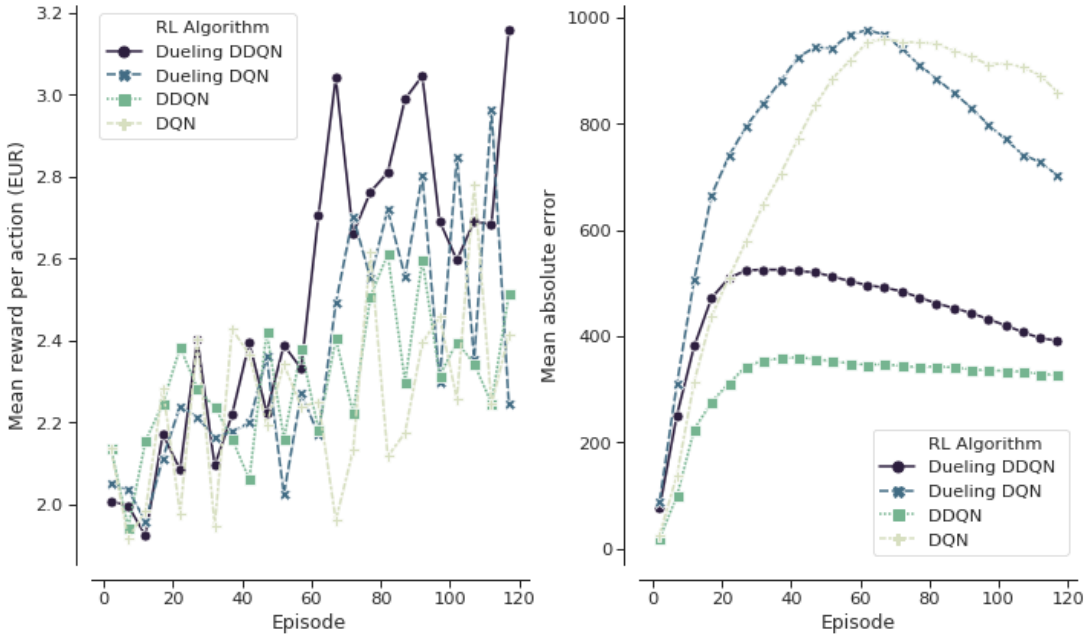


Figure 4: Comparison of the learning performance between the proposed RL algorithm and the other three simpler algorithms, averaged over 5 training attempts. Each training period is performed in 1.5 years simulation time with real world data. The dueling DDQN algorithm (dark blue line) learns faster, and achieves better end results than prior algorithms.

The experiment shows that the dueling DDQN algorithm learns the fastest and shows a large increase in mean reward per action after roughly 60 episodes (about 227 days of simulation time) of training. The dueling DDQN algorithm shows the largest reward increase and highest reward per action after the whole training period, which makes it the best algorithm to solve the charging problem. Despite that it still has a larger mean absolute error than the DDQN algorithm, indicating that it is more likely to cause imbalances with the dueling architecture than without. None of the algorithms determined a policy that never caused imbalances after training on the full 1.5 years of simulation time (about one hour computing time). In other words, without prior training with existing data the RL agent would need more than one and a half years to learn to avoid imbalances. A possible explanation is the problem of learning from long delayed rewards, first discussed by ? (?). Long delayed rewards increase the difficulty of RL problems, since the agent needs to connect occurring decision outcomes to specific actions way back in the past. In the case of the presented controlled fleet charging

problem, this effect is especially pronounced because a negative reward signal (caused imbalances) can occur up to 672 timesteps (one week) after the agent decided on the bidding quantity for the balancing market, whereas the reward signals from the intraday market occur almost immediately after 2 timesteps (30 minutes).

In summary, both experiments show that our approach is able to learn a profit-maximizing bidding strategy under varying circumstances, without using any a priori information about the EV rental patterns. The proposed control mechanism improves existing approaches and the RL agent can successfully optimize the VPP portfolio strategy by estimating the risk that is associated with bidding on the markets.

1.4 Sensitivity Analysis

The ability to accurately forecast the available fleet charging power plays an important role in determining the optimal bidding quantity to submit to the markets. If the fleet controller is certain about the number of connected EVs that it can use for VPP activities in the future, it can aggressively trade the available charging power on the markets, without being concerned about turning away customers or facing the risk of not being able to charge the committed amount of electricity. In our previous experiments, we assumed a fixed prediction accuracy that we set to an estimate of what modern mobility demand forecasts algorithms can achieve. In order to test the robustness of the results and their dependence on the prediction accuracy, we conducted a sensitivity analysis. Therefore, we tested the previously introduced RL approach with increasing levels of predecition accuracy, from 50% to 100% accuracate forecasts 7 days and 30 minutes ahead.

In Table 5 the results of the sensitivity analysis are presented. The left plot shows the effect of the prediction accuracy on the total gross profit increase, whereas the right plot shows the effect on the learned risk factors of the RL agent. Intuitively, the realized profit increases with rising accuracy of the forecasts, while the estimated risk factors decrease with more accurate forecasts.

Interestingly, the RL agent does not always estimate higher risks for bidding on the balancing market than on the intraday market, despite lower accuracy levels for predicting the available charging power 7 days ahead than 30 minutes ahead. This result indicates that the RL performance underlies some variations. The agent succesfully learns to avoid imbalances first by estimating a high total risk and only later learns to optimize the portfolio by fine-tuning the risk factors of both markets. We are confident that the agent's limited amount of training steps is the reason of these variations in learning success and expect to achieve

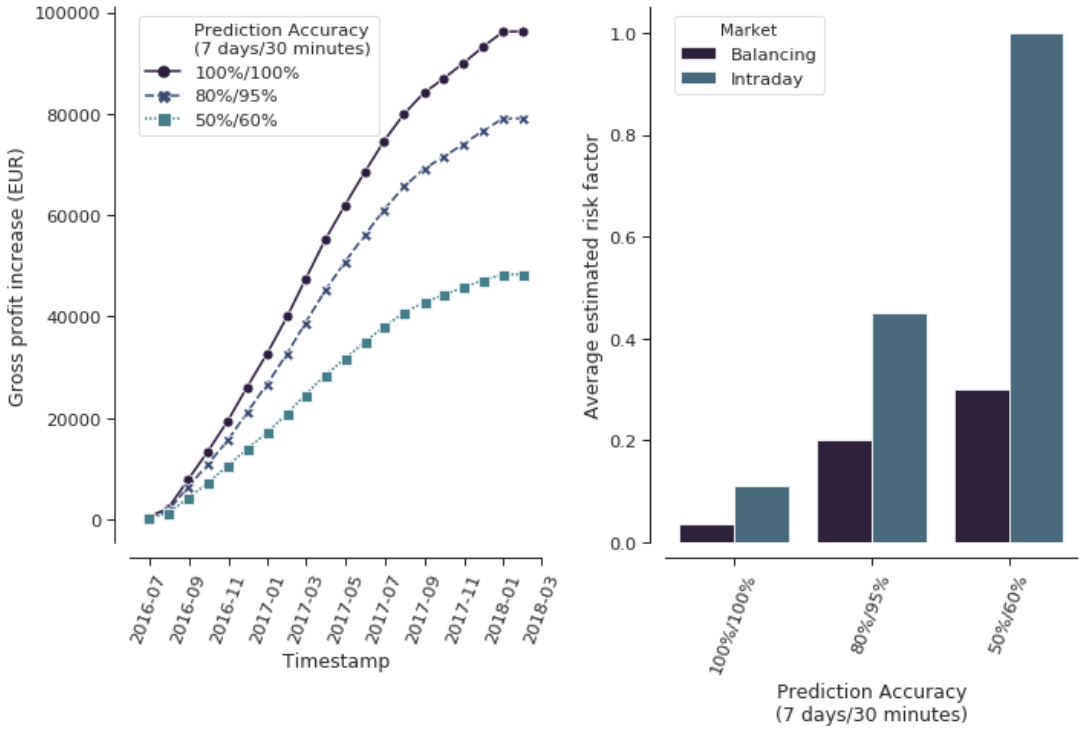


Figure 5: Sensivity Analysis: Prediction Accuracy

more robust results with a increased training time.

Also remarkable is the magnitude of the prediction accuracy's effect on the profit increase. After 1.5 years of simulation time, a RL agent that can rely on perfect predictions (100% accuracy) generates almost twice as much profit from trading electricity than an agent that can only rely on predictions with 50% and 60% accuracy, 7 days and 30 minutes ahead respectively. It is striking that the prediction accuracy has a larger effect (99.13% profit increase between lowest and highst accuracy) on the realized profit than the type of bidding strategy (73.10% profit increase between worst and best strategy), which we examined in the previous sections, leaving room for future research.

2 Conclusion

Integrating volatile renewable energy sources into the electricity system imposes challenges on the electricity grid. In order to ensure grid stability and avoid blackouts, balancing power is needed to match electricity supply and demand. Balancing power can be provided by VPPs that generate or consume energy within a short period of time and offer these services on the electricity markets (?, ?). EV fleet operators can utilize idle vehicles to form VPPs and offer available EV battery capacity as balancing power to the markets. The fleet can offer balancing services directly via tender auctions on the balancing market or via continuous trades on the intraday market, where participants procure or sell energy to self-balance their portfolios (?, ?). Both markets have complementary properties in terms of price levels and lead times to delivery, which motivates the creation of a VPP portfolio to profitably participate in both markets and extend the business model of the fleet.

However, there are certain risks associated with this business model extension. EVs can only be allocated to a VPP portfolio if they are connected to a charging station and have sufficient free battery capacity available, information which is unknown to the fleet at the time of market commitment. This uncertainty makes it difficult for fleet operators to estimate the size of the VPP and calculate the optimal bidding quantities. Moreover, if fleet operators offer more balancing power than they can provide, they face high imbalance penalties from the markets. For a sustainable business model fleet operators also need to balance VPP activities with their primary offering, customer mobility. Denying customer rentals to ensure that the fleet can fulfill the market commitments, results in opportunity costs of lost rentals that compromise the profitability of the fleet.

In the following chapter, we will summarize the conducted research of this thesis to address the aforementioned challenges. Furthermore, the contribution of this work is outlined and the main results are presented and discussed. Finally, we list specific limitations of this study and give insights on further areas of research.

2.1 Contribution

The core contributions of this thesis are the following: First, we conceptualized a DSS for controlled EV charging under uncertainty. The DSS constitutes the core of a business model extension for fleet operators to manage a VPP portfolio of EV batteries. A controlled charging problem has been mathematically formulated and a control mechanism introduced that aims solve the proposed problem. Second, we developed an event-based simulation platform that facilitates fleet management research with real-world data. We evaluated various baseline bid-

ding strategies within the simulation platform and tested out the behavior of intelligent agents in the context of controlled EV charging. Third, we proposed a novel integrated bidding strategy that offers balancing services of a VPP portfolio to multiple electricity markets simultaneously. Instead of submitting only conservative amounts of battery capacity to a single market, like previous studies did (?, ?, ?), our proposed strategy aims to maximize profits by procuring electricity from the multiple markets to the greatest extent possible without causing imbalances. Forth, we proposed the usage of a RL agent that learns to optimize the composition of the VPP portfolio and the associated risks of bidding on the markets. The agent dynamically determines risk factors, depending available and predicted fleet information and market conditions. Therefore, we formulated a MDP that is designed for agents to work in previously unknown environments and uncertain conditions. The RL approach was developed with the focus on real-world applicability, fast convergence rates and generalizability. We expect the proposed approach to work in a variety of different settings, with different kinds of EVs (e.g., electronic bikes), independent of the geographical location of the fleet.

The proposed method was evaluated in a series of expriments with real-world carsharing data from Car2go in Stuttgart and German electricity market data from June 2017 until January 2018. The results show that the developed method improves existing approaches and would increase the gross profits of the fleet by roughly 78000 € over a 1.5 year period. Since free float carsharing has inherently uncertain demand patterns, better results are expected with other kinds of EV fleets. We found that the integrated bidding strategy generates 49%-54% more profits, compared to the naive benchmark strategies. Fleet controllers following the integrated bidding strategy, can mitigate risks and increase profits by exploiting market properties of both markets. Further, we showed that the proposed RL agent can optimize the integrated bidding strategy under uncertainty by roughly 15%, almost [how much?] reaching the optimal solution with full information available. The agent was able to successfully estimate the bidding risk, avoid imbalances and keeping lost rentals to a minimum. [Balancing power, stabilized the grid] Additionally, we tested and compared the performance of modern RL algorithms and found that recent advances in deep RL do improve the robustness and convergence rates in real-world applications. Despite these improvements, all RL approaches needed more than 1.5 years of simulation time to learn to avoid imbalances penalties, which makes RL agents unsuitable to deploy in unseen fleet environments without prior training data. A sensitivity analysis of the prediction accuracy of the fleet balancing power showed that the accuracy has large impact on the fleet profitability. If the controllers predictions are very uncertain, the

RL agent estimates high bidding risks and can only offer conservative amounts of balancing power on the markets. Surprisingly, the effect of prediction accuracy on the total gross profit increase is larger than the choice of bidding strategy or RL algorithm architecture, which leaves room for future research.

- Integrated bidding strategy exploit multiple markets properties
- Online learning algorithm (difference? – real data?)
- Advantage over forecast like Kahlen: General, unseen, etc..
  Impact on grid: Balancing power, mention balancing power.
  Discuss:
- Compare to other studies!
  - Fleet Charging
    - No uncertainty
    - Only one market
    - No sensitivity on accuracy prediction (We found very important)
  - Other approaches (VPP, stochastic) (?, ?) (no imbalance costs!) (?, ?)
- Discuss results?
  - Gross profits small - Mention before?
  - Policy
  - Investment
  - V2G?
  - Market Setup

2.2 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
  - Deny rentals only in the same market period (More deny, less imbalance)
- RL: See (?, ?) conclusion for limitations.
  - Training time in real time. Generalization to other cities?

2.3 Future Research

- This result shows that leaving promising room for future research of highly accurate mobility demand algorithms.
- Model:

  - Investigate modern/current market design, that changed their bidding mechanisms to to better integrate renewable energy generators.
    * Daily/Day-ahead tenders with 4 hour market periods.
    Mischpreisverfahren
    i.e. daily w/ 4h slots. German "Mischpreisverfahren"
- RL: Long-delayed rewards, different reward structure, memory based
- Prediction Algorithms improvement, reference to sensitivity analysis