Forecasting What Matters: DecisionFocused RL for Controlled EV Charging with Unknown Departure Times

Giuseppe Gabriele
IDLab
Ghent University - imec
Ghent, Belgium
giuseppe.gabriele@ugent.be

Fabio Pavirani
IDLab
Ghent University - imec
Ghent, Belgium
Fabio.Pavirani@UGent.be

Seyed Soroush Karimi Madahi
IDLab
Ghent University - imec
Ghent, Belgium
SeyedSoroush.KarimiMadahi@UGent.be

Chris Develder
IDLab
Ghent University - imec
Ghent, Belgium
Chris.Develder@UGent.be

Abstract

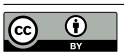
The recent growth of EV adoption poses challenges for power systems, including increased peak demand and potential grid instability. Smart control of EV charging — e.g., based on reinforcement learning (RL) — can alleviate these issues by learning temporal and contextual patterns from historical data. Yet, in real-world scenarios, key features, such as departure time, often are unavailable. This, in turn, makes it harder for an RL agent to learn and execute an effective charging policy. To mitigate this uncertainty, a trained forecaster can approximate the unknown features from available data. However, since these forecasting models are typically trained for accuracy (rather than their impact on a downstream agent's decision quality), their errors may propagate and hinder the overall performance of a controller that is using the forecasts. To avoid this, we propose a decision-focused RL (DF-RL) framework in which the forecaster is trained *end-to-end*, i.e., with feedback from the charging policy actions taken by the RL agent. Such joint training of both the forecaster and controller ultimately results in higher-quality actions: our proposed DF-RL method yields superior charging decisions compared to other baselines, achieving up to a 14% improvement in total reward and a 55% reduction of unsupplied energy (i.e., charging that failed to happen because the EV already left), relative to the RL method without departure time forecasting.

CCS Concepts

• **Computing methodologies** → **Reinforcement learning**; Neural networks; • **Hardware** → **Smart grid**; • **Applied computing** → *Forecasting*.

Keywords

Decision-Focused Learning, Electric Vehicle, Reinforcement Learning, Smart Charging



This work is licensed under a Creative Commons Attribution 4.0 International License.
E-Energy '26, Banff, AB, Canada
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2011-6/26/06
https://doi.org/10.1145/3744255.3811736

ACM Reference Format:
Giuseppe Gabriele, Fabio Pavirani, Seyed Soroush Karimi Madahi, and Chris Develder. 2026. Forecasting What Matters: DecisionFocused RL for Controlled EV Charging with Unknown Departure Times. In *The 17th ACM International Conference on Future and Sustainable Energy Systems (E-Energy '26), June 22–25, 2026, Banff, AB, Canada.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3744255.3811736

1 Introduction

Electric vehicle (EV) charging scheduling is an increasingly important problem for the development of a cleaner and more resilient energy ecosystem, driven by the rapid growth in EV adoption [15]. By coordinating EVs — i.e., deciding when and how vehicles are charged — energy demand can be balanced, costs can be reduced, and renewable power can be used more effectively. Optimizing EV charging thus is key in reducing overall energy expenses for both users and charging service providers.

Two prevalent approaches exist in the literature for EV charging scheduling. The first is *Model Predictive Control (MPC)*, which optimizes decisions over a short future horizon and updates them iteratively as new information becomes available. In the EV context, it forecasts future demand, electricity prices, and vehicle availability, then computes the best charging plan while enforcing operational constraints [1, 23]. The second widely used method is *Reinforcement Learning (RL)*. Thanks to its ability to learn optimal policies through interaction with the environment, RL has demonstrated strong performance for a wide range of EV charging problems [11, 24]. Various objectives are considered, including load flattening [14], provision of grid services [16], and, as considered in our work, overall cost minimization [20]. Since RL is data-driven and model-free, it can overcome the challenges faced by MPC in dealing with uncertainties and nonlinear dynamics that are hard to model accurately.

A critical challenge for both RL and MPC is *incomplete or unknown information*: e.g., the real departure time of the vehicle is typically unknown beforehand. This lack of information further complicates the problem, as the time flexibility available for EVs to shift their charging to off-peak hours is unknown. Moreover, also user satisfaction and charging constraints are directly influenced by the departure time. Only a limited number of works explicitly address this challenge: most studies assume access to exact departure

time or perfect foresight state information. In [13], a hybrid machine learning model is developed to forecast EV session durations, along with arrival times and energy demand. These forecasts are then fed into an MPC to minimize charging costs. Also [19] proposed a machine learning-based framework to predict EV session duration and energy consumption. In [22] a semi-decentralized real-time charging scheduling scheme is put forward: a central operator uses a chance-constrained model to estimate aggregate charging energy and models uncertainties in EV users' behavior using a Gaussian mixture model. Within RL-based EV charging methods, augmented Lagrangian methods have been proposed to handle uncertainty and operational constraints by embedding constraint violations directly into the reward function [3, 4], rather than feeding the uncertain (predicted) departure time as part of the RL state. Another relevant work is [21], which incorporates a lookup table that stores historical EV information, from which departure times of each vehicle are inferred based on previously observed patterns. Yet, this approach is limited to direct interpolation over past observations, and this lack of flexibility compromises generalization.

When dealing with predicting unknown quantities for the sake of control, *decision-focused* (DF) learning offers an effective paradigm for integrating prediction and optimization. Unlike traditional approaches that prioritize predictive accuracy in isolation, DF learning accounts for the impact of predictions on the downstream decision-making problem. DF-based solutions enhance the learning process by guiding predictions based on the quality of the resulting decisions rather than just prediction accuracy, allowing the control agent to further improve task performance [12, 18]. Yet, only few studies have been conducted in EV charging control literature which use DF methods for predicting cost and energy charging demand [7, 8, 10]. All these works formulate the EV charging problem within an optimization framework and leverage decision-focused learning for predicting electricity price or energy demand. The main challenge in integrating DF learning with optimization-based methods is to calculate the gradient of task-aware loss functions with respect to the forecaster parameters, which can be highly non-trivial. To enable backpropagation through the downstream optimization problem, implicit differentiation can be applied to its Karush-Kuhn-Tucker (KKT) conditions (as in the papers cited above). Nevertheless, the inherent discontinuities in linear program outputs can pose difficulties for this approach.

Here, we address the challenge of dealing with a priori unknown EV departure times, while avoiding the complexity of and integrating DF learning with MPC and optimization-based methods. We contribute a **decision-focused RL (DF-RL) framework** to effectively minimize EV charging costs under **uncertain EV departure times**. We enrich the RL agent's state representation with a neural network-based forecast of the EV session duration. In contrast to decision-focused MPC methods, our RL pipeline allows the forecaster to be trained directly from optimizing the control objective: the forecaster and (RL) control agent are thus updated in a *decision-focused* way and trained *end-to-end*. To the best of our knowledge, the application of the decision-focused learning in RL is unique in the context of EV charging scheduling.

2 Problem Formulation

We formalize the EV charging control problem as a Markov Decision Process (MDP) [17]: $\mathcal{M} = (\mathcal{S}, \mathcal{A}, \mathcal{P}, \rho)$, where $\mathcal{S}$ defines the observable state space, $\mathcal{A}$ the action space, $\mathcal{P}$ the state transition function (mapping current state and action to the next state), and ρ the reward function. We consider sequential EV charging sessions, each one defined by arrival time, departure time, and energy demand (to fully charge the battery; in kWh). We adopt a discrete-time setting, with a fixed time resolution Δt (30 min in our study).

The goal is to find a policy $\pi : \mathcal{S} \to \mathcal{A}$ that gives the action to perform, based on a given state as input, to maximize the expected cumulative reward. For such a given policy π, we consider the Q-function Q^π that quantifies the long-term value of taking action a_t when in state s_t (with immediate reward $\rho(s_t, a_t)$) and continuing to follow the policy π (i.e., $a'_t = \pi(s'_t)$ from $t' = t + 1$ onwards) [2]:

$$Q^\pi(s_t, a_t) = \mathbb{E}_\pi \left[\rho(s_t, a_t) + \sum_{k=1}^{K} \gamma^k \, \rho(s_{t+k}, a_{t+k}) \right] \tag{1}$$

Specifically, the agent's **state** comprises two parts:

$$s_t = (m_t, \hat{d}_t) \in \mathcal{S}. \tag{2}$$

The first component m_t comprises known and measurable variables: $m_t = (p_t, E^{\mathrm{u}}_t, h_t)$, where p_t is the electricity price (€/kWh) for that timestep, and E^{u}_t is the energy (kWh) that the EV still needs to be fully charged. Note that we assume to know the total energy that the EV needs to be charged with upfront, and denote it as E^{req} (thus $E^{\mathrm{u}}_t = E^{\mathrm{req}}$ at the time or arrival of the EV). Finally, $h_t \in \mathcal{H} \triangleq \{0, \Delta t, 2\Delta t, ..., 24 - \Delta t\}$ represents the time of the day.

The second component $\hat{d}_t$ represents the remaining charging session duration, which in practice we do not know exactly. To estimate it, we forecast the total session duration $\hat{T}_t$ at each time step (Eq. (3)). Forecasting at every timestep allows $\hat{T}_t$ to take into account the most recent information about the unmet energy. The remaining session duration is then obtained using Eq. (4).

$$\hat{T}_t = f_\psi(x_t) \tag{3}$$

$$\hat{d}_t = \hat{T}_t - T^{\mathrm{elapsed}}_t \tag{4}$$

Our forecaster $f_\psi(.)$ takes the following inputs:

$$x_t = (z_t, E^{\mathrm{c}}_t, E^{\mathrm{u}}_t), \tag{5}$$

where z_t is relevant time information (e.g., including the hour of the day, day of the week, season),[1] and E^{c}_t is the energy the EV has been charged with so far at time t.

We consider a binary **action** space: the charging action $a_t \in \mathcal{A}$ is to either charge at full power ($a_t = P_c$, where P_c is a priori fixed charging power) or to remain idle ($a_t = 0$).

In our problem, the state **transition function** includes some stochasticity, mainly through the time-varying price p_t and the session duration. Part of the transition function is defined by Eq. (6), which shows the update of the remaining energy for the session

$$E^{u}_{t+1} = E^{\mathrm{u}}_t - E_t, \tag{6}$$

[1] In our experiments, based on data analysis of the real-world EV charging sessions, we set $z_t = (h_t, w_t)$ where h_t is the time of the day, and $w_t \in [0, 1]$ a binary variable indicating whether it is weekend or not.

where E_t indicates the energy delivered within the slot, defined as:

$$E_t = \min(a_t \cdot \Delta t, E_t^{\mathrm{u}}).$$

We define the **reward** $r_t \triangleq \rho(s_t, a_t)$ for training and evaluation:

$$r_t = -p_t \cdot E_t \; - \; C_t \cdot \mathbb{I}\{\text{EV left at timestep } t\}, \tag{7}$$

where $\mathbb{I}\{.\}$ is the indicator function, and C_t an extra penalty cost, which through the indicator function only applies upon departure:

$$C_t = \begin{cases} \left(\lambda \cdot E_t^{\mathrm{u}}\right)^2 + C_{\mathrm{fix}}, & \text{if } E_t^{\mathrm{u}} > 0 \text{ (EV not fully charged)}, \\ 0, & \text{otherwise.} \end{cases} \tag{8}$$

The first term of Eq. (7) captures the charging cost incurred at time step t. The second term, adds a penalty cost when a vehicle is not fully charged when it leaves. That penalty comprises a fixed cost C_{fix} and a squared penalty for the unmet charging demand upon departure (i.e., E_t^{u}), regularized by a constant λ.

3 Methodology

In the proposed EV charging scheduling framework, as defined formally by the MDP stated above, we use Soft Actor Critic (SAC) [9] as the RL algorithm. SAC is an off-policy actor-critic algorithm [5] that encourages exploration by maximizing not only expected return, but also the entropy of the policy. That trade-off between reward maximization and policy entropy, is controlled by a temperature parameter α, which is learned online via automatic entropy tuning based on the current policy entropy.

In our **proposed DF-RL method**, RL incorporates a regression forecaster model f_ψ, which augments the state representation by providing additional predictive information (in our case the EV's session duration). Following the DF concept, the forecaster is trained jointly with the RL objective, to not only focus on accurate forecasts, but also optimizing the RL policy's obtained reward. Specifically, the forecaster is updated during training by minimizing the following loss function:

$$\mathcal{L}_{\mathrm{total}} = \beta\, \mathcal{L}_R \; + \; (1-\beta)\, \mathcal{L}_{DF}, \qquad \beta \in [0,1]. \tag{9}$$

Here, the first term, $\mathcal{L}_R$, updates the forecaster based on a regression loss function $\ell(\cdot)$ for the deviation between the forecasted session duration $f(x_t)$ and the real session duration T_{tot} of the EV. [2]

The second term, $\mathcal{L}_{\mathrm{DF}}$, corresponds to the decision-focused component and leverages the actor's decision to guide the update of the forecaster:

$$\mathcal{L}_{\mathrm{DF}} = \mathbb{E}_s \left[\sum_{a \in \mathcal{A}} \pi(a \mid s) \left(\alpha \, \log \pi(a \mid s) - A(s,a) \right) \right]. \tag{10}$$

In this formulation, $A(s,a)$ denotes the advantage function obtained from the standardized Q-value [6], i.e., the action-normalized Q-value, and its policy-weighted expectation.[3]

Within this framework, the parameters of the DF-RL algorithm are updated at the end of each charging session, including the forecaster parameters, optimized via the DF loss defined in Eq. (9).

Note that through the weight parameter β in Eq. (9), we can modulate the degree of updating the forecaster based on the DF objective: setting $\beta = 0$ results in a purely DF-based update, whereas $\beta = 1$ corresponds to a standard regression-based update with respect to the target value.

Now that we explained our DF-RL framework, in the next section, we investigate the benefits of a DF forecaster (and thus the resulting DF-RL policy) over the forecaster trained using a regression loss and integrated with RL.

4 Experiments

Our experiments are based on a real-world dataset of EV charging sessions, from which we extract vehicle arrival times, session durations, and charging requirements. The sessions are collected from an EV parking station located at a hospital[4]. From this dataset, the 20 most frequent EV users are considered to model EV features using statistical distributions to generate realistic simulation samples. We thus construct a training dataset of 350 sessions, and will test on $N = 120$ sessions to report results. For all sessions, we assume a fixed charging power of $P_c = 6.5\,\mathrm{kW}$. Since we focus on time departure uncertainty, we assume a single deterministic time-varying price profile for all sessions, shown in Appendix A.

To assess the performance of DF-RL, we compare several variants, ranging β in [0,1]. We further pretrain the forecaster purely based on minimizing regression loss, separate from the training data used for the RL agent. To assess the relevance of the RL agent knowing the session duration (or not), we further provide a baseline where it does not have any information ('no forecast'; i.e., $\hat{d}_t$ is omitted from the state) or has perfect information ('real departure'). To gauge the performance of these RL-based solutions, we also provide a business-as-usual (BAU) baseline where the EV starts charging immediately upon arrival. Finally, we note that the RL agents are trained for 200 episodes (an episode being one pass over the entire training set), the discount factor is set to $\gamma = 0.99$, and cost parameters are $C_{\mathrm{fix}} = 2.1$, $\lambda = 0.041$.

To evaluate the performance of these DF-RL approaches and mentioned baselines, we report metrics relevant for the EV users: (i) the total charging cost of all EVs ('charging cost'), (ii) number of sessions with incomplete charging ('unsatisfied EVs'), and (iii) the fraction of the unmet charging load ('unmet demand'). The latter *unmet demand* is more specifically calculated as the percentage of kWh that were requested during the EV session (E^{req}), but were eventually not supplied by the time it departs (E^{unmet}), averaged over all N sessions:

$$\text{Unmet demand} = \frac{1}{N} \sum_{i=1}^{N} \frac{E_i^{\mathrm{unmet}}}{E_i^{\mathrm{req}}}. \tag{11}$$

To provide insight in the RL algorithm's operation, we further list their reward and penalty costs, i.e., (i) the cumulative reward defined in Eq. (7), over all test charging sessions, and (ii) the total penalty cost, C_t, again summed over all sessions.

The resulting metrics are reported in Table 1. We notice that all vanilla RL-based methods outperform the BAU solution in terms of total reward, primarily due to the high charging costs incurred by the BAU strategy. Focusing on the vanilla RL methods, we note the expected benefit of having departure time information: RL with the conventional forecaster ($\beta = 1$) almost halves the unmet demand,

[2] In our implementation, we use Mean Squared Error (MSE) loss.

[3] Note that here we assume the policy function $\pi(s,a)$ to return the probability to take action a when in given state s.

[4] This was collected at a public parking facility at a hospital, in the frame of the European H2020 project RENergetic https://cordis.europa.eu/project/id/957845

Table 1: Comparison of our proposed decision-focused RL (DF-RL) approach vs. conventional RL and business-as-usual (BAU) baselines. Metrics are averages ± standard deviation over 4 experiment runs, on a test set of $N = 120$ EV charging sessions.

	User-focused metrics			RL algorithm metrics	
Algorithm	Charging Cost (€)	# Unsatisfied EVs	Unmet Demand (%)	Total Reward	Penalty Cost
BAU (immediate charging)	168.43	0.0	0.0	−168.43	0.0
RL (no forecast)	60.94 ± 14.06	33 ± 4.95	14.7 ± 9.5	−166.88 ± 22.20	105.94 ± 36.26
RL w/ Forecaster ($\beta = 1$)	93.42 ± 14.06	33 ± 11.67	7.3 ± 5.2	−151.00 ± 22.20	57.58 ± 36.26
RL w/ Real Departure	102.17 ± 3.66	5 ± 4.71	0.6 ± 0.6	−115.48 ± 11.09	13.31 ± 8.91
DF-RL $\beta = 0$	55.70 ± 13.90	48 ± 1.50	24.9 ± 1.3	−177.46 ± 9.01	121.76 ± 4.89
DF-RL $\beta = 0.2$	93.79 ± 12.36	21 ± 8.14	6.3 ± 4.3	−142.85 ± 6.59	51.16 ± 16.87
DF-RL $\beta = 0.4$	91.90 ± 14.46	24 ± 13.32	6.6 ± 5.6	−142.85 ± 14.02	50.96 ± 26.04
DF-RL $\beta = 0.6$	94.69 ± 15.45	24 ± 12.42	6.6 ± 5.6	−142.90 ± 12.61	48.21 ± 27.73
DF-RL $\beta = 0.8$	92.62 ± 14.36	27 ± 10.26	7.3 ± 4.0	−144.14 ± 6.82	51.52 ± 20.99

compared to the RL without any forecaster. Still, the forecast errors imply that there is still a meaningful performance gap compared to having perfect information ('RL w/ Real Departure'), where unmet demand is virtually eliminated.[5]

Looking at our decision-based forecast methods (DF-RL), while they still use the same inputs as conventional forecasting, we note that DF-RL manages to further reduce unmet demand, while also achieving a lower charging cost (see $\beta = 0.2$ or 0.4): out of the 120 EVs in the test set, an additional ~10 are fully charged compared to RL using conventional forecasts. We note that lower β (e.g., 0.2) results in slightly higher costs — similar to $\beta = 1$ — but fewer unmet charging demands, while higher β (e.g., 0.8) reduces cost at the expense of more unsatisfied EVs. This reflects the trade-off: lower β prioritizes meeting energy demand, increasing electricity use. Additionally, higher β leads to greater cost variability, whereas $\beta = 0.2$ produces more stable outcomes. In terms of the RL algorithm's internal metrics, using DF-RL increases total reward by 15%, 13%, and 5%, compared to, respectively, BAU, RL without departure time forecasts, and RL with conventional (non-DF) forecasting. Overall, all DF-RL variants with $0 < \beta < 1$ obtain promising results.[6]

As already noted in Table 1, the main benefit of our proposed DF-RL approaches is that it reduces the number of EVs that fail to complete charging ('unsatisfied EVs'). This suggests that our DF forecasts do generate signals that are more informative for the RL agent's policy learning (while not necessarily producing more accurate predictions). A particular example of this behavior is illustrated in Fig. 1, which compares the same EV charging session under DF-RL with $\beta = 0.4$ and RL with the traditional forecaster ($\beta = 1$). We note that the decision-focused forecaster predicts a slightly shorter session duration (ending around 11:10) than the true one (around 11:20). This conservative prediction effectively encourages the RL agent to start charging early enough (at 8:30), allowing the EV to complete charging before its actual departure. In contrast, the traditional forecaster overestimates the available session duration (predicting its end around 12:05), leading to delayed

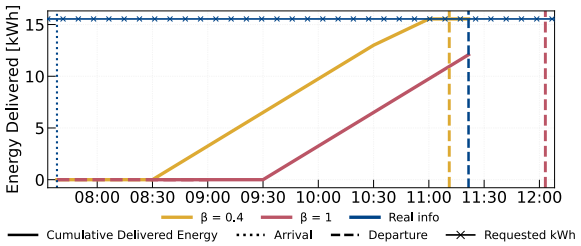


Figure 1: Exemplary charging session for DF-RL (with $\beta = 0.4$) compared to traditional RL (non-DF forecaster, with $\beta = 1$).

charging and an incomplete charging session. Overall, these results thus demonstrate our proposed DF-RL approach's effectiveness.

5 Conclusions and Future Work

We introduced a decision focused reinforcement learning (DF-RL) framework, to solve the problem of dealing with unknown information (focusing particularly on departure times), in coordinated EV charging scenarios. Our DF-RL solution introduces a forecaster that augments the RL agent's state and is jointly trained using feedback from that agent. In the considered EV charging control problem, the forecaster predicts the anticipated EV's departure time (learned from historical behavior). In our DF-RL approach, that prediction is optimized to improve charging decisions, rather than solely focusing on prediction accuracy (based on MSE minimization). We benchmarked our DF-RL approach against RL with such conventional forecasting (as well as boundary baselines comprising a business-as-usual "charge-when-plugged" scenario, and an RL agent without any departure time information). This comparison demonstrated the benefit of our DF-RL in terms of a 5% improvement on the total reward, and 14% less of the total unmet energy, compared to the model with the conventional forecaster.

Future work will focus on extending the proposed DF-RL framework to a more realistic scenario with different daily price profiles, incorporating more expressive forecasting models capable of capturing temporal patterns such as time-of-day effects and seasonality. In addition, we plan to investigate cooperative extensions of the

[5]Even then it is not 0, since we use a penalty factor for constraint satisfaction, applying a soft penalty rather than enforcing a hard constraint on the agent.
[6]For $\beta = 0$, which amounts to neglecting forecasting accuracy (see Eq. (9)), DF-RL unexpectedly performs poorly: without forecaster loss guidance, the forecaster and RL agent still converge but requires more exploration and additional training iterations.

proposed framework, where multiple vehicles are controlled to better handle grid constraints and system-level limitations, ultimately enabling a scalable and more efficient charging solution.

Acknowledgments

This research was partly funded by the Flemish Government through the "Onderzoeksprogramma Artificiële Intelligentie (AI) Vlaanderen" programme, and by the EU through the Horizon Europe project BlueBird (https://bluebirdproject.eu/ – grant agreement no. 101192452)

References

[1] Ali Saadon Al-Ogaili, Tengku Juhana Tengku Hashim, Nur Azzammudin Rahmat, Agileswari K. Ramasamy, Marayati Binti Marsadek, Mohammad Faisal, and Mahammad A. Hannan. 2019. Review on Scheduling, Clustering, and Forecasting Strategies for Controlling Electric Vehicle Charging: Challenges and Recommendations. *IEEE Access* 7 (2019), 128353–128371. https://doi.org/10.1109/ACCESS.2019.2939595

[2] Lucian Busoniu, Robert Babuska, Bart De Schutter, and Damien Ernst. 2017. *Reinforcement learning and dynamic programming using function approximators*. CRC press.

[3] Guibin. Chen and Xiaoying. Shi. 2022. A Deep Reinforcement Learning-Based Charging Scheduling Approach with Augmented Lagrangian for Electric Vehicle. arXiv:2209.09772 [cs.AI] https://arxiv.org/abs/2209.09772

[4] Guibin Chen, Lun Yang, and Xiaoyu Cao. 2025. A deep reinforcement learning-based charging scheduling approach with augmented Lagrangian for electric vehicles. *Applied Energy* 378 (2025), 124706.

[5] Ivo Grondman, Lucian Busoniu, Gabriel AD Lopes, and Robert Babuska. 2012. A survey of actor-critic reinforcement learning: Standard and natural policy gradients. *IEEE Transactions on Systems, Man, and Cybernetics, part C (applications and reviews)* 42, 6 (2012), 1291–1307.

[6] Ivo Grondman, Lucian Busoniu, Gabriel A. D. Lopes, and Robert Babuska. 2012. A Survey of Actor-Critic Reinforcement Learning: Standard and Natural Policy Gradients. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 42, 6 (2012), 1291–1307. https://doi.org/10.1109/TSMCC.2012.2218595

[7] Chengyang Gu, Yuxin Pan, Ruohong Liu, and Yize Chen. 2024. Learning and Optimization for Price-based Demand Response of Electric Vehicle Charging. arXiv:2404.10311 [eess.SY] https://arxiv.org/abs/2404.10311

[8] Chengyang Gu, Yuxin Pan, Ruohong Liu, and Yize Chen. 2024. Learning and Optimization for Price-Based Demand Response of Electric Vehicle Charging. In *2024 American Control Conference (ACC)*. 3625–3630. https://doi.org/10.23919/ACC60939.2024.10644254

[9] Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel, and Sergey Levine. 2018. Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor. arXiv:1801.01290 [cs.LG] https://arxiv.org/abs/1801.01290

[10] Wenxian Hao, Jingxiang Wang, and Zhaojian Wang. 2025. Day-Ahead V2G Station Arbitrage Scheduling: A Decision-Focused Approach. In *2025 37th Chinese Control and Decision Conference (CCDC)*. 1531–1537. https://doi.org/10.1109/CCDC65474.2025.11090572

[11] Seyed Soroush Karimi Madahi, Giuseppe Gabriele, Bert Claessens, and Chris Develder. 2025. Scalable Attention-based Reinforcement Learning Method for Multi-asset Control. In *ICML 2025 CO-BUILD Workshop on Computational Optimization of Buildings*. https://openreview.net/forum?id=3h0v1Ht73L

[12] Jayanta Mandi, James Kotary, Senne Berden, Maxime Mulamba, Victor Bucarey, Tias Guns, and Ferdinando Fioretto. 2024. Decision-focused learning: Foundations, state of the art, benchmark and future opportunities. *Journal of Artificial Intelligence Research* 80 (2024), 1623–1701.

[13] Graham McClone, Avik Ghosh, Adil Khurram, Byron Washom, and Jan Kleissl. 2023. Hybrid machine learning forecasting for online mpc of work place electric vehicle charging. *IEEE Transactions on Smart Grid* 15, 2 (2023), 1891–1901.

[14] Hesam Mosalli, Saba Sanami, Yu Yang, Hen-Geul Yeh, and Amir G Aghdam. 2025. Dynamic Load Balancing for EV Charging Stations Using Reinforcement Learning and Demand Prediction. In *2025 IEEE International systems Conference (SysCon)*. IEEE, 1–7.

[15] Matteo Muratori, Marcus Alexander, Doug Arent, Morgan Bazilian, Pierpaolo Cazzola, Ercan M Dede, John Farrell, Chris Gearhart, David Greene, Alan Jenn, et al. 2021. The rise of electric vehicles—2020 status and future expectations. *Progress in Energy* 3, 2 (2021), 022002.

[16] Keonwoo Park and Ilkyeong Moon. 2022. Multi-agent deep reinforcement learning approach for EV charging scheduling in a smart grid. *Applied energy* 328 (2022), 120111.

[17] Martin L Puterman. 1990. Markov decision processes. *Handbooks in operations research and management science* 2 (1990), 331–434.

[18] Sanket Shah, Kai Wang, Bryan Wilder, Andrew Perrault, and Milind Tambe. 2022. Decision-focused learning without decision-making: Learning locally optimized decision losses. *Advances in Neural Information Processing Systems* 35 (2022), 1320–1332.

[19] Sakib Shahriar, Abdul-Rahman Al-Ali, Ahmed H Osman, Salam Dhou, and Mais Nijim. 2021. Prediction of EV charging behavior using machine learning. *Ieee Access* 9 (2021), 111576–111586.

[20] Muddsair Sharif and Huseyin Seker. 2024. Smart EV charging with context-awareness: Enhancing resource utilization via deep reinforcement learning. *IEEE Access* 12 (2024), 7009–7027.

[21] Felix Tuchnitz, Niklas Ebell, Jonas Schlund, and Marco Pruckner. 2021. Development and evaluation of a smart charging strategy for an electric vehicle fleet based on reinforcement learning. *Applied Energy* 285 (2021), 116382.

[22] Weilun Wang and Lei Wu. 2024. A semi-decentralized real-time charging scheduling scheme for large EV parking lots considering uncertain EV arrival and departure. *IEEE Transactions on Smart Grid* 15, 6 (2024), 5871–5884.

[23] Lei Yang, Xinbo Geng, Xiaohong Guan, and Lang Tong. 2024. EV Charging Scheduling Under Demand Charge: A Block Model Predictive Control Approach. *IEEE Transactions on Automation Science and Engineering* 21, 2 (2024), 2125–2138. https://doi.org/10.1109/TASE.2023.3260804

[24] Jin Zhang, Liang Che, and Mohammad Shahidehpour. 2023. Distributed training and distributed execution-based Stackelberg multi-agent reinforcement learning for EV charging scheduling. *IEEE Transactions on Smart Grid* 14, 6 (2023), 4976–4979.

A Price Profile

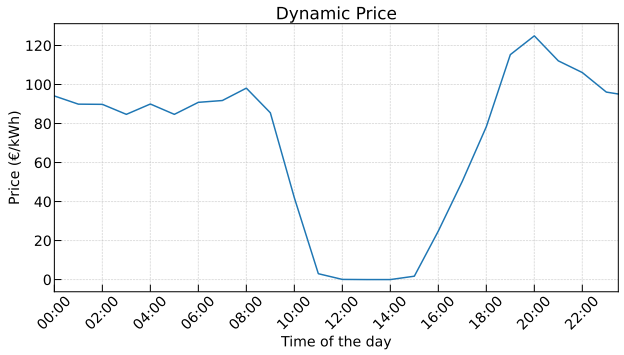


Figure 2: Daily price profile considered in our experiments.