energies

MDPI

Article

Optimal Management for EV Charging Stations: A Win–Win Strategy for Different Stakeholders Using Constrained Deep Q-Learning

Athanasios Paraskevas [1], Dimitrios Aletras [1], Antonios Chrysopoulos [1,2], Antonios Marinopoulos [3] and Dimitrios I. Doukas [1,*]

[1] NET2GRID BV, Krystalli 4, 54630 Thessaloniki, Greece; thanos@net2grid.com (A.P.); aletras@net2grid.com (D.A.); antonios@net2grid.com (A.C.)
[2] School of Electrical and Computer Engineering, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece
[3] European Climate, Infrastructure and Environment Executive Agency (CINEA), European Commission, B-1049 Brussels, Belgium; antonios.marinopoulos@ec.europa.eu
* Correspondence: dimitrios@net2grid.com

Abstract: Given the additional awareness of the increasing energy demand and gas emissions' effects, the decarbonization of the transportation sector is of great significance. In particular, the adoption of electric vehicles (EVs) seems a promising option, under the condition that public charging infrastructure is available. However, devising a pricing and scheduling strategy for public EV charging stations is a non-trivial albeit important task. The reason is that a sub-optimal decision could lead to high waiting times or extreme changes to the power load profile. In addition, in the context of the problem of optimal pricing and scheduling for EV charging stations, the interests of different stakeholders ought to be taken into account (such as those of the station owner and the EV owners). This work proposes a deep reinforcement learning-based (DRL) agent that can optimize pricing and charging control in a public EV charging station under a real-time varying electricity price. The primary goal is to maximize the station's profits while simultaneously ensuring that the customers' charging demands are also satisfied. Moreover, the DRL approach is data-driven; it can operate under uncertainties without requiring explicit models of the environment. Variants of scheduling and DRL training algorithms from the literature are also proposed to ensure that both the conflicting objectives are achieved. Experimental results validate the effectiveness of the proposed approach.

Keywords: dynamic pricing; EV charging station; pricing and scheduling; reinforcement learning; deep Q-learning; demand response

Citation: Paraskevas, A.; Aletras, D.; Chrysopoulos, A.; Marinopoulos, A.; Doukas, D.I. Optimal Management for EV Charging Stations: A Win–Win Strategy for Different Stakeholders Using Constrained Deep Q-Learning. *Energies* **2022**, *15*, 2323. https://doi.org/10.3390/en15072323

Academic Editor: Hugo Morais

Received: 2 March 2022
Accepted: 15 March 2022
Published: 23 March 2022

1. Introduction

There has been increasing concern about global warming and climate change due to gas emissions [1]; at the same time, the energy demand is rapidly increasing [2,3], and for the most part it is satisfied through fossil-fuel energy sources [1]. Fossil fuel combustion and carbon dioxide (CO_2) emissions are significantly contributing to environmental pollution and global warming [1,4]. Therefore, the decarbonization of the transportation sector has naturally arisen a potential partial solution. In particular, the adoption of electric vehicles (EVs) is a promising option because of their benefits over standard fossil-fuel vehicles and their sustainable qualities [5,6]. The report of the International Energy Agency (IEA) [7] mentions that EVs are developing at a rapid pace, indicating that the global EV fleet exceeded 5.1 million in 2018 and that the there may be 250 million units by 2030.

To that end, establishing public charging infrastructure is a critical task that could lead to widespread EV adoption [8,9]. However, for public EV charging stations, there is a need to develop new business models and tackle additional challenges. For example, sub-optimal

scheduling decisions when charging multiple EVs could result in high waiting times, while also significantly changing the demand profile of the utility by causing increased electricity demand, particularly at peak times [10,11]. Therefore, it is crucial to investigate different optimization strategies for EV charging stations while at the same time taking into consideration the perspectives of different stakeholders.

A literature survey revealed that most published research focuses on single-charger settings. According to [12], solving the global optimization problem is impractical in the absence of the distributions of future EV arrivals, charging duration, and base loads. Traditional approaches formulate the EV charging scheduling problem as a sequential decision-making task [13,14] and solve it using dynamic programming [15]. However, these conventional approaches require modeling the uncertainties, which is not necessarily feasible under real-life conditions.

On the other hand, reinforcement learning (RL) can be leveraged to solve problems formulated as Markov Decision Processes (MDP) [16]. In particular, deep reinforcement learning (DRL) methods have proven to be highly effective at complex tasks, outperforming human experts [17]. The main advantage of model-free DRL methods is that they are data-driven; i.e., the agents learn directly from experience without requiring explicit models of their environment. Note that different RL reward definitions lead to different optimization objectives, such as maximizing the EV owners' profits, focusing on the EV charging stations' profits, prioritizing the distribution system operator's needs, or reducing waiting times [18].

In [19], kernel density estimation, was used to model the joint probability distribution of the arrival times and charging duration of EVs at a public charger. Then, a deep Q network (DQN) agent was trained to decide the charging/discharging rate in each time slot by choosing from a discrete number of levels. The observation space consists of the 24 h electricity price history, the remaining energy until the EV is fully charged, and the remaining time until departure. At the same time, the optimization objective takes into account minimizing charging costs and satisfying charging demands. In [20], arrival and departure times, and charging demand at a single charger, were modeled as truncated normal distributions. The observation space was similar to the one of [19], and a long-short-term-memory network was used to predict future electricity prices based on historical data. A modified deep deterministic policy gradient algorithm, called control deep deterministic policy gradient, allows the agent to choose charging/discharging rates from a continuous interval, aiming at maximizing the EV owner's profit and satisfying the charging demand.

The proposed solution in [21] uses a combination of two networks, one for extracting representative features on the electricity price time series, and a DQN agent to control the EV real-time charging/discharging actions. The rewards' definition considers both the charging costs and a penalty component proportional to the amount of uncharged energy, representing the "range anxiety" factor. The state information comprises the presence of the EV at home or not, the remaining battery energy, and the price time series for the past 24 h. Reference [22] introduces an EV charging station environment model and an admission system, where different types of EVs are modeled (simulating different customer profiles) and are presented with charging prices, accordingly to demand. The RL agent decides the amount of energy to purchase (which will be used to charge some of the parked EVs) and the price to announce to new EVs that arrive in each time slot. The models are trained using a variant of the well-known state–action–reward–state–action (SARSA) algorithm [16], called Hyperopia SARSA. The state information includes residual charging demands and parking times, and the reward is modeled towards optimizing the profit of the charging station.

The scope of this paper is to present an intelligent agent that optimally decides, in real-time and under uncertainties (such as the distribution of future EV arrivals and the electricity price), the pricing and scheduling actions needed to maximize a particular EV charging station's profit. Simultaneously, the EV owners' expectations and needs are taken into account. The main contributions of this paper are:

- In contrast to prior strategies [19–22], the proposed strategy is a win–win for both stakeholders, i.e., the EV owners and the EV charging station operators. Fulfilling charging demands under agreed conditions is prioritized, and profit maximization from the charging station operator's perspective follows.
- Although direct bench-marking against pre-published literature is difficult because of the different operating conditions and data used, the financial benefit that is achieved for the charging station herein is considerable and comparable to the profit achieved in the literature [22].
- A new training scheme is proposed for the Q-learning algorithm. The constraints imposed guarantee customer satisfaction, which is removed from the optimization objective to allow the RL agent to maximize EV charging station profit.
- The proposed strategy is easy to adjust, and a different balance/prioritization between stakeholders needs can be selected (see Equation (13)).
- The strategy takes into account real-time conditions and data. In contrast, some of the implementations already proposed in the literature [18,23] do not do so, and they mainly focus on the day-ahead time window.

The remainder of the paper is structured as follows: Section 2 presents the environment that was developed to represent the operations of an EV charging station, with regard to the pricing and scheduling decisions that are made. The problem is formulated as an MDP. Section 3 describes the proposed solution that is able to both decide the optimal sequence of actions and ensure that customers' demands are being fulfilled. Furthermore, the architecture of the DRL agent is detailed in that section, along with the training algorithm used. Section 4 details the datasets on which the proposed approach was trained, and the settings of the experiments carried out. Section 5 presents the results of the training experiments that validate the effectiveness of the proposed approach. The agent's decision-making ability is analyzed, and implications are discussed in the context of two case studies. Finally, Section 6 concludes the paper by stating the primary findings of this work.

2. System Model

2.1. EV Charging Station Environment

We use a DRL-based approach to tackle the problem, since it is data-driven and does not require explicit modeling of the uncertainties, as mentioned previously.

The entity of interest and the basis of the proposed model is an EV charging station environment. The environment is observed in discrete time slots (indexed by t). The length (duration) of each time slot in minutes is denoted by t_{len}. The notation and formulation that follows are based on [22].

At the beginning of each time slot t (meaning during the entire slot $t-1$), a set of EVs $\mathcal{I}_t$ arrives at the station. We denote by $\mathcal{J}_t$ the set of EVs that are already parked in the station before time slot t and have not yet finished charging. Thus, the EVs that require charging at time slot t are denoted by $\mathcal{K}_t := \mathcal{I}_t \cup \mathcal{J}_t$.

Each EV $i \in \mathcal{I}_t$ that arrives at the beginning of time slot t is presented with the price rate r_t (determined by the charging station and measured in the currency/kWh) and accordingly responds with its charging demand, d_i and maximum desired waiting time p_i. The following assumptions are made:

- EVs are price-sensitive; i.e., they adjust their charging demands based on the value of r_t provided by the station. Thus, $d_i = D_i(r_t)$, where $D_i(\cdot) : \$/\text{kWh} \rightarrow \text{kWh}$ is the demand–response function of EV i. Obviously, if EV i decides not to accept the presented rate, then $d_i = 0$. Additionally, note that the demand–response function is EV-specific in the general case.
- The price rate r_t presented to $\mathcal{I}_t$ will be constant for each EV in $\mathcal{I}_t$ during its parking time.
- There is a fixed and finite number of individual chargers at the station, N. Thus, for all time slots t, $|\mathcal{K}_t| \leq N$, which means that at any given time, at most N EVs are parked at the station. Suppose the number of EVs, $|\mathcal{I}_t|$, that arrive at the station overflow the

available chargers. In that case, a subset of $\mathcal{I}_t$ is selected, in a first-come-first-served manner, to meet the parking capacity of the station.

It directly follows from the above that if (t_i^a, p_i, d_i) denote the arrival time, parking time, and charging demand of EV $i \in \mathcal{I}_t$, then d_i must be fulfilled before the departure of the EV at time $t_i^a + p_i$.

In time slot t, the station also determines the charging rate $x_{i,t}$ at which each EV $i \in \mathcal{K}_t$ will be charged during the time slot.

Let $x_{\max}$ be the maximum individual charging rate (limited by the specifications of every single charger) and $e_{\max}$ be the maximum total charging rate for the charging station. (It is assumed that the charging rate limit of the EV itself is always higher than the charger limit).

The following constraints hold:

$$0 \leq x_{i,t} \leq x_{\max}, \quad t = 1, 2, \ldots, \forall i \in \mathcal{K}_t \tag{1}$$

$$\sum_{i \in \mathcal{K}_t} x_{i,t} \leq e_{\max}, \quad t = 1, 2, \ldots \tag{2}$$

$$\alpha \sum_{t=t_i^a}^{t=t_i^a+p_i} x_{i,t} \geq d_i, \quad \forall i \tag{3}$$

where the coefficient $\alpha := \frac{t_{\text{len}}}{60}$ converts the charging rate $x_{i,t}$ (kW) assigned to each EV for the current time slot t to the total amount of energy (kWh) that it will have received by the end of the time slot.

Equations (1) and (2) follow by definition. Equation (3) ensures that the charging demand of each EV i is fulfilled by the time it is set to leave the station. In general, the optimal pricing and scheduling policy might result in an EV not being charged at all for several time slots (when the electricity price is expected to be increased, for example), though of course, still being charged in the end. As explained in [24], these idle times could potentially negatively affect the charging infrastructure in terms of its availability, sizing, and cost. That aspect is not studied in the context of this work.

Finally, for each time slot t, the set of newly arrived EVs $\mathcal{I}_t$ pay a total of:

$$\sum_{i \in \mathcal{I}_t} r_t D_i(r_t) \tag{4}$$

to the charging station, according to the price rate r_t and the requested charge $d_i = D_i(r_t)$ of each EV i (of course, this is not valid unless the charging demand is actually satisfied by the departure time).

At the same time, in order to charge EVs in each time slot t, the charging station pays an electricity bill of

$$c_t \sum_{i \in \mathcal{K}_t} \alpha x_{i,t} \tag{5}$$

where c_t is the electricity price (\$/kWh). It is assumed that c_t varies under the real-time pricing scheme [25].

The interactions between the different components of the charging station environment can be seen in Figure 1.

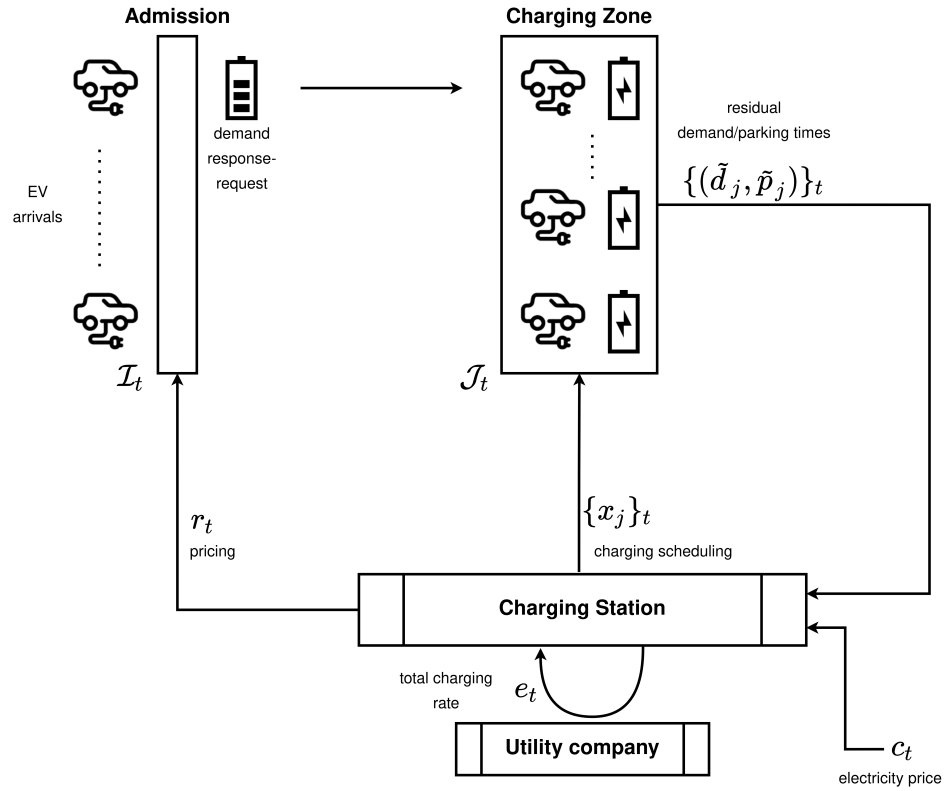

Figure 1. The RL environment for an EV charging station.

2.2. Problem Formulation Using the MDP Framework

The MDP definition [16] provides the basic framework on which RL agents are formally developed.

In particular, at each time step t, the environment is at state S_t; the agent interacts with the environment by selecting an action A_t; the environment responds by transitioning to the next state S_{t+1}, which is returned to the agent along with the reward R_{t+1}. The latter is a scalar signal that depends on the environment and the selected action A_t. In turn, the agent uses the information of S_{t+1}, R_{t+1} to decide the next action A_{t+1}, so the above steps are repeated. This process is illustrated in Figure 2.

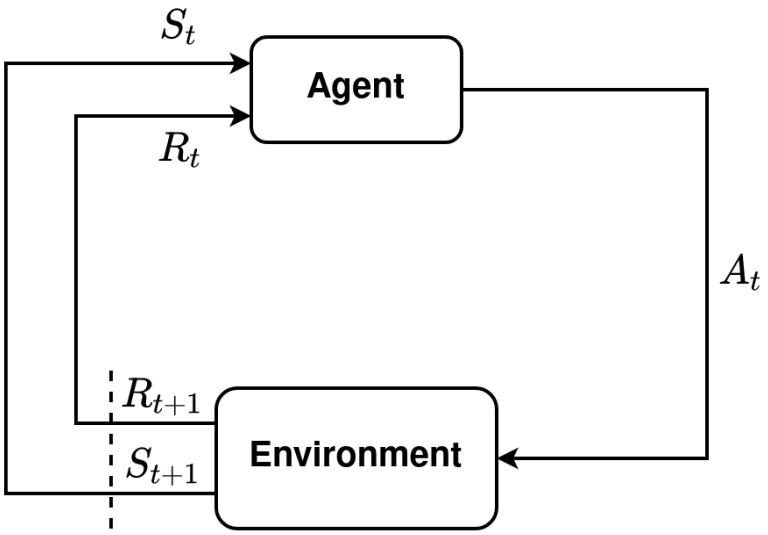

Figure 2. Interactions between an agent and its environment in an RL setting.

The optimization objective of an RL algorithm is to train an agent that selects a series of actions that maximize the total expected return. Equivalently set, the optimization criterion is:

$$\max \mathbb{E}\left[\sum_t \gamma^t R_t\right] \tag{6}$$

where $\gamma \in [0, 1)$ is the *discount rate*, which is used to decrease the importance of distant future rewards, compared to immediate ones. We proceed to formulate the problem of optimal real-time scheduling and pricing in EV charging stations using the MDP framework.

State/Observation Space

The system state at time slot t is defined by:

$$\begin{aligned} S_t = \Big(& \mathcal{J}_t, \{\tilde{d}_j^t\}\Big|_{j\in\mathcal{J}_t}, \{\tilde{p}_j^t\}\Big|_{j\in\mathcal{J}_t}, \\ & \mathcal{I}_t, \\ & \boldsymbol{c}_{t:t-t_{24\,\mathrm{h}}} \Big) \end{aligned} \tag{7}$$

and includes:

- The EVs that are parked at the station $\mathcal{J}_t$, along with the residual charging demand $\tilde{d}_j^t$ and parking time $\tilde{p}_j^t$ for each EV $j \in \mathcal{J}_t$
- The newly arrived EVs, $\mathcal{I}_t$
- The last 24 h of values of the electricity price time series. Under the assumption that electricity price changes every Δt slots, the 24 h historical values can be represented by:

$$c_t, c_{t-\Delta t}, c_{t-2\Delta t}, \ldots, c_{t-M\Delta t} \tag{8}$$

where $M\Delta t = 24\frac{60}{t_{\mathrm{len}}}$. Equivalently, the number of samples M is given by:

$$M = 24\frac{60}{t_{\mathrm{len}}\Delta t} \tag{9}$$

Action Space

At each time slot t, the action to be determined by the agent is the tuple

$$A_t = (r_t, e_t); \tag{10}$$

that is, the price rate for new EVs that arrive at the station, and the total charging rate $e_t := \sum_{i\in\mathcal{K}_t} x_{i,t}$ to be distributed among parked EVs.

As proved in [22], under certain conditions it is sufficient to determine, at each time slot t, the value of e_t instead of the individual charge amounts $x_{i,t}$. In turn, those can be found by applying the least laxity first (LLF) algorithm.

The laxity $l_{i,t}$ of EV i at time slot t is defined as:

$$l_{i,t} := \tilde{p}_i^t - \frac{\tilde{d}_i^t \cdot 60}{x_{\max}} \tag{11}$$

$\tilde{d}_i^t$ is multiplied by 60 so as to convert the energy measured in kWh to kW·min, which in turn is divided by the maximum individual charging rate, $x_{\max}$ measured in kW. Intuitively, $l_{i,t}$ represents the "headroom" between the remaining parking time and the minimum charging time required to fulfill the remaining demand.

Having determined the value of e_t, LLF schedules the values of $x_{i,t}$ by assigning higher priority to those EVs presenting the least laxity. In other words, according to LLF, the station should first charge those EVs that are most urgent to finish charging. For more details on

the LLF algorithm, the reader is referred to [22]. An improved implementation of the LLF algorithm, called constrained LLF, is described in Section 3.1.

Reward Modeling

The definition of the reward function is related to the optimization objective of the desired solution. In this work, the problem is studied from the points of view of the EV charging station and the EV owners; thus, the first objective is to maximize the station's profit. Taking into account Equations (4)–(6), the reward at each time slot t is defined as the total payment the station collects from new EVs minus the cost for charging all parked EVs:

$$R_t := \sum_{i \in \mathcal{I}_t} r_t D_i(r_t) - \alpha c_t e_t \tag{12}$$

Equation (12) is valid only as long as each EV $i \in \mathcal{I}_t$ is indeed fully charged with its required demand. Otherwise, the difference between the requested charging $D_i(r_t)$ and the actual charge provided should be introduced in the calculations. It should be noted that the use of the constrained total charging rate e'_t that is obtained via the constrained LLF algorithm, presented in the next section, ensures that the laxity of each EV remains positive, leading to a successful charge.

3. Proposed Solution

3.1. Constrained Least Laxity First

As the agent can freely choose the total charging rate e_t, a situation could arise in which the EVs have not been adequately charged. Equivalently, the residual charging demand $\tilde{d}_i^t$ of some EVs would not reach zero by the time they are set to leave the station ($\tilde{p}_i^t = 0$).

In that case, the constraint mentioned in Equation (3) is not satisfied, and EV owners could be discontent with the amount of energy they received during a charging session, thereby violating the second objective set in the previous section. Note that this case is not explicitly handled in [22].

It can be observed that, if at any time slot t, the laxity of an EV i, $l_{i,t}$, is negative, that EV can no longer be satisfied in its initial energy demand.

Constraining the total charging rate e_t could prevent such an event from occurring. With that in mind, a lower bound for e_t is introduced when applying the LLF algorithm, as described in Algorithm 1. Essentially, the *constrained LLF* algorithm first charges EVs with the least laxity with the maximum individual charging rate ($x_{\max}$) until the total charging rate e_t is distributed. The algorithm then constrains the total charging rate if needed to prevent any negative laxities from occurring in the next time slot. It should be noted that the use of constrained LLF requires that the charging station is capable of charging all EVs at the maximum charging rate concurrently; i.e., $e_{\max} = N \cdot x_{\max}$.

It should also be mentioned that the constraints of the LLF algorithm could be relaxed, allowing the agent to slightly undercharge EVs, with the aim of improving the charging station profit. Specifically, each laxity could be allowed to reach sub-zero levels, meaning that the residual demand of some EVs might not be met by the end of a charging session. The logical expression for the residual demand given the statement in Algorithm 1 would be written as:

$$l_{i,t+1} < \xi \tag{13}$$

where $\xi < 0$ is the relaxation coefficient. The maximum amount of residual demand that could potentially be unfulfilled is analogous to $|\xi|$.

Algorithm 1 Constrained least laxity first.

Require: Total charging rate e_t
Require: Total number of chargers N
Require: Residual demand $\tilde{d}_i^t$, $i \in \mathcal{K}_t$
Require: Residual parking time $\tilde{p}_i^t$, $i \in \mathcal{K}_t$
  Initialize remaining total charging rate $\tilde{e}_t \leftarrow e_t$
  for i = 1, N **do**
    Initialize $x_{i,t} \leftarrow 0$
    Calculate laxity $l_{i,t} \leftarrow \tilde{p}_i^t - \frac{\tilde{d}_i^t \cdot 60}{x_{\max}}$
    Initialize $l_{i,t+1} \leftarrow l_{i,t}$
  end for
  while $\tilde{e}_t > 0$ **do**
    Find EV $\hat{\imath}$ with the least laxity that has $x_{\hat{\imath},t} = 0$
    Update charging rate of EV $\hat{\imath}$: $x_{\hat{\imath},t} \leftarrow \min\left(\tilde{e}_t, x_{\max}, \tilde{d}_{\hat{\imath}}^t \cdot \frac{1}{\alpha}\right)$
    Calculate laxity of EV $\hat{\imath}$ for next time slot $t+1$: $l_{\hat{\imath},t+1} \leftarrow l_{\hat{\imath},t} + \frac{x_{\hat{\imath},t} \cdot t_{\text{len}}}{x_{\max}} - t_{\text{len}}$
    Update remaining total charging rate $\tilde{e}_t \leftarrow \tilde{e}_t - x_{\hat{\imath},t}$
  end while
  for i = 1, N **do**
    if $l_{i,t+1} < 0$ **then**
      Constrain charging rate of EV i: $x_{i,t} \leftarrow \min\left(x_{\max}, \tilde{d}_i^t \cdot \frac{1}{\alpha}\right)$
    end if
  end for
  Calculate constrained total charging rate $e'_t \leftarrow \sum_{i=1}^{N} x_{i,t}$

3.2. Agent Architecture

The agent is modeled as a deep neural network, whose architecture is shown in Figure 3. The state information (Equation (7)) is provided as input to the agent. In particular, the network has:

- N input nodes, each of which is the laxity of an EV at charger i, $l_{i,t}$.
- M input nodes corresponding to the values of the electricity price over the last 24 h, according to Equations (8) and (9).
- One node corresponding to the number of EV arrivals observed at the admission zone of the station.

The network's output approximates the total expected return for each action that the agent can choose in the current state. Since the total expected return per action provides information about the value of each action, it is called the action value. As will be described in Section 3.3, the training objective of the deep neural network is based on the *Q-learning* algorithm [26].

Most straightforward DRL algorithms, including standard deep Q-learning [17], operate on discrete action spaces; i.e., the set of all available actions $\mathcal{A}_t = \{A_t\}, \forall t$ is countable and finite. However, by definition (Equation (10)), the action space in this case is continuous. Therefore, actions are discretized as shown below:

- Let $w_r = \{w_{r,1}, w_{r,2}, \ldots, w_{r,L}\}$ be the L discrete price rate levels.
- Let $w_e = \{w_{e,1}, w_{e,2}, \ldots, w_{e,K}\}$ be the K discrete charging rate levels.
- Then, the action space is :

$$\mathcal{A}_t = w_e \times w_r = \{(w_{r,1}, w_{e,1}), \ldots, (w_{r,L}, w_{e,1}), (w_{r,1}, w_{e,2}), \ldots, (w_{r,L}, w_{e,K})\}, \quad (14)$$

i.e., the Cartesian product of the discrete level sets, with cardinality $|\mathcal{A}_t| = L \cdot K$

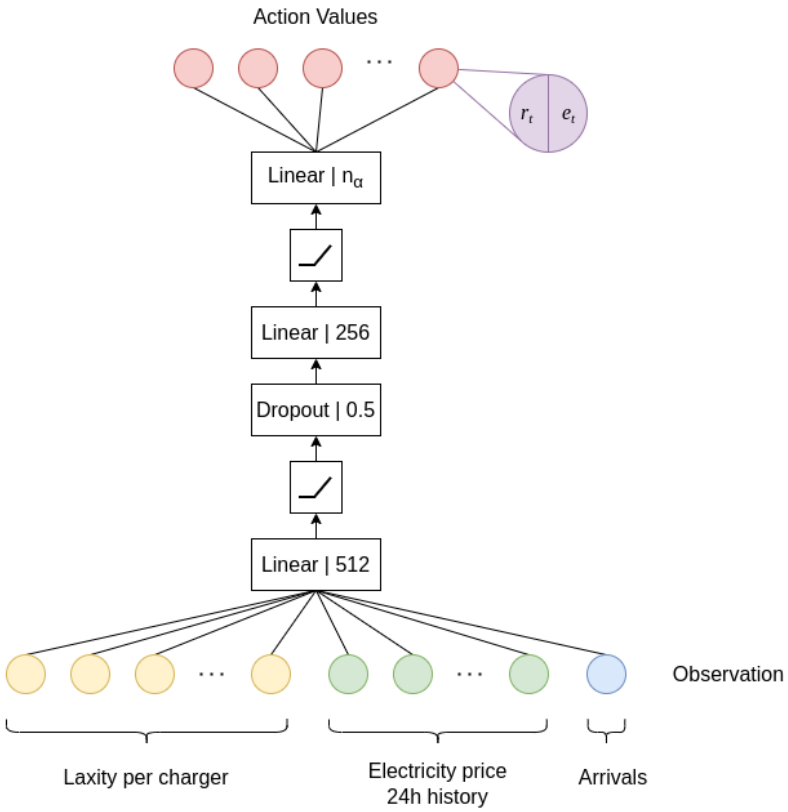

Figure 3. DQN agent architecture.

A limitation of discretizing continuous action spaces is that the number of discrete actions could potentially explode. Therefore, the exploration phase of the algorithm and evaluating all individual actions become impractical [27]. Proper discrete levels should be selected that reflect the solution boundaries for the selected datasets/parameters.

3.3. Training Approach

During training, the agent consists of two identical deep neural networks: a policy and a target network, as explained in [28]. The target network copies the weights of the policy network every few updates of the latter, lagging behind a few episodes, to improve stability. The agent plays through episodes while storing experiences (observations, actions taken, rewards gained, and new observations) in a replay buffer. This buffer is then sampled at every step, and a batch is used for (continuously) training the networks.

A behavior policy is used to explore the environment while collecting data to prevent the agent from adhering to a sub-optimal policy due to the local minima of the loss function. The ϵ-greedy policy is commonly used to achieve such goals. According to the ϵ-greedy policy, the agent selects the greedy action that maximizes reward with probability $1 - \epsilon$ and a random action with probability ϵ. As training progresses, the probability ϵ decays to ensure convergence.

As mentioned in Section 2.2, the charging rate e_t selected by the agent should be above a lower bound in order to satisfy the problem formulation constraints. However, the action space is discretized, as described in Section 3.2. Thus, if the constrained charging rate e'_t obtained by the constrained LLF algorithm is higher than the selected e_t, then the agent is forced to select the discrete charging level $w_{e,i}$ which is closest to e'_t and satisfies the inequality $w_{e,i} \geq e'_t$. On the other hand, the selected price rate r_t does not have any constraints, so it is left unchanged.

Furthermore, it was experimentally found that gradually increasing the episode duration helps the agent grasp the EVs' charging cycle. Specifically, the small initial episode duration provides the agent with data of the environmental state when the station is not yet busy. As the agent learns to charge a small number of EVs at the start of the episode, the duration increases, allowing the agent to apply the knowledge gained to charge many EVs and schedule charging concurrently.

The training approach is summarized in the Algorithm 2.

Algorithm 2 Constrained deep Q-learning.

Require: Episode length schema function, h
Require: Exploration rate schema, l
Initialize replay memory D to capacity N
Initialize action-value $Q \equiv Q(s, a, ; \boldsymbol{\theta})$ parametrized with random weights $\boldsymbol{\theta}$
Initialize target action-value $\hat{Q}$ with weights $\boldsymbol{\theta}^{-}$
for episode = 1, E **do**
  Initialize state s_1
  Get current episode duration $T = h(\text{episode})$
  for t = 1, T **do**
    Get exploration rate $\epsilon = l(\text{episode}, t)$
    With probability ϵ select a random action a_t, otherwise select $a_t = \arg\max_a Q(s, a\,; \boldsymbol{\theta})$
    Constrain a_t using the Constrained LLF algorithm
    Execute a_t and observe reward R_t and next state s_{t+1}
    Store transition (s_t, a_t, R_t, s_{t+1})
    Sample random minibatch of transitions (s_i, a_j, R_j, s_{j+1})
    Set target

$$y_j = \begin{cases} R_j, & \text{if } s_{j+1} \text{ final state} \\ R_j + \gamma \max_{a'} \hat{Q}(s_{j+1}, a'; \boldsymbol{\theta}^{-}), & \text{otherwise} \end{cases}$$

    Perform a gradient descent step on $\left(y_j - Q(s_j, a_j; \boldsymbol{\theta})\right)^2$
    Every C steps copy policy network weights to target network weights $\boldsymbol{\theta}^{-} = \boldsymbol{\theta}$
  end for
end for

4. Evaluation Methodology

4.1. Datasets

Two datasets were used to model the EV charging station environment: one for the EV arrivals at the station and one for the hourly electricity price the station pays to the utility company for the energy purchased during each hour.

The dataset for the EV arrivals was provided by [22], which is an open-source code repository from the author of [22]. It contains vehicle arrivals per 30 s for Richards Ave station near downtown Davis. The EVs are divided into three types, namely, (a) emergent, (b) normal, and (c) residential. Each type has different demand preferences and available parking time, which are described in Section 4.2. The following preprocessing steps were performed on the data points to match these data to a realistic charging station scenario:

- They were upsampled to 60 min intervals.
- They were scaled by a factor of $\frac{1}{100}$ and rounded to the closest integer.
- They were undersampled to 1 min intervals, by randomly distributing the 1 h samples to intermediate minutes using a uniform distribution.

An overview of the average number of EV arrivals per hour of the day for the different charging profiles can be observed in Figure 4. The averaging was performed for every hour separately, for all the days that are included in the dataset.

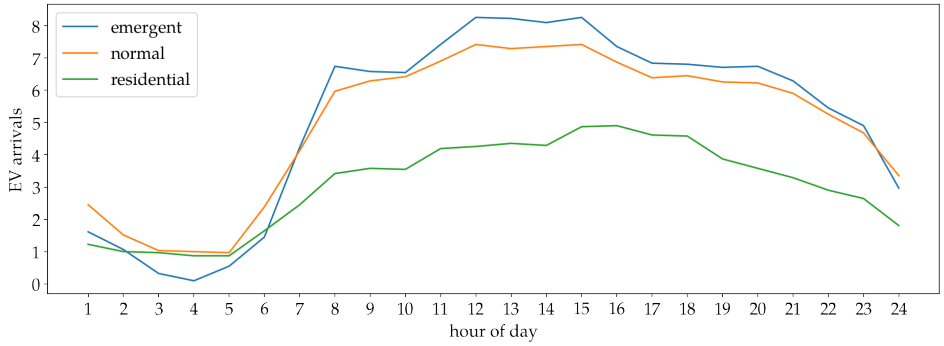

Figure 4. Average EV arrivals per hour of day per charging profile.

For the electricity price, a dataset from the Korean grid [29] was utilized, which is publicly available. It contains hourly prices per kWh of energy purchased from the grid. The dates of the observations range from 1 July 2021 to 31 July 2021, matching the month of the dataset for the arrivals mentioned above. Initially, a currency conversion was realized, and the price was scaled to achieve greater variance during a day and challenge the agent to adapt to intraday fluctuations. The conversion function is $f(x) = \frac{0.00084x^2}{3}$, where x is price in Korean ₩ and $f(x)$ is price in US $. An overview of the average electricity price per hour of the day can be observed in Figure 5. The averaging was performed in a similar manner as explained for the EV arrivals dataset.

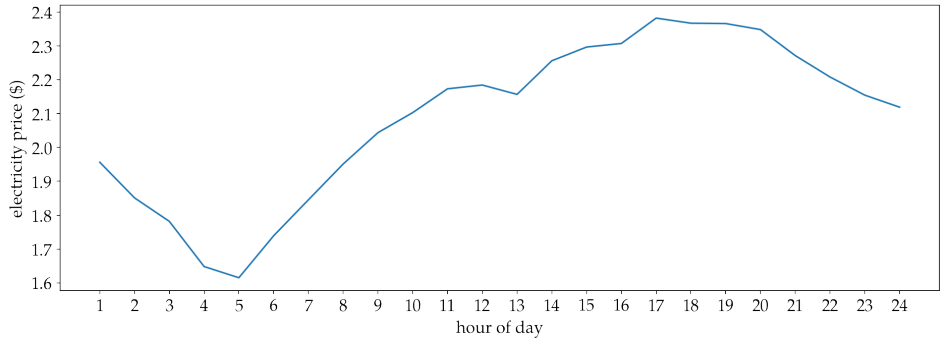

Figure 5. Average electricity price per hour of day.

4.2. Experimental Setup

Before conducting the experiments, a set of hyperparameters were selected. These include the following:

- Chargers of the station: $N = 20$.
- Maximum charging rate per charger: According to the U.S. Department of Energy (https://afdc.energy.gov/fuels/electricity_infrastructure.html, accessed on 15 February 2022), most EVs on the road today are not capable of charging at rates higher than 50 kW. Thus, a more conservative approach of 30 kW was selected. Note that 22 kW is the closest standard charging rate (i.e., Level 2 EV charging), but the purpose of this work is to present a more general approach.) $x_{\max} = 30$ kW
- Maximum total charging rate: $e_{\max} = N \cdot x_{\max} = 600$ kW.
- Time slot length: $t_{\text{len}} = 5$ min.
- Episode duration: 1 day or 1440 min or 288 time slots.
- Discrete price rate levels: $\{1, 2, 3, 4, 5, 6\}$$.
- Discrete charging rate levels: $\{0, 60, 120, 180, 240, 300, 360, 420, 480, 540, 600\}$ kW.
- Cardinality of action space: $|\mathcal{A}_t| = 66$.

Demand–Response Function

The demand–response function is modeled as a linear equation of the form:

$$D_i(r) = \beta_1 r + \beta_2 + \mathcal{N}\left(0, \sigma^2\right) \tag{15}$$

where β_1, β_2, and σ are the parameters of each EV i; and $\mathcal{N}\left(0, \sigma^2\right)$ is Gaussian noise with mean $\mu = 0$ and standard deviation σ. Following the type division of the EV arrivals dataset, EVs are grouped into three different types, each with specific parameters, which are presented in Table 1. These parameters were adopted from the related work in [22]. The respective plot of the demand–response functions is illustrated in Figure 6. As can be seen in [30], the potential charging demands are in line with the battery capacities of some of the latest EV models. The dotted lines show the Gaussian Noise's variance by adding one standard deviation σ to each demand–response function. These also provide an approximate limit to the maximum price that the customers of each type are willing to pay to the station.

Table 1. Demand–response function parameters and parking time for each EV type.

EV Type	Standard Deviation σ	β_1 [kWh/$]	β_2 [kWh]	Parking Time
Emergent	4.47	−1	6	30
Normal	3.96	−4	15	120
Residential	2.63	−25	100	720

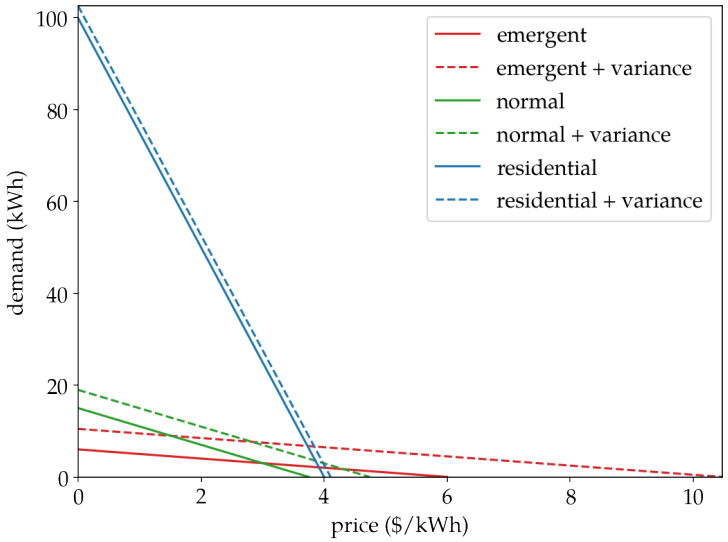

Figure 6. Plot of demand–response function for each EV type.

ϵ-Greedy Policy

The decaying probability ϵ of the ϵ-greedy policy is calculated by the equation:

$$\epsilon = \epsilon_{\text{end}} + (\epsilon_{\text{start}} - \epsilon_{\text{end}}) \cdot \exp\left\{-\frac{x}{\epsilon_{\text{decay}}}\right\} \tag{16}$$

where x is the episode number; $\epsilon_{\text{start}} = 0.9$ and $\epsilon_{\text{end}} = 0.05$ are the initial and final probabilities of a random action (for $x = 0$ and $x \to \infty$, respectively); and $\epsilon_{\text{decay}} = 200$ is the rate of decay for ϵ. A plot of the above equation can be observed in Figure 7. Essentially, the probability ϵ converges to its final value after 800 episodes.

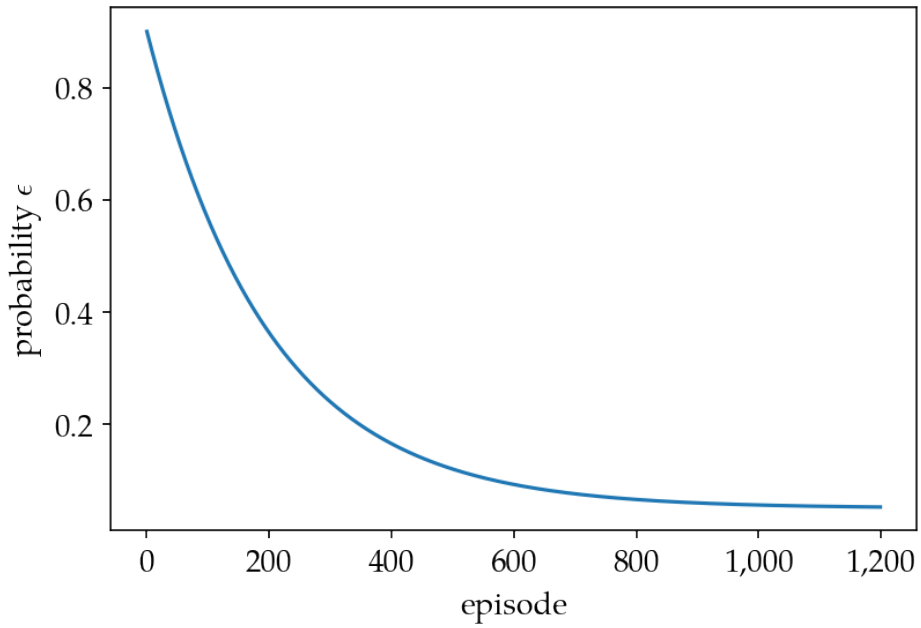

Figure 7. Plot of random probability for ϵ-greedy policy during training.

Episode Duration

The episode duration starts from 10 timeslots and increases by one timeslot every two episodes, up to 288 timeslots (a complete day cycle). Figure 8 shows the plot of the episode duration for each episode during training.

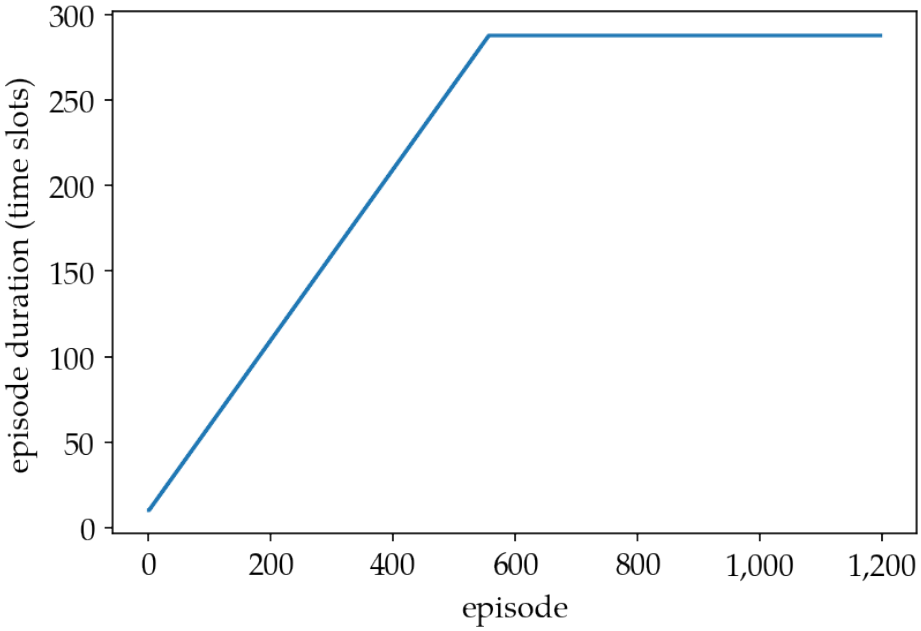

Figure 8. Plot of incremental episode duration during training.

5. Results

5.1. Training Results

Training is performed over 1200 episodes and is repeated five times to ensure consistency. For each episode, a day is selected randomly from the EV arrivals and electricity price datasets, which include 31 days in total. The training curves are presented in Figure 9, where the accumulated reward and the invalid actions per episode are plotted. An invalid

action refers to a selected charging rate e_t that was constrained to a higher charging level due to insufficient charge. The best training run is highlighted, and in Figure 10, the results from that run are averages over a moving window of 50 episodes.

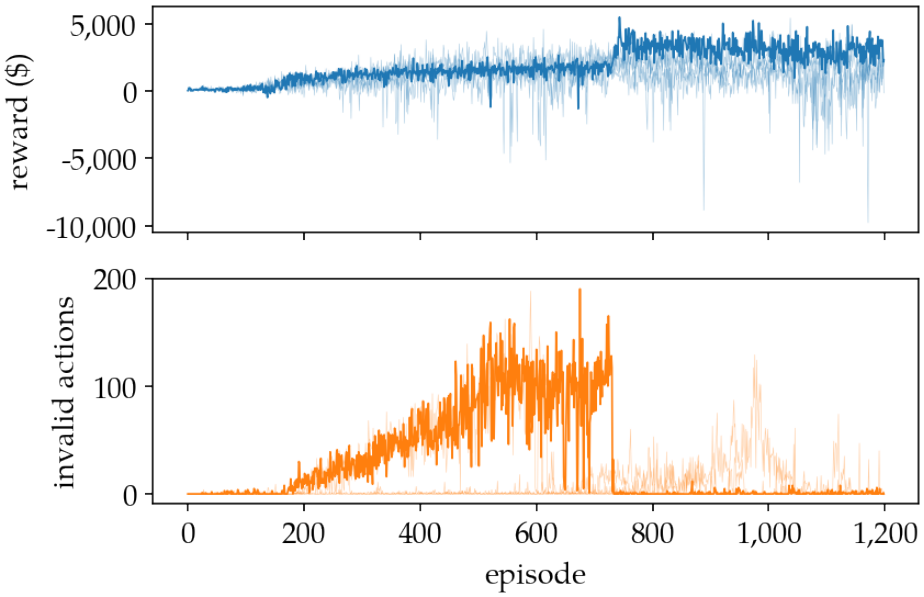

Figure 9. Training curves of the proposed model, with the best of five runs being highlighted.

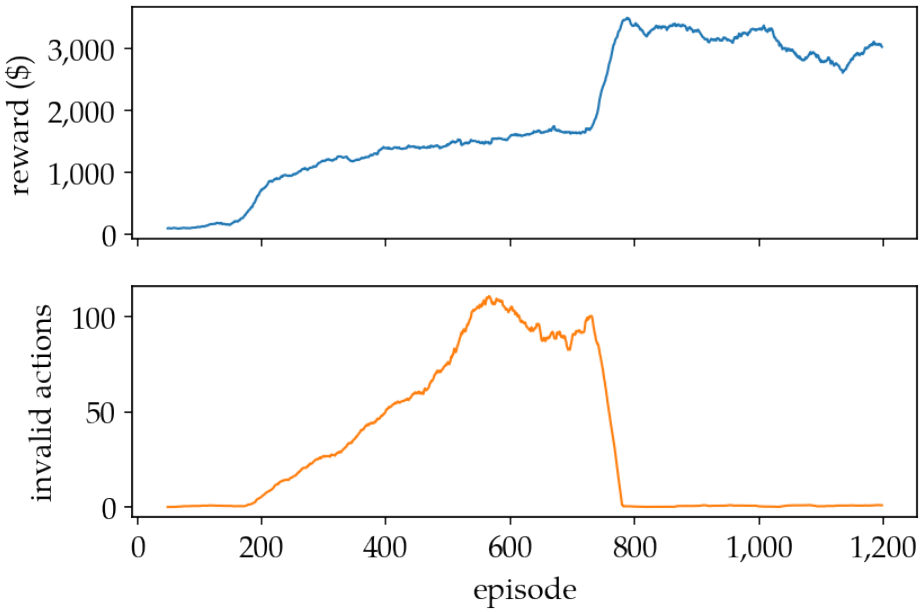

Figure 10. Training curves of the best run of the proposed model, averaged over a moving window of 50 episodes.

It can be observed that the model gradually achieves better total reward per episode, but it also increases invalid actions taken up to a certain point. During the algorithm's exploration phase, the agent is mostly choosing random actions and observes the rewards accumulated. Then, during the exploitation phase, it minimizes invalid actions and further increases reward. At that point, the agent mostly takes deterministic actions based on the values calculated for each observation–action pair and tries to find the sequence of actions that yields the best reward.

The maximum reward achieved over the five training runs was 5403 $. The mean value of the maximum reward per run was 4692 $, which indicates that training reaches a high accumulated reward consistently.

5.2. Policy Analysis

In this subsection, a trained model is examined for its policy during the day. The agent's actions are monitored in response to electricity price and residual demand. From Figure 11, it can be concluded that the optimal policy that the agent follows dictates keeping prices provided to customers at a constant level for most of the time slots. Furthermore, Figure 12 indicates that the EVs are charged at maximum charging rates most of the time, since an increasing total residual demand increases the charging rate. Hence, the optimal policy could be summarized as, "Keep the price stable at 3 $ and charge as much as possible".

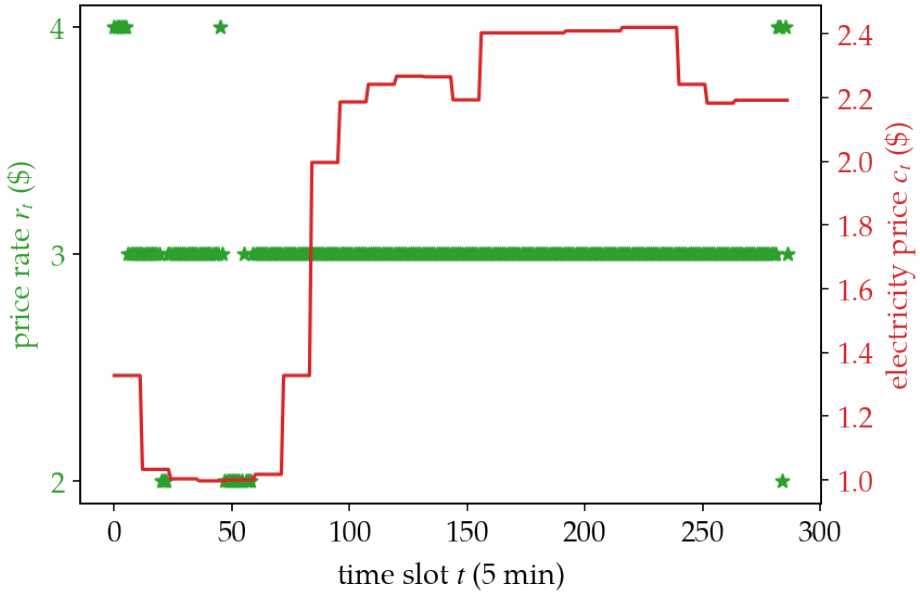

Figure 11. Price announced to customers vs. electricity price paid to the utility company.

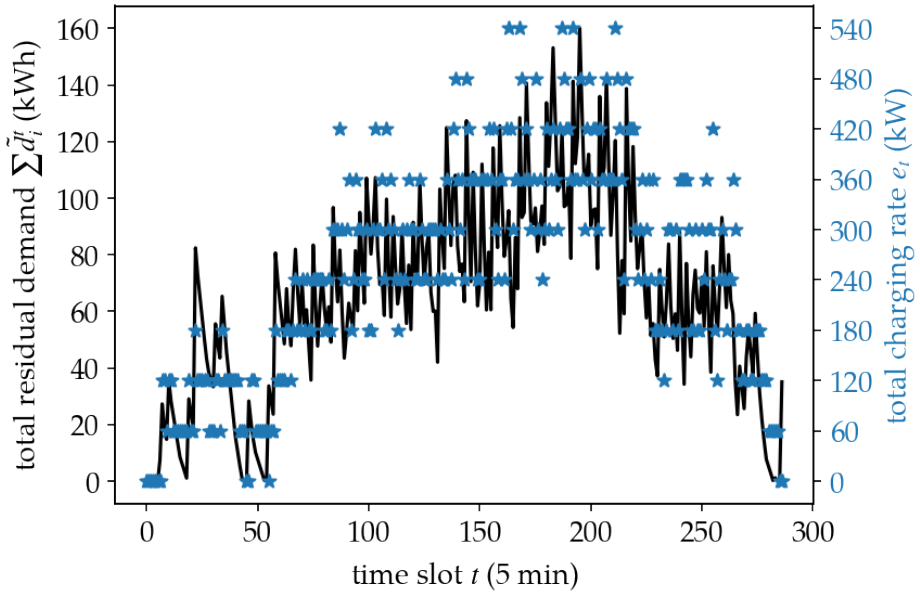

Figure 12. Total residual demand vs. total charging rate.

Figures 13 and 14 present an overview of the residual demand per charger. The agent keeps demands under a threshold and satisfies them as soon as possible. The very few flat lines show this, implying that an EV is not being charged for some time slots.

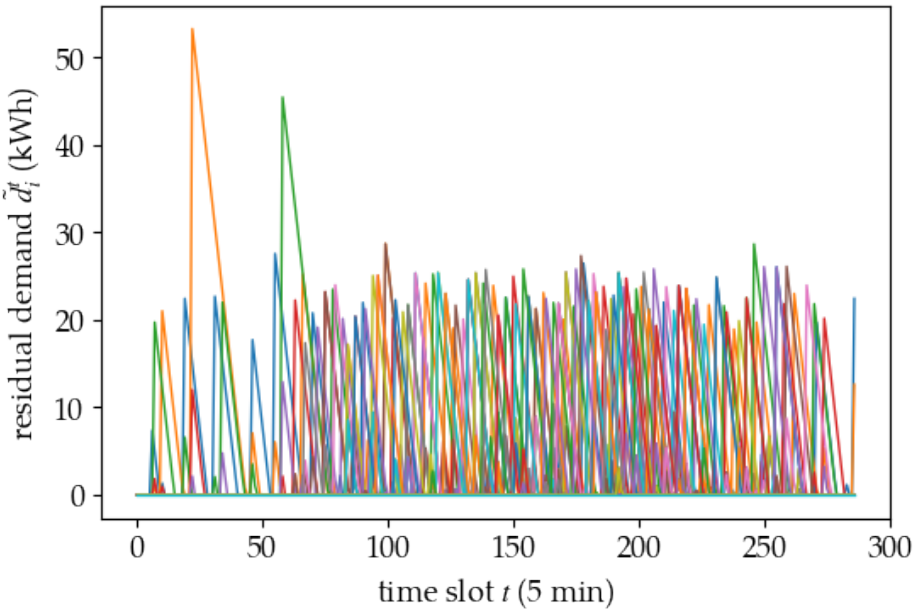

Figure 13. Residual demand per charger (20 in total).

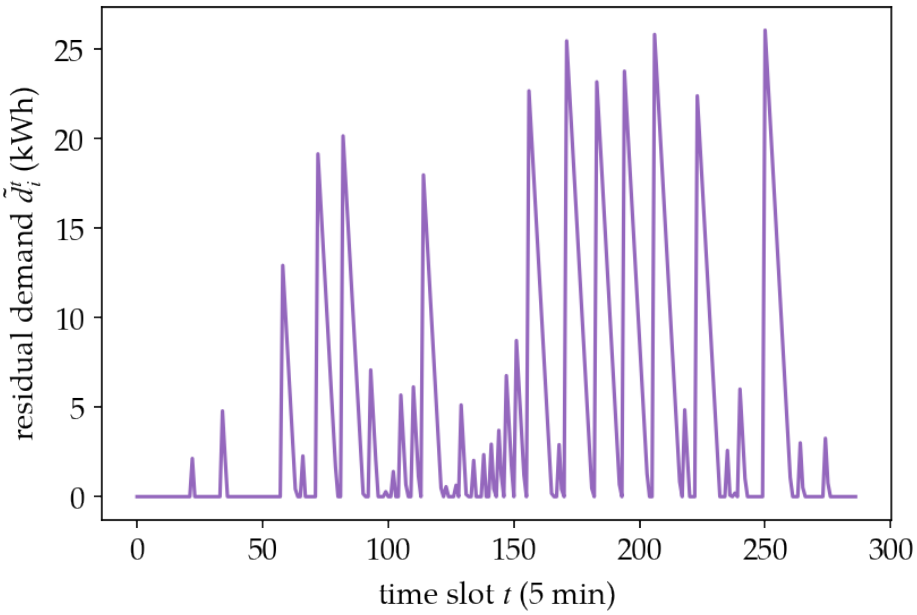

Figure 14. Residual demand for a single charger.

5.3. Case Study: Increasing Episode Time Horizon

It could be argued that, due to the smaller number of arrivals and lower electricity prices during night h, the agent should take actions that charge more conservatively during the end of the day to maximize profit. A three-day episode of training and testing was conducted with that in mind. The episode duration was incremented in the same manner as in the one-day maximum duration and is illustrated in Figure 15. Due to the higher number of episodes needed to reach maximum episode duration (i.e., 864 timeslots), the ϵ_{decay} from the ϵ-greedy policy was adjusted to 600 to compensate for the more extensive exploration

phase and enable the agent to choose actions while episode duration still increases randomly. The plot of the adjusted probability of random action can be observed in Figure 16.

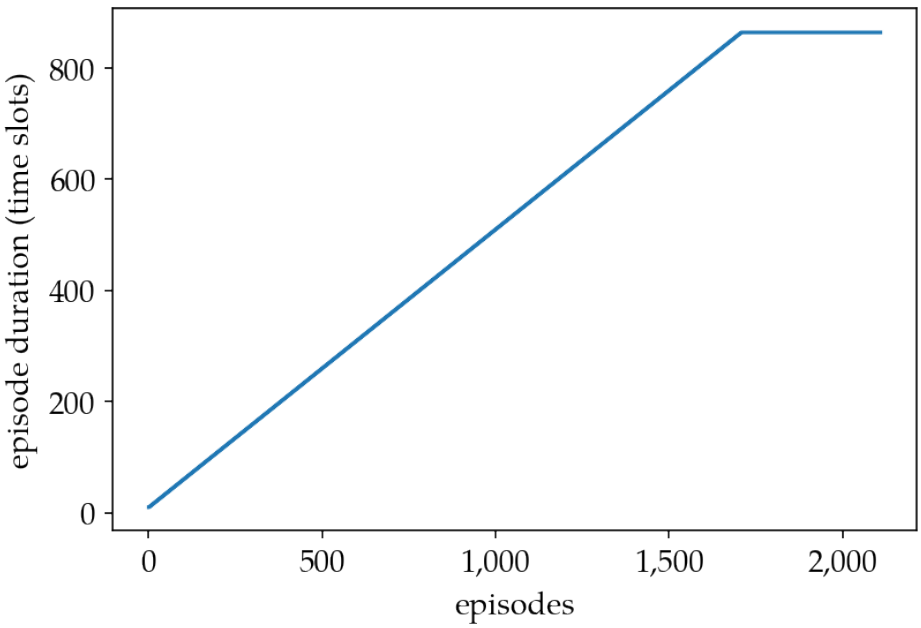

Figure 15. Plot of incremental episode duration during training (three-day episode duration).

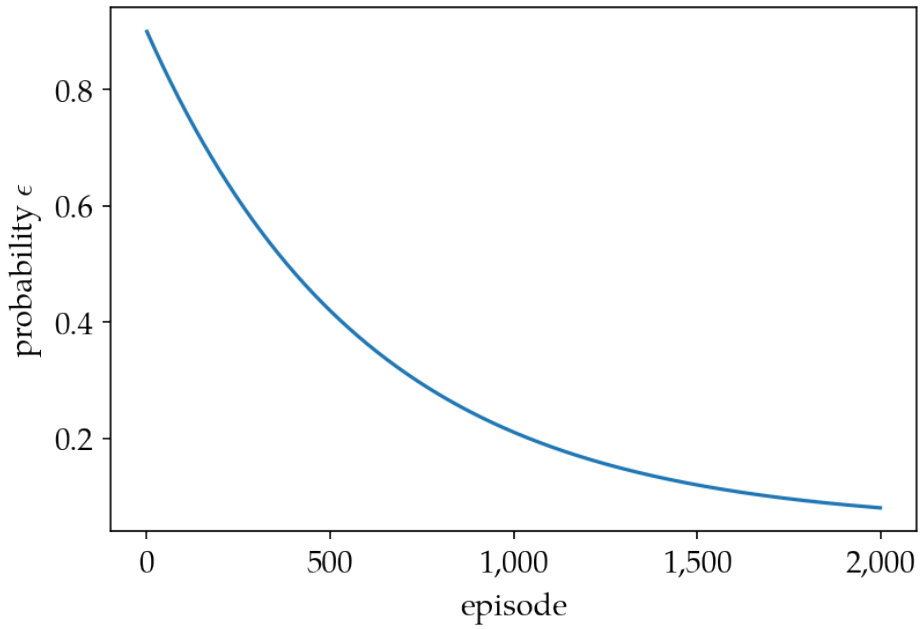

Figure 16. Plot of random probability for ϵ-greedy policy during training (three-day episode duration).

Figures 17–22 show the results in a similar manner as in the one-day experiment. Regarding the price, the agent seems to have a similar optimal policy, which is to keep it constant at \$3, according to Figure 19. There are also some \$2 actions when the price is dropping, suggesting that the agent attempts to receive extra energy demands to fulfill during low price time slots. On the other hand, the charging rates do not exceed 240 kW during peak demand times, as seen in Figure 20, contrary to the 540 kW maximum charging rate for one-day episode duration, as illustrated in Figure 12. This means that the agent adapts to the expanded episode duration and attempts to stall charging EVs when close to

a spike in electricity price. Another indication of that is evident in Figures 21 and 22, since flat lines can be observed for EVs with high demands during time slots 100 to 300.

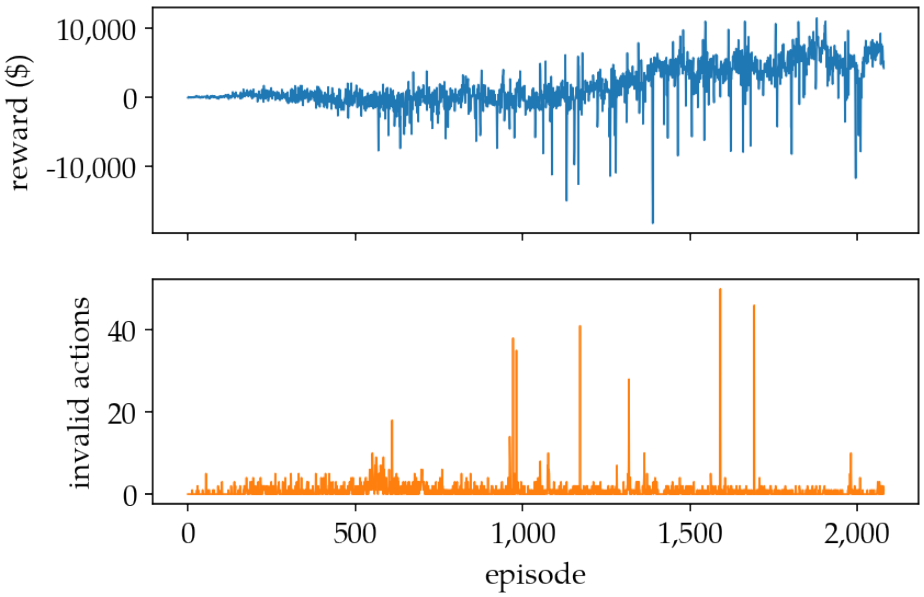

Figure 17. Training curves of the proposed model (three-day episode duration).

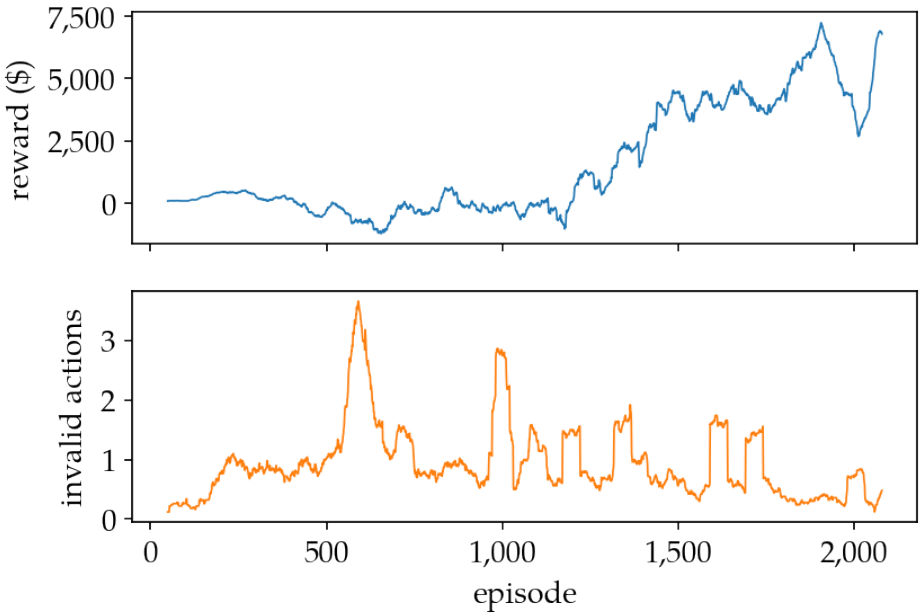

Figure 18. Training curves of the proposed model, averaged over a moving window of 50 episodes (three-day episode duration).

The behavior mentioned above negatively impacts the actual reward for the selected price parameters. The accumulated reward for three-day episodes is a little over double the accumulated reward for one-day ones, which can be deduced from observing Figure 17 in comparison to Figure 9 for the final episodes of training. However, it should be noted that one-day episodes may avoid charging costs at the end of each day, since not all EVs are charged when an episode ends. Three-day episodes include those costs to the accumulated reward for the two nights between the three days.

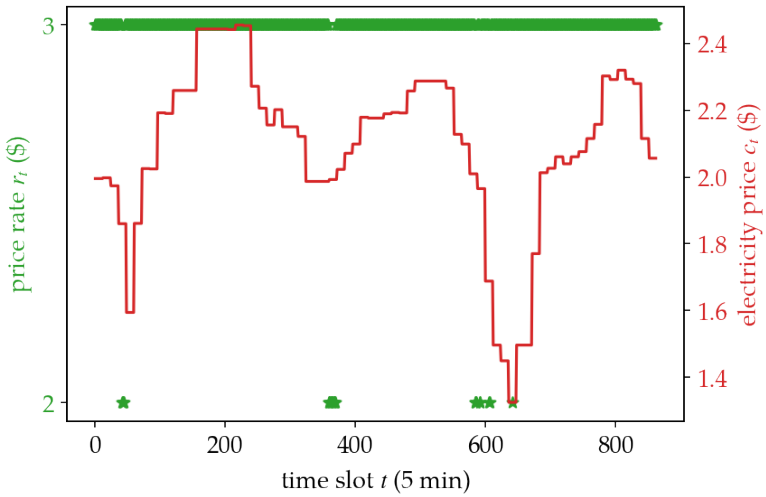

Figure 19. Price announced to customers vs. electricity price paid to utility company (three-day episode duration).

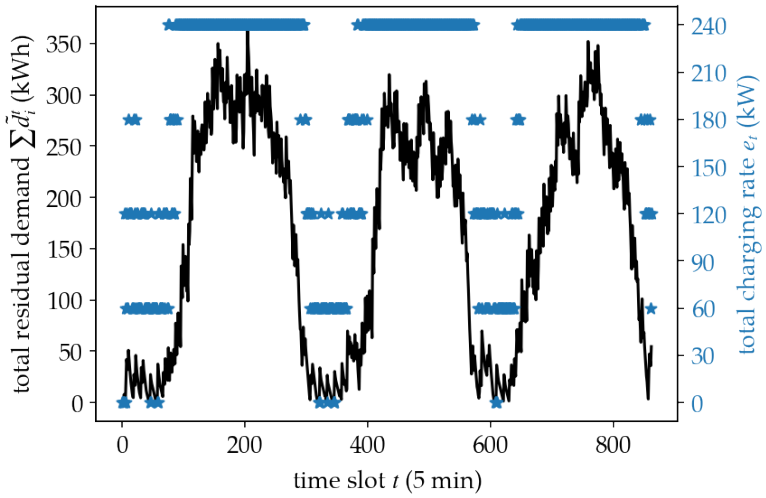

Figure 20. Total residual demand vs. total charging rate (three-day episode duration).

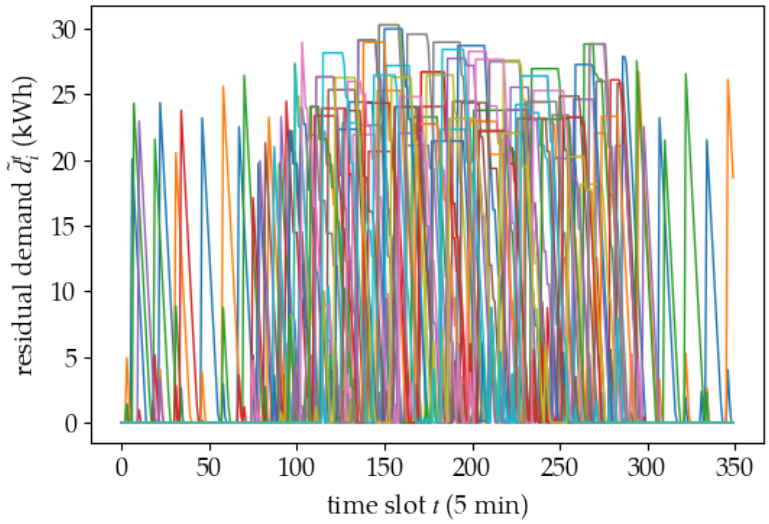

Figure 21. Residual demand per charger for the first 350 time slots (three-day episode duration).

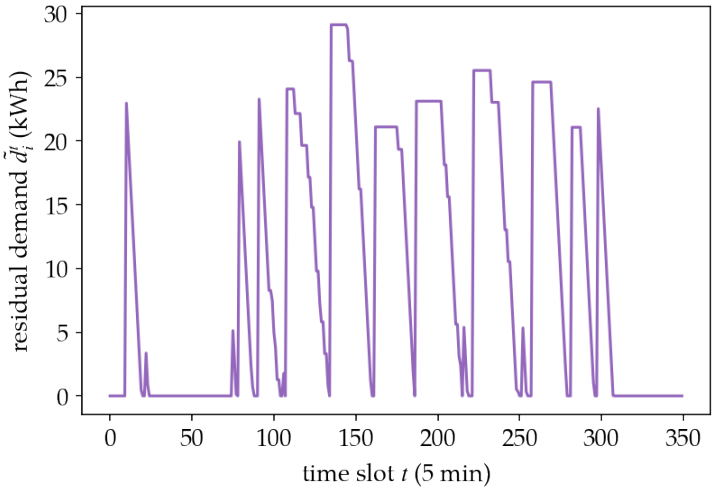

Figure 22. Residual demand for a single charger for the first 350 time slots (three-day episode duration).

5.4. Case Study: Removing Constraints

An experiment with no constraints was conducted to test the efficiency of the constraining mechanism and provide a way of comparing the proposed method with models of the respective literature, such as [22]. The method used for the experiment is similar to the proposed one, with the following key difference:

The LLF algorithm was used with no constraints. Specifically, EVs with the least laxity were charged with the maximum individual charging rate $x_{\max}$ until the total charging rate e_t selected by the agent was distributed. No further checks concerning negative laxities were performed, introducing the possibility of an EV reaching its departure time with unfulfilled demand. Whenever this occurred during an episode, the unfulfilled demand amount (kWh) was monitored, and the accumulated unfulfilled demand is presented at the end.

The training curves of the unconstrained model are presented in Figures 23 and 24. The unconstrained model achieved a maximum reward of 4044 \$; however, this reward was achieved with most EVs leaving the station with unfulfilled energy demands (a total of 1177 kWh). Furthermore, as the accumulated reward increased, the accumulated unfulfilled demand increased proportionately. This observation stems from the fact that the agent was unconstrained and had no measure of customer dissatisfaction included in its reward, leaving the single optimization objective of maximizing charging station profit. Thus, the agent learned to charge customers money for electricity that it never provided, as the total charging rate e_t that it selected for each time slot t was almost always zero. In contrast, the electricity price r_t was non-zero.

In [22], a constraining method is proposed and mathematically proven to provide a feasible solution that satisfies all demands. This method is non-trivial and to be implemented with an online agent. It requires future information about the EV arrivals at the station, which is not available. In the present context, the constraining mechanism ensures customer satisfaction while optimizing the charging station profit.

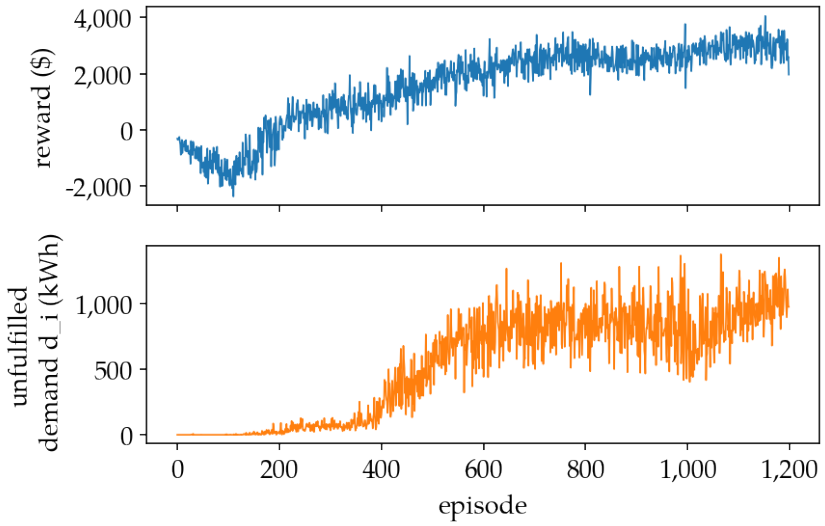

Figure 23. Training curves of the unconstrained model.

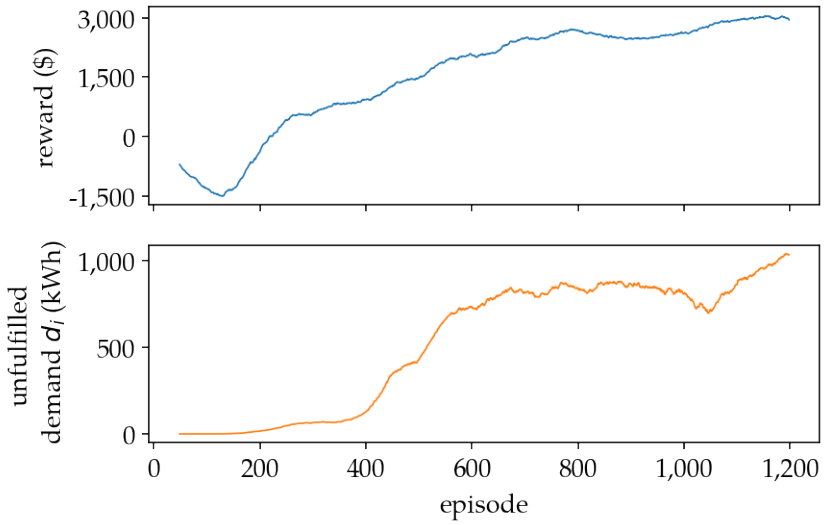

Figure 24. Training curves of the unconstrained model, averaged over a moving window of 50 episodes.

6. Conclusions

In this paper, a DRL-based approach was developed to solve optimal pricing and scheduling in an EV charging station under a dynamic, varying electricity price scheme. The proposed approach is model-free, which means that the DRL agent can operate under uncertainties, such as EV arrivals and their charging demands, and stated parking times, without explicit knowledge about the randomness. Instead, it can learn directly from underlying patterns present in real-world data. In addition, a charging scheduling algorithm was proposed, and the standard deep Q-learning algorithm was modified that ensures that EVs are adequately charged. Experimental results validated the effectiveness of the proposed solution in two ways: on the one hand, the trained agent managed to follow a policy that maximizes the profit of the charging station; at the same time, EV owners' charging demands were successfully fulfilled. Finally, it directly follows from the above analysis that the proposed system can make online decisions in real-time or near real-time by setting appropriate values for the duration of each slot.

The work presented in this study can be extended in many different directions. Some of them are listed below:

- As a first step, the technique of constraining the estimated charging rate could be incorporated into different DRL training algorithms that would operate on continuous action spaces, thereby lifting the need for discretizing scheduling and pricing actions.
- This work could serve as the basis for different formulations that consider more stakeholders, e.g., the grid operators and the corresponding constraints.
- Furthermore, the assumption was made that the total charging rate requested by the charging station is constrained only by the number of individual chargers. Consequently, potentially all parked EVs can be scheduled to charge during each slot; respecting additional constraints placed by the grid operator is an aspect that naturally arises as a potential future extension.
- In addition, more financial tools can be considered in modeling the relationship between different stakeholders.
- Finally, a more automated version of such a system can also be tailor-made for real-time EV detection, on a non-intrusive load monitoring (NILM) basis [31].

Author Contributions: Conceptualization, methodology, software, visualization, formal analysis, and writing—original draft preparation, D.A. and A.P.; validation, investigation, and writing—review and editing, D.A., A.P., A.C. and D.I.D.; resources, D.A., A.P. and A.C.; data curation, D.A. and A.P.; supervision, A.M. and D.I.D.; project administration, funding acquisition, A.M. and D.I.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research has been co–financed by the European Regional Development Fund of the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH–CREATE–INNOVATE (project: T2EDK-03898).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The publicly available datasets for EV arrivals and electricity prices have been used in this study.

Conflicts of Interest: The information and views set out in this article are those of the authors and do not necessarily reflect the official opinion of the European Commission.

Nomenclature

t	The time slot index
t_{len}	The length (duration) of each time slot
$\mathcal{I}_t$	The set of EVs that have arrived at the station at the beginning of time slot t
$\mathcal{J}_t$	The set of EVs that are already parked in the station before time slot t
$\mathcal{K}_t$	The set of EVs that require charging at time slot t
r_t	The price rate announced to the customers at time slot t
t_i^a	The arrival time of EV i
d_i	The charging demand of EV i
p_i	The maximum desired parking time of EV i
$D_i(\cdot)$	The demand–response function of EV i
β_1, β_2, σ	The parameters of the demand–response function
N	The total number of chargers in the station
$x_{i,t}$	The charging rate at which EV i will be charged during time slot t
$x_{\max}$	The maximum individual charging rate for every charger
e_t	The total charging rate at time slot t
e'_t	The constrained total charging rate at time slot t
$e_{\max}$	The maximum total charging rate for the charging station
α	The charging rate to energy conversion coefficient
c_t	The electricity price that the charging station pays to the utility company
$(S_t, A_t, R_{t+1}, S_{t+1})$	The 4-tuple of elements of the Markov decision process
γ	The discount rate

$\tilde{d}_i^t$	The residual charging demand for EV i at time slot t
$\tilde{p}_i^t$	The residual parking time for EV i at time slot t
$l_{i,t}$	The laxity of EV i at time slot t
ξ	The relaxation coefficient
$\mathcal{A}_t$	The set of all available actions
w_r	The set of discrete price rate levels
L	The number of discrete price rate levels
w_e	The set of discrete charging rate levels
K	The number of discrete charging rate levels
ϵ	The probability of a random action of the ϵ-greedy policy
$\epsilon_{\text{start}}, \epsilon_{\text{end}}, \epsilon_{\text{decay}}$	The parameters of the ϵ-greedy policy

References

1. Azam, A.; Rafiq, M.; Shafique, M.; Yuan, J. Towards Achieving Environmental Sustainability: The Role of Nuclear Energy, Renewable Energy, and ICT in the Top-Five Carbon Emitting Countries. *Front. Energy Res.* **2021**, *9*, 804706. [CrossRef]
2. Shafique, M.; Azam, A.; Rafiq, M.; Luo, X. Evaluating the Relationship between Freight Transport, Economic Prosperity, Urbanization, and CO_2 Emissions: Evidence from Hong Kong, Singapore, and South Korea. *Sustainability* **2020**, *12*, 664. [CrossRef]
3. Shafique, M.; Azam, A.; Rafiq, M.; Luo, X. Investigating the nexus among transport, economic growth and environmental degradation: Evidence from panel ARDL approach. *Transp. Policy* **2021**, *109*, 61–71. [CrossRef]
4. Shafique, M.; Luo, X. Environmental life cycle assessment of battery electric vehicles from the current and future energy mix perspective. *J. Environ. Manag.* **2022**, *303*, 114050. [CrossRef]
5. Yilmaz, M.; Krein, P.T. Review of the Impact of Vehicle-to-Grid Technologies on Distribution Systems and Utility Interfaces. *IEEE Trans. Power Electron.* **2013**, *28*, 5673–5689. [CrossRef]
6. Shafique, M.; Azam, A.; Rafiq, M.; Luo, X. Life cycle assessment of electric vehicles and internal combustion engine vehicles: A case study of Hong Kong. *Res. Transp. Econ.* **2021**, 101112. [CrossRef]
7. International Energy Agency. Global EV Outlook. In *Scaling-Up the Transition to Electric Mobility*; IEA: London, UK, 2019.
8. Statharas, S.; Moysoglou, Y.; Siskos, P.; Capros, P. Simulating the Evolution of Business Models for Electricity Recharging Infrastructure Development by 2030: A Case Study for Greece. *Energies* **2021**, *14*, 2345. [CrossRef]
9. Almaghrebi, A.; Aljuheshi, F.; Rafaie, M.; James, K.; Alahmad, M. Data-Driven Charging Demand Prediction at Public Charging Stations Using Supervised Machine Learning Regression Methods. *Energies* **2020**, *13*, 4231. [CrossRef]
10. Moghaddam, V.; Yazdani, A.; Wang, H.; Parlevliet, D.; Shahnia, F. An Online Reinforcement Learning Approach for Dynamic Pricing of Electric Vehicle Charging Stations. *IEEE Access* **2020**, *8* , 130305–130313. [CrossRef]
11. Ghotge, R.; Snow, Y.; Farahani, S.; Lukszo, Z.; van Wijk, A. Optimized Scheduling of EV Charging in Solar Parking Lots for Local Peak Reduction under EV Demand Uncertainty. *Energies* **2020**, *13*, 1275. [CrossRef]
12. He, Y.; Venkatesh, B.; Guan, L. Optimal Scheduling for Charging and Discharging of Electric Vehicles. *IEEE Trans. Smart Grid* **2012**, *3*, 1095–1105. [CrossRef]
13. Tang, W.; Zhang, Y.J. A Model Predictive Control Approach for Low-Complexity Electric Vehicle Charging Scheduling: Optimality and Scalability. *IEEE Trans. Power Syst.* **2017**, *32*, 1050–1063. [CrossRef]
14. Zhang, L.; Li, Y. Optimal Management for Parking-Lot Electric Vehicle Charging by Two-Stage Approximate Dynamic Programming. *IEEE Trans. Smart Grid* **2017**, *8*, 1722–1730. [CrossRef]
15. Bellman, R. Dynamic Programming. *Science* **1966**, *153*, 34–37. [CrossRef] [PubMed]
16. Sutton, R.S.; Barto, A.G. *Reinforcement Learning: An Introduction*, 2nd ed.; The MIT Press: Cambridge, MA, USA, 2018.
17. Mnih, V.; Kavukcuoglu, K.; Silver, D.; Graves, A.; Antonoglou, I.; Wierstra, D.; Riedmiller, M.A. Playing Atari with Deep Reinforcement Learning. *arXiv* **2013**, arXiv:1312.5602.
18. Abdullah, H.M.; Gastli, A.; Ben-Brahim, L. Reinforcement Learning Based EV Charging Management Systems—A Review. *IEEE Access* **2021**, *9*, 41506–41531. [CrossRef]
19. Lee, J.; Lee, E.; Kim, J. Electric Vehicle Charging and Discharging Algorithm Based on Reinforcement Learning with Data-Driven Approach in Dynamic Pricing Scheme. *Energies* **2020**, *13*, 1950. [CrossRef]
20. Zhang, F.; Yang, Q.; An, D. CDDPG: A Deep-Reinforcement-Learning-Based Approach for Electric Vehicle Charging Control. *IEEE Internet Things J.* **2021**, *8*, 3075–3087. [CrossRef]
21. Wan, Z.; Li, H.; He, H.; Prokhorov, D. Model-Free Real-Time EV Charging Scheduling Based on Deep Reinforcement Learning. *IEEE Trans. Smart Grid* **2019**, *10*, 5246–5257. [CrossRef]
22. Wang, S.; Bi, S.; Zhang, Y.A. Reinforcement Learning for Real-Time Pricing and Scheduling Control in EV Charging Stations. *IEEE Trans. Ind. Inform.* **2021**, *17*, 849–859. [CrossRef]
23. Chis, A.; Lunden, J.; Koivunen, V. Reinforcement Learning-Based Plug-in Electric Vehicle Charging with Forecasted Price. *IEEE Trans. Veh. Technol.* **2016**, *66*, 36740–3684. [CrossRef]
24. Lucas, A.; Barranco, R.; Refa, N. EV Idle Time Estimation on Charging Infrastructure, Comparing Supervised Machine Learning Regressions. *Energies* **2019**, *12*, 269. [CrossRef]

25. Deng, R.; Yang, Z.; Chow, M.Y.; Chen, J. A Survey on Demand Response in Smart Grids: Mathematical Models and Approaches. *IEEE Trans. Ind. Inform.* **2015**, *11*, 570–582. [CrossRef]
26. Watkins, C.J.C.H. Learning from Delayed Rewards. Ph.D. Thesis, King's College, Cambridge, UK, 1989.
27. Pazis, J.; Lagoudakis, M.G. Reinforcement learning in multidimensional continuous action spaces. In Proceedings of the 2011 IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning (ADPRL), Paris, France, 11–15 April 2011; pp. 97–104. [CrossRef]
28. Mnih, V.; Kavukcuoglu, K.; Silver, D.; Rusu, A.A.; Veness, J.; Bellemare, M.G.; Graves, A.; Riedmiller, M.A.; Fidjeland, A.; Ostrovski, G.; et al. Human-level control through deep reinforcement learning. *Nature* **2015**, *518*, 529–533. [CrossRef] [PubMed]
29. Exchange, K.P. System Marginal Price. Data Retrieved from Electric Power Statistics Information System. 2022. Available online: http://epsis.kpx.or.kr/epsisnew/selectEkmaSmpShdGrid.do?menuId=040202&locale=eng (accessed on 8 February 2022).
30. Al-Saadi, M.; Olmos, J.; Saez-de Ibarra, A.; Van Mierlo, J.; Berecibar, M. Fast Charging Impact on the Lithium-Ion Batteries' Lifetime and Cost-Effective Battery Sizing in Heavy-Duty Electric Vehicles Applications. *Energies* **2022**, *15*, 1278. [CrossRef]
31. Athanasiadis, C.L.; Papadopoulos, T.A.; Doukas, D.I. Real-time non-intrusive load monitoring: A light-weight and scalable approach. *Energy Build.* **2021**, *253*, 111523. [CrossRef]