

Article

Not peer-reviewed version

Smart Pricing for Smart Charging: A Deep Reinforcement Learning Framework for Residential EV Infrastructure

[Christos Pergamalis](#) , Eleftherios Tsampasis , [Panagiotis K. Gkonis](#) ^{*} , Charalampos N. Ilias

Posted Date: 26 March 2026

doi: 10.20944/preprints202603.2101.v1

Keywords: electric vehicle charging; dynamic pricing; reinforcement learning; proximal policy optimization; residential infrastructure; multi-objective optimization



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

Smart Pricing for Smart Charging: A Deep Reinforcement Learning Framework for Residential EV Infrastructure

Christos Pergamalis ¹, Eleftherios Tsampasis ¹, Panagiotis K. Gkonis ^{2*} and Charalampos N. Ilias ¹

¹ Department of Digital Arts and Cinema, National and Kapodistrian University of Athens, Dirfies Messapies, 34400, Greece

² Department of Digital Industry Technologies, National and Kapodistrian University of Athens, Dirfies Messapies, 34400, Greece

* Correspondence: pgkonis@uoa.gr

Abstract

The growing use of electric vehicles in the residential building sector presents new challenges in the management of the charging infrastructure, especially in deciding how to best price the use of it to balance operator revenue and user satisfaction with grid stability. Traditional pricing methods like fixed pricing rates and time-of-use tariffs cannot accommodate the dynamic nature of charging demand, which fluctuates depending on temporal patterns and weather conditions as well as user behavior. This limitation means that the resources are not used optimally and the revenue opportunities are lost during periods of high demand. To overcome this issue, we propose a reinforcement learning framework for dynamic pricing for residential electric vehicle charging stations. The framework models the pricing problem as a Markov Decision Process and uses Proximal Policy Optimization to learn a policy for setting optimal prices of private and shared charging stations according to real-time conditions. The state representation is done using ten features such as temporal indicators, current loading on the grid, grid status, traffic volume, and weather data. A multi-objective reward function is an approach to balance four objectives - revenue maximization, station utilization, grid stability, and user satisfaction. The system is trained on actual charging data from a residential complex in Trondheim, Norway. 6878 charging sessions during a 13-month period are used for training. We compare the learned policy with three baseline technologies: fixed pricing, time-of-use pricing and rule-based pricing. Experimental results show that the proposed approach reaches an overall score of 0.569, which is 32.9% and 48.9% improvements in comparison to fixed pricing and time-of-use pricing, respectively. The learned policy is able to successfully adjust the prices based on different conditions and sustain a balanced performance for all the goals. The main contributions include a custom reinforcement learning environment for residential EV charging pricing, a multi-objective reward formulation, and empirical evidence that learned policies outperform traditional pricing approaches.

Keywords: electric vehicle charging; dynamic pricing; reinforcement learning; proximal policy optimization; residential infrastructure; multi-objective optimization

1. Introduction

The shift towards electric vehicles (EVs) is a fast-paced development in personal transportation that is supported by environmental concerns, government incentives and improvements in battery technology. Global EV sales have increased significantly in the last decade, and projections show that electric vehicles will make up a majority of new car sales by the year 2030 in many developed markets [1]. While this transition promises to have tremendous advantages in helping to mitigate greenhouse gas emissions and air pollution at the urban level, there are new challenges in the management of the

electrical grid and in the operation of the charging infrastructure. This need for smarter management becomes even more critical as battery electric mobility scales up, since infrastructure readiness, energy supply pathways, and system efficiency remain open challenges in comparison with alternative zero-emission technologies [2]. Residential buildings, especially apartment complexes where parking spaces are shared, are becoming key places for the deployment of the EV charging infrastructure. In these environments, building managers and charging operators have the challenge of setting the price of charging services in a manner that ensures sufficient revenue, grid stability and resident satisfaction with the service. The pricing problem is complicated by the variable nature of charging demand, which varies according to the time of day, the weather and the individual user's schedule.

Dynamic pricing has been identified as a viable tool for managing demand for electricity in different applications. Unlike fixed pricing schemes that charge different rates under the same conditions, dynamic pricing varies rates according to factors such as current demand levels, capacity constraints of the grid and day time of day. The electricity industry has also long adopted time-of-use (ToU) electricity tariffs that charge higher rates in peak hours to encourage time shift of demand from peak to off-peak [3]. More sophisticated approaches involve real-time pricing, where rates fluctuate continually dependent on wholesale electricity market conditions, and critical peak pricing where much higher rates are charged during times of grid stress [4]. The application of dynamic pricing to the RI of EV charging is an especially relevant topic as charging sessions may be flexible in terms of timing. Unlike electricity demand that is immediately required like lighting or appliances, often a few hours delay or advance of EV charging may not have a significant effect to the user, as long as the vehicle is sufficiently charged before the next trip. The price elasticity of demand for EV charging has been estimated in a number of studies, with estimates generally being in the range of -0.2 to -0.5, showing that demand for charging is meaningfully responsive to changes in its price, but relatively inelastic compared to discretionary purchases [5].

Early studies on EV charging pricing paid attention to optimization-based methods that represented the pricing problem as a mathematical program that had explicit objective functions and constraints. These methods usually take known demand functions as given and solve for prices that maximize operator revenue or social welfare subject to the limits in the capacity of the grid. While optimization approaches provide theoretical insights and optimal solutions under idealized conditions these approaches require accurate models of user behavior which may be difficult to get in practice. Rule-based systems are another form of pricing systems that involve the application of predetermined logic to adjust prices according to observable conditions. For example, a rule-based system could raise the prices when the occupancy of the stations reaches a certain level or the load on the grid is close to reaching capacity limits [6]. The advantage of rule-based approaches is that they are interpretable and easy to implement, however their performance depends highly on how good the rules are, which have to be designed manually based on domain expertise. Statistical and machine learning techniques have also been adopted for the EV charging issues, but mostly focused on demand instead of price decisions. Regression models, neural networks, and ensemble methods have been employed to forecast the charging load at different time horizons, which can be used to develop pricing strategies [7]. The forecasting method divides the prediction and decision-making process, which simplifies each component, but may not be entirely comprehensive of the interaction between the prices and demand.

Reinforcement learning (RL) has become a promising area for sequential decision-making problem where an optimal action is determined with the current state and has an impact on future outcomes. In the setting of energy systems, RL has been adopted to the energy administration of buildings, battery storage function and demand reaction programs [8]. The great benefit of RL is its capability of learning proper policies through communication with the endeavor with no knowledge of an explicit model for system dynamics. Several recent studies have been done to explore RL for EV charging applications, but a large number of these studies have focused on scheduling the charging process, rather than pricing. Wang et al. [9] proposed a deep reinforcement learning solution for

collaborative charging schedule of multiple vehicles to reduce impact on the grid. Q-learning as a driver perspective-based method for optimizing the selection of charge station has been applied by Wan et al. [10]. In the area of dynamic pricing, RL has been applied to ride-sharing services, cloud computing resources and retail electricity markets [11,12]. These applications illustrate the possibilities of learning about pricing policies that can adapt to changing conditions and compromise multiple objectives. The combination of reinforcement learning and EV charging pricing is a relatively under-explored area with large potential to have a practical impact.

Despite the increasing body of research studies on the area of EV charging management, there are still a number of gaps in the ongoing literature. First, most pricing studies have been done on public charging stations or commercial fleets as opposed to residential settings that have very different characteristics such as predictable user populations, overnight charging opportunities, and integration with building electrical systems. Second, many approaches optimize for one single objective, such as revenue maximization or grid stability, without explicitly taking into account the multi-objective nature of the problem with user satisfaction being an important factor for long-term viability of the service. Third, currently existing RL applications to EV charging have focused mostly on scheduling and coordination problem and little attention has been paid to the actual pricing decision. Fourth, the dynamic pricing from a practical deployment point of view needs systems that are able to give the price recommendations on demand depending on current conditions instead of following some fixed schedules or that require long computation for optimization. These gaps are the drive for developing new approaches that meet the specific needs of residential EV charging price.

The work presented here fills these gaps by proposing a reinforcement learning-based dynamic pricing strategy for the charging stations of electric vehicles in homes. The pricing problem is formulated as a Markov Decision Process in the framework while Proximal Policy Optimization (PPO) algorithm is used to learn a pricing policy that balances revenue generation, station utilization, grid stability and user satisfaction. The system is designed to work on demand, taking information on the current state as input, and outputting recommended prices for private and shared charging stations. Training is done with actual charging data from a real residential apartment complex in Trondheim, Norway and a demand response model based on known price elasticity parameters. The learned policy is compared with traditional pricing approaches including the fixed pricing, time of use pricing, and rule-based pricing. Experimental results show that the proposed method has better performance on the aggregate score with all four objectives, which has over 30% improvement on fixed pricing baselines.

To this end, the key contributions of this work can be summarized as follows: (1) a customized reinforcement learning environment designed to the residential EV charging pricing problem with a multi-objective reward function; (2) state representation where temporal, demand, infrastructure, and environmental features are taken into consideration; (3) an empirical comparison that shows the benefits of learned pricing policies over traditional ones; and (4) a practical system design, which will allow real-time price recommendations based on the current conditions.

The remainder of this paper is organized as follows. Section 2 is the background of reinforcement learning and the PPO algorithm. Section 3 covers the dataset, problem formulation, proposed framework as well as evaluation methodology. The experimental results in terms of exploratory data analysis, learned policy behavior and comparison with baseline methods are presented in Section 4. Section 5 summarizes the paper with findings and indications for future research.

2. Materials and Methods

2.1. Dataset Description

The data used in this study comes from a publicly available data set on residential electric vehicle charging behavior collected from Trondheim, Norway. Distribution of the dataset was done by the research team of Sorensen et al. (2021) and available via the Kaggle platform with the name "Residential EV Charging Data". Data collection was done in apartment buildings that had both

private and shared charging stations between 21 December 2018 and 31 January 2020. The dataset consists of the 6878 individual charging sessions from multiple users from different garages across the residential complex. Each record of a session contains the beginning and ending time stamps, user ID, station type (private or shared) and energy used in kilowatt-hours. The temporal resolution of the aggregated load data is one hour enabling detailed analysis of the charging patterns throughout the day. Weather data for the same location and time has been obtained from a meteorological station, and contains daily data on temperature, precipitation, wind speed, and cloud cover. Traffic data from five measurement points located near the residential area give data about local vehicle movement patterns at hourly intervals. The combination of these data sources is allowing for a rich feature set that captures the various factors in influence of charging demand to be constructed.

The final dataset for unification is composed of 9,772 hourly records after merging and preprocessing all the data sources. Private charging stations make up about 79% of the combinational charging sessions recorded, with the shared charging stations representing the remaining 21% of the recorded. The average duration of the session lasts 9.2 hours with a standard deviation of 7.8 hours, which shows a large variation in user charging behavior. The average energy consumption per session is 14.3 kWh which corresponds to a partial charge of the majority of electric vehicles. From exploratory analysis, it can be noticed that there are distinct daily patterns with maximum charging activity between 16:00 and 20:00 on weekdays, during which residents return home from work. Weekend charging patterns have a much more spread nature without peaks. These characteristics make the dataset useful for the purpose of developing and testing dynamic pricing strategies that will be forced to adapt to varying demand conditions.

2.2. Problem Formulation

The dynamic pricing problem for residential EV charging is formulated as a Markov Decision Process (MDP) which offers a mathematical model for sequential decision making under uncertainty. At each time step, a pricing agent takes into account the current situation of the system and chooses prices for both the private and the shared charging stations. The environment then changes to a new state according to the user responses to the selected prices and the agent receives a reward signal, which measures multiple objectives, such as revenue generation, station utilization, grid stability, and user satisfaction. The task of the agent is to learn a pricing policy to maximize the expected cumulative reward over time. The MDP formulation is suitable for this problem because the demand for charging at any time of the day is not only dependent on the history of past states, but on the current circumstances, and thus satisfies the Markov property. The randomness of user behavior and external factors like weather conditions make the problem have uncertainty which the agent must learn by experience.

Formally, the pricing problem is defined as an MDP tuple:

$$M = (S, A, P, R, \gamma) \quad (1)$$

where S is the state space, A is the action space, $P(s_{t+1} | s_t, a_t)$ is the state transition probability, $R(s_t, a_t)$ is the reward function, and γ in $[0,1]$ is the discount factor.

At each discrete time step t , the agent observes a state s_t in S , selects an action a_t in A , receives a reward r_t , and the system transitions to the next state s_{t+1} . In the present problem, the state includes current temporal, demand, infrastructure, and environmental information, while the action corresponds to the choice of charging prices for private and shared stations:

$$a_t = (p_t^{priv}, p_t^{shared}) \quad (2)$$

The Markov property implies that the next state depends only on the current state and action, and not on the full past history.

$$P(s_{t+1} | s_t, a_t, s_{t-1}, a_{t-1}, \dots, s_0, a_0) = P(s_{t+1} | s_t, a_t) \quad (3)$$

The reward received after taking action a_t in state s_t is $R(s_t, a_t)$

$$r_t = R(s_t, a_t) \quad (4)$$

and is designed to capture the multiple operational objectives of the system. The agent seeks to learn a policy $p(a|s)$ that maximizes the expected discounted return

$$G_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k} \quad (5)$$

so that the optimal policy satisfies

$$\pi^* = \arg \max_{\pi} \mathbb{E}_{\pi} \left[\sum_{t=0}^{\infty} \gamma^t r_t \right] \quad (6)$$

Under this formulation, the dynamic pricing problem becomes a sequential optimization problem in which prices must be selected not only for immediate benefit but also for their long-term effect on future demand, utilization, and grid conditions.

2.3. Proposed Reinforcement Learning Framework

We propose a RL framework for the problem of real-time dynamic pricing of residential EV charging stations. The framework has three main components: the environment model, the pricing agent and the demand response simulator. The environment model encapsulates the state model and transition model of the charging system. The pricing agent is an implementation of a deep neural network policy with the pricing agent mapping the states to the actions of pricing. The demand response simulator predicts the user's change in charging behavior as a result of price changes using predetermined parameters from the literature to determine price elasticity. The combining of these components allows the agent to train pricing through simulation without the need for real-world experimentation which could have an adverse impact on real users. The framework is designed to work on demand, which means it can give price recommendations at any time when queried with current system conditions.

Formally, the pricing agent learns a stochastic policy $\pi_{\theta}(a|s)$ parameterized by neural network parameters θ , which maps a system state s to a probability distribution over pricing actions a . The objective of reinforcement learning is to maximize the expected discounted return

$$J(\theta) = \mathbb{E}_{\pi_{\theta}} \left[\sum_{t=0}^{\infty} \gamma^t r_t \right] \quad (7)$$

where r_t is the reward obtained at time step t and $\gamma \in [0,1]$ is the discount factor that determines the relative importance of future rewards.

Policy optimization is performed using gradient-based methods that update the policy parameters in the direction of increasing expected return. A deep neural network is adopted in this research to produce the probability distribution of discrete pricing actions as the policy. The Proximal Policy Optimization (PPO) algorithm is used to optimize the parameters of the network by increasing the stability of training through limiting the size of policy changes.

2.3.1. State Space Design

The state space is a representation of the information available to the pricing agent at each of the decision points. We design a ten-dimensional state vector in terms of temporal, demand, infrastructure, and environmental features. The temporal part is the hour of the day as well as the day of week, both represented in sine and cosine functions in order to maintain their cycle phase. The hour encoding is based on $\sin(2\pi h/24)$ and $\cos(2\pi h/24)$ where h = the hour of the day and the day of week encoding is based on $\sin(2\pi d/7)$ and $\cos(2\pi d/7)$ where d = the day index. Demand component is comprised of the existing charging load at the private stations and the existing charging load at the shared stations normalized using z-score standardization based on the statistics of training data. The infrastructure part captures the current grid load that is read by smart meters installed in the garage. The environmental component comprises the traffic volume of the adjacent points of measurement and the ambient temperature, which is normalized in the same way. The state representation is completed by a binary indicator for weekend days.

Table 1. State Space Components.

Feature	Description	Encoding
Hour of day	Current hour (0-23)	Sine/cosine transformation
Day of week	Current day (0-6)	Sine/cosine transformation
Private load	Current charging load at private stations (kW)	Z-score normalization
Shared load	Current charging load at shared stations (kW)	Z-score normalization
Grid load	Total grid consumption from smart meter (kW)	Z-score normalization
Traffic volume	Vehicle count from nearby sensors	Z-score normalization
Temperature	Ambient temperature (°C)	Z-score normalization
Weekend indicator	Binary flag for Saturday/Sunday	Binary (0 or 1)

2.3.2. Action Space Design

The action space describes the space of pricing decisions that the agent can make. We use a discrete action space that represents a combination of the price level of private and shared charging stations. The price range for private stations is set between 0.15 and 0.50 euro per kW/h and the price range for shared stations is set between 0.20 and 0.60 euro per kW/h. These ranges here are based on normal values for the Norwegian electricity market with corresponding margins for infrastructure and service costs. Each range is subdivided into ten discrete levels for a total possible number of actions (10 prices (private) times 10 prices (shared)) of 100 actions. The discrete formulation makes the learning problem more tractable than continuous action spaces while still providing enough granularity for making practical decisions in pricier areas. The steps or increments between each level of price are around 0.039 EUR/kWh for private stations and 0.044 EUR/kWh for shared stations: Small enough to enable fine-grained price adjustments. Following this approach, the agent learns how to choose the suitable price combination according to the state that is observed.

2.3.3. Reward Function Design

The reward function is a measure of the quality of pricing decisions and provides an indication for the learning process towards desirable behavior. We design a multi-objective reward function balancing four important aspects of charging stations operation. The first component is revenue which is calculated as the sum of price multiplied by adjusted demand for both types of stations. The second component is station utilization as measured by actual charging load divided by maximum capacity. The third part is grid stability which is defined as a penalty that is increased when the total load is approaching the limit set by the grid capacity. The fourth component is user satisfaction, assumed in the form of a declining function of price versus acceptable price range. The full reward function is a weighted sum of these components with parameters of 0.25 for each one, which puts equal importance to all of the objectives. The mathematical formulation is expressed as $R = 0.25 \times R_{\text{revenue}} + 0.25 \times R_{\text{utilization}} + 0.25 \times R_{\text{grid}} + 0.25 \times R_{\text{satisfaction}}$, where each component is normalized to the range [0, 1] before combination. The equal weighting scheme produces a balanced policy that does not excessively favor any single objective at the expense of others.

2.3.4. Demand Response Model

Since we know that there is no price information in the original dataset, we use a demand response model to simulate user reaction to different price levels. The model is based on the concept of price elasticity of demand, which is a measure of the percentage change in quantity demanded in response to some percentage change in price. Following the literature on EV charging behavior, we choose values of elasticity of -0.35 for private station users and -0.45 for shared station users [5]. The higher degree of negative elasticity for shared stations reflects the higher price sensitivity of users who have other charging options. The formula for demand adjustment is $D_{\text{adjusted}} = D_{\text{base}} \times (P /$

P_{base}^{ϵ} , where D_{base} is the base demand from the dataset, P is the selected price, P_{base} is the reference price (i.e., midpoint of the price range) and ϵ is the coefficient of elasticity. The sign of negative elasticity makes sure that demand is reduced with increase in price and vice-versa. Gaussian noise with standard deviation of 5% is added to adjusted demand to model stochastic changes of user behavior. The demand model allows the agent to learn about the price and charging behavior relationship by simulation, which is widely used case when there is no historical price information.

2.4. Learning Algorithm

We are using Proximal Policy Optimization (PPO) as the reinforcement learning algorithm for training the pricing agent. PPO is a class of policy gradient algorithms and has shown good performance in a large set of continuous control and discrete decision problems [14]. The algorithm maintains a neural network policy which is a direct output of action probabilities given an input state. During training PPO looks ahead to collect batches of experience by running the current policy in the environment and then updating the policy parameters to maximize expected returns while being constrained in how large policy updates are. The constraint is incorporated via a clipped surrogate objective, which avoids making too large changes to the policy at a time: The clipping parameter is set to 0.2, which implies that the ratio of probability between new and old policies has values in the range [0.8-1.2]. The policy network is two hidden layers of 64 neurons each and has hyperbolic tangent as the activation function. The value function network is of the same architecture and is used to measure state values that are used to estimate advantage values.

The length of the training process is 100,000-time steps, each of which corresponds to an hour in the simulated environment. The agent collects 2048 steps of experience before each policy update, with collected data being processed in mini-batches of 64 samples. The batches of experience are each used for 10 epochs of gradient descent updates. The learning rate is set at 3×10^{-4} and the discount factor γ is set to 0.99, indicating a strong preference for long-term rewards. The parameter, Generalized Advantage Estimation (GAE) parameter, λ was set to 0.95. An entropy bonus coefficient of 0.01 helps to explore during training by penalizing too deterministic policies. The training episodes are of fixed length of 168 steps or one week of operation. At the beginning of each episode the first state is chosen at random from the possible data points to introduce the agent to various starting conditions. Evaluation is done every 5000 training stages using 10 test episodes and deterministic action selection. The model that delivers the best performance on evaluation is saved to be used later for deployment.

2.5. Baseline Methods

To evaluate the performance of the proposed method, three baseline pricing strategies commonly used in practice are implemented. The first baseline is Fixed Pricing, where a constant tariff is applied throughout the evaluation horizon, equal to the midpoint of the allowable range for each station type (0.325 EUR/kWh for private and 0.40 EUR/kWh for shared). The second baseline is Time-of-Use (ToU) pricing, which applies higher tariffs during peak hours (07:00-09:00 and 16:00-20:00) and lower tariffs during off-peak periods. Peak prices are set to 85% of the maximum allowable value, while off-peak prices are set 10% above the minimum allowable tariff in order to remain within feasible operational bounds. The third baseline is Rule-Based Pricing, where tariffs are adjusted according to predefined thresholds of the observed load. When the private station load exceeds the 75th percentile of historical values, the price is increased to 90% of the maximum allowed level. When the load falls below the 25th percentile, the price is set close to the lower bound, specifically at 1.10 times the minimum permitted price, avoiding unrealistic negative or zero tariffs. An additional rule introduces a weekend discount when high-load conditions are detected. All baseline strategies are evaluated within the same environment and demand-response model as the PPO agent to ensure a fair comparison.

2.6. Evaluation Metrics

The performance of each pricing method is evaluated with the help of 4 metrics corresponding to the part of the reward function. Revenue is measured as the total revenue from charging services over one week (evaluation period) as a sum of the product of prices and adjusted demands over all hourly time steps. Station utilization is the ratio of the actual charging load divided by the maximum recorded load and is an indicator of how well the existing infrastructure is being utilized. Grid stability is measured by the mean value of stability component, and the higher the value, the farther the machine is operated from the grid capacity limit. User satisfaction is an average value of satisfaction component, which are prices remaining in the lower range of acceptable range. In addition to these individual metrics, we also calculate an Overall Score as a weighted combination of all four of these metrics and equal weights. The Overall Score is one summary measure of balanced performance across various objectives. Each of the metrics is normalized in order to enable comparison and aggregation. The number is re-evaluated for 20 independent episodes with different random starting points and we report the mean and standard deviation for each metric.

2.7. Implementation Details

The proposed framework is implemented on Python using Stable-Baselines3 library for reinforcement learning algorithms. The custom environment conforms to the Gymnasium interface specification that has compatibility with standard RL tools and allows easy integration with different learning algorithms. Data preprocessing and analysis is done using the Pandas and NumPy library. Visualization of results are made by Matplotlib and Seaborn libraries with customization for publication quality figures. All experiments are done on a desktop computer with an Intel Core processor and 16 GB of RAM. The training process for 100,000-time steps takes about 15 minutes. The trained model is stored in a compressed form which can be used for inference without any waiting for the original training data. The inference time of one price prediction is less than 10 milliseconds, which is acceptable for real-time applications. The source code and trained models are made available in a public repository in order to finish reproducibility of the results reported.

3. Results and Discussion

3.1. Exploratory Data Analysis

Before proceeding with developing the pricing model, we performed an exploratory analysis of the charging data, to understand the patterns and relationships underlying the charging data. The analysis was concentrated on charging patterns over time and in the course of a session, as well as correlation with outside factors. These results were used to design the state space and reward function for the reinforcement learning agent. The following subsections present the main observations from each of the aspects of the exploratory analysis.

3.1.1. Temporal Charging Patterns

The heatmap visualization in Figure 1 reveals association of certain charging patterns for different hours of the day and days of the week. Private charging stations show a pronounced peak between 16:00 and 22:00 on weekdays with the interpretivity at a maximum on Monday through Thursday evening. The average load during these peak hours reaches about 17.5 kW as against less than 2.5 kW during early morning hours. Weekend patterns are also very different to weekdays with a more distributed charging profile, without one predominant peak. Saturday and Sunday have moderate amounts of charging activity throughout the afternoon and evening hours. In comparison, shared charging stations exhibit much lower load levels at all time periods. The maximum mean load at communal stations hardly exceeds 5 kW and the temporal profile is much less pronounced than at private stations. The difference in patterns between types of stations reflects the different behaviors of use by residents who have dedicated private chargers and those which obtain a charge from shared infrastructure. Private station users seem to have a

predictable routine of plugging in their vehicles when they come home from their jobs, whereas shared station users seem to have more variable charging schedules.

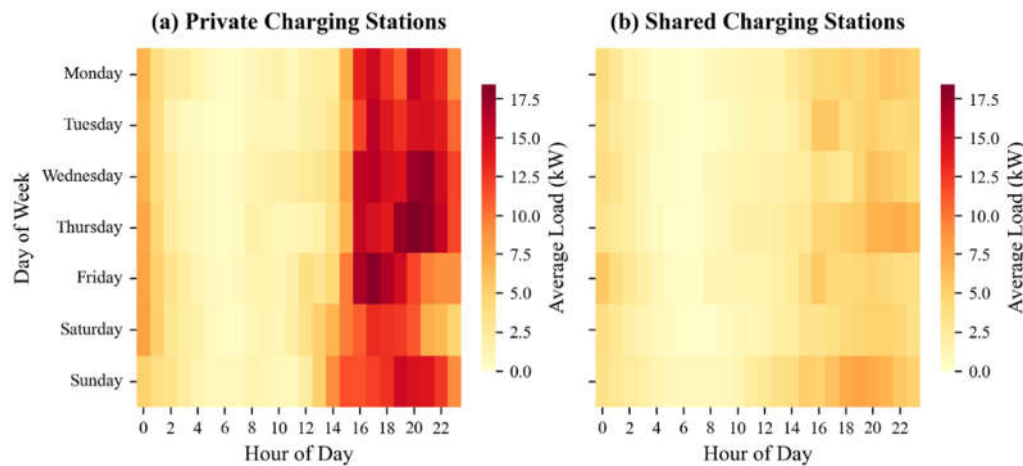


Figure 1. Heatmap of average charging load (kW) by hour of day and day of week for (a) private charging stations and (b) shared charging stations. The color intensity represents the average load, with darker colors indicating higher demand.

A more detailed view of the hourly charging patterns averaged over all days in the data is shown in the daily load profile of Figure 2. Private stations exhibit a very distinct trough between 05:00 and 08:00 hours when the average load is less than 1 kW with a gradual rise throughout the day. The load accelerates very quickly after 14:00 and peaks around 17:00 with an average of around 15 kW. The shaded area around the curve is the 95% confidence interval and the variability is relatively low during the off-peak hours and relatively high during peak hours. Shared stations have a similar daily profile but to a much lower extent, with peak loads of around 6kW, occurring slightly later on in the evening around 20:00. The evening peak period from 16:00 to 20:00 is particularly relevant for dynamic pricing since it is the time period in which the demand management interventions can have the most impact on both the revenue generation and balancing the load on the grid. The observed patterns confirm the existence of predictable temporal structure which a learning-based pricing agent can exploit.

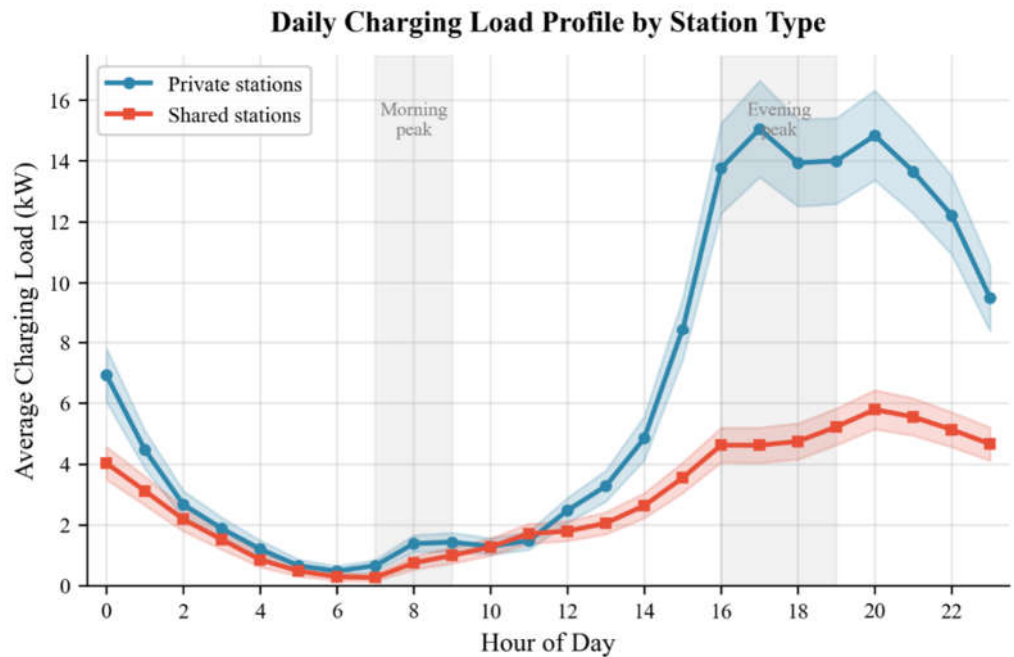


Figure 2. Daily charging load profile showing average hourly load for private and shared stations. The shaded regions represent the 95% confidence intervals. Gray vertical bands indicate morning (07:00-09:00) and evening (16:00-19:00) peak periods.

3.1.2. Session Characteristics

The distribution of the characteristics of charging sessions gives information about the user behavior and energy requirements. Figure 3(a) depicts the distribution of the session duration for private and shared stations. Private station sessions have a bimodal distribution with a peak at short durations ranging from 1 to 3 hours and a peak for longer durations ranging from 12 to 16 hours. The short duration peak is associated with rapid top-up charges and the longer duration peak is associated with overnight charging sessions (where users plug-in their vehicles in the evening and unplug it the next morning). Shared station sessions exhibit a unimodal distribution focused at shorter durations with the bulk of sessions being shorter than 5 hours. The data set includes 5318 private sessions and 1399 shared sessions, meaning that private charging is about 79% of the total recorded sessions. The distribution of energy consumption in Figure 3(b) indicates that most sessions consume between 5 and 20 kWh regardless of the type of station. Private sessions have a peak of around 5 to 10 kWh and a long right tail up to 50 kWh and shared sessions have a similar distribution, albeit with higher concentration in the 10 to 20 kWh range. The average energy consumption per session is 14.3 kWh which corresponds to about 50 to 70 km of driving range for a typical electric vehicle. These characteristics imply that users are doing partial charges instead of full recharge of batteries in the vehicle, which is consistent with the convenience-oriented charging behavior observed in residential settings.

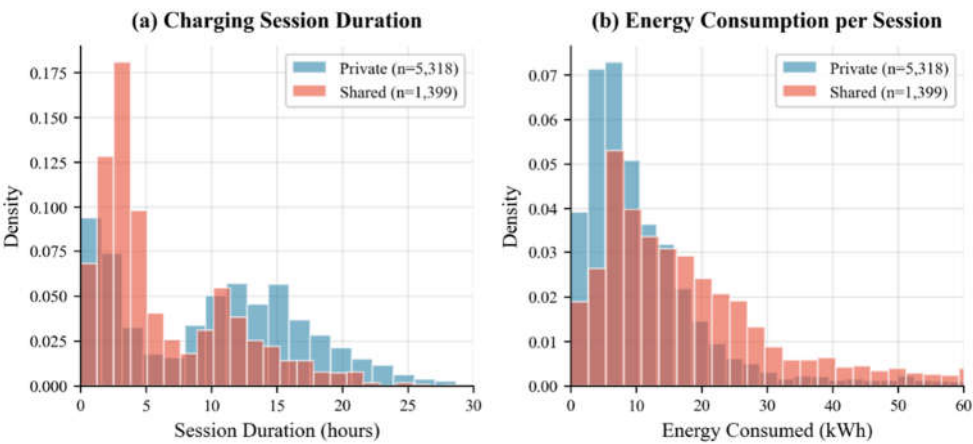


Figure 3. Distribution of charging session characteristics: (a) session duration in hours and (b) energy consumption per session in kWh. Blue bars represent private stations (n=5,318) and red bars represent shared stations (n=1,399).

3.1.3. Correlation with External Factors

Understanding the relationship between charging demand and external factors is important in order to design an effective state representation for the pricing agent. Figure 4 shows the correlation analysis between the charging load and the different environmental and contextual variables. Panel (a) presents the relationship between traffic volume and charging load, which reveals that there is a weak positive correlation ($\rho = 0.117$). Higher traffic volumes tend to coincide with slightly higher charging loads, which is not surprising as both variables are at their highest during the commuting hours when residents travel to and from work. Panel (b) shows the relationship between the temperature and the charging load, and there is a weak negative correlation ($\rho = -0.139$). Lower temperatures are linked to an increase in charging demand, probably because cold weather decreases battery efficiency and increases energy demand for cabin heating in electric vehicles. Panel (c)

compares the normalized daily profiles of the charging load and traffic volume. The traffic pattern is highest around 14:00 to 16:00, whereas the charging pattern is higher at a later time, i.e., 17:00 to 20:00. The temporal lag of about two to three hours between traffic and charging peaks is due to the time it takes the residents to travel home and plug in their vehicles. Panel (d) shows the correlation matrix of all the examined variables. The correlations between the charging load and individual external factors are relatively weak, and the absolute value is below 0.15 in all cases. The weak individual correlations mean that charging demand is affected by several factors at once as opposed to one variable driving the demand. The problem is multivariate in nature, so the use of a machine learning approach that can capture complex interactions among features is appropriate.

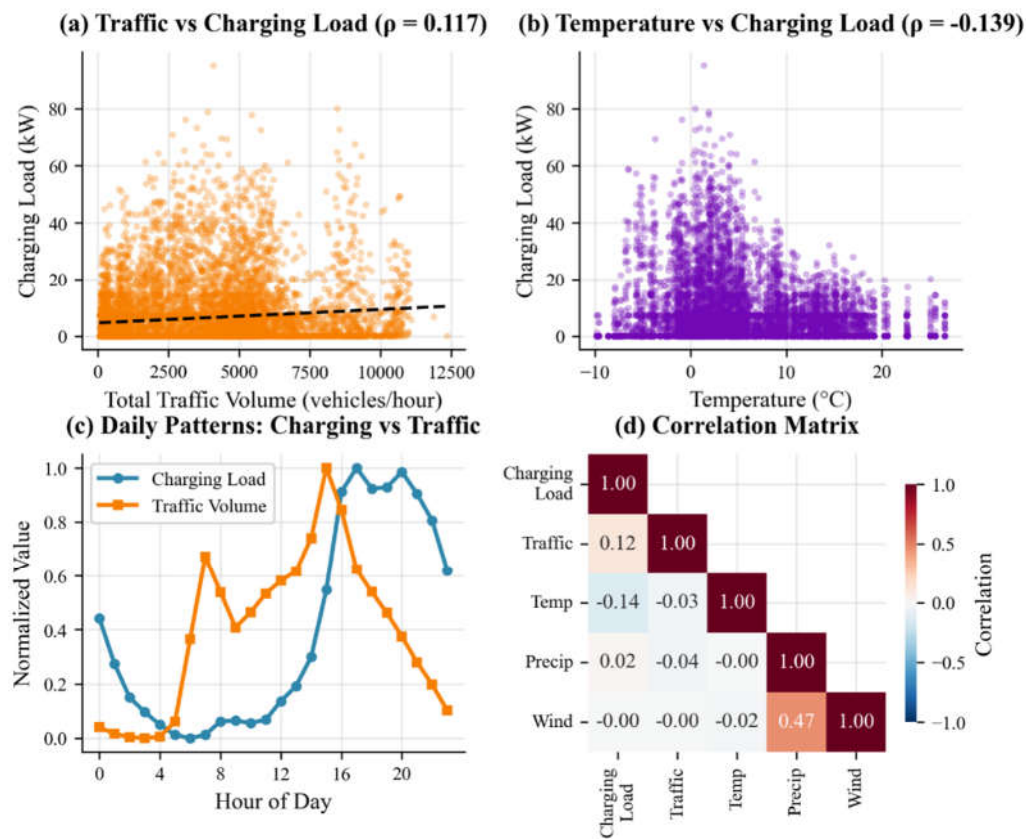


Figure 4. Correlation analysis between charging load and external factors: (a) scatter plot of traffic volume versus charging load, (b) scatter plot of temperature versus charging load, (c) comparison of normalized daily patterns for charging and traffic, and (d) correlation matrix showing pairwise Pearson correlation coefficients.

3.2. Learned Pricing Policy Behavior

After training the PPO agent for 100,000-time steps, we analyzed the behavior of the learned pricing policy over a representative week time. Figure 5 shows the dynamic change of prices based on the trained model as well as demand and revenue effects. Panel (a) illustrates the optimal prices set by the policy used for private and shared charging stations over 168 consecutive hours. The policy has specific pricing patterns which vary within and across days. Private station prices are in the range of about 0.18-0.50 EUR/kWh while shared station prices are in the range of about 0.22-0.60 EUR/kWh. The vertical dashed lines in the figure show the boundaries between days, and we can see that the policy has learned day-specific pricing strategies, as opposed to learning a uniform pattern for all days. During the times of anticipated high demand, the policy tends to charge higher rates for sharing stations with lower rates for private stations. During times of low demand, both prices fall to promote charging activity.

Panel (b) of Figure 5 shows the adjusted charging demand arising from the chosen prices. The demand patterns exhibit distinct peaks at times when the policy charges a lower price, thus demonstrating the responsiveness of the simulated user behavior to pricing signals. Private station demand peaks at around 6 kW during good pricing periods, whereas shared station demand peaks at up to 10 kW during some time intervals. The intermittent nature of the demand is the sum effect of the underlying charging patterns from the dataset and the price-induced adjustments from the demand response model. Panel (c) shows the hourly revenue obtained by the pricing policy. The revenue profile has peaks that represent periods of high demand coupled with not too low prices. The higher hourly revenues around 4 to 6 EUR are achieved in the evening hours because the demand for charging is naturally higher. The policy is effective in generating revenue through the week without suffering from periods of zero income. The behavior that is observed shows that the PPO agent has learned to optimize the trade-off between setting high prices that maximizes per-unit revenue and setting moderate prices that allows for an adequate demand volume.

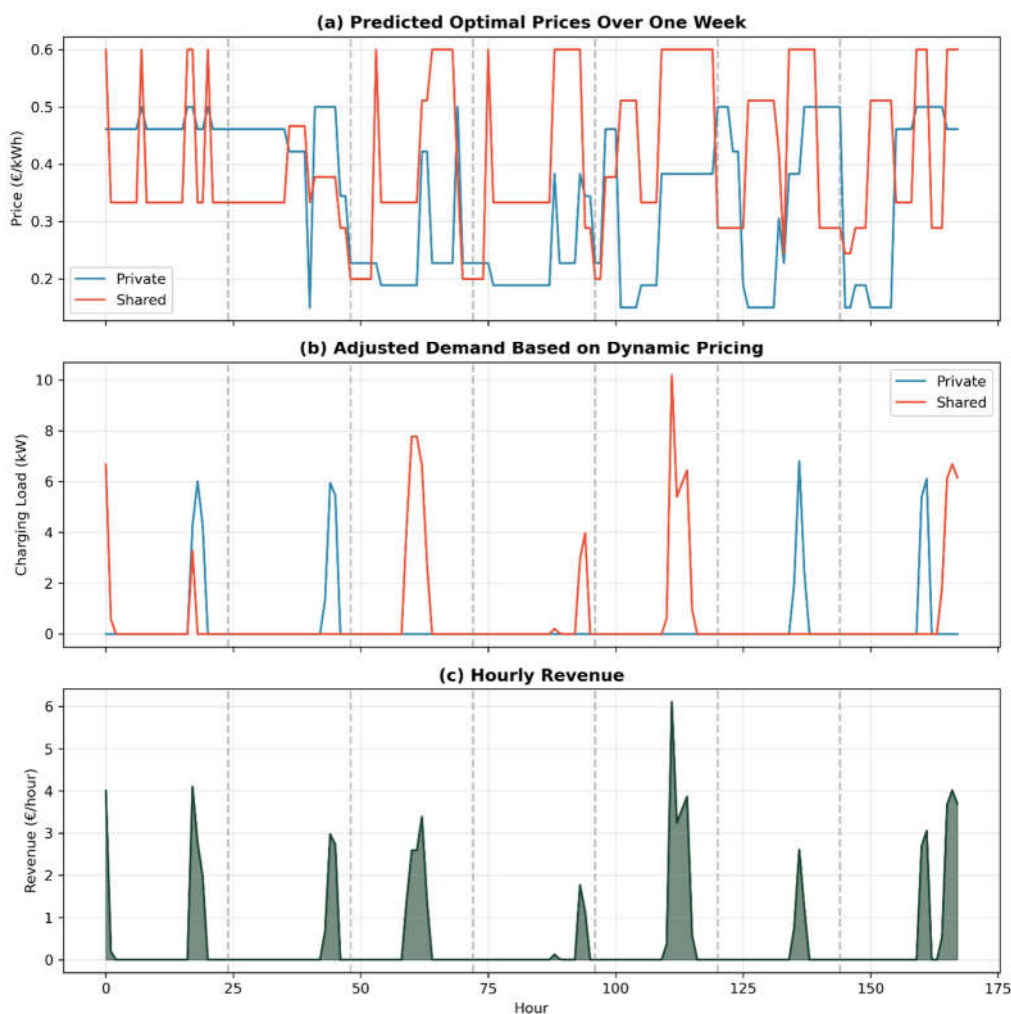


Figure 5. Behavior of the learned PPO pricing policy over one week (168 hours): (a) optimal prices for private and shared stations in EUR/kWh, (b) adjusted charging demand in kW resulting from the selected prices, and (c) hourly revenue in EUR. Vertical dashed lines indicate day boundaries.

3.3. Comparison of Pricing Methods

The proposed PPO-based pricing method was compared with three baseline approaches: Fixed Pricing, Time of Use (ToU) Pricing and Rule Based Pricing. Each method was tested through 20 independent episodes of one week duration and the results are summarized in Figure 6. There are four performance metrics taken into account for the evaluation which are weekly revenue, station

utilization, grid stability, and user satisfaction. An aggregate metric called the Overall Score is generated by adding these four components with equal weights in order to give a balanced measure of each method.

Figure 6(a) shows the weekly revenue obtained by each of the pricing methods. The best-revenue generating method proposed is the PPO method with the highest average revenue of about 680 euro per week, which is followed by Fixed Pricing with 500 euro per week, Rule-Based Pricing with 365 euro per week and ToU Pricing with 370 euro per week. The PPO method provides about 36% increase in revenue compared to the Fixed Pricing baseline. The reason for the higher revenue is the fact that the learned policy can detect periods and situations in which higher prices can be charged without a significant drop of demand. Figure 6(b) presents the results of the utilization rates of the stations, where the highest utilization rate of around 0.09 is obtained with the proposed method, meaning that on average 9% of the maximum observed capacity is being utilized. While this absolute value seems low, it is the nature of residential charging where the vehicles are present and charging for only a fraction of every day. The PPO method helps to improve utilization by about 50% over the ToU baseline, which has the lowest utilization due to its static peak pricing, which discourages charging during high demand periods.

Figure 6(c) shows grid stability scores of each method. ToU Pricing has the highest grid stability score of around 0.77 closely followed by Fixed Pricing with 0.74. The grid stability score of the proposed PPO method is 0.72, slightly less than that of the traditional methods but a reasonable range. The slightly lower grid stability is a side-effect of the higher utilization gained from PPO because more charging activity naturally puts a higher load on the infrastructure of the electrical grid. Rule Based pricing demonstrates widest variable grid stability performance with large Standard Deviation, reflecting its threshold-based logic that can lead to inconsistent behavior in response to specific conditions of the load encountered. Figure 6(d) shows the user satisfaction scores based on the pricing levels of each method. ToU Pricing has the highest level of satisfaction at around 0.69 since it charges at a low rate during off-peak hours when the majority of its charging takes place. Fixed Pricing and the proposed PPO method have similar satisfaction scores of about 0.50, whereas Rule-Based Pricing has the lowest satisfaction score of around 0.46. The similar levels of satisfaction under Fixed and PPO pricing suggest that the policy that the learner has learned to make does not derive its revenue benefits from excessive increases in prices but rather from clever timing of price increases.

Figure 6(e) shows the decomposition of the Overall Score in its four component contributions for each method. The stacked bar visualization indicates that the PPO method gets large contributions from all four contributors, with especially good performance on revenue and utilization. The Fixed Pricing method shows balanced but moderate contributions in all the components. The ToU method has significant contributions from grid stability and satisfaction and low contributions from revenue and utilization. Rule-Based Pricing has the most not so uniform distribution with high variation in its contributions. Figure 6(f) shows the aggregate Overall Score that incorporates all the four metrics with equal weight. The proposed PPO method is the best performed with the Overall Score of 0.569, which is better than Fixed Pricing (0.428), ToU Pricing (0.382), and Rule-Based Pricing (0.336). The improvement of 32.9% compared to Fixed Pricing and 48.9% compared to ToU Pricing shows that reinforcement learning approach is able to learn a pricing policy that balances multiple competing objectives. The results prove that the proposed method is able to simultaneously increase revenue generation without excessive levels of grid stability and user satisfaction.

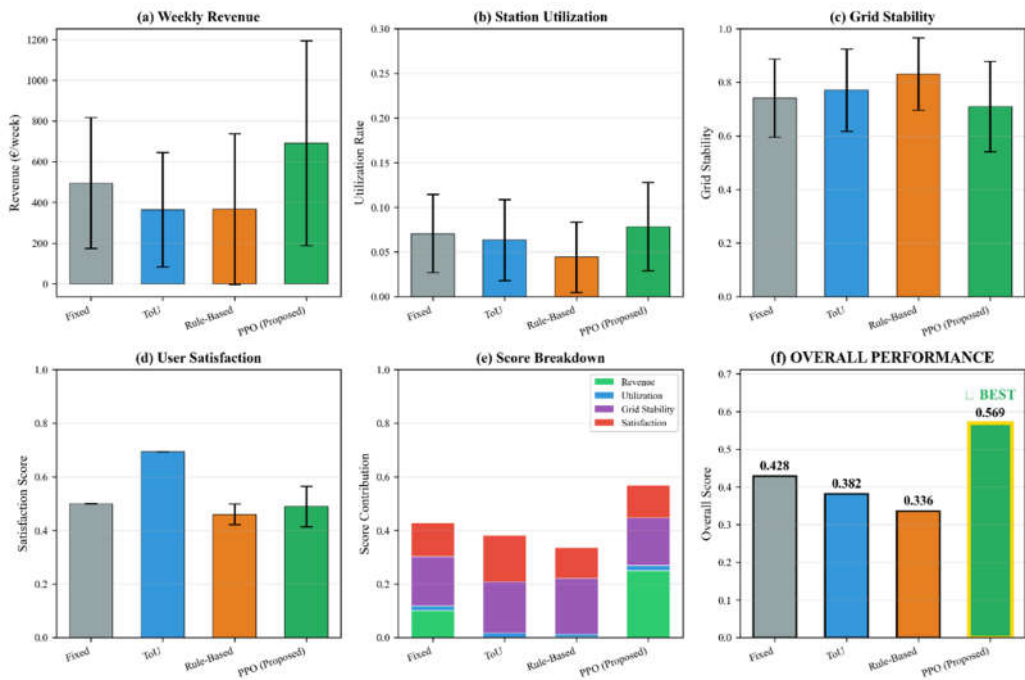


Figure 6. Comparison of pricing methods across multiple performance metrics: (a) weekly revenue in EUR, (b) station utilization rate, (c) grid stability score, (d) user satisfaction score, (e) breakdown of Overall Score components, and (f) aggregate Overall Score. Error bars represent one standard deviation over 20 evaluation episodes. The proposed PPO method achieves the highest Overall Score of 0.569.

3.4. Discussion

The experimental results show that reinforcement learning is an effective algorithm for solving the dynamic pricing problem for residential EV charging. The proposed PPO-based method is better than traditional pricing methods on the aggregate performance measure and has competitive or better results on individual measures. The success of the learned policy may be attributed to a number of factors. First, the ability to observe system state continuously enables the agent to make pricing decisions based on the current state of the system and not predetermined schedules or rules. Second, the process of trial-and-error learning allows the agent to learn the effective pricing strategies that would be hard to design manually. Third, the multi-objective reward function steers the agent to balanced policies that are not paid-up in one objective to the other.

The comparison with baseline methods reveals interesting trade-offs between different philosophies of pricing. Fixed Pricing offers dependable and predictable results and cannot adjust to changing demand situations, thus leaving potential revenue unharvested in high demand periods. ToU Pricing is successful in achieving good user satisfaction and grid stability, by providing incentives for off-peak charging, but as a rigid scheduling in the latter, it does not consider variations in day-to-day demand patterns, nor does it consider external factors. Rule-Based Pricing attempts to react to the current situation and is based on thresholds manually specified not necessarily generalizing well to different situations as can be seen by its high-performance variability. The learned PPO policy adjusts its behavior according to the observed state which allows more nuanced responses to the complex and variable nature of charging demand.

Several limitations of this study should be noted. The demand response model applied in our experiments is based on assumed values of price elasticity from the literature rather than actual observed responses of prices from actual users. Real user behavior may deviate from simulated response and there would be the need for field validation prior to practical deployment. The dataset comes from a single residential complex in Norway, and the policy learned might need retraining or fine-tuning for it to be used in other geographical or demographic contexts. The discrete action space containing ten levels of prices per type of station offers some reasonable granularity but may not

capture optimal prices that fall between the predetermined prices. The continuous action spaces or finer discretization can be considered in future work to overcome this drawback. The balanced preference of the objectives by equating the reward function is to say we have balanced these preferences but different operators may have different priorities that would require us to alter these weights accordingly.

4. Conclusions

The present work proposes an approach of reinforcement learning for dynamic pricing of electric vehicle charging station in residential buildings. The proposed framework addresses the need for intelligent pricing systems, which can adapt to changing conditions, taking into account multiple stakeholder objectives at the same time. Unlike traditional methods that use fixed schedules or manually designed rules, the reinforcement learning agent uses the environment to learn effective pricing strategies by discovering patterns and relationships that would be difficult to encode explicitly. The framework is constructed around a custom environment, that models the residential charging context with appropriate state and action spaces, is accompanied by a reward function that reflects the multi-faceted nature of the pricing problem.

We applied the proposed framework with the Proximal Policy Optimization and trained the pricing agent using real data from apartment buildings in Trondheim, Norway. The evaluation compared the learned policy with fixed pricing, time of use pricing and rule-based pricing using four performance metrics. The obtained results reveal that the PPO-based approach achieves the best aggregate performance with an Overall Score of 0.569, which is superior to all of the baseline approaches. The learned policy raises roughly 36% more revenue than the fixed pricing without much difference in the user satisfaction. The analysis of policy behavior shows that the agent is trained to dynamically adjust prices on the basis of temporal patterns and demand conditions, increasing prices during the peak periods, and reducing prices when demand is anticipated to be low. The comparison with the baseline methods confirms the advantages offered by the adaptive nature of the learned policy, as opposed to static or rule-based methods which cannot adapt to the complex and variable patterns that exist in the residential charging data.

This work is an important step towards intelligent management of a residential EV charging infrastructure. As electric vehicle adoption continues to increase, the need for effective pricing mechanisms will become increasingly important for both infrastructure operators and grid managers. The reinforcement learning approach shown in this study provides a means to getting to pricing systems that are capable of learning and adapting without programmed rules for decision-making. Future research directions involve validation against actual user responses to dynamic prices, extension to larger multi-building cases, coupling with renewable energy availability signals, and examining fairness issues in pricing algorithms. The framework that has developed in this work offers a basis for these extensions, and contributes to the larger objective of establishing the basis for sustainable and efficient electric mobility in residential environments.

Author Contributions: All authors have equally contributed to the final version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this study are publicly available on the Kaggle platform.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

EV	Electric Vehicle
RL	Reinforcement Learning
ToU	Time -of- Use
PPO	Proximal Policy Optimization
MDP	Markov Decision Process

References

1. Tiwari, S.; Jamwal, P.S.; Jaiswal, S. Electric vehicle global market survey. In *Development of Electric Vehicles in Smart Grid Concepts*; Elsevier: Amsterdam, The Netherlands, 2026; pp. 219-236.
2. Pergamalis, C.; Tsampasis, E.; Dedes, I.C.; Elias, C. Hydrogen fuel cell electrical vehicles (FCEV)-battery electric vehicles (BEV)-comparison and future challenges. In *Proceedings of the 2024 13th Mediterranean Conference on Embedded Computing (MECO)*, June 2024; pp. 1-5.
3. Faruqui, A.; Sergici, S. Household response to dynamic pricing of electricity: A survey of 15 experiments. *J. Regul. Econ.* 2010, 38, 193-225.
4. Borenstein, S. The long-run efficiency of real-time electricity pricing. *Energy J.* 2005, 26, 93-116.
5. Hardman, S.; Jenn, A.; Tal, G.; Axsen, J.; Beard, G.; Daina, N.; et al. A review of consumer preferences of and interactions with electric vehicle charging infrastructure. *Transp. Res. Part D Transp. Environ.* 2018, 62, 508-523.
6. Garcia-Villalobos, J.; Zamora, I.; San Martin, J.I.; Asensio, F.J.; Aperribay, V. Plug-in electric vehicles in electric distribution networks: A review of smart charging approaches. *Renew. Sustain. Energy Rev.* 2014, 38, 717-731.
7. Zhu, J.; Yang, Z.; Guo, Y.; Zhang, J.; Yang, H. Short-term load forecasting for electric vehicle charging stations based on deep learning approaches. *Appl. Sci.* 2019, 9, 1723.
8. Zhang, J.; Hou, L.; Zhang, B.; Yang, X.; Diao, X.; Jiang, L.; Qu, F. Optimal operation of energy storage system in photovoltaic-storage charging station based on intelligent reinforcement learning. *Energy Build.* 2023, 299, 113570.
9. Wang, S.; Bi, S.; Zhang, Y.A. Reinforcement learning for real-time pricing and scheduling control in EV charging stations. *IEEE Trans. Ind. Inform.* 2019, 17, 849-859.
10. Wan, Z.; Li, H.; He, H.; Prokhorov, D. Model-free real-time EV charging scheduling based on deep reinforcement learning. *IEEE Trans. Smart Grid* 2018, 10, 5246-5257.
11. Chirita, M.; Chirita, G. A comprehensive overview of deep learning for algorithmic pricing in ride-sharing platforms. *Econ. Appl. Inform.* 2024, 1, 177-181.
12. Li, H.; Wan, Z.; He, H. Constrained EV charging scheduling based on safe deep reinforcement learning. *IEEE Trans. Smart Grid* 2019, 11, 2427-2439.
13. Sorensen, A.L.; Lindberg, K.B.; Sartori, I.; Andresen, I. Residential electric vehicle charging datasets from apartment buildings. *Data Brief* 2021, 36, 107105.
14. Wang, Y.; He, H.; Tan, X. Truly proximal policy optimization. In *Proceedings of the Uncertainty in Artificial Intelligence*, August 2020; pp. 113-122.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.