



**HAL**  
open science

## Multi-Agent DDPG Based Electric Vehicles Charging Station Recommendation

Khalil Bachiri, Ali Yahyaouy, Hamid Gualous, Maria Malek, Younes Bennani, Philippe Makany, Nicoleta Rogovschi

► **To cite this version:**

Khalil Bachiri, Ali Yahyaouy, Hamid Gualous, Maria Malek, Younes Bennani, et al.. Multi-Agent DDPG Based Electric Vehicles Charging Station Recommendation. *Energies*, 2023, 16 (16), pp.6067. 10.3390/en16166067 . hal-04366014

**HAL Id: hal-04366014**

**<https://hal.science/hal-04366014v1>**

Submitted on 5 Jul 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.




L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Article

# Multi-Agent DDPG Based Electric Vehicles Charging Station Recommendation

Khalil Bachiri <sup>1,2,\*</sup>, Ali Yahyaouy <sup>2</sup>, Hamid Gualous <sup>3</sup>, Maria Malek <sup>1</sup>, Younes Bennani <sup>4</sup>, Philippe Makany <sup>3</sup> and Nicoleta Rogovschi <sup>5</sup>

<sup>1</sup> ETIS Laboratory, CNRS, ENSEA, CY TECH, CY Cergy Paris University, 95011 Cergy, France; maria.malek@cyu.fr

<sup>2</sup> LISAC Laboratory, Sidi Mohammed Ben Abdellah University, Fez 30000, Morocco; ali.yahyaouy@usmba.ac.ma

<sup>3</sup> LUSAC Laboratory, University of Caen Normandie, 14032 Caen, France; hamid.gualous@unicaen.fr (H.G.); philippe.makany@unicaen.fr (P.M.)

<sup>4</sup> LIPN Laboratory—CNRS UMR 7030, La Maison des Sciences Numériques, University of Sorbonne Paris Nord, 93000 Paris, France; younes@lipn.univ-paris13.fr

<sup>5</sup> LIPADE Laboratory, University of Paris Descartes, 75006 Paris, France; nicoleta.rogovschi@parisdescartes.fr

\* Correspondence: khalil.bachiri@usmba.ac.ma

**Abstract:** Electric vehicles (EVs) are a sustainable transportation solution with environmental benefits and energy efficiency. However, their popularity has raised challenges in locating appropriate charging stations, especially in cities with limited infrastructure and dynamic charging demands. To address this, we propose a multi-agent deep deterministic policy gradient (MADDPG) method for optimal EV charging station recommendations, considering real-time traffic conditions. Our approach aims to minimize total travel time in a stochastic environment for efficient smart transportation management. We adopt a centralized learning and decentralized execution strategy, treating each region of charging stations as an individual agent. Agents cooperate to recommend optimal charging stations based on various incentive functions and competitive contexts. The problem is modeled as a Markov game, suitable for analyzing multi-agent decisions in stochastic environments. Intelligent transportation systems provide us with traffic information, and each charging station feeds relevant data to the agents. Our MADDPG method is challenged with a substantial number of EV requests, enabling efficient handling of dynamic charging demands. Simulation experiments compare our method with DDPG and deterministic approaches, considering different distributions and EV numbers. The results highlight MADDPG's superiority, emphasizing its value for sustainable urban mobility and efficient EV charging station scheduling.

**Keywords:** EV charging station recommendation; reinforcement learning; deep learning; MADDPG; multi-region; smart city



**Citation:** Bachiri, K.; Yahyaouy, A.; Gualous, H.; Malek, M.; Bennani, Y.; Makany, P.; Rogovschi, N. Multi-Agent DDPG Based Electric Vehicles Charging Station Recommendation. *Energies* **2023**, *16*, 6067. <https://doi.org/10.3390/en16166067>

Academic Editor: Tek Tjing Lie

Received: 27 July 2023

Revised: 16 August 2023

Accepted: 17 August 2023

Published: 19 August 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Smart cities have emerged as a vision where citizens can securely access, manage, and exchange data concerning various aspects of their daily lives in a sustainable and pervasive manner [1,2]. With the increasing prominence of electric vehicles (EVs) in urban transportation and logistics [3,4], the transition towards greener mobility has become a global focus. In 2020, there were already 10 million EVs on the road worldwide, with over 2.61 million in China alone [5]. Projections indicate that the number of EVs will reach 40 million by 2030. The European Union's commitment to significantly reduce greenhouse gas emissions by 2050 has further accelerated the shift towards EVs in the automotive industry [6]. As EV adoption grows in smart cities, the establishment of robust public charging infrastructure becomes imperative [7,8].

However, EVs come with inherent limitations, including restricted battery range, long charging times, and potential waiting times for charging [9]. Charging an EV often takes

considerably longer than refueling a conventional vehicle, ranging from 30 minutes to several hours. Moreover, many EV owners lack the convenience of charging their vehicles at home, which is primarily due to the high power requirements of EVs. Scaling up EV charging infrastructure can pose challenges to the power distribution system, necessitating an increased reliance on charging stations.

The recommendation of electric vehicle charging stations (EVCS) in the context of future smart cities presents significant challenges. To optimize the charging experience for EV drivers and improve charging solutions, an intelligent system is needed. Current approaches, which tend to be deterministic [10–12], overlook the uncertainty of ever-changing traffic conditions, leading to suboptimal recommendations. Devising EVCS recommendations in uncertain future states is a complex problem. Renewable-integrated standalone charging stations provide sustainable EV charging. With growing EV adoption, integrating smart grid architecture becomes essential. Standardizing power distribution protocols is critical, and comprehensive proposals for EV charging via current infrastructure are now available [13–16].

Reinforcement learning (RL) [17] offers a viable solution to tackle intricate decision-making problems and can be applied for intelligent EVCS recommendations [18]. RL enables agents to learn from repeated trial and error to maximize long-term objectives. However, the large-scale setting involving millions of EVs and thousands of charging stations poses challenges for conventional Q-learning approaches. Additionally, most RL methods utilize single-agent reinforcement learning, which has limitations, whether centralized or decentralized.

To address these challenges, deep reinforcement learning (DRL) has gained prominence as it excels in complex sequential decision-making tasks, even surpassing human performance in various domains [17,19]. The deep deterministic policy gradient (DDPG) [20] is a DRL technique that has shown promise in addressing continuous action problems, outperforming Deep Q-Learning (DQL) in real-time control tasks. DDPG leverages policy gradients, iteratively updating the policy's parameter values to maximize the predicted cumulative return.

This paper aims to achieve optimal EV charging station recommendations by adopting the principles of multi-agent deep reinforcement learning (MADRL) [21]. We employ a framework based on Markov games and evaluate the learning process using all agents' joint actions [22–27]. To depict the optimal joint action, we apply the steady state concept from game theory. Specifically, we consider each region of EV charging stations in the city as an individual agent and formulate the EV charging recommendations as a multi-agent deep deterministic policy gradient (MADDPG) task, using a multi-agent actor-critic framework. Our approach emphasizes coordination and collaboration among all agents within the large-scale system.

The results of our experiments demonstrate that the proposed approach outperforms benchmark strategies, including single-agent and conventional methods, as well as state-of-the-art approaches. Furthermore, our approach effectively minimizes total travel time under dynamic traffic conditions, accommodating different EV distributions and numbers of EVs.

The following is a summary of this paper's significant contributions.

- A novel approach employing the multi-agent deep deterministic policy gradient (MADDPG) approach, specifically designed for optimal EV charging station recommendation in a city with multiple regions. By considering dynamic traffic conditions, our method concurrently minimizes the total travel time in a stochastic environment. This innovative approach addresses the challenges of recommending charging stations in a rapidly changing urban landscape.
- To facilitate efficient cooperation and communication among agents, a multi-agent system is devised, featuring a centralized learning model and decentralized execution. This distinctive amalgamation empowers agents to collaboratively make optimal EV

charging station recommendations, thereby augmenting the overall efficiency and performance of the system.

- By adopting a suitable framework for the study of multi-agent decision processes, we formulate the EV charging station recommendation problem as a Markov game. This modeling selection adeptly manages the uncertainties and dynamic nature of the environment, thereby enabling more precise and robust recommendations.
- Rigorous performance experiments are conducted to assess the effectiveness of our proposed methodology. By comparing our MADDPG against the DDPG algorithm and conventional deterministic methods, we demonstrate the superiority of our methodology. Additionally, we consider two different distributions of EVs to validate the robustness and generalizability of our approach under varying scenarios.

The remainder of this paper is structured as follows: Section 2 presents the research background in this field. In Section 3, the key materials and methods of the EV charging system architecture considered in this paper are introduced. The EV charging station recommendation problem is then formulated as a sequential decision-making challenge using Markov games in Section 4. Section 5 outlines our proposed multi-agent deep deterministic policy gradient algorithm for EVCS recommendation. The simulation results and analyses are presented in Section 6. Finally, the paper concludes in Section 7, summarizing our significant contributions.

## 2. Research Background

The charging and discharging of electric vehicles (EVs) present several challenges due to the uncertainties arising from dynamic and random charging requests. As a result, substantial research efforts have been devoted to designing EV charging recommendations and scheduling strategies to address congestion issues and optimize charging operations. Deep reinforcement learning has emerged as a promising approach to tackle these challenges.

In [28], the authors proposed a planning model based on stochastic simulation for EV charging stations (EVCSs) to reduce investment and operational expenses. Another study by Kim et al. [29] utilized the RL Q-learning algorithm and Markov decision process to model an energy management system for a smart building, incorporating PV generation, electrical storage, and a vehicle-to-grid station, leading to reduced overall energy costs. Zhang et al. [30] addressed congestion control in charging station allocation through the combination of communication technology using Q-learning. In a different context, Dabbaghjamanesh et al. [31] proposed a Q-learning method based on recurrent neural networks for estimating loads of plug-in hybrid electric vehicles under various conditions.

Chics et al. [32] considered the scheduling complexity of EV charging and discharging and estimated the value-action function using a Q-value table by discretizing power prices and EV charging activities, aiming to lower long-term electricity costs. Wan et al. [33] formulated the electric vehicle charging scheduling problem using a Markov decision process (MDP) based on deep reinforcement learning (DRL), engaging with the environment to teach the ideal policy and handling the randomness in an EV's arrival and start charge time. However, their approach did not account for traffic congestion and other realistic restrictions. These directions center on electric vehicles, encompassing emission and traffic simulation models. Electric vehicles have distinct traffic traits due to their exclusive torque and acceleration attributes [34]. This necessitates precise calibration of traffic simulation models to suit these vehicles.

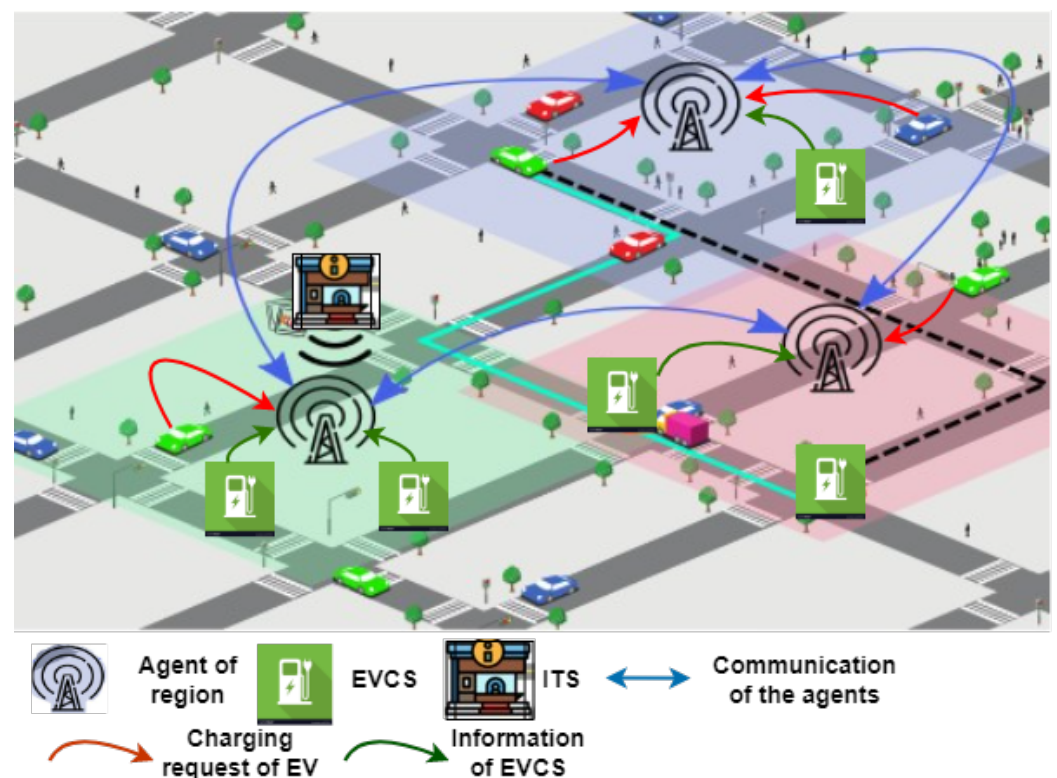
In another work [35], a combined routing and charging behavior approach based on network equilibrium and an equilibrium-based charging station location problem was suggested. In [36], a stochastic model for bidirectional smart charging of EVs was considered, taking into account the interests of EV customers. Chen et al. [37,38] proposed a stochastic model predictive control method for energy management and traffic control, leveraging reinforcement learning and a DQN algorithm to calculate the best action-value function, involving an RL-controller in the optimization process. Meanwhile, in [39], deep

reinforcement learning (DRL) was used to overcome the impact of physical restrictions in the stability properties model to optimize outcomes.

The research [40] presents a smart charging algorithm for electric vehicles (EVs) that incorporates real-time data on energy prices, grid demand, and charging station availability. By utilizing deep reinforcement learning (DRL) techniques, the algorithm dynamically optimizes EV charging schedules to take advantage of lower energy prices and reduce strain on the grid during peak hours. This not only leads to cost savings for EV owners but also contributes to grid load balancing. The importance of multi-agent system (MAS) applications, where agents must compete and collaborate to achieve the best overall results, was emphasized. To address the challenges of high-dimensional environments and the limitations of Q-learning in real-world scenarios, several studies have combined reinforcement learning with deep neural networks. A multi-agent deep reinforcement learning approach is employed in this study to make sequential decisions in a multi-agent environment, where single-agent DRL methods may not suffice due to increasing complexity in real-world problems [22–26]. In such situations, multi-agent system (MAS) applications become crucial, as agents must both compete and collaborate to achieve optimal results.

### 3. Materials and Methods

In this section, we introduce the key concepts and assumptions of the EV charging system architecture considered in this paper. The number of waiting EVs and the number of charging electric vehicles may be obtained from each EV charging station, and the traffic data may be obtained from an intelligent transport system (ITS) as illustrated in Figure 1.



**Figure 1.** Overall architecture for the multi-region of EV charging station.

#### 3.1. Network Model

The city is partitioned into  $n$  regions, with each region containing EV charging stations. The city's network topology is represented as a graph  $G = (V, E)$ , where  $V = \{1, 2, \dots, n\}$  denotes the set of vertices, and  $E = \{l_{ij} \mid i, j \in V\}$  denotes the set of edges representing the road connections between nodes  $i$  and  $j$ .

To efficiently manage the traffic flow, the system utilizes an intelligent transport system (ITS) to gather information on the number of waiting EVs and the charging electric vehicles at each charging station. The traffic data is collected and analyzed to obtain the average road velocity  $v_{ij}^t$  between nodes  $i$  and  $j$  at time step  $t$ . This information enables the calculation of driving times for each link  $l_{ij}$  in the network, defined as:

$$\tau_{ij}^t = \frac{d_{ij}}{v_{ij}^t} \quad (1)$$

where  $\tau_{ij}^t$  represents the driving time to travel through link  $l_{ij}$ , with  $d_{ij}$  denoting the distance between nodes  $i$  and  $j$ , and  $v_{ij}^t$  being the average road velocity at time step  $t$ .

To optimize the EV charging process, the city's topological graph is divided into  $n$  regions, each containing an EV charging station. Utilizing information from the intelligent transportation system and EV charging requests, the system manages traffic matrices and network topology. To estimate the arrival time, charging time, and charge quantity of EVs, the Dijkstra algorithm is employed to find the shortest time path from the EV's current position to each charging station in the region [41,42]. This allows the system to calculate the estimated arrival time to each charging station based on the identified route. Since future traffic conditions are uncertain, the estimates are based on current traffic conditions.

The path to a specific EV charging station can also be used to estimate the charge required for the EV. A featured state is extracted for each region based on the current time information received from the EV's charging request. As a result, the network model for each region is continuously updated and maintains real-time traffic data acquired from the ITS center, providing a set of estimated arrival times to each charging station as input to its respective agent.

The estimated arrival time from the origin position to EV charging station  $k$  in region  $i$  is calculated as follows:

$$\tau_{EAT}^{i,k} = \sum_{\forall l \in L^{i,k}} \frac{d_l}{v_l^i}, \quad \forall k \in K, \quad \forall i \in n \quad (2)$$

where  $L^{i,k}$  is the set of links from the origin to EV charging station  $k$  in region  $i$ . This information is vital for the intelligent charging recommendation process, enabling the system to efficiently manage EV charging stations and optimize the overall charging experience for EV drivers.

### 3.2. EV Charging Station

The city is equipped with EV charging stations distributed across different regions. Each EV charging station consists of parking spaces and charging poles. Regular updates on the charging status, such as the current cost of charging, the number of EVs charging, and the number of EVs waiting, are sent by all EVCSs in a region to the respective agent handling that region. Whenever an EV makes a charging request, this data is refreshed to ensure accurate estimations of charging and waiting times for all EV charging stations, which serves as the input for the corresponding agent.

To predict the waiting time for upcoming EV charging requests, the agent manages the charging reservations. The EVs are distributed within the EVCS based on the "First Come First Serve" (FCFS) policy. The expected waiting time for an EV at EVCS  $k$  can be calculated as:

$$\tau_{WaitT}^{i,k} = \max\{0, T_{start}^{i,k} - \tau_{EAT}^{i,k}\} \quad (3)$$

where  $T_{start}^{i,k}$  represents the time when the EV starts charging at EVCS  $k$  in region  $i$ . If an EV can be charged as soon as it arrives at a charging pole, it does not need to wait, and  $\tau_{WaitT}^{i,k} = 0$  based on the EV's arrival time. Otherwise, the EV has to wait until a charging pole becomes available.

The time when an EV starts charging at EVCS  $k$  in region  $i$  is determined using the formula:

$$T_{start}^{i,k} = \min\{T_{start}^{N_{ev}-N_{slot},i,k} + \tau_{ch}^{i,k}, T_{start}^{N_{ev}-N_{slot}+1,i,k} + \tau_{ch}^{i,k}, \dots, T_{start}^{N_{ev}-1,i,k} + \tau_{ch}^{i,k}\} \quad (4)$$

where  $N_{ev}$  is the number of EV requests for charging, and  $N_{slot}$  is the number of charging slots at EV charging station  $k$ . The earliest charging time for an EV is determined by considering EVs  $[N_{ev} - N_{slot}, N_{ev} - N_{slot} + 1, \dots, N_{ev} - 1]$  based on the first-come-first-served principle.

The charging time  $\tau_{ch}^{i,k}$  at EVCS  $k$  in region  $i$  is estimated using the equation:

$$\tau_{ChT}^{i,k} = \frac{(SOC_{req} - SOC_{arr}^{i,k})}{\eta \mu^{i,k}} \times C_{max} \quad (5)$$

where  $C_{max}$  is the maximum battery capacity of the EVs,  $\mu^{i,k}$  is the charging power of EV charging station  $k$ , and  $\eta$  is the charging efficiency.  $SOC_{arr}^{i,k}$  represents the state of charge of an EV when it arrives at EVCS  $k$  in region  $i$ , and it is given as:

$$SOC_{arr}^{i,k} = SOC_{cur} - SOC_{cons}^{i,k}, \quad 0 < SOC_{arr}^{i,k} < SOC_{req} \quad (6)$$

where  $SOC_{cur}$  is the current state of charge (starting position) of the EV.

Based on the previously defined path from the EV's position to each charging station in each region, along with the data received from the ITS, we can calculate the energy consumption  $e_l$  used to represent the EVCS statuses along with traffic conditions. Consequently, the state of charge consumption is calculated as:

$$SOC_{cons}^{i,k} = \varepsilon \sum_{\forall l \in L_f^{i,k}} \frac{d_l}{C_{max}}, \quad \forall k \in K, \quad \forall i \in n \quad (7)$$

where  $\varepsilon$  is the energy consumption rate (kWh/km). This information is crucial for the intelligent charging recommendation process, allowing the system to efficiently manage the EV charging stations and optimize the overall charging experience for EV drivers.

### 3.3. Communication

Effective communication is at the heart of our multi-agent approach for optimal EV charging station recommendation. In a complex urban environment, seamless data exchange between electric vehicles (EVs), EV charging stations, the intelligent transportation system (ITS), and agents within each city region is paramount to ensure efficient charging recommendations as illustrated in Figure 1. Various wireless technologies can be harnessed to facilitate this dynamic data flow and enable real-time decision-making. These technologies are chosen based on factors such as communication range, data transfer rates, and latency requirements.

- **Communication Channels:** each region in the city is equipped with a dedicated agent functioning as a central system. This agent acts as a communication hub, orchestrating interactions between all relevant entities. This includes EVs, EV charging stations, and the intelligent transportation system (ITS). Additionally, agents within a specific region communicate with each other to exchange vital information.
- **Agent Communication:** within a region, agents collaborate through wireless communication to ensure coordination and synergy. This communication allows agents to share insights about charging station occupancy, current traffic conditions, and pending EV charging requests. The collective intelligence of agents enhances the accuracy of charging station recommendations.
- **EV to Agent Communication:** EVs within each region establish communication with the corresponding agent responsible for that region. This communication can be facilitated through wireless technologies such as Wi-Fi, cellular networks, or Bluetooth.

EVs share essential information such as battery state of charge (SoC), current location, and destination.

- **ITS:** the intelligent transportation system (ITS) plays a pivotal role by providing real-time traffic information to the central agents. Communication between agents and the ITS can occur through various wireless technologies, ensuring up-to-date traffic insights.
- **Charging Station Interaction:** charging stations are equipped with communication modules that allow them to share real-time information about charging station occupancy, charging rates, and availability with the central agent. Wireless technologies such as cellular networks or dedicated short-range communications (DSRC) can facilitate this exchange.
- **Recommendation:** through ongoing data exchange, agents gather inputs from EVs, charging stations, and the ITS. This data fusion enables the agents to make informed decisions using our multi-agent deep deterministic policy gradient (MADDPG) algorithm. Agents collaborate to recommend optimal charging stations for EVs, considering travel time, charging station availability, and real-time traffic conditions based on Algorithm 1.

In conclusion, the communication framework in our methodology is designed to harness the power of wireless technologies for seamless data exchange in a multi-agent environment. This facilitates effective decision-making in recommending EV charging stations, considering the uncertainty of future charging requests and dynamically changing traffic conditions. By integrating wireless communication between EVs, charging stations, ITS, and agents, our approach optimizes the overall charging experience for EV drivers and contributes to the efficient utilization of charging infrastructure in a multi-region city setting. We deal with the details of the Markov game formulation of EV charging control thoroughly in Section 4 and with the training process of MADDPG algorithm for EV charging station recommendation in Section 5.

#### 4. Markov Game Formulation of EV Charging Control

This section presents the electric vehicle charging station recommendation using a multi-agent reinforcement learning framework. In this framework, each region with EV charging stations is considered an agent, and these agents collaborate and compete to maximize the overall reward. Unlike single-agent reinforcement learning, where an agent interacts with the environment described by a Markov decision process to receive immediate rewards, multi-agent reinforcement learning involves agents impacted by both the environment and other agents. As a result, the environment can no longer be described solely by a Markov decision process, leading to the formulation of a Markov game [26], which combines elements of both MDP and game theory [43,44].

**Markov Game.** An  $n$ -agent Markov game ( $n \geq 2$ ) is represented as a tuple:

$$\langle S, A^1, \dots, A^n, r^1, \dots, r^n, P \rangle$$

where  $n$  is the number of agents,  $S$  is the set of environment states, usually referring to the joint state of all agents,  $A^i$  is the set of actions available to the  $i$ -th agent, and the joint-action space is defined as  $A^1 \times \dots \times A^n$ .

$O^i : S \mapsto \mathbb{R}^d$  is the set of observations for agent  $i$ , returning a  $d$ -dimensional observation, and  $O^1, \dots, O^n$  is the set of joint observations. Each agent  $i$  utilizes a stochastic policy  $\pi_{\theta_i} : O^i \times A^i \mapsto [0, 1]$  to select actions.

The transition probability function  $P : S \times A^1 \times \dots \times A^n \mapsto P(S)$  describes the probability distribution from the current state  $s_t \in S$  to a new state  $s_{t+1} \in S$  when action  $a \in A$  is executed. Additionally, each agent  $i$  receives its own immediate rewards based on the state and agent's action, given the reward function  $r^i : S \times A^1 \times \dots \times A^n \mapsto \mathbb{R}$ .

The deep reinforcement learning (DRL) agents engage in sequential interactions with electric vehicles (EVs), EV charging stations in each region, and the intelligent transport



system center (ITSC) within discrete-time steps to determine the optimal charging station recommendation in a stochastic environment. Considering the multi-regional environment, each region participates in simultaneous cooperation and competition with other regions, leading to a mixed-cooperative-competitive setting. As depicted in Figure 2, each region uses observations of the environment state to select an action  $a^i \in A^i$  from its action space  $A^i$  based on the current policy  $\pi_{\theta_i}$  at time step  $t$ . The goal of each agent  $i$  is to maximize its own total expected return  $R^i = \sum_{t=0}^T \gamma^t r_i^t$ , where  $T$  is the time horizon and  $\gamma$  is a discount factor.

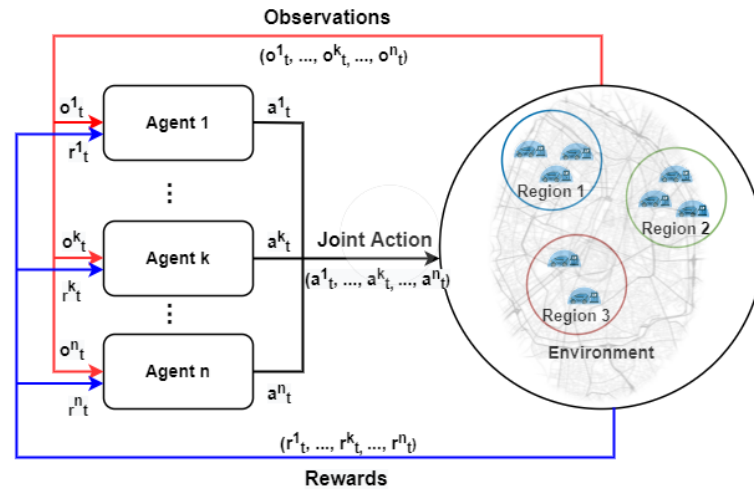


Figure 2. Multi-agent reinforcement learning.

#### 4.1. Agent

In our approach, each region containing EV charging stations in the city is treated as an individual agent. These agents are responsible for providing timely recommendations and decisions regarding the optimal and suitable EV charging stations for a sequence of charging requests. These decisions are based on real-time information received from requesting EVs, charging stations, and transportation systems, with the objective of achieving long-term optimization goals.

#### 4.2. Observation

The observation  $o_t^i$  at time step  $t$  for agent  $i$  consists of various features related to the arrived electric vehicle charge request ( $CR_t$ ) and the EV charging stations located in region  $i$ . It can be formulated as follows:

$$o_t^i = \{CR_t, \tau_{ETA}^{i,k}, \tau_{WaitT}^{i,k}, \tau_{ChT}^{i,k}, \mu^{i,k}\} \tag{8}$$

where  $\tau_{ETA}^{i,k}$ ,  $\tau_{WaitT}^{i,k}$ ,  $\tau_{ChT}^{i,k}$ , and  $\mu^{i,k}$  are sets of features representing the estimated time of arrival from the starting position to each EV charging station  $k$  in agent  $i$ , the waiting time, the charging time, and the charging power of each EV charging station  $k$  in agent  $i$ , respectively.

The charge request  $CR_t$  includes various parameters such as the time state  $T$ , the current state of charge  $SOC_{cur}$ , the required state of charge  $SOC_{req}$ , the maximum battery capacity of the required EV  $C_{max}$ , the starting position  $P_s$ , and the destination position  $P_d$ .

$$\begin{aligned} CR_t &= \{T, SOC_{cur}, SOC_{req}, C_{max}, P_s, P_d\} \\ \tau_{ETA}^{i,k} &= \{\tau_{ETA}^{i,1}, \tau_{ETA}^{i,2}, \dots, \tau_{ETA}^{i,K}\} \\ \tau_{WaitT}^{i,k} &= \{\tau_{WaitT}^{i,1}, \tau_{WaitT}^{i,2}, \dots, \tau_{WaitT}^{i,K}\} \\ \tau_{ChT}^{i,k} &= \{\tau_{ChT}^{i,1}, \tau_{ChT}^{i,2}, \dots, \tau_{ChT}^{i,K}\}, \\ \mu^{i,k} &= \{\mu^{i,1}, \mu^{i,2}, \dots, \mu^{i,K}\} \end{aligned} \tag{9}$$

Additionally, we define  $s_t = \{o_t^1, o_t^2, \dots, o_t^n\}$  as the state of all agents at step  $t$ . It should be noted that the observation is just a partial perspective of the environment, as it excludes details about the other agents.

#### 4.3. Action

Based on the observation of the environment, each agent  $i$  jointly determines the appropriate offer for the charging request. Specifically, each agent makes a recommendation for the charging request by selecting the optimal charging station located in its region. The joint action is defined as  $a_t = \{a_t^1, a_t^2, \dots, a_t^n\}$ .

#### 4.4. Transition Probability

At time step  $t$ , each agent  $i$  takes an action  $a_t^i \in A^i$ , which induces a transition in the environment from the current charging request  $CR_t$  to the next charging request  $CR_{t+j}$  after it is finished. This transition is governed by the state transition function  $\mathcal{P}(o_{t+j}^i | o_t^i, a_t^i)$ .

#### 4.5. Reward Function

The EV charging station recommendation in region  $i$  is based on the minimization of the total travel time expected for EV charging requests, which can be expressed as:

$$\tau_{travel}^{i,k} = \tau_{ETA}^{i,k} + \tau_{WaitT}^{i,k} + \tau_{ChT}^{i,k} \quad (10)$$

In our system, the total reward is defined as the difference between the actual travel time detected at the end of the EV charging and the estimated travel time derived by selecting the appropriate action for the EV charging station recommendation. If the actual travel time cannot be determined due to current traffic circumstances and other EV charging patterns, the reward is specified as the estimated travel time.

$$r_t = \begin{cases} -[\tau_{travel,act}^{i,k} - \tau_{travel,estm}^{i,k}], & \text{Si EV terminal} \\ -\tau_{travel,estm'}^{i,k} & \text{Otherwise} \end{cases} \quad (11)$$

The objective of MADRL is to maximize total rewards, where reducing travel time is specified as a negative value. Directly extending single-agent reinforcement learning methods, such as the deep Q-network (DQN) algorithm [19] or deep deterministic policy gradient (DDPG) algorithm [20], to learn synchronized policies by training each agent individually presents several challenges. Training in a multi-agent structure is more complicated and intricate than in a single-agent case. Each agent's policy changes constantly, and the environment appears non-stationary from the agent's perspective due to interactions with other agents, preventing the simple use of experience replay, which is critical for a DQN to learn consistency. Recently, a solution called multi-agent deep deterministic policy gradient (MADDPG) [45], which is an extension of DDPG to multi-agent systems, has been developed. This approach is based on the centralized training and decentralized execution architecture.

### 5. MADDPG Algorithm for EV Charging Station Recommendation

MADDPG is an actor-critic algorithm designed for continuous actions and states, capable of learning policies for complex coordination among multiple agents, considering the actions of other agents. In our study, the city is divided into multiple regions, each housing EV charging stations, creating a mixed-cooperative-competitive environment where agent policies and the environment are unstable and non-stationary.

Our approach encourages agents to learn coordinated policies and addresses the non-stationary environment problem. To achieve this, agent policies consist of two separate deep neural networks with different parameters: an actor-network (policy-network) with parameters  $\theta_i^M$ , which generates an action distribution to approximate policy  $\mu_{\theta_i^M}(o_i)$ , and a critic network with parameters  $\theta_i^Q$ , which predicts discounted future returns to approximate

the state-value function  $Q_i^{\theta_i^Q}(s_i, a)$ . We also include time-delayed copies  $\theta_i^{\mu'}$  and  $\theta_i^{Q'}$  of these networks to act as targets, enabling more stable training.

This framework of centralized training and decentralized execution helps us achieve our goal of efficient EV charging station recommendations. The centralized training phase allows the multi-agents to exchange additional information, depicted by dotted lines in Figure 3, such as observations and actions of other agents, as well as local observations of the environment, making the training process more effective. On the other hand, the decentralized execution mechanism ensures speed and flexibility in real-time recommendation as it operates independently and does not require comprehensive information during execution.

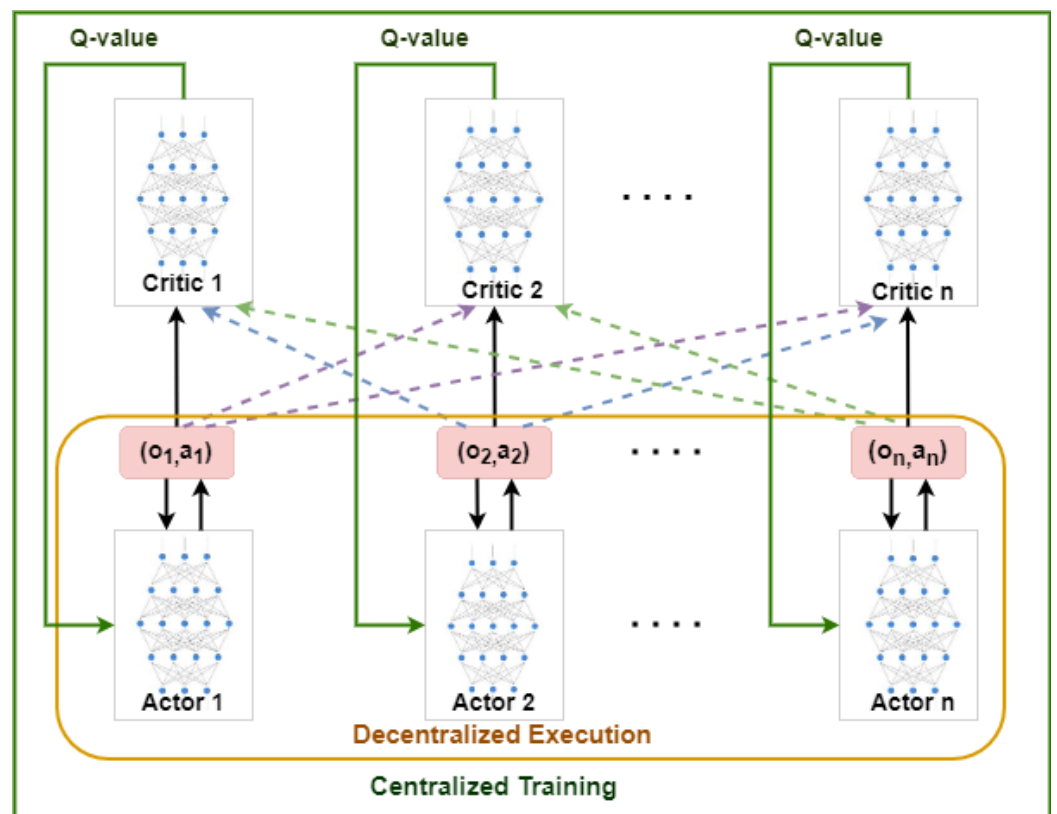


Figure 3. Centralized training and decentralized execution of the MADDPG algorithm.

The actor creates an exploratory policy such that  $a_t^i = \mu_{\theta_i^\mu}(o_t^i) + \mathcal{N}_t$  at time step  $t$  by deterministically transferring a local observation  $o_t^i$  to a particular continuous action  $\mu_{\theta_i^\mu}(o_t^i)$  and then adding a random noise process  $\mathcal{N}$ . In order to use the experience in the subsequent decision-making steps, i.e., to improve sampling efficiency and training stability, the agent’s transition, as well as the supplemental information, the joint actions of all agents  $a_t = \{a_t^1, a_t^2, \dots, a_t^n\}$ , and returns the immediate reward  $r_t^i$  and the next observation  $o_{t+1}^i$  to the respective agents, are stored in a replay buffer  $\mathcal{D}$  in the form of  $e_t^i = (s_t^i, a_t, r_t^i, s_{t+1}^i)$  of the agent  $i$ .

A random mini-batch of  $S$  samples  $(s^j, a^j, r^j, s'^j)|_{j=1}^S$  is extracted from the replay buffer  $\mathcal{D}$  to train the main networks, denoting a sample index by  $j$ , and the critic for each agent is updated by minimizing the loss function:

$$L(\theta_i) = \frac{1}{S} \sum_j (y^j - Q_i^\mu(s^j, a_1^j, \dots, a_N^j))^2 \tag{12}$$

where  $y^j$  is the target value expressed as

$$y^j = r_i^j + \gamma Q_i^\mu(s^j, a_1^j, \dots, a_N^j)|_{a_k = \mu_k^j(o_k^j)}$$

More precisely, by applying the gradient, the parameters of the critic network  $\theta_i^Q$ , are modified by (12) so that:

$$\theta_i^Q \leftarrow \theta_i^Q - \beta_Q \nabla_{\theta_i^Q} L(\theta_i^Q)$$

where  $\beta_Q$  is the learning rate of the critic network. On the other hand, the actor-network updates its parameters to optimize agent  $i$  predicted long-term discounted reward return,  $J_{\theta_i^\mu}(\mu_i) = \mathbb{E}[R_i]$ , according to

$$\theta_i^\mu \leftarrow \theta_i^\mu - \beta_\mu \nabla_{\theta_i^\mu} J_{\theta_i^\mu}(\mu_i)$$

where  $\beta_\mu$  represents the actor network's learning rate, and  $\nabla_{\theta_i^\mu} J_{\theta_i^\mu}(\mu_i)$  is the deterministic policy gradient, can be written as:

$$\nabla_{\theta_i^\mu} J_{\theta_i^\mu}(\mu_i) \approx \frac{1}{S} \sum_j \nabla_{\theta_i^\mu} \mu_i(o_i^j) \nabla_{a_i} Q_i^\mu(s^j, a_1^j, \dots, a_i, \dots, a_N^j)|_{a_i = \mu_i(o_i^j)} \quad (13)$$

Soft updates are applied to update the target parameters in both actor and critic networks as described below:

$$\theta'_i \leftarrow \tau \theta_i + (1 - \tau) \theta'_i \quad (14)$$

where  $\tau$  is a constant close to zero.

The DDPG algorithm is used to train learning agents over many episodes, each with several time steps. The pseudocode of the DDPG algorithm's specifics is described in Algorithm 1.

---

**Algorithm 1** Multi-agent deep deterministic policy gradient for n agents

---

- 1: **for** episode = 1 to M **do**
  - 2:   Initialize a random process  $\mathcal{N}$  for action exploration.
  - 3:   Receive initial state  $s$ .
  - 4:   **for** t = 1 to max-episode-length **do**
  - 5:     **for** each agent  $i$ , select action
 
$$a_i = \mu_{\theta_i}(o_i) + \mathcal{N}_t$$
  - 6:     w.r.t. the current policy and exploration
  - 7:     Execute actions  $a = (a_1, \dots, a_N)$  and observe reward  $r$  and new state  $s'$
  - 8:     Store  $(s, a, r, s')$  in replay buffer  $\mathcal{D}$
  - 9:      $s \leftarrow s'$
  - 10:     **for** agent  $i = 1$  to n **do**
  - 11:       Sample a random mini-batch of  $S$  samples  $(s^j, a^j, r^j, s'^j)$  from  $\mathcal{D}$
  - 12:       Update critic by minimizing the loss given by (12).
  - 13:       Update actor using the sampled policy gradient given by (13).
  - 14:     **end for**
  - 15:     Update target network parameters for each agent  $i$  to (14).
  - 16: **end for**
-

## 6. Performance Evaluation

In this section, we present a detailed implementation of our proposed MADRL-based approach and determine the training parameters for simulations to illustrate its efficacy.

### 6.1. Experimental Setup

In our case study, the transportation network in our simulation is represented using a topological graph that captures the connectivity and relationships between different city regions. The city center is divided into three distinct regions, each covering a  $4 \times 4 \text{ km}^2$  area. Within these regions, roadways, intersections, and travel routes are defined based on a realistic urban layout. This network topology serves as the foundation for EV movement and travel time calculations. The charging station locations are strategically positioned within the city's regions. Specifically, the first and second regions are equipped with three charging stations each, while the third region houses two charging stations. These charging stations are geographically dispersed to ensure equitable coverage and accessibility for EV users across the city. Additionally, we consider that each charging pole at each charging station offers a different charging power. To realistically replicate dynamic traffic conditions, we introduce traffic variations that change throughout the simulation's 24-h duration. The fluctuating traffic conditions include factors such as congestion, road closures, and varying travel speeds during peak and off-peak hours. These conditions impact EV travel times and charging station availability, adding a layer of realism to our evaluation.

To evaluate our approach, the simulated EV context encompasses a diverse set of factors that mirror real-world EV behavior. The number of EVs involved in our simulation varies based on the specific scenario under investigation. For instance, scenarios with different numbers of EV charging requests arriving within a day are generated using both uniform and normal distributions, as detailed in Table 1. Our approach takes into account a dynamic scenario where EVs are in constant movement, and their battery state of charge (SoC) is regularly monitored. Users' transportation behavior is treated as random variables, as described by [46]. Each EV's behavior is modeled using random variables, such as the current state of charge (SoC), required SoC, and estimated arrival time (ETA), which are sampled from normal distributions. We assume EVs with a battery capacity of  $C_{max} = 60 \text{ kWh}$ . The EVs' driving patterns are derived from a range of average speeds randomly generated between 2.7 m/s and 16.6 m/s. This variation in speeds reflects the diversity of driving styles and conditions encountered in urban environments. Furthermore, the EVs' interactions with the transportation network database enable them to navigate through the city, dynamically adapting to changing traffic conditions. This variety of scenarios allows us to demonstrate the efficiency and versatility of the proposed algorithm.

**Table 1.** User transportation behavior as random variables.

	Distribution	Boundary
Battery current SoC	$SoC_{cur} \sim N(1, 0.1^2)$	[0.2, 0.4]
Battery required SoC	$SoC_{req} \sim N(0.2, 0.1^2)$	[0.8, 0.9]
ETA	$\tau_{ETA}^{i,k} \sim N(10, 1^2)$	[12, 22]

The MADDPG-based approach utilizes two neural networks for each agent: an actor network and a critic network. The networks consist of fully connected layers with ReLU activation functions, comprising 128 nodes per layer for the actor and 256 nodes per layer for the critic. The actor network's outputs employ a sigmoid activation function. During training, we use the Adam optimizer with a learning rate of 0.01 for the critic network and 0.001 for the actor network. To stabilize the learning process, target networks are updated every 1000 samples, with a discount factor of 0.99 and a replay buffer size of 20,000. We use a mini-batch size of 512 for training efficiency.

The implementation is in Python using TensorFlow, and the training was performed on a computer equipped with an Intel Core i5-4860 CPU. The MADDPG approach demonstrates convergence after 16,000 episodes of training.

### 6.2. Performance Evaluation

The performance of the proposed MADDPG-based algorithm approach is compared with two other methods: the DDPG-based one-agent algorithm and a conventional benchmark strategy based on the deterministic environment.

- The DDPG implementation uses a centralized control mechanism where a single agent oversees all regions in the city concurrently and recommends EV charging stations for dynamically arriving charging requests.
- On the other hand, the benchmark strategy in the deterministic environment does not consider the uncertainty arising from constantly changing traffic conditions and the presence of other electric vehicles entering the charging stations at the time of EV charging station recommendation.

For a fair comparison, all methods are performed with identical parameters, and the number of EV charging requests arriving in one day is set to 100. Figure 4 illustrates the experimental results of the training process.

As depicted in Figure 4, the average cumulative reward for MADDPG surpasses that of the DDPG algorithm, although at the initial stages of training, the effects may not be immediately noticeable as the agents require some time to learn. However, after around 5000 episodes, the proposed algorithm's average reward value starts to increase significantly, reaching a value of approximately  $-20$  after around 8000 episodes and stabilizing at around 10,000 training episodes. On the other hand, the DDPG algorithm reaches a maximum average reward of  $-29.7$ . The MADDPG approach maintains remarkable stability after reaching its peak performance, in stark contrast to the DDPG algorithm, which is unstable and challenging to converge.

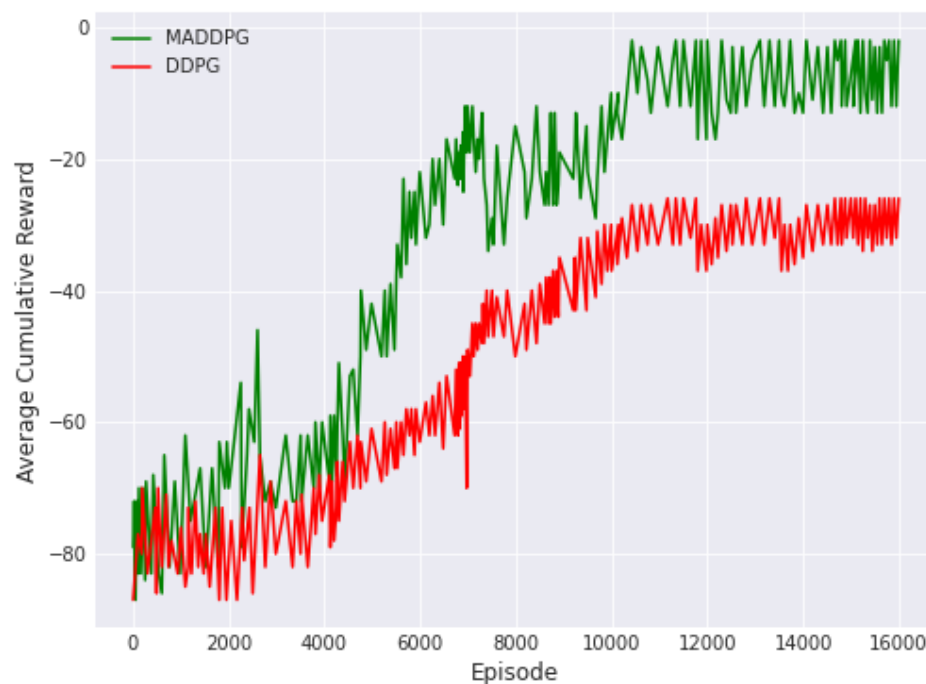


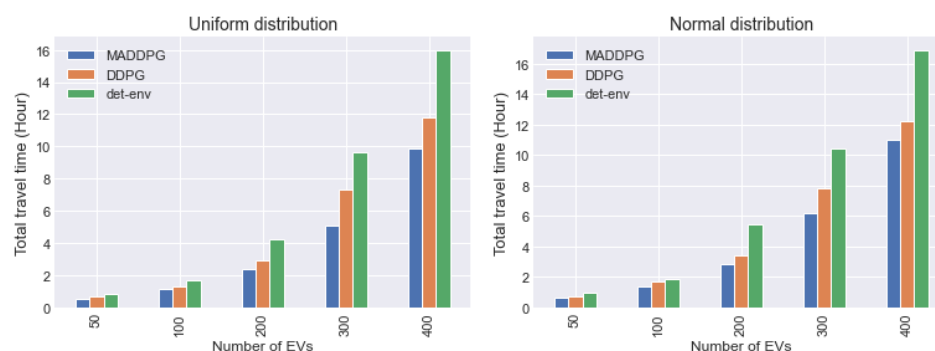
Figure 4. Agent reward on different algorithms during the process of training.

In Figure 4, MADDPG exhibits a clear convergence trend, with the average cumulative reward steadily increasing over time. This indicates that the agents are consistently improving their performance as they learn from their interactions with the environment.

The algorithm's stability is evident as the reward curve stabilizes after a certain number of training episodes, showcasing its capacity to converge to a consistent and optimal policy. In contrast, DDPG, while initially improving, shows fluctuations and instability in the learning curve. This behavior aligns with the challenge of training a single agent to oversee all regions concurrently, leading to difficulties in finding an optimal policy consistently.

To demonstrate the efficiency and versatility of the proposed approach, we conducted experiments with different numbers of EV charging requests arriving in one day, specifically "50", "100", "200", "300", and "400" based on both uniform and normal distributions. The results were compared with the DDPG algorithm and an approach based on the deterministic environment (det-env). The proposed approach selects the EV charging station for each EV to minimize the travel time.

In Figure 5, we compared the total travel time of our MADDPG algorithm with DDPG and the deterministic environment methods for different numbers of EV charging demands arriving using two distributions, uniform and normal. As shown, our MADDPG approach consistently outperforms all other methods, and as the number of EV charging requests increases, the total travel time for all methods also increases. For instance, with 400 EVs, MADDPG exhibits approximately 16.1% and 38% performance improvements compared to DDPG and the deterministic environment strategy under the uniform distribution, and 9.8% and 34.9% performance enhancements under the normal distribution. In this scenario, EV charging requests spike at random times, arriving dynamically.



**Figure 5.** Comparison of total travel time by the hour of MADDPG, DDPG, and the deterministic environment methods count distributed uniformly and normally with different numbers of EVs.

MADDPG demonstrates clear superiority in recommending charging stations, particularly when many charging requests arrive simultaneously, while effectively handling high-dimensional states and uncertainties in the environment. In this Markov game, all agents are treated as common components of the environment, resulting in a static invariance of the environment. This enables effective cooperation between agents to make optimal decisions in a stochastic environment with unknown future requests.

The impact of our approach goes beyond performance enhancement. Our algorithm's dynamic adaptation to real-time traffic data has a direct positive influence on traffic congestion reduction. By intelligently distributing EVs to strategically located charging stations, our approach mitigates unnecessary congestion caused by inefficient routing. Furthermore, our method's ability to optimize energy distribution across charging infrastructure is of paramount importance for long-term sustainability. Our results exhibit that through optimal EV charging station assignments, we achieve improved energy utilization and reduced overall energy consumption. When compared to other methods, our MADDPG approach displays superior adaptability to dynamic traffic conditions and EV charging demand. The multi-agent nature of MADDPG facilitates efficient coordination among charging stations and EVs, leading to better utilization of resources and enhanced system performance. This comparison underlines the long-term advantages of employing a collaborative multi-agent approach such as MADDPG in complex urban mobility scenarios. Comparing our approach to other deep learning methods, MADDPG's multi-agent nature grants it superior

adaptability to dynamic traffic conditions and EV charging demands. The collaborative decision-making between charging stations and EVs, facilitated by MADDPG, results in better resource utilization and system performance. This comparison underscores the long-term benefits of employing a collaborative multi-agent approach such as MADDPG in complex urban mobility scenarios.

## 7. Conclusions

This study introduces a methodology for optimal EV charging station recommendation within a multi-region city, based on the multi-agent deep deterministic policy gradient (MADDPG) algorithm. The primary objective is the minimization of total travel time in a stochastic environment, accounting for the dynamic and random arrival of charging requests. The city is divided into multiple regions, each supervised by an agent responsible for recommending EV charging stations and collaborating with other agents to manage charging requests efficiently. The agents obtain traffic information from an intelligent transportation system and charging station information within their regions. By learning a central critic based on collective agent observations and actions, optimal decisions for each EV charging request are achieved. Through simulation experiments with different numbers of EVs and two distribution scenarios (uniform and normal), we compared our proposed MADDPG approach with DDPG and the deterministic environment method. The results demonstrated the superiority of MADDPG in minimizing total travel time, indicating its effectiveness and consistency in a complex multi-agent environment with unknown future requests. As part of our future work, we aim to extend the algorithm to consider multiple charging stations with diverse objectives in formulating the reward functions. Additionally, we plan to implement the proposed algorithm in real-world scenarios using more complex data for validation and practical application. One notable challenge lies in overcoming the computational complexity entailed in training and coordinating multiple agents. As agent numbers rise, the training process could grow time-consuming and resource-intensive, necessitating a nuanced balance between agent count and computational efficiency. Furthermore, the efficacy of our approach hinges significantly on the precision and accessibility of real-time traffic data and charging station information. Discrepancies or inaccuracies within these data streams could undermine recommendation quality, highlighting the imperative of a dependable data pipeline and data quality management to ensure algorithmic robustness. This would further establish the robustness and applicability of MADDPG in addressing real-world EV charging station recommendation challenges.

**Author Contributions:** All authors equally contributed to the conception, the writing, and the revision of the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** Not applicable

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

## References

1. Anthony Jnr, B. Smart city data architecture for energy prosumption in municipalities: concepts, requirements, and future directions. *Int. J. Green Energy* **2020**, *17*, 827–845. [[CrossRef](#)]
2. Heo, T.; Kim, K.; Kim, H.; Lee, C.; Ryu, J.H.; Leem, Y.T.; Jun, J.A.; Pyo, C.; Yoo, S.M.; Ko, J. Escaping from ancient Rome! Applications and challenges for designing smart cities. *Trans. Emerg. Telecommun. Technol.* **2014**, *25*, 109–119.
3. Han, J.; Liu, H.; Zhu, H.; Xiong, H.; Dou, D. Joint Air Quality and Weather Prediction Based on Multi-Adversarial Spatiotemporal Networks. *Proc. Aaai Conf. Artif. Intell.* **2021**, *35*, 4081–4089. [[CrossRef](#)]
4. Savari, G.F.; Krishnasamy, V.; Sathik, J.; Ali, Z.M.; Abdel Aleem, S.H.E. Internet of Things based real-time electric vehicle load forecasting and charging station recommendation. *Isa Trans.* **2020**, *97*, 431–447. [[CrossRef](#)] [[PubMed](#)]
5. Zheng, Y.; Shao, Z.; Zhang, Y.; Jian, L. A systematic methodology for mid-and-long term electric vehicle charging load forecasting: The case study of Shenzhen, China. *Sustain. Cities Soc.* **2020**, *56*, 102084. [[CrossRef](#)]
6. Leonard, M.; Pisani-Ferry, J.; Shapiro, J.; Tagliapietra, S.; Wolff, G. *The Geopolitics of the Policy Contribution*; Issue n°04/21 February 2021; Bruegel AISBL: Brussel, Belgium, 2021.



7. Oda, T.; Aziz, M.; Mitani, T.; Watanabe, Y.; Kashiwagi, T. Mitigation of congestion related to quick charging of electric vehicles based on waiting time and cost–benefit analyses: A Japanese case study. *Sustain. Cities Soc.* **2018**, *36*, 99–106. [[CrossRef](#)]
8. Han, P.; Wang, J.; Han, Y.; Li, Y. Resident Plug-In Electric Vehicle Charging Modeling and Scheduling Mechanism in the Smart Grid. *Math. Probl. Eng.* **2014**, *2014*, e540624. [[CrossRef](#)]
9. Wang, G.; Zhang, Y.; Fang, Z.; Wang, S.; Zhang, F.; Zhang, D. FairCharge: A Data-Driven Fairness-Aware Charging Recommendation System for Large-Scale Electric Taxi Fleets. *Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol.* **2020**, *4*, 28:1–28:25. [[CrossRef](#)]
10. Tan, J.; Wang, L. Real-Time Charging Navigation of Electric Vehicles to Fast Charging Stations: A Hierarchical Game Approach. *IEEE Trans. Smart Grid* **2015**, *8*, 846–856. [[CrossRef](#)]
11. Yang, H.; Deng, Y.; Qiu, J.; Li, M.; Lai, M.; Dong, Z.Y. Electric Vehicle Route Selection and Charging Navigation Strategy Based on Crowd Sensing. *IEEE Trans. Ind. Inform.* **2017**, *13*, 2214–2226. [[CrossRef](#)]
12. Pan, L.; Yao, E.; Yang, Y.; Zhang, R. A location model for electric vehicle (EV) public charging stations based on drivers' existing activities. *Sustain. Cities Soc.* **2020**, *59*, 102192. [[CrossRef](#)]
13. Campaña, M.; Inga, E. Optimal Planning of Electric Vehicle Charging Stations Considering Traffic Load for Smart Cities. *World Electr. Veh. J.* **2023**, *14*, 104. [[CrossRef](#)]
14. Ali, A.; Shakoor, R.; Raheem, A.; Muqet, H.A.u.; Awais, Q.; Khan, A.A.; Jamil, M. Latest Energy Storage Trends in Multi-Energy Standalone Electric Vehicle Charging Stations: A Comprehensive Study. *Energies* **2022**, *15*, 4727. [[CrossRef](#)]
15. Sica, L.; Deflorio, F. Estimation of charging demand for electric vehicles by discrete choice models and numerical simulations: Application to a case study in Turin. *Green Energy Intell. Transp.* **2023**, *2*, 100069. [[CrossRef](#)]
16. Singh, P.P.; Wen, F.; Palu, I.; Sachan, S.; Deb, S. Electric Vehicles Charging Infrastructure Demand and Deployment: Challenges and Solutions. *Energies* **2023**, *16*, 7. [[CrossRef](#)]
17. Sutton, R.S.; Barto, A.G. *Reinforcement Learning, Second Edition: An Introduction*; Google-Books-ID: uWV0DwAAQBAJ; MIT Press: Cambridge, MA, USA, 2018.
18. Xu, P.; Zhang, J.; Gao, T.; Chen, S.; Wang, X.; Jiang, H.; Gao, W. Real-time fast charging station recommendation for electric vehicles in coupled power-transportation networks: A graph reinforcement learning method. *Int. J. Electr. Power Energy Syst.* **2022**, *141*, 108030. [[CrossRef](#)]
19. Mnih, V.; Kavukcuoglu, K.; Silver, D.; Rusu, A.A.; Veness, J.; Bellemare, M.G.; Graves, A.; Riedmiller, M.; Fidjeland, A.K.; Ostrovski, G.; et al. Human-level control through deep reinforcement learning. *Nature* **2015**, *518*, 529–533. [[CrossRef](#)] [[PubMed](#)]
20. Lillicrap, T.P.; Hunt, J.J.; Pritzel, A.; Heess, N.; Erez, T.; Tassa, Y.; Silver, D.; Wierstra, D. Continuous control with deep reinforcement learning. *arXiv* **2019**. arXiv:1509.02971.
21. Nguyen, T.T.; Nguyen, N.D.; Nahavandi, S. Deep Reinforcement Learning for Multiagent Systems: A Review of Challenges, Solutions, and Applications. *IEEE Trans. Cybern.* **2020**, *50*, 3826–3839. [[CrossRef](#)]
22. Bowling, M.; Veloso, M. Multiagent learning using a variable learning rate. *Artif. Intell.* **2002**, *136*, 215–250. [[CrossRef](#)]
23. Greenwald, A.; Hall, K. Correlated-Q learning. In Proceedings of the Twentieth International Conference on International Conference on Machine Learning, ICML '03, Washington, DC, USA, 21–24 August 2003; pp. 242–249.
24. Littman, M. Friend-or-Foe Q-learning in General-Sum Games. In Proceedings of the ICML '01: Proceedings of the Eighteenth International Conference on Machine Learning, San Francisco, CA, USA, 28 June–1 July 2001.
25. Hu, J.; Wellman, M. Nash Q-Learning for General-Sum Stochastic Games. *J. Mach. Learn. Res.* **2003**, *4*, 1039–1069. [[CrossRef](#)]
26. Littman, M.L. Markov games as a framework for multi-agent reinforcement learning. In *Machine Learning Proceedings 1994*; Cohen, W.W., Hirsh, H., Eds.; Morgan Kaufmann: San Francisco, CA, USA, 1994; pp. 157–163. [[CrossRef](#)]
27. Qiu, D.; Wang, Y.; Zhang, T.; Sun, M.; Strbac, G. Hybrid Multiagent Reinforcement Learning for Electric Vehicle Resilience Control Towards a Low-Carbon Transition. *IEEE Trans. Ind. Inform.* **2022**, *18*, 8258–8269.
28. Wang, S.; Dong, Z.Y.; Luo, F.; Meng, K.; Zhang, Y. Stochastic Collaborative Planning of Electric Vehicle Charging Stations and Power Distribution System. *IEEE Trans. Ind. Inform.* **2018**, *14*, 321–331.
29. Kim, S.; Lim, H. Reinforcement Learning Based Energy Management Algorithm for Smart Energy Buildings. *Energies* **2018**, *11*, 2010.
30. Zhang, L.; Gong, K.; Xu, M. Congestion Control in Charging Stations Allocation with Q-Learning. *Sustainability* **2019**, *11*, 3900.
31. Dabbaghjamesh, M.; Moeini, A.; Kavousi-Fard, A. Reinforcement Learning-Based Load Forecasting of Electric Vehicle Charging Station Using Q-Learning Technique. *IEEE Trans. Ind. Inform.* **2021**, *17*, 4229–4237.
32. Chiş, A.; Lundén, J.; Koivunen, V. Reinforcement Learning-Based Plug-in Electric Vehicle Charging With Forecasted Price. *IEEE Trans. Veh. Technol.* **2017**, *66*, 3674–3684.
33. Wan, Z.; Li, H.; Prokhorov, D. Model-Free Real-Time EV Charging Scheduling Based on Deep Reinforcement Learning. *IEEE Trans. Smart Grid* **2018**, *10*, 5246–5257. [[CrossRef](#)]
34. Maźziel, M.; Campisi, T. Energy Consumption of Electric Vehicles: Analysis of Selected Parameters Based on Created Database. *Energies* **2023**, *16*, 1437.
35. Li, J.; Xie, C.; Bao, Z. Optimal en-route charging station locations for electric vehicles: A new modeling perspective and a comparative evaluation of network-based and metanetwork-based approaches. *Transp. Res. Part Emerg. Technol.* **2022**, *142*, 103781. [[CrossRef](#)]
36. Salmani, H.; Rezazade, A.; Sedighzadeh, M. Stochastic peer to peer energy trading among charging station of electric vehicles based on blockchain mechanism. *Iet Smart Cities* **2022**, *4*, 110–126. [[CrossRef](#)]

37. Chen, Z.; Hu, H.; Wu, Y.; Zhang, Y.; Li, G.; Liu, Y. Stochastic model predictive control for energy management of power-split plug-in hybrid electric vehicles based on reinforcement learning. *Energy* **2020**, *211*, 118931. [[CrossRef](#)]
38. Yan, F.; Wang, J.; Du, C.; Hua, M. Multi-Objective Energy Management Strategy for Hybrid Electric Vehicles Based on TD3 with Non-Parametric Reward Function. *Energies* **2023**, *16*, 74.
39. Sun, X.; Qiu, J. A Customized Voltage Control Strategy for Electric Vehicles in Distribution Networks With Reinforcement Learning Method. *IEEE Trans. Ind. Inform.* **2021**, *17*, 6852–6863.
40. 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.
41. Dijkstra, E.W. A note on two problems in connexion with graphs. *Numer. Math.* **1959**, *1*, 269–271. [[CrossRef](#)]
42. Eklund, P.; Kirkby, S.; Pollitt, S. A dynamic multi-source Dijkstra’s algorithm for vehicle routing. In Proceedings of the 1996 Australian New Zealand Conference on Intelligent Information Systems. Proceedings. ANZIIS 96, Adelaide, Australia, 18–20 November 1996, pp. 329–333. [[CrossRef](#)]
43. Hu, J.; Wellman, M.P. Multiagent Reinforcement Learning : Theoretical Framework and an Algorithm. In Proceedings of the Fifteenth International Conference on Machine Learning (ICML ’98), Madison, WI, USA, 24–27 July 1998.
44. Buşoniu, L.; Babuška, R.; De Schutter, B. Multi-agent Reinforcement Learning: An Overview. In *Innovations in Multi-Agent Systems and Applications-1*; Srinivasan, D., Jain, L.C., Eds.; Studies in Computational Intelligence; Springer: Berlin/Heidelberg, Germany, 2010; pp. 183–221. [[CrossRef](#)]
45. Lowe, R.; Wu, Y.; Tamar, A.; Harb, J.; Pieter Abbeel, O.; Mordatch, I. Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Curran Associates, Inc.: Red Hook, NY, USA, 2017; Volume 30.
46. Ren, H.; Zhang, A.; Wang, F.; Yan, X.; Li, Y.; Duić, N.; Shafie-khah, M.; Catalão, J.P.S. Optimal scheduling of an EV aggregator for demand response considering triple level benefits of three-parties. *Int. J. Electr. Power Energy Syst.* **2021**, *125*, 106447. [[CrossRef](#)]

**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.