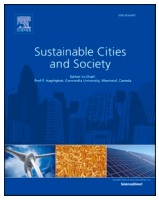
Sustainable Cities and Society 113 (2024) 105567

Contents lists available at ScienceDirect

Sustainable Cities and Society

journal homepage: www.elsevier.com/locate/scs



Optimal planning for electric vehicle fast charging stations placements in a city scale using an advantage actor-critic deep reinforcement learning and geospatial analysis

Jae Heo, Soowon Chang *

Smart and Sustainable Human-Urban-Building Interactions Laboratory (S2-HUB), School of Construction Management Technology, Purdue University, 363 N. Grant St., West Lafayette, IN 47907, USA

ARTICLE INFO

Keywords:
Electric vehicle fast charging station
Infrastructure planning
Advantage actor-critic
Deep reinforcement learning
Geospatial analysis

ABSTRACT

The transition to Electric Vehicles (EVs) for reducing urban greenhouse gas emissions is hindered by the lack of public charging infrastructure, particularly fast-charging stations. Given that electric vehicle fast charging stations (EVFCS) can burden the electricity grid, it is crucial for EVFCS to adopt sustainable energy supply methods while accommodating the growing demands of EVs. Despite recent research efforts to optimize the placement of renewable-powered EV charging stations, current planning methods face challenges when applied to a complex city scale and integrating with renewable energy resources. This study thus introduces a robust decision-making model for optimal EVFCS placement planning integrated with solar power supply in a large and complex urban environment (e.g., Chicago), utilizing an advantage actor-critic (A2C) deep reinforcement learning (DRL) approach. The model balances traffic demand with energy supply, strategically placing charging stations in areas with high traffic density and solar potential. As a result, the model is used to optimally place 1,000 charging stations with a random starting search approach, achieving total reward values of 74.30 %, and estimated the capacities of potential EVFCS. This study can inform the identification of suitable locations to advance the microgrid-based charging infrastructure systems in large urban environments.

1. Introduction

With the growing environmental concerns related to internal combustion engine (ICE) vehicles, the transition to electric vehicles (EV) has emerged as a means for reducing greenhouse gas (GHG) emissions in the transportation sector (Sanguesa et al., 2021). According to the United States Department of Energy (DOE) (McLaren et al., 2016), nationwide carbon dioxide (CO2) emissions per vehicle are reported to be 4.5 times higher in conventional ICE vehicles (12,594 pounds of CO2 in 2021) in all EVs (2817 pounds of CO2 in 2021) annually. EVs running only on electricity have zero emissions of GHGs (e.g., CO2, methane (CH4), nitrous oxide (N2O), etc.) during driving.

The dissemination of EVs has been challenged due to the lack of public charging stations (Singh et al., 2022). EV users have expressed concern about potential roadside standing, a consequence of limited access to charging infrastructure. The development of a widespread charging infrastructure is pivotal to ensuring driving feasibility for EVs and promoting adoption (Nie et al., 2016). The current charging infrastructure is mostly the low-speed public charging stations (i.e., level-2) (87.75 % in the United States) (U.S. Department of Renewable Energy, 2023), which require 4 - 10 h to 80 % charge from empty. However, these slow charging stations cannot accommodate the growth of high charging demand efficiently. Thus, the installation of direct-current (DC) fast charging stations (i.e., level-3) in an appropriate place has been studied to support time-efficient long-distance travels. DC fast charging stations incur high installation fees, exceeding $20,000. A misplacement of expensive fast charging station, lacking the investigation of potentially high charging demand area, may result in significant economic losses. In this respect, it is crucial to make optimal planning for EV fast charging stations (EVFCS) that can consider growing charging demand.

Increasing demand of the EV charging infrastructure imposes substantial pressure on the national power electricity grid (Pareek et al., 2020). The burdens may introduce uncertainties in the reliable supply of electricity to other consumers. The current capacity of transmission lines cannot be affordable for the upcoming surge in demand from EV users

* Corresponding author.
E-mail addresses: heo27@purdue.edu (J. Heo), chang776@purdue.edu (S. Chang).

https://doi.org/10.1016/j.scs.2024.105567
Received 23 February 2024; Received in revised form 13 May 2024; Accepted 31 May 2024
Available online 19 July 2024

(Rodriguez-Calvo et al., 2017). In addition, the reliance on fossil fuels for electricity generation is still challenged to justify the promotion for EV adoption. With the growing EV adoption and increasing need of EVFCS, it is needed to transition the electricity resources to renewable energy. As a sustainable and least-invasive way of supplying the increasing needs of electricity, decentralized energy resources (e.g., solar) can be used. Solar photovoltaic (PV) power systems can be a viable and promising solution to complement the demand capacity of the grids as well as address environmental issues by reducing greenhouse gases (GHG) emissions during electricity generation.

Recent research on renewable-powered EVCS placement problems has focused on optimal locations, where balancing high charging demand with the distribution of renewable energy generation (Mozafar et al., 2017). Considering the limited electricity produced by solar PV systems, optimal EVFCS sites need to be identified based on reliable estimations of capacity and the number of charging stations powered by solar PV (Huang et al., 2019). Notably, as the generation of renewable energy can be spread across vast areas, the problem of placing renewable-powered EVCS needs to be extended to examine coverage throughout the entire city. Despite the demand for exploring potential EVFCS locations at a large scale, considering the increase in the number of charging stations and alternative sites, current planning approaches (e.g., heuristic-based path planning, multiple-criteria decision-making approach (MCDM)) may face challenges when applying them to solve renewable-powered EVFCS placement problems due to a lack of scalability and applicability. The existing planning approach may not effectively solve the optimization problem for a large number of EVCS placements, as it requires substantial computational resources that increase with the size of the problem (i.e., the number of charging stations) (Nandy et al., 2023; Wang et al., 2021). The previous approach has also been examined to a limited extent in a small-scale case study area, without covering the entire city (Lam et al., 2014). For example, the MCDM approach evaluates optimal sites from several candidate sites (Rani & Mishra, 2021) and the path planning approach cannot consider outside road networks to investigate high charging demand (Deb et al., 2018).

To address the knowledge gaps, this study aims to propose a robust decision-making model for the optimal planning of EVFCS powered by solar power using an advantage actor-critic (A2C) deep reinforcement learning (DRL) approach. A reinforcement learning-based decision-making model is devised to learn the complex urban environment using geospatial data (e.g., land use, potential electricity produced by PVs, traffic flow, existing EVCS vector maps). Without a pre-determined decision process, such as initial weights for criteria (i.e., parameters) or a deterministic model, the DRL model learns and optimizes decision-making parameters (i.e., policy and value) over episodes. In addition, by refining the model in testing various strategies, this research proposes the optimal planning for EVFCS in a city level. It also estimates potential charging stations' capacities from a large charging infrastructure network model. This research broadens the body of knowledge in optimal energy infrastructure planning on a city scale considering sustainable development, with a specific focus on fast-charging stations connected with renewable power systems. It also provides insight into how capacities of charging stations are balanced with charging demand and electricity supply. The results of this research can contribute to exploring EVFCS planning options in a complex urban environment using a comprehensively learned model and promoting electric transit for urban sustainability.

The remainder of this paper is organized as follows. The literature review section reviews previous studies on EVCS placement problems explaining the knowledge gaps of (1) a constrained search area (e.g., district level), (2) a lack of robustness in geospatial analysis-based decision-making approaches, (3) consideration of installing a small number of charging stations, and (4) a deficiency in integrating the distribution of solar energy with site selection of EVFCS. The methodology section describes the research objective, introducing the investigation of optimal placements of public electric vehicle fast charging stations while considering complex spatial patterns in map-type data at an urban scale. The results section presents a case study and the optimal sites for EVFCS by examining two strategies and six scenarios. The discussion section analyzes the results and discusses any limitations identified, Lastly, the conclusion section summarizes the major findings in our work, outlines contributions, and suggests future works.

2. Literature review

Previous studies for EVFCS site selection approaches were reviewed in this section. Key decision-making models from these studies were summarized in Table 1, presenting decision models, the number of charging stations, range of target area, utilization of renewable resources, and criteria or data used.

2.1. Multi-criteria decision-making approach for EVCS planning

MCDM approach has often been used to identify optimal places for EVCS. For example, Rani et al. (2021) (Rani & Mishra, 2021) stated that EVCS location selection problems can be solved by prioritizing candidate sites. The priority was evaluated by analyzing the interrelation among various parameters (e.g., environment, economy, social, and technology criteria). Interrelation analysis is calculated depending on the evaluation of weight of criteria (i.e., parameters), using various MCDM approaches, such as analytic hierarchy process (AHP) (Guler & Yomralioglu, 2020), fuzzy AHP (Tripathi et al., 2021), linguistic entropy weight (LEW) method and fuzzy axiomatic design (FAD) (Feng et al., 2021).

For example, Rane et al. (2023) implemented EVCS placement suitability problem using the integration of GIS and MCDM techniques (e.g., multi influencing factor (MIF) weights and technique for order preference by similarity to ideal solution (TOPSIS)). They estimated the priority between factors that distance to roads, proximity to commercial offices, and distance from bus depot are the most significant features in mapping the site suitability classes (not suitable to very high suitability). Yu et al. (2022) proposed the optimal EVCS placement in urban road networks using the TOPSIS decision-making model with traffic, charging demand, and land price. They suggested specific areas for an optimal solution through the road segment technique, differing from previous MCDM approaches, which is mapping the suitability. Panah et al. (2022) investigated the best MCDM techniques among normal pairwise comparison (equal weighting (EW)), hesitant fuzzy independent (HFIJ), AHP, and hesitant fuzzy AHP, using crow search algorithm (CSA). They selected three candidates on the IEEE 69-bus test system, consequently four different MCDM techniques showed disparate results depending on the experts' opinions and the weighting coefficients. These approaches can be used to assess both quantitative and qualitative parameters. However, there are three critical challenges of previous MCDM approaches for a robust site selection of EVCS: (1) limited search area: only a small number of charging stations in candidate sites (i.e., alternative) were explored as a case study, neglecting a myriad of potential locations on a large city scale, (2) lack of robustness: the solutions highly depend on initial criteria and weights, often collected from subjective judgments of experts, and (3) lack of model transferability: MCDM-based site selection approach is difficult to apply to different cities due to complexities of requiring a wide variety of data.

2.2. Path planning for placement

A path planning approach has frequently been utilized to investigate optimal electric vehicle charging facilities. The goal of path planning for solving charging station placement problems (CSPP) is to investigate the optimal sites where the charging demand is maximized (Kuby & Lim, 2007). The charging demand depends on various factors such as EV

Table 1
Summary of previous studies on EVFCS placement.

Author	Model	The number of charging station	Range of target area	Utilization of renewable resources	Criteria or Data used
(Erbaş et al., 2018)	GIS-based MCDM Fuzzy AHP and TOPSIS	12 EVCSs	8 districts in Ankara (Turkey)	No	• Environment: distance to vegetation, water resources and landslide risk, slope, possibility of expansion, and earthquake risk; • Urbanity: service area population, and proximity to junctions, main roads, the substation, petrol station and other EVCS; and • Economic: land cost, EV ownership in the service area, and distance to power cut
(Liu et al., 2018)	MCDM DEMATEL and UL-MULTIMOORA	9 EVCSs	4 districts in Shanghai (China)	No	• Environment: destruction degree on vegetation and water, waste discharge, and air pollutants reduction; • Economy: construction cost, annual operation and maintenance cost; • Society: harmonization of EVCS with the development planning of urban road network and power grid, traffic convenience, service capability, and adverse impact on people's lives
(Kaya et al., 2020)	GIS-based MCDM AHP, PROMETHEE, and VIKOR	100 EVCSs	38 districts in Istanbul (Turkey)	No	• Economic: EV numbers, Number of vehicle, land cost, and household income; • Geographical: forest, water resources, landslide, earthquake, and slope; • Energy; current EVCS, petrol station, solar energy potential, and substation; • Social/Environmental: air quality, service area population, and social areas; and • Transportation: road, junction, and parking lot
(Zhao & Li, 2016)	MCDM FDM, fuzzy GRA-VIKOR	5 EVCSs	5 districts in Tianjin (China)	No	• Economy: total construction cost, internal rate of return, and annual operation and maintenance cost; • Society: impact on living quality in service area, service capacity, traffic convenience, and coordinate level of EVCS with urban development planning; • Environment: deterioration on soil and vegetation, atmospheric particulates emission reduction, and GHG emission reduction; and • Technology: substation capacity permits, power quality influence, and power grid security implications
(Liu et al., 2020)	MCDM FDM, GRA-BWM, and EWM	5 EVCSs	–	No	• Economic: consumption level, public facilities, return on investment, construction investment cost, and operating and management costs; • Environmental: emissions of GHG, Destruction of soil and vegetation, garbage handling convenience, population intensity, and substation capacity; and • Traffic: distance to the substation, power quality, terrain advantage, road patency, service capacity, number of roads, main road number, and service radius
(Kuby & Lim, 2005)	Path-planning FRLM, MIL, heuristic	25 EVCSs	Main roads in Florida (USA)	No	Traffic flow volume
(Riemann et al., 2015)	Path-planning AC-PC FRLM, MNL	Less than 12 EVCSs	–	No	Traffic flow, drivers' routing choice behavior
(Pan et al., 2020)	Path-planning	5 EVCSs	Main roads in Beijing (China)	No	EV drivers' existing activities, home and public charging availability, range anxiety, and the energy consumption of remaining trips, drivers' trips, and travel demand
(Petratos et al., 2021)	Agent learning DQN	Maximum 26 EVCSs in one county	3 counties in New York (USA)	No	Existing charging station location, traffic volume, and POI information
(Jordán et al., 2022)	Agent learning GA	50, 100, 200 EVCSs	Valencia (Spain)	No	Population, traffic, social network, cost station, cost charger, cost distance energy, energy radius, influence radius, and station role in SimFleet software
(Liu et al., 2023)	Path-planning & Agent learning PPO (DRL)	3, 13, 29, 30, 113 EVCSs	5 main roads in Stanford and Culver city (USA), Queenstown city (Singapore), Cambridge city (UK), and Rouen (France	No	Road network coordinates, road types, and number of lanes; existing charging station information: location, number of chargers, charging capacity, charging cost and charger prices

users' charging behavior, location of the charging station, and charging speed (e.g., fast, slow) (Chaudhari et al., 2018). To investigate the charging locations based on the estimation of charging demand, researchers have used the path planning-based charging station location model (e.g., flow-capturing location model (FCLM) (Kuby & Lim, 2005), flow-refueling location model (FRLM) (Lim & Kuby, 2010; Riemann et al., 2015), agent-based model (García-Magariño et al., 2018)). Path planning model assumes that the refueling demands occur within vehicle flows on the way from origin to destination. Kuby & Lim (2005) applied the FCLM proposed by Hodgson (1990) to investigate optimal location of refueling spots for alternative vehicles (e.g., fuel cells, natural gas). Also, they extended FCLM, which assumes the driver will stop only once along their path, to FRLM, which can consider multiple refueling behaviors on longer trips with a combination of facilities along

the shortest path in origin-destination pairs. Wu et al. (2017) proposed a stochastic flow-capturing location model to estimate the number of EVs in a trip chain, which represents a sequence of trips per day. They determined the placement of charging station where EV charging demands are maximized. Moreover, researchers have enhanced these path planning models by aggregating the users' charging behavior and integrating them with other optimization algorithms (e.g., clustering approach, agent-based algorithm). Pan et al. (2020) proposed the coverage location model for public EVCS where missed trips, the driver fails to accomplish a trip, are minimized using a genetic algorithm (GA) based optimization model. To compute the optimization problem of minimization of missed trips, they constrained charging behavior and energy consumption. Specifically, as a case study, their model has been focused on 395 traffic analysis zone TAZs (i.e., average size of 0.76 km^2) in Beijing Municipality and suggested three spots among 50 to 200 potential spots as the best places for the public EVCS. Li et al. (2021) implemented the public EVCS location problem using improved GA considering the investment of CS operators and the travel cost in the city of Shanghai. They simplified searching the area to the road network, which assumes that candidate CSs are distributed to each node of the road network, considering 26 to 44 EVs with estimating the size of CSs.

Although extensive studies of path planning-based CSPP have been conducted to estimate charging demand and determine optimal locations where greater demand is expected, previous path planning based CSPP models limit the search range by depending on a network/node-based heuristic architecture. While their model can easily identify areas with high charging demands, it encounters difficulty when exploring areas beyond the network/nodes during investigations. In addition, previous studies, focused on the estimation of charging demand based on charging behavior, have not fully considered other crucial factors (e.g., geospatial factors).

2.3. Agent-based optimal placement of EVCS

Agent-based approaches (e.g., reinforcement learning, game theory) have recently been used to investigate the optimal placement of EVCS by optimizing the distribution of charging stations with finding the best number of charging stations, and coverage service of charging stations in the district scale (Petratos et al., 2021; Jordán et al., 2022). For example, Petratos et al. (2021) proposed supervised learning with Deep Q-Network (DQN) reinforcement learning (RL) to predict EV charging demand using existing charging stations data, traffic data, and points of interest (POIs) in the state of New York. Their algorithm examined placing a charger (maximum installation of 10 in one episode) in one grid cell (26×26, one grid cell is 250 by 250 m) estimating the charging demand of a potential CS and potential CS's service coverage. However, their agent could not find the areas including the highest charging demand, which constrained them to extend their model to larger regions. Also, they suggested an extension of their model by considering a DC Fast charging station. Jordán et al. (2022) explored the most suitable locations for EVCS using a genetic' algorithm. They simulated the impact of the placement of EVCS on the city at the district level while changing the number of charger points (e.g., 50, 100, 200) involved in the charging station (the maximum number of chargers per station is 10). Their model distributed the chargers by learning the balance of weights between the population (0.2), traffic (0.4), and network activity (0.4). Yang et al. (2020) optimized a charging plan by placing charging stations in the road network, which satisfies the condition that the benefit function (e.g., charging station capacity, influential radius, and coverage) is maximized and the cost function is minimized. Liu et al. (2023) applied proximal policy optimization (PPO)attention model to solve CSPP, which is similar to the path-planning approach, with estimating profit, cost, and fairness values. These values are computed relying on the influential radius of the CSs to EVs. Specifically, they constructed the environment of the road network data as graph-structured data format, expressed as nodes and edges to calculate three values, as well as to predict charging behaviors (e.g., waiting time, travel time, and charging time).

Most the agent-based approaches have focused on charging planning (charging scheduling) (Zhang et al., 2020; Li et al., 2022), not on CSPP, and have extended the path-planning method to consider more complex environments and to optimize the problem with more charging stations than previous path-planning method (Yang et al., 2020; Liu et al., 2023). Specifically, their common data format (i.e., graph-structured data) has been limited to investigating the optimal sites for public EVCS within the road network, but optimal sites are also outside the road network while satisfying other significant factors (e.g., solar power, land use). Also, although agent-based approaches can estimate the suitable areas in the high charging demand and sufficient potential CS's service coverages, which means that potential CS does not overlay with other potential or existing CSs, no one does estimate the potential charging stations' capacities and investigate the optimal sites, integrating the distribution of renewable resources (e.g., solar energy), produced by PV plants located near the target area, with it.

3. Methodology

This study designs and deploys a site selection decision model using deep reinforcement learning (DRL) to investigate optimal placements of public electric vehicle fast charging stations. The proposed optimal planning model considers complex spatial patterns for charging station placement in map-type data without predefined models or instructions. To estimate the capacities of charging stations and the number of charging stations, and to explore their placement across all areas at a city scale, we employ Advantage Actor Critic (A2C) Deep Reinforcement Learning (DRL) algorithms. The A2C-based optimal planning does not require a high computational load to solve the CSPP compared to other algorithms (e.g., general decision-making approach, path planning approach, deep Q-learning). A2C algorithm can be constructed by learning optimal decisions by themselves without heuristic or intuitive decisions, and it parameterizes the policy directly (on-policy) without requiring extensive experiences, such as a replay buffer. In this respect, this research proposes optimal decisions for the placement of public EVFCS with estimates of capacities and the number of charging stations in a complex environment.

Fig. 1 illustrates the overall procedure for optimal site selection of public EVFCS using A2C DRL algorithms, built using PyTorch package in Python. In the data generation and conversion step, four datasets are collected on Geographic Information System (GIS), ArcGIS Pro 10.3 presented by Esri: (1) land use, (2) existing charging infrastructure, (3) traffic flow, and (4) potential electricity produced by solar power. The spatial dataset is used to construct the environment for DRL. With the constructed environment, the proposed A2C DRL model learns geospatial patterns, enabling the agent (i.e., all public EVFCS in a city) to take the best actions (i.e., change in locations and capacities) in a direction of increasing reward values. It estimates the locations of public EVFCS sites, which can cover potential charging demands in all areas at the city scale. Finally, the site selection results are visualized using GIS to visually inspect the proposed model effectively learned spatial patterns through trial-and-error processes.

3.1. Model environment in a case study

3.1.1. Data generation

The city of Chicago, located in Cook County, Illinois, United States, was selected as the case study area and represented its environment in RL. Chicago is situated in the Midwest region of the United States, near the southwestern shores of Lake Michigan making it a central point for transportation between other cities and states, as shown in Fig. 2.

Chicago is the third-most populous in the United States and second most congested city, experiencing high levels of transportation-related CO2 emissions due to heavy traffic congestion and car dependency

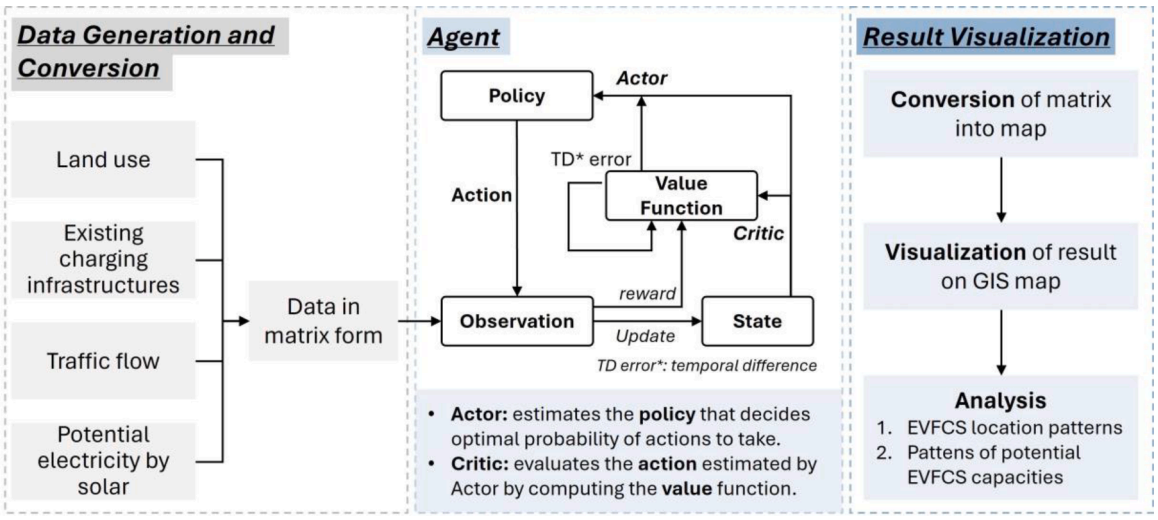


Fig. 1. Three-step framework for optimal site selection of EVFCS: data generation and conversion for input, A2C RL modeling, and result visualization on GIS.

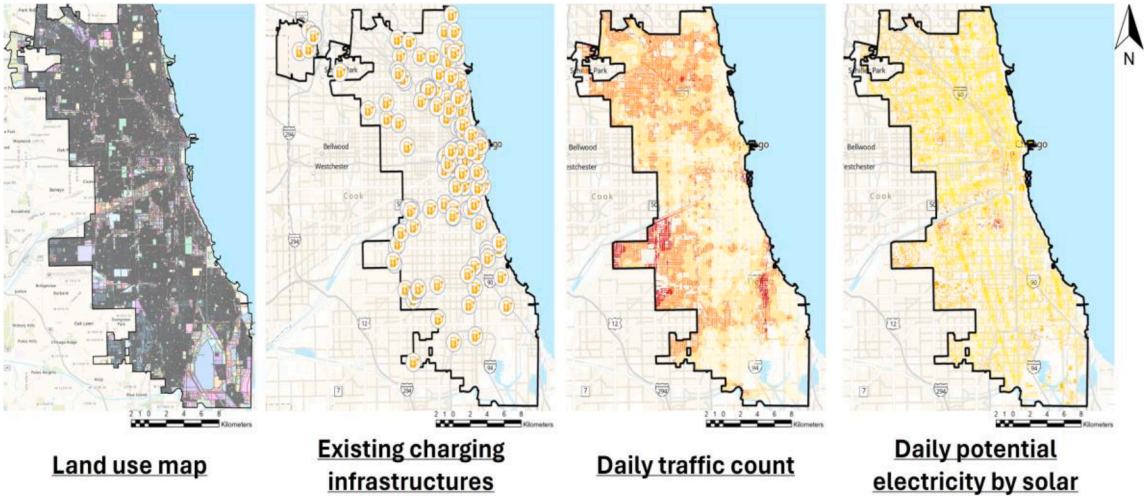


Fig. 2. Illustration of map data generation in case study area: Chicago, IL.

(Chris Gilligan, U.S. News & World Report, 2023). As an alternative way to alleviate air pollution and environmental emissions by vehicles, Chicago dwellers have transitioned to EVs. Despite a consistent increase in the number of alternative vehicles registered in Chicago and Illinois over the past six years, the city has been faced with limited charging infrastructure. Specifically, there are fewer DC fast chargers and public Level-2 chargers per million population (less than 30) compared to the size of the city (Bui et al., 2020). In addition, 70 % of Chicago dwellers live in multi-unit buildings without sufficient spaces for installing chargers, leading to persistent challenges in finding convenient and regular charging spots. Accordingly, this study explores the city of Chicago to determine the optimal locations for installing EVFCS using proposed approach without relying on training labels in datasets.

To construct the environment in the reinforcement learning, this study utilizes four types of geospatial data: land use map, existing charging infrastructure, daily traffic count information, and daily potential electricity map. All data were digitized into a spatial map type using a uniform coordinate system (North American Datum (NAD) 1983 Universal Transverse Mercator (UTM) 16 N) and the same resolution (10 $\times$ 10 m^2).

Previous studies have presented various socio-technical factors such as distance to vegetation, deterioration on soil, and vegetation in environment criteria, construction, operation, and maintenance cost in economic and proximity to the main road network, and service capacity in society, urbanity, transportation, and traffic criteria, as shown in Table 1. Among these factors, this study selected four data types by calculating the frequency of use in EVCS site selection papers. In previous studies, land use considerations included vegetation, water resources, green areas, and available spaces (such as parking lots) for installing charging stations within the environmental criteria. Also, existing Electric Vehicle Charging Stations (EVCS) and charging demand were factors considered, taking into account service levels, traffic convenience, population density, and intensity as referenced in studies (Erbaş et al., 2018; Liu et al., 2018, 2020; Zhao & Li, 2016; Rane et al., 2023), and (Feng et al., 2021). Charging demand was calculated based on historical Electric Vehicle (EV) numbers from studies (Erbaş et al., 2018; Kaya et al., 2020), and (Hisoglu et al., 2023), along with traffic conditions and traffic count data from references (Kahraman & Gündogdu, 2021) and (Kaya et al., 2020). Although potential solar energy was not considered in the previously reviewed studies, it presents a promising source to meet the increasing demand for EV charging while reducing CO_2 emissions from electricity generation without overburdening the national electricity grid. In addition, other important factors, not explicitly collected during data generation, were

incorporated into decision-making within our reinforcement learning-based model, such as proximity to main road junctions and service areas.

First, land use is utilized to identify suitable regions for charging station installation based on land use codes, excluding areas deemed unsuitable, such as miscellaneous (water resources), agricultural (vegetation), institutional, and limited government (military) zones (Liu et al., 2020). For example, Erbaş et al. (2018) and Liu et al. (2020) indicate that distance to vegetation and water resources may have been adversely affected by the installation of EVCS, and thus its installation should be far away from natural resources. The location of EVCS should be closer to commercial centers due to mobility and sociality (Kaya et al., 2020). In this study, the land use map data were obtained from (Chicago Metropolitan Agency for Planning (CMAP) 2018) in polygon-based GIS format. The polygon-based land use data was converted into point map to align it with all other data as grid format. With the goal of investigating optimal locations for public fast-charging stations, residential areas were excluded as potential sites in this study.

Existing charging infrastructure may have an adversary effect to determine the installation of new EVCS (Erbaş et al., 2018). Considering the existence of existing charging stations can prevent the division of charging demands which might result in a deterioration of profitability (Zhao et al., 2020). The existing location data were obtained from U.S. Department of Renewable Energy (2023) in point data format.

High-charging demands have a direct impact on the investigation of the optimal placement of EVCS because EV users may visit potential EVCS sites to charge their vehicles (Kaya et al., 2020). This research assumes that numerical value of charging demand is represented using daily traffic count data. Current traffic count can serve as an indication of future EV usages. These data were obtained from City of Chicago (2006) in point map format. To represent the traffic flow data aligned with the road, point format data is expanded to polyline format using the buffer function in GIS.

Lastly, high capability of energy supply directly influences the placement of EVCS because EVCS, especially EVFCS, requires a large amount of electricity to charge EVs rapidly (Kaya et al., 2020). Specifically, it is highly important to install EVFCS at which high renewable resources are available to afford the increased electricity load by new charging demand (Ali et al., 2022). This potential electricity data generated from rooftops of buildings were obtained from Google, Google Project Sunroof (2019) in point map format, organized by a census tracts. These point map-based data are interpolated into raster images, which are then extracted to building shape. Overall, the four datasets: land use, existing charging infrastructure, traffic flow, and potential electricity by solar were overlaid onto a grid format, uniformly constrained to the size of Chicago city.

Overall, geospatial data encompasses time-based characteristics of information or features along with location. This data is used to construct an environment and interact with the reinforcement learning model. The environment is constituted by combining all four geospatial data types into a high-dimensional space. The deep reinforcement learning model analyzes this environment by extracting spatial features, which capture correlations among the data or discover hidden information within the dataset.

3.1.2. Data conversion

During computation within a reinforcement learning environment, the latitude and longitude coordinates of data points are transformed into a rows-and-columns matrix coordinate format through two steps: (1) converting latitude and longitude coordinates to x, y extent (from vector format to tabular format), and (2) converting x, y extent to row and column matrix coordinates, as depicted in Fig. 3.

Initially, raw datasets in vector format are represented using NAD 1983 UTM 16 N, which contains latitude and longitude coordinates. Latitude-longitude coordinate systems are defined as angular units calculated by the intersection of a line orthogonal to the Earth's surface at a specific point and the plane of the Equator. However, using angular units makes it challenging to display the data on a two-dimension (2D) or three-dimensional (3D) map. In this research, the latitude-longitude coordinate system is converted into x, y extent coordinates. The (x, y) coordinates describe the data as points with meter/mile units on a geographical boundary, allowing the map to be bounded like a square frame. For example, the data (444,161, 4,636,747) with a value of 6676 (kWh) represents electricity usage of 6676 kWh located at (444,161, 4,636,747) meter (m) on the map. These (x, y) coordinates are useful for overlaying multiple layers with the same coordinates and sizes. The converted data is stored in tabular format, including (x, y) coordinate values and variable values (e.g., potential electricity, traffic count). Then, to create the environment as a grid-based Markov game in reinforcement learning, the (x, y) coordinates in tabular format are converted into matrix format using Eq. (1), in the GeoPandas package, Python.

$$(x,y) = \begin{cases} Row: \ ((\max x - \ \max y) - \text{point}(y)) \\ Column: \ (\text{point}(x) - \ \min x) \end{cases} \quad (1)$$

where point (x, y) denotes the data coordinate, and (min x, max x, min y, max y) denotes the boundary of the map.

In more detail, the distance between the maximum and minimum values of x corresponds to the length of a column and the distance between the maximum and minimum values of y corresponds to the length of a row. For example, the upper left corner of the boundary map is

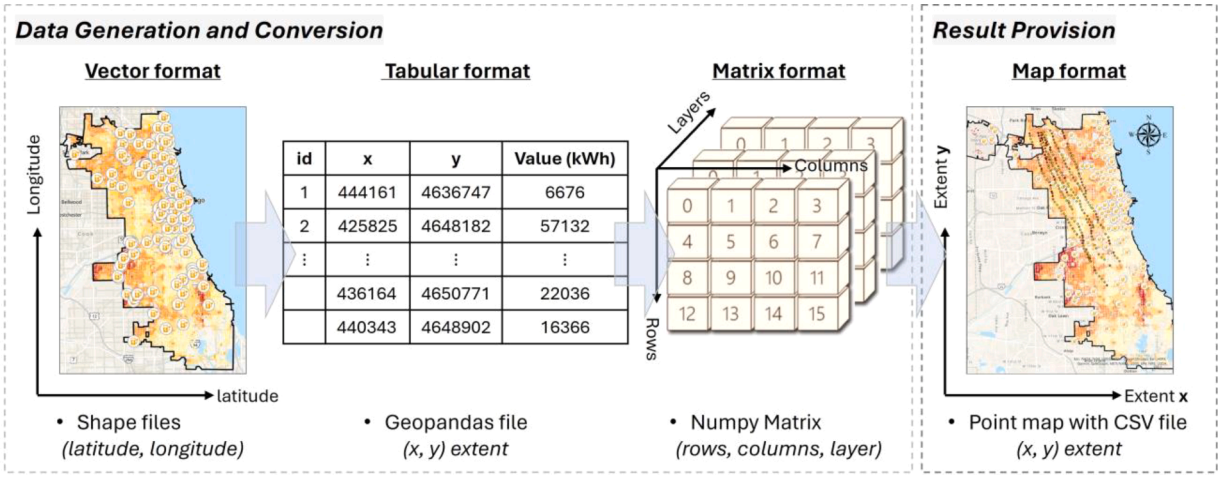


Fig. 3. Map data conversion process for the reinforcement learning environment.

represented as (0,0), and the lower right corner is represented as (4207, 3420) using Eq. (1). Moreover, the results from the DRL with converted (row, column) coordinates are transformed back to their original coordinates (x, y) to visualize the results in a GIS environment.

3.2. Reinforcement learning architecture

As a subfield of machine learning, reinforcement learning refers to learning to optimal decision-making to control an environment (Szepesvári, 2022). In classic reinforcement learning, an agent interacts with an environment over a series of discrete steps t (Sutton & Barto, 2018). At any time step t, the agent (charging station) receives a state s_t (the current position of charging station (x, y), the current capacity of charging station, land use, existing charging infrastructure locations, traffic density information, and potential electricity information as matrix format) and selects an action at (next position of potential charging station (x, y), and the capacity of potential charging station) from a set of possible actions A (action space), based on policy π_a, a probability distribution over actions given states. In return, the agent receives the next state s_{t+1}, following the transition dynamics (P_s), refers to the probability of an agent transitioning from one state to another state, and receives a scalar reward r_t, it is detailed in sub-Section 3.2.4 Reward function, until the agent reaches a terminal state or goal. The return G_t is the total accumulated reward sequences R_{t+k+1} at time step k, with a discount factor γ, a real value $\in (0,1]$, defined as $\sum_{k=0}^{\infty}\gamma^k r_{t+k+1}$ in each episode. The agent's goal is to maximize the expected return from each s_t, computed using Markov decision process (MDP) $\mathbb{E}[R_{t+1}|\{S, A, P_s, R_t, \gamma\}, \pi]$.

Reinforcement learning algorithms are assessed for their performance and precision, in terms of future rewards and expected return using a value function. The value function is to measure potential future rewards from being in state, defined by the state-value function $v^{\pi}(s_t) = \mathbb{E}\left[\sum_{k=0}^{\infty}\gamma^k r_{t+k+1}\middle| s_t\right]$ as the expected cumulative reward in a specific policy achieved by agent in state s_t and the action-value function $Q^{\pi}(s_t, a_t) = \mathbb{E}\left[\sum_{k=0}^{\infty}\gamma^k r_{t+k+1}\middle| s_t, a_t\right]$ where $\mathbb{E}_{\pi}[\cdot]$ denotes the expected return given that the agent follows a policy π and at any time step t. However, many of the tasks are combinatorial and highly compositive. In such cases, traditional learning algorithms, summarized in Table 2, cannot estimate to find an optimal policy or the optimal value function in the limit of time and data (Arulkumaran et al., 2017). Specifically, it cannot appropriately handle the state accurately that is not included in the dataset.

As the environment has grown in complexity with the increasing sophistication of models, recent reinforcement learning approaches have incorporated the policy gradient method in conjunction with neural networks. Deep reinforcement learning (DRL) approximates large state and action spaces to generalize it by effectively extracting features in a complex environment. In the mathematical description, DRL is to estimate action-value function by using the Bellman equation $Q^*(s,a) = \mathbb{E}_{s\sim\epsilon}\left[r + \gamma \underset{a'}{max} Q^*(s',a')\middle| s,a\right]$ that takes the best action (a') maximizing the expected cumulative return value ($r + \gamma Q^*(s',a')$) (Mnih et al., 2013). Specifically, DRL can be trained by updating the probability distribution of actions (i.e., minimizing a sequence of loss function at each episode) that can estimate higher expected rewards, expressed as Eq. (2)

$$J(\theta) = \mathbb{E}\left[\sum_{t=0}^{T-1}\nabla_{\theta}\log\pi_{\theta}(a_t|s_t)\sum_{t'=t+1}^{T}\gamma^{t'-t-1}R_{t'}\right] \tag{2}$$

The DQN model updates the loss function of stochastic gradient descent, which computes by comparing inputs and targets, expressed as Eq. (3).

$$\nabla_{\theta_i}L_i(\theta_i) = \mathbb{E}_{s,a\sim\rho(\cdot),s'\sim\epsilon}\{[r + \gamma \mathrm{max}Q^*(s',a'|\theta_{i-1})) - Q(s,a|\theta_i)]\nabla_{\theta_i}Q(s,a|\theta_i)\} \tag{3}$$

Specifically, Mnih et al. (2015) developed a novel agent (deep Q-network, DQN) that can address the fundamental instability problem in traditional DRL, combining Q-function and deep neural network. To solve the instability problem due to updating the full trajectory of state and action pairs at a time, the DQN utilizes experience replay (i.e., replay buffer), which samples the agent's training experience at any time step $e_t = (s_t, a_t, r_t, s_{t+1})$, to perform model updates. After sampling from the replay buffer, the agent takes an action according to an ϵ-greedy policy. Although DQN has successfully done simple tasks using the policy gradient method with neural network, there are some limitations (1) The DRL can lead to overestimation of action-value function, resulting in the inaccuracy in the policy (Mnih et al., 2016), (2) DRL cannot handle continuous action spaces due to limitation in architecture framework (Yang et al., 2017), and (3) experience replay-based learning requires heavy computation load to optimize the learning model (Liu & Yu, 2018).

3.2.1. Advantage actor critic (A2C) algorithm

To overcome these challenges in traditional DRL and DQN, recent studies have employed the actor-critic method with a baseline b_t, which compares the cumulative reward in the policy gradient as expressed by Eq. (4)

$$J(\theta) = \mathbb{E}\left[\sum_{t=0}^{T-1}\nabla_{\theta}\log\pi_{\theta}(a_t|s_t)(G_t - b(s_t))\right],\ G_t = \sum_{t'=0}^{T-1}\gamma^{t'}R_{t'} \tag{4}$$

The actor-critic is a hybrid architecture combining a policy-based method in the actor, which controls how the agent behaves, and a value-based method in the critic, which estimates what the best action taken by the Actor is. In the actor-critic method, the policy gradient ($J(\theta)$, actor loss) is represented by the subtraction of cumulative reward with a baseline ($G_t - b(s_t)$) in Eq. (4). From the Bellman optimal equation, Eq. (4) can be rewritten as the advantage function (A_t), which compares taking specific action to the average action at the given state using the Q value and V value, as shown in Eq. (5).

$$\nabla_{\theta}J(\theta) = \sum_{t=0}^{T-1}\nabla_{\theta}\log\pi_{\theta}(a_t|s_t)A_t,\ A_t = r_{t+1} + \gamma V_{\phi}(s_{t+1}) - V_{\phi}(s_t) \tag{5}$$

Table 2
Comparison of reinforcement learning methods: (1) traditional RL (MDP, Q-learning), (2) deep RL (REINFORCE), (3) off-policy (DQN), and (4) on-policy (A2C).

Methods	Feature	Advantages	Shortcomings
Traditional RL (MDP, Q-learning)	Model-based	Simplicity and interpretability	Limited scalability to large state/action spaces and requires explicit model
Deep RL (REINFORCE)	Policy gradients	Scalability for high-dimensional action/state spaces and handle continuous actions	High variance in training and less stable training compared to Q-learning
Off-policy (DQN)	Q-learning	Efficient of learning process from historical data using experience replay	Requires more computation load due to experience replay, can be sensitive to hyperparameters, and difficulty to find balance between exploration and exploitation
On-policy (A2C)	Actor-Critic	Balance between exploration and exploitation and more sample-efficient than other learning method	Slower convergence compared to off-policy method

Moreover, the A2C is approximated toward minimizing the loss $L(\theta)$, which is a combination of Actor and Critic losses for optimizing the model. Actor loss is calculated by policy gradient and the critic loss estimates the expected returns (G_t) for q-values using the mean squared error (i.e., squared L2 norm) expressed by Eq. (6)

$$L_{critic} = \left(\sum_{t'=0}^{T-1} \gamma^{t'} r_{t'} - V_{\phi}(s_t) \right)^2 \tag{6}$$

In contrast to typical deep learning approaches where the loss function compares predictions with ground truth values using error metrics (e.g., L1 loss, L2 loss) in deep learning, the loss function in the DRL optimizes not the reward function directly but the product of the estimated value function and the probability of the action.

The critic loss represents the estimation of the value of the state to minimize the temporal difference (TD) error, which is the difference between s_t and s_{t+1}, using gradient descent Eq. (7)

$$\phi = \phi + \alpha \delta \nabla_{\phi}(s) \tag{7}$$

where w denotes the set of parameters of the value function (critic network), α denotes the learning rate, δ denotes the TD error, and $\nabla_w V(s, w)$ denotes the gradient of the critic network.

Also, the actor loss determines the probability of the action selected by $\pi_{\theta(a|s)}$, as expressed by Eq. (8)

$$\theta = \theta + \alpha \delta \nabla_{\theta}(a|s) \tag{8}$$

where θ denotes the set of parameters of the actor network, and $\nabla_{\theta} \log \pi_{\theta}(a|s)$ denotes the probability of selecting action a by policy π_{θ}.

Advantage Actor Critic (A2C) Algorithm is presented in pseudo-code Algorithm 1.

The core advantage of the A2C method is that it is affordable to train deep reinforcement learning for continuous state and action space with model-free and on-policy. It can also resolve a massive and complex range of problems with lower variance and stability in performance because it is less affected by hyperparameters than traditional DRL.

3.2.2. Agent and action

The A2C DRL approach was adopted to address the challenge of placing EVFCS in complex spatial patterns within a large grid environment. The core elements in the A2C DRL are agent, action, state, and reward. The agent represents the locations and capacities of all public EVFCS in the city of Chicago. The agent interacts with the environment by taking optimal actions and responding to new situations (i.e., state) in the environment. The A2C DRL model explores the environment to decide on optimal sites for installing EVFCS and determining their sizes. The model is trained from the agent's experiences to maximize the cumulative rewards. The agent selects three types of continuous action on rows, columns, and capacities of charging stations. The row and column values are represented to the range from −50 to +50 where the agent moves to the next state (s_{t+1}) after investigating the site in the current state (s_t). For example, when the agent selects an action row of −30 and the action column of +30 at the current position of (2000, 3000), the agent moves into a new position of (1970, 3030) and investigates the sites at the position. The agent predicts the capacities of potential charging stations at any time step t, which ranges from 0 to 30,000 (kW). This research assumes that fast charging stations are equipped with V3, Telsa SuperCharger, which have maximum power of 250 kW and EVs are Tesla Model 3, which have a battery capacity of 65.6 kWh (80 % of full capacity). For example, when the agent predicts the capacity of EVFCS to be 30,000 kW, potential charging stations can accommodate a maximum of 457 electric fleets per day and are equipped with at least five chargers.

Although classic DRL algorithms can handle discrete action spaces that yield one output node for each action in the neural network, which simply takes the highest Q-value or a distribution probability of taking action, they face challenges when dealing with continuous action spaces. Naive discretization of the action space, attempted to mitigate the curse of dimensionality, may discard potentially crucial information needed to solve the problem, and result in inaccurate estimates (Lillicrap et al., 2015). On the other hand, the A2C algorithm implements a parameterized stochastic policy by sampling random actions from a parameterized Gaussian probability distribution, achieved through finding the definite integral within a range of actions. Therefore, the action space should be symmetrically normalized to [−1,1] to increase training speed and prevent divergence. This research adopts a normalized range of −1 to 1 for its three defined actions, as shown in Table 3.

Algorithm 1
A2C Algorithm.

//Assume parameter vectors: θ and ϕ
Input: A policy $\pi_{\theta}(a|s)$, a value function $V_{\phi}(s)$
Initialize: step counter: $t \leftarrow 1$
Initialize: episode counter: $E \leftarrow 1$
Initialize: random parameters θ, ϕ
Initialize: network gradients $\leftarrow 0$
For E = 1, M **do** ▷ $E_{max} = 1000$
  Get state s_t ▷ $t_{max} = 500, 1000, 1500$
  While not done **do**
    Take action a_t, according to policy $\pi_{\theta}(a_t|s_t)$
    Receive reward r_t, and new state s_{t+1} ▷ r_t computed from **Table 1**
    $t \leftarrow t + 1$
  End while terminal s_t or $t - t_0 = = t_{max}$
  $R = \begin{cases} 0, & \text{for terminal } s_t \\ V_{\phi}(s_t, \phi), & \text{for non} - \text{terminal } s_t \end{cases}$
  $R \leftarrow r_i + \gamma R$
  Set TD target: $Q_n(s_t, a_t) = \sum_{k=0}^{T-1} \gamma^k r_{t+k} + \gamma^n V_{\phi}(s_t)$ ▷ n-step target
  Advantage: $A_n = Q_n(s_t, a_t) - V_{\phi}(s_t)$
  Descent Advantage loss $\phi = \phi - \alpha \cdot \nabla_{\phi} \sum_t (A_n(s_t, a_t))^2$
  Ascent policy gradient: $\theta = \theta + \alpha \cdot \sum_t [A_n(s_t, a_t) \cdot \nabla_{\theta} \log \pi_{\theta}(a_t|s_t)]$
  Move to $a_t \leftarrow a_{t+1}$, and $s_t \leftarrow s_{t+1}$
  $E \leftarrow E + 1$
End for $E > E_{max}$

3.2.3. State

A state in this model represents the current environment that the agent is in. Our environment encompasses vast areas in Chicago. To reduce the computational load and prevent out-of-memory errors during model training, we defined the state by extracting the environment space that retains important geospatial information around the agent, affecting the decision-making for the site selection of new charging stations. Specifically, the state space is cropped to include a region with 50 grids (e.g., 1 grid 100 × 100 m) in rows-column coordinates around the target EVFCS at each time step as shown in Fig. 4.

For example, the position of the agent (e.g., (1328, 1773) in the environment) is mapped to (49, 49) in the state, which we refer to as the window (i.e., state). The new window state is utilized to compute the weight parameters in the actor and critic neural network and to calculate the reward values, as explained in the next Section 3.2.4.

3.2.4. Reward function

The reward function is a core component in reinforcement learning for creating a robust decision-making model. The reward factors are defined as shown in Table 4.

The term R_1 evaluates potential sites based on land use classifications

Table 3
Description of action characteristics of action name, raw values and normalized values.

Action name	Raw values	Normalized values
Rows	−50 to 50	−1 to +1
Columns	−50 to 50	−1 to +1
Capacities	0 to 30,000	−1 to +1

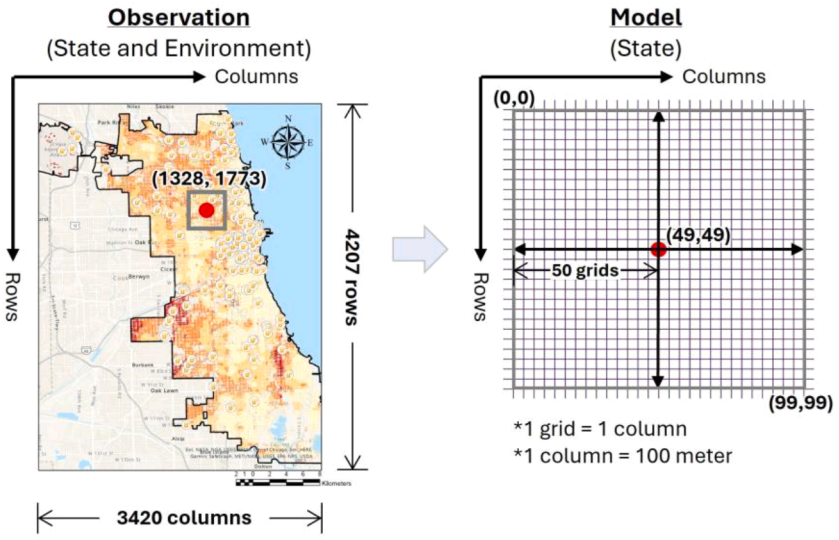


Fig. 4. Concept of state segmentation from the observation spaces (i.e., environment).

Table 4
The definition and equation of rewards.

Rewards	Definition	Reward function
R_1	Available land use type for EVFCS installation	$R_1 = \begin{cases} \text{available site} : +2 \\ \text{unavailable site} : -4 \end{cases}$
R_2	Existence of existing charging infrastructures	$R_2 = \begin{cases} \text{null} : -1 \\ 0 < \text{distance} < 50 : +2 \\ \text{distance} \geq 50 : 0 \end{cases}$
R_3	Traffic density	$R_3 = \begin{cases} \text{high traffic flow } (0.66 \sim 1) : +2 \\ \text{medium traffic flow } (0.33 \sim 0.66) : +1 \\ \text{low traffic flow } (0 \sim 0.33) : -1 \end{cases}$
R_4	The balance of energy between charging capacity and potential solar power supply	$R_4 = \begin{cases} \text{capacity} \geq \text{energy supply } (\sim 250m) : +2 \\ \text{capacity} \geq \text{energy supply } (\sim 2km) : +1 \\ \text{otherwise} : -2 \end{cases}$
R_5	Prevention of duplicating the coverage areas of each EVFCS	$R_5 = \begin{cases} \text{distance between CS} \geq 250\text{m} : +2 \\ \text{otherwise} : -2 \end{cases}$

extracted from the land use map to determine if charging stations can be installed (Heo & Chang, 2024). For instance, when the agent is located in unavailable areas for charging station installation such as miscellaneous, agricultural, residential, institutional, and government-owned zones, R_1 incurs a penalty of -4. Selecting available areas such as commercial, industrial, and vacant areas, results in R_1 of $+2$. The term R_2 involves assessing the proximity of existing charging stations to potential sites by calculating the distance between the agent's position and the positions of existing charging stations. If existing charging stations are within a 5 km radius of the agent's position, R_2 receives $+2$, otherwise, it incurs a penalty of -1 as the solutions can substitute existing charging stations. R_2 does not receive any rewards (0) when no existing charging station is nearby potential areas. The term R_3 computes relative traffic flow using Eq. (9) to compare it with the average of traffic volume in Chicago.

$$\text{Relative traffic flow} = \frac{\text{Traffic flow near to potential CS}}{\text{Average of traffic flow at Chicago}} \tag{9}$$

The agent receives R_3 of $+2$ for potential areas close to high traffic flow ($0.66 \sim 1$), $+1$ for medium traffic flow ($0.33 \sim 0.66$), and -1 for low traffic flow ($0 \sim 0.33$). R_4 addresses the balance of energy between the predicted capacity of charging stations taken by action and the potential supply, which is the sum of electricity generated from rooftop PVs within a 250 m radius of the agent's location. Tapping electricity from farther locations results in a voltage drop of up to 3.7 % over 275 m without considering the inverter [61, 62]. R_4 receives positive rewards ($+2$) when the charging capacity provided by potential charging stations is less than the potential supply produced by adjacent rooftop PV systems. If electricity production is sufficient compared to the predicted capacities of charging stations, the investigation area expands to 2 km $\times$ 2 km to calculate the summation of electricity production. R_4 then receives $+1$ if the charging capacity is still lower than the potential supply; otherwise, it receives -2. Finally, the term R_5 penalizes duplicate installations of charging stations. As the agent investigates new potential areas over s_t, the challenge arises when new charging stations are installed in the same areas that satisfy all other factors (e.g., R_1 to R_4). R_5 receives a penalty of -2 when the distance between new charging stations is less than 250 m.

4. Results

Before training the model, we explored various combinations reward functions to determine the best reward functions. We initially calculated the sum of all rewards without weighting. However, this calculation resulted in underestimation in the training model, because all rewards do not contribute equally in the constitution of the reward function. To resolve this issue, the reward function is designed to be like a regression model. The best regression-shaped reward function was found manually by trial and error, as a result, Eq. (10) outperformed other combinations.

$$\text{Reward function} = R_1 - 0.1 * R_2 + 2 * R_3 - R_4 + R_5 \tag{10}$$

It implies that distance to existing charging stations (R_2) does not have a significant impact on EVFCS placement problems, while traffic density (R_3) greatly influences site selection investigation. Specifically, R_2 negatively impacts the total reward, which performs better during the training process. This indicates that the proposed model prevents the charging stations from overlapping their service areas with each other. In addition, the energy balance between potential electricity production and the capacities of potential charging stations (R_4) negatively impacts the training model, suggesting that the demand of charging station capacities should exceed electricity production to make the placement of charging stations attractive.

The trained model is employed to explore EVFCS locations under two strategies. The first strategy involves the selection of "starting points". The position of the starting point influences how the agent explores

suitable regions for the installation of EVFCS in a vast and unknown environment. For instance, if the starting point is in high-parameter areas (e.g., near heavy traffic density), resulting in high reward values, the model can be trained more easily compared to starting in low-parameter areas. This study employed two types of starting point methods: (1) fixed starting point and (2) random starting point. The fixed starting point is positioned at specific positions (i.e., the center of the environment space), where the installation of charging stations can be available. This study defined the coordinate of the fixed starting point as (1982, 1196) in matrix coordinate format. On the other hand, the starting point is randomly changed at the beginning of each training episode. In the random starting point method, the starting point is selected from available areas extracted from the land use layer in the environment. This method offers the advantage of self-validating the training model without additionally testing the model.

The second strategy involves the "number of charging stations". The charging stations obtain electricity from nearby solar PV systems which can potentially be installed on the rooftops of buildings. As the number of charging stations increases, the demand for electricity also rises. In particular, when charging stations are too close together, it can overload the district distributional system. This study investigated an optimization model by testing different numbers of stations (e.g., 500, 1000, 1500), which in turn alters the time steps of each episode in the training model.

Overall, six scenarios were tested by simulating 1000 episodes, varying the combination of starting points and the number of charging stations, as shown in Table 5. The performance of the scenarios, measured by training rewards (total cumulative rewards), is presented in Fig. 5.

In addition, the proportion of satisfied rewards from all charging stations (e.g., 500, 1000, 1500) was computed, as presented in Table 6. The proportion was calculated as the ratio of each episode that satisfies the criteria of each reward explained in the reward function section at all EVFCS (i.e., 500, 1000, 1500 as time steps). For example, 100 % of R_1 in the fourth scenario represents that all of the investigated charging stations (i.e., 500) were installed in available areas.

The results demonstrate that our algorithm effectively solves EVFCS placement problems, regardless of the starting point. For example, the training rewards of Scenarios 4, 5, and 6, based on random starting points were slightly lower than those of Scenarios 1, 2, and 3, based on fixed starting points. It implies that our model may be stable even when trained randomly over episodes. Although more charging stations might provide the model with more opportunities to learn and explore, potentially improving learning, this study observed that the stability of learning decreases when there are too many charging stations (1500). In Fig. 5, it is shown that the algorithm lost previous experience in Scenarios 3 and 6. This is because this study designed the agent to represent "all public EVFCS in a city". In this respect, introducing more charging stations could increase the complexity of the environment, leading to more variability of the agent's experience. This increased variability could potentially make learning less stable if the agent struggles to generalize effectively from its experience.

The composite reward function is used to compute the expected discounted return for exploring the optimal policy. However, this approach may pose challenges for RL users in understanding which reward factors influence the decision-making process. To analyze how the model balances reward factors, our research decomposes the composite reward function into individual components represented by R_1 to R_5 factors, as shown in Table 6.

Table 5
Six training scenarios by the combination of two strategies: (1) starting point and (2) the number of charging stations.

Scenario	Starting point	The number of CS (t)
1	Fixed point	500
2	Fixed point	1000
3	Fixed point	1500
4	Random point	500
5	Random point	1000
6	Random point	1500

The satisfaction percentage of reward factor R_1 is consistently the highest (e.g., 100 % in all Scenarios) among the reward factors. This indicates that the agent successfully selects viable sites for EVFCS such as commercial areas or parking lots, while avoiding unsuitable locations like vegetation or military zones. In contrast, the satisfaction percentage of reward factor R_2 is the lowest (e.g., 14 % in Scenario 5) among all scenarios. This suggests that existing charging infrastructures may not significantly influence site selection, possibly conflicting with other reward factors or negatively correlating with potential fast charging station locations. Regarding reward factor R_3, the agent prioritizes available sites that satisfy high traffic density (i.e., charging demand). For R_4, the proximity to potential electricity generation within a 2 km radius of the target charging station should exceed the predicted charging station capacities evaluated in the third action. Lastly, duplication of coverage areas for each potential EVFCS is prevented through reward factor R_5, except in the sixth scenario. This implies that our model exhibits a bias where the total rewards acquired by densely installing 1500 charging stations surpass those obtained by preventing duplication between potential EVFCS. This bias issue did not arise in a fixed starting point-based learning model, which can repeatedly train in the same regions over 1000 episodes without generating such biases.

Overall, the use of random starting points and 1000 charging stations (Scenario 5) generally yields higher rewards for all reward factors except for R_2 compared to Scenarios 4 and 6. Although Scenario 3 (fixed starting point with 1500 charging stations) outperformed Scenario 5 in terms of the highest point of cumulative rewards, we selected the model from Scenario 5 as the best scenario. This decision was based on the higher stability of the model, characterized by less variability in the agent's experience, compared to Scenario 3. Scenario 5 consistently offered stable as well as superior performance when applied to any region without model validation, as shown in Fig. 5 and Table 6.

Scenario 5, identified as the best scenario based on the above results, is utilized to evaluate the learning performance by visualizing the training actor and critic losses over episodes, as shown in Fig. 6(a) and (b).

As shown in Fig. 6, the loss graphs depict an observation wherein the action loss consistently surpasses the critic's estimated loss, implying a positive error in the actor loss. Such disparity suggests that the actions chosen by the actor are significantly deviating from what the critic thinks would lead to higher rewards. This situation indicates that the actor's policy does not well align with the critic's evaluation of the state-action values. To improve the learning process and encourage the actor to select better actions, it is necessary to reduce the actor loss throughout the episodes. In our learning process, we observed a singular loss point where the loss function experiences a significant change or anomaly around episodes 750 ~ 800. This anomaly suggests the possibility of encountering a local minimum during optimization. Given that our learning problem is an unconstrained optimization problem, with multiple potential solutions, such singular points are expected. Therefore, to mitigate the risk associated with getting stuck in a local minimum, we opted to select the model from Scenario 5 at episode 720 as the best option. We anticipate that this selected model will effectively address potential convergence issues and uphold the robustness of the model's performance

To visually investigate the results of potential charging stations from the proposed model, we examine the solutions along with important factors such as traffic flow in Fig. 7(a), potential electricity in Fig. 7(b), and estimated capacity in Fig. 7(c).

Fig. 7(a) shows the potential EVFCS placement with traffic flow categorized into five travel density classes. The color scheme denotes varying traffic densities, with light yellow representing lower flow (700 ~ 1500) and dark brown indicating higher flow (74,000 ~ 165,000) on

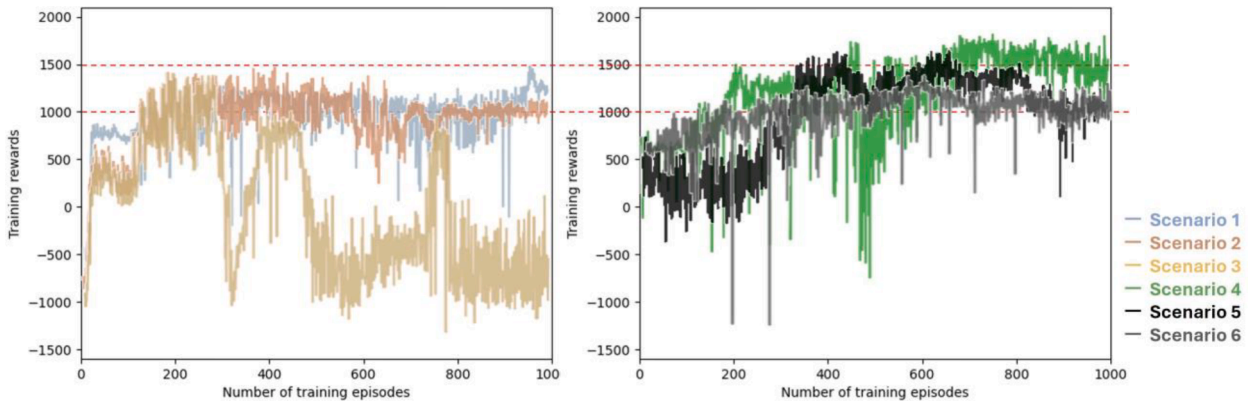


Fig. 5. Training results of six scenarios.

Table 6
Learning performance of each reward factor for six training scenarios.

Scenario	R_1	R_2	R_3	R_4	R_5	$R_{\text{all except for } R_2}$
1	100.00 %	19.40 %	82.80 %	76.40 %	83.20 %	61.80 %
2	100.00 %	6.00 %	83.80 %	67.60 %	68.20 %	48.90 %
3	100.00 %	7.33 %	99.00 %	96.00 %	84.33 %	81.33 %
4	100.00 %	9.20 %	98.00 %	70.00 %	89.90 %	68.10 %
5	100.00 %	14.00 %	87.00 %	76.30 %	90.40 %	74.30 %
6	100.00 %	6.87 %	100.00 %	29.47 %	0.33 %	0.33 %

a GIS environment. In Fig. 7(b), the spatial distribution of EVFCS is depicted alongside potential solar power availability, delineated by a color gradient ranging from high values (i.e., dark red (142 MW) to low values (i.e., light red (0.7 MW)). Fig. 7(c) represents the estimated capacities of potential EVFCS sites, one of the actions taken by the agent in the A2C DRL model. These capacities are classified into five classes, with light blue representing low capacity (14 ~ 20 MW) and dark blue indicating higher capacity (29 ~ 30 MW). The results highlight a spatial tendency of potential charging locations, particularly in the southeast region, characterized by high traffic flow and solar power output. This concentration in the southeast contrasts with the dispersed distribution observed in regions where either parameter is high, such as the southwest or northwest. These observations underscore the significant influence of traffic flow density and solar power availability on the determination of optimal EVFCS placement and capacity estimation.

Furthermore, the EVFCS placement map was analyzed to understand the spatial patterns of potential charging stations across six scenarios (Fig. 7 and Appendix A), which helps support the interpretation of learning performance (Table 6). The results revealed that the northern region of Chicago (scenarios 1, 2, and 4) could be another optimal site for EVFCS placement, characterized by high traffic flow but relatively lower energy supply. This suggests that alternative sites with high charging demand should also be considered. However, as depicted in Fig. A.3 and Fig. A.5, when considering a higher number of EVFCS installations compared to the charging demand and energy supply, installations tend to concentrate in specific areas rather than being distributed across suitable locations throughout the city. This finding underscores the importance of balancing installation numbers with demand and supply considerations to achieve more effective and equitable coverage of charging infrastructure.

The visualization map was further analyzed to understand the training progress over episodes, which can show the difference in results between an underfitted model Fig. 8(a), optimized model Fig. 8(b), and overfitted model (i.e., local minima problem) Fig. 8(c).

Over the episodes, the decision-making pattern regarding potential EVFCS placement evolves from clustering in specific areas (Fig. 8(a)) to spreading across general areas with a distinct pattern, demonstrating an inclination towards accommodating high parameters such as traffic flow density and potential electricity from solar (Fig. 8(a)). However, a singularity, observed between 750 and 800 (Fig. 6) leads to erratic behavior in the decision pattern, potentially resulting from overfitting induced by local minima.

Table 7 presents a statistical summary including minimum, mean, median, maximum, and standard deviation computed for further analysis of potential capacities in the fifth scenario. The distribution of

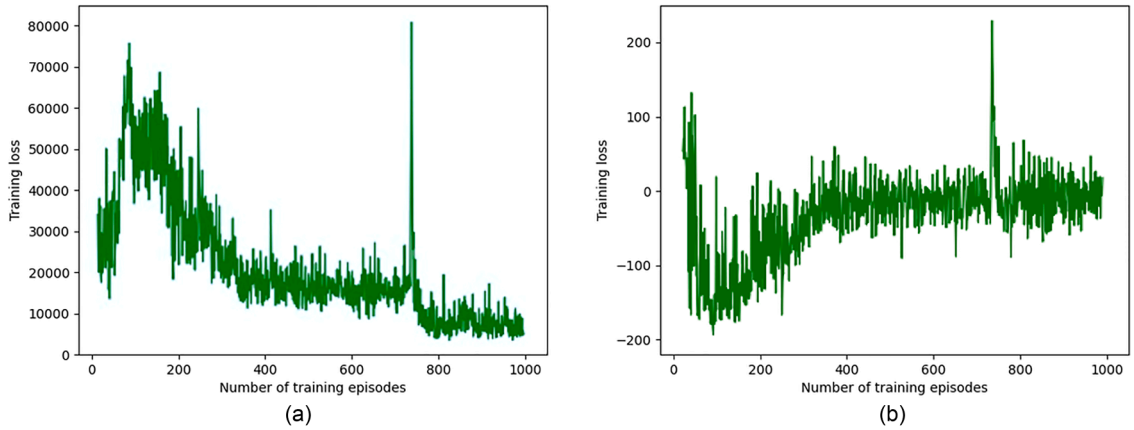


Fig. 6. Two loss graphs over the course of training for EVFCS placement problems: (a) actor loss graph and (b) critic loss graph.

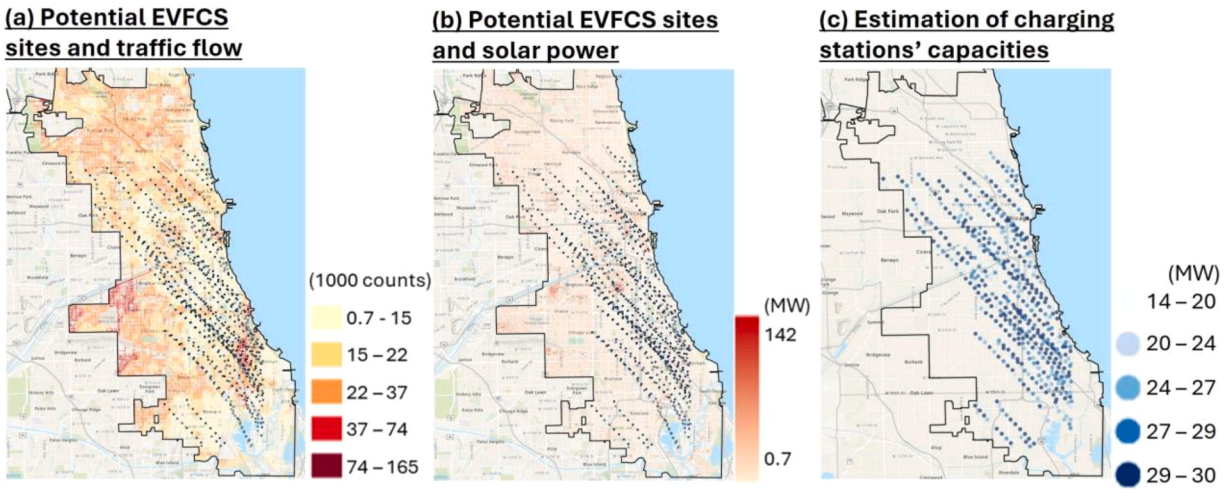


Fig. 7. Potential EVFCS locations and capacities plotted alongside (a) traffic flow, (b) potential solar power, and (c) estimated capacity.

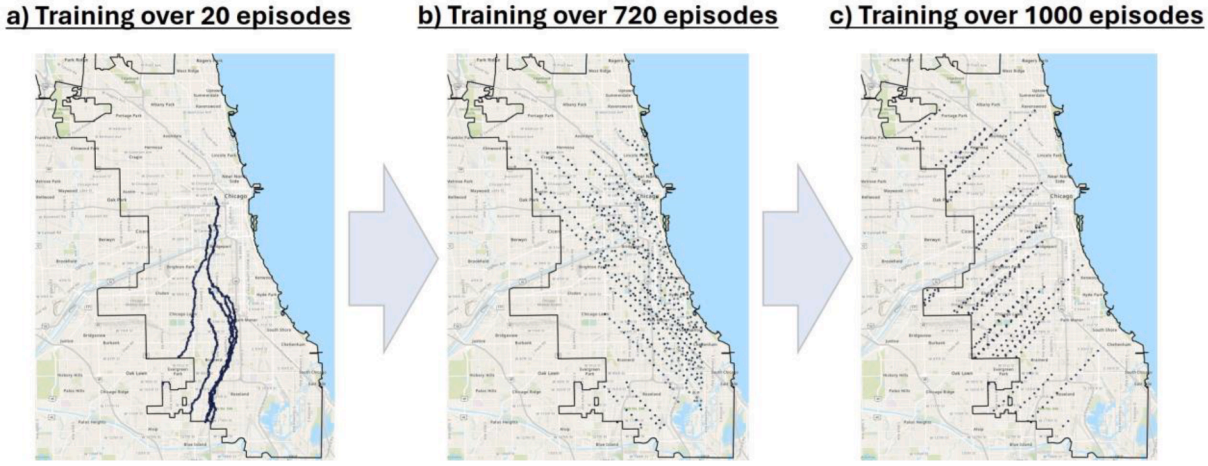


Fig. 8. Evolution of potential EVFCS placement over episode in Scenario 5.

Table 7
The statistics summary of evaluated capacities over episodes in the fifth scenario.

Episodes	Minimum	Mean	Median	Maximum	Standard deviation
20	11,179 (kW)	16,516 (kW)	16,500 (kW)	21,385 (kW)	1564 (kW)
700 (all result)	14,054 (kW)	27,242 (kW)	28,369 (kW)	30,000 (kW)	3127 (kW)
700 (except R_2)	14,054 (kW)	27,805 (kW)	28,993 (kW)	30,000 (kW)	3014 (kW)
1000	18,369 (kW)	27,678 (kW)	28,626 (kW)	30,000 (kW)	2681 (kW)

evaluated capacities may closely follow a Gaussian distribution under untrained conditions (Episode 20). This occurs because the agent's actions were normalized to a Gaussian distribution prior to learning. Therefore, during early training, the evaluation of capacities, one of the actions, followed this normalization while the policy was not yet properly trained. By the trained condition (Episode 700), the distribution of evaluated capacities shows no significant difference in statistical values compared to the original output, which includes all potential charging stations, versus the modified output that only considers satisfaction of all reward factors except for R_2. However, towards the end of training, an overfitting issue becomes apparent as the distribution of evaluated capacities becomes left-skewed towards the maximum value (30,000 kW). This suggests that the model struggles to take optimal actions and exhibits a biased decision-making process.

5. Discussion

The optimal planning of EVFCS is greatly influenced by geospatial information covering the city. This research proposes an A2C DRL model capable of making optimal decisions regarding the installation of EVFCS by evaluating the capacities of charging stations. Specifically, A2C DRL model can reflect elements from the previous methods discussed in the literature review: (1) geospatial analysis-based optimal decision-making in Multi-criteria Decision-Making (MCDM) techniques; (2) Evaluation of charging station capacities by estimating charging demand, which often addressed through path planning methods; and (3) Optimal planning of EVCS installation in complex environments addressed using agent-based optimal placement of EVCS. In contrast to path-planning placement approaches limited to road network investigations (node-based data structures), the geospatial analysis-based optimal decision-making approach allows for exploration of vast areas within complex city environments for EVFCS installation. This approach enabled our model to identify optimal sites for EVFCS across the entire city, outperforming similar research models (Petratos et al., 2021; Jordán et al., 2022) that

achieve faster convergence to optimization (in 700 episodes compared to 3500 episodes in Petratos et al. (2021)).

Furthermore, our model integrates the optimal EVFCS placement with distributed energy resources powered by solar PV plants, a novel aspect not studied in previous research. Unlike previous studies focused on correlating stationary spatial factors for optimal site selection, our model addresses the non-stationary EVFCS placement challenge in complex environments where fast charging stations utilize potential electricity generated by solar PV plants near target areas. This electricity utilization varies across episodes, significantly increasing the complexity of the learning process. To address this, our model controls EVFCS capacities as one of the agent's actions, balancing them with potential electricity generated by solar PV plants and traffic density (charging demand) as described in the fourth reward factor. For example, as shown in Fig. 7(c), our model devises optimal plans to install more charging stations or increase capacities in areas with high charging demand and energy supply. Our model achieves distributed installation of EVFCS in suitable areas by controlling the fifth reward factor. In other cases, our model installs low capacity charging stations in areas with high charging demand but low potential energy supply, where surrounding electricity production from solar PV plants cannot meet the demand for charging.

The strategies involving the starting point of the agent (charging station) and the number of charging stations can play a critical role in solving the optimal decision-making problem in a vast and unknown complex environment. Through experimentation, it was found that our model can make optimal decisions by approximating a partially observable state in each episode, even if the agent encountered the state the first time. The feature allows the use of a random starting point-based model achieved stable performance among all other scenarios. Although the fixed point-based model showed the highest learning performance (third scenario), it is not validated in other regions where the environment is significantly different from the training areas, which may result in the issue of overfitting. It can be inferred that the random starting point-based model serves a role similar to the cross-validation technique, helping to avoid overfitting issues by adapting to varied environments in each episode. It also implies that the random starting point method can assist the model in constructing a robust probability distribution of the state, which is used to estimate the value function in our model because the agent can experience various states during training.

Furthermore, the number of charging stations plays a crucial role in the decision-making process within our model, presenting both advantages and disadvantages. On the positive side, regions with high charging demand (traffic density) and potential energy supply (electricity availability) can accommodate concentrated installation of a large number of charging stations, provided they meet additional conditions such as avoiding duplication with existing stations and ensuring available installation areas. However, selecting sites where either the charging demand or potential energy supply is low may lead to biased decision-making or loss of previous learning in our model. For example, in the sixth scenario, the model prefers that many charging stations be closely installed in specific regions, resulting in the opposite results compared to the other five scenarios. The reasons for these results can be attributed to two reasons: (1) the biased learning outcomes resulting from the total reward values obtained by duplicating the installation of multiple charging stations in suitable areas significantly outweigh the penalty in the fifth reward factor (R_5), which aims to the prevention of duplicating installation in suitable areas; and (2) the biased learning outcomes stemming from the estimated total capacities of charging stations exceeding the amount of electricity generated by solar PV plants in the fourth reward factor (R_4). To address these challenges, our proposed method explores optimal combinations, such as deploying 1000 charging stations with a random starting point and optimizes the distribution of charging station placements. This approach carefully evaluates the capacities of potential charging stations while adhering to the best scenario (i.e., the fifth scenario). This strategy aims to enhance the effectiveness and efficiency of our decision-making process in urban EV infrastructure planning.

Future studies should address the following limitations. First, while the proposed model effectively explores optimal sites for EVFCS throughout a city, it is less effective in exploring boundary areas. For example, in Fig. 7(a), the eastern area with high charging demand and potential energy supply remains underexplored due to the boundary area's shape, which hinders the model's ability to navigate the next target site. Second, this study solely on the city of Chicago as a case study. Although the proposed model demonstrated generalizability by exploring optimal sites from random starting points, further investigation is needed to apply the model in other cities with distinct geospatial features. To overcome these limitations, future research can expand the current site selection model by training it in diverse cities and utilizing meta-learning techniques to enhance the model's adaptability across multiple tasks and environments. This approach will enable broader applicability and robustness of the proposed EVFCS installation optimization model.

6. Conclusion

The transition to Electric Vehicle (EVs) is seen as a promising solution for reducing greenhouse gas emissions in cities. However, the lack of availability of public charging infrastructure, particularly fast-charging stations, poses a challenge to widespread EV adoption. Also, their inadequate placement could result in significant economic losses. In addition, the increasing demand for EV Fast Charging Stations (EVFCS) strains the national electricity grid, demanding a shift to renewable energy sources like solar power.

Despite recent efforts to optimize the placement of renewable-powered EV charging stations, current planning methods face scalability issues. In this respect, this study proposes a robust decision-making model for the optimal planning of EVFCS powered by solar power using a geospatial map-based advantage actor-critic (A2C) deep reinforcement learning (DRL) approach. The model can recognize the spatial features and patterns of a vast city environment. These learning patterns are utilized to investigate the optimal sites for the installation of EVFCS and to evaluate the capacities of EVFCS. Conventional approaches to site selection of EVCS (Table 1) primarily focus on the investigation of optimal sites for EVCS at a small scale (considering less than 100 charging stations), best refueling locations with estimating charging demands, or considering grid connected charging stations. This may not always integrate the optimal planning of EVFCS placement with distributed energy resources (DERs) powered by solar Photovoltaic (PV) plants, covering the vast and complex city environment. Unlike such approaches, the A2C DRL-based decision-making model can find the optimal sites, satisfying various geospatial factors, which can highly affect the placement of the charging stations, while estimating the capacities of its charging stations.

As a result, the proposed model was tested in the city of Chicago and achieved a total reward value percentage of 74.30 % except for R2 in the scenario involving 1000 charging stations and random starting points, fifth scenario. Based on the results, the major findings and implications are summarized as follows: (1) The A2C DRL-based decision-making model allowed for the mapping of geospatial patterns in a vast complex city environment; (2) the combination of the starting point and the number of charging stations strongly affected stable learning performance (e.g., total reward values ranging from 0.33 % to 81.33 %) (Table 6) and proposed unbiased decision-making process; (3) the evaluation of capacities of potential EVFCS is strongly affected by the balance of energy supply and charging demand (Fig. 7); and (4) the combination of reward factors was also a critical factor affecting the nonstationary environment-based decision-making model. The proposed learning model can be used to investigate suitable EVFCS placements within a vast and complex urban environment. It can also

evaluate the capacities of potential EVFCS by balancing charging demand (traffic density) and energy supply (potential electricity generation). The model is designed to pinpoint detailed locations for EVFCS at optimal sites such as commercial areas, ensuring no duplication with existing charging infrastructures and appropriate spacing between potential charging stations. In addition, the proposed learning model can facilitate charging services for potential EV users by strategically distributing a large number of charging stations across urban environment.

This study contributes to broadening the body of knowledge in large-scale infrastructure planning and optimization, particularly enhancing real-world applicability. By leveraging geospatial data and employing the proposed A2C DRL-based geospatial analysis decision-making model, this research offers a comprehensive framework for EVFCS infrastructure planning, thereby accelerating the adoption of EVs and promoting sustainable urban mobility. However, despite the higher scalability and complexity of the proposed learning model, future studies should further develop it by incorporating advanced search methods to investigate unsmoothed regions (i.e., boundary regions) and by applying it to other cities with dissimilar geographic, climate, and urban policy characteristics compared to the case study city (Chicago).

CRediT authorship contribution statement

Jae Heo: Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation. **Soowon Chang:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

The author is an Editorial Board Member/Editor-in-Chief/Associate Editor/Guest Editor for *Sustainable Cities and Society* and was not involved in the editorial review or the decision to publish this article.

Data availability

Data will be made available on request.

Appendix A

Figure A: Best maps of Potential electric vehicle fast charging stations in A.1 Scenario 1; (A.2) Scenario 2; (A.3) Scenario 3; (A.4) Scenario 4; (A.5) Scenario 6.

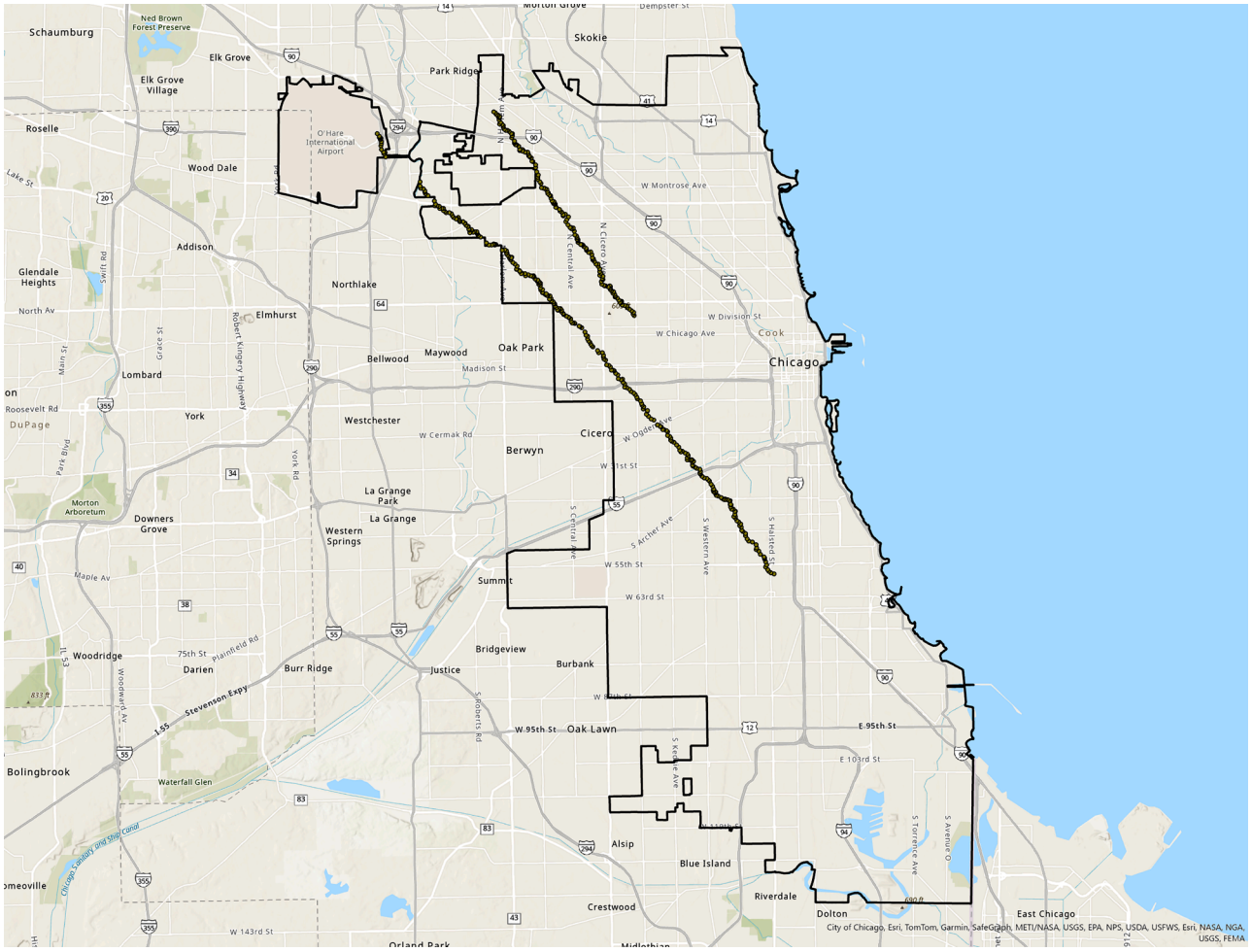


Fig. A.1. Fixed starting point and 500 charging stations in Scenario 1.

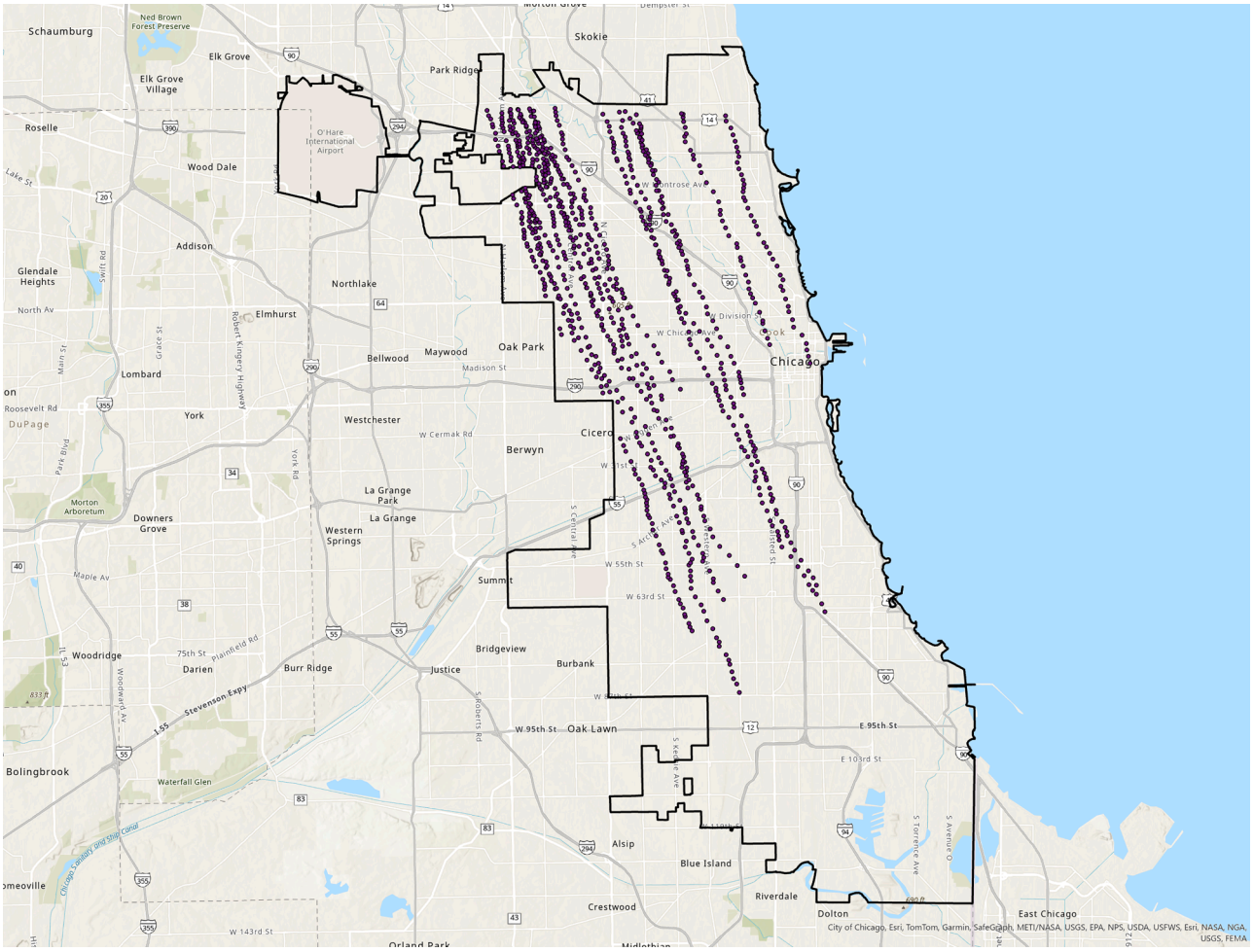


Fig. A.2. Fixed starting point and 1000 charging stations in Scenario 2.

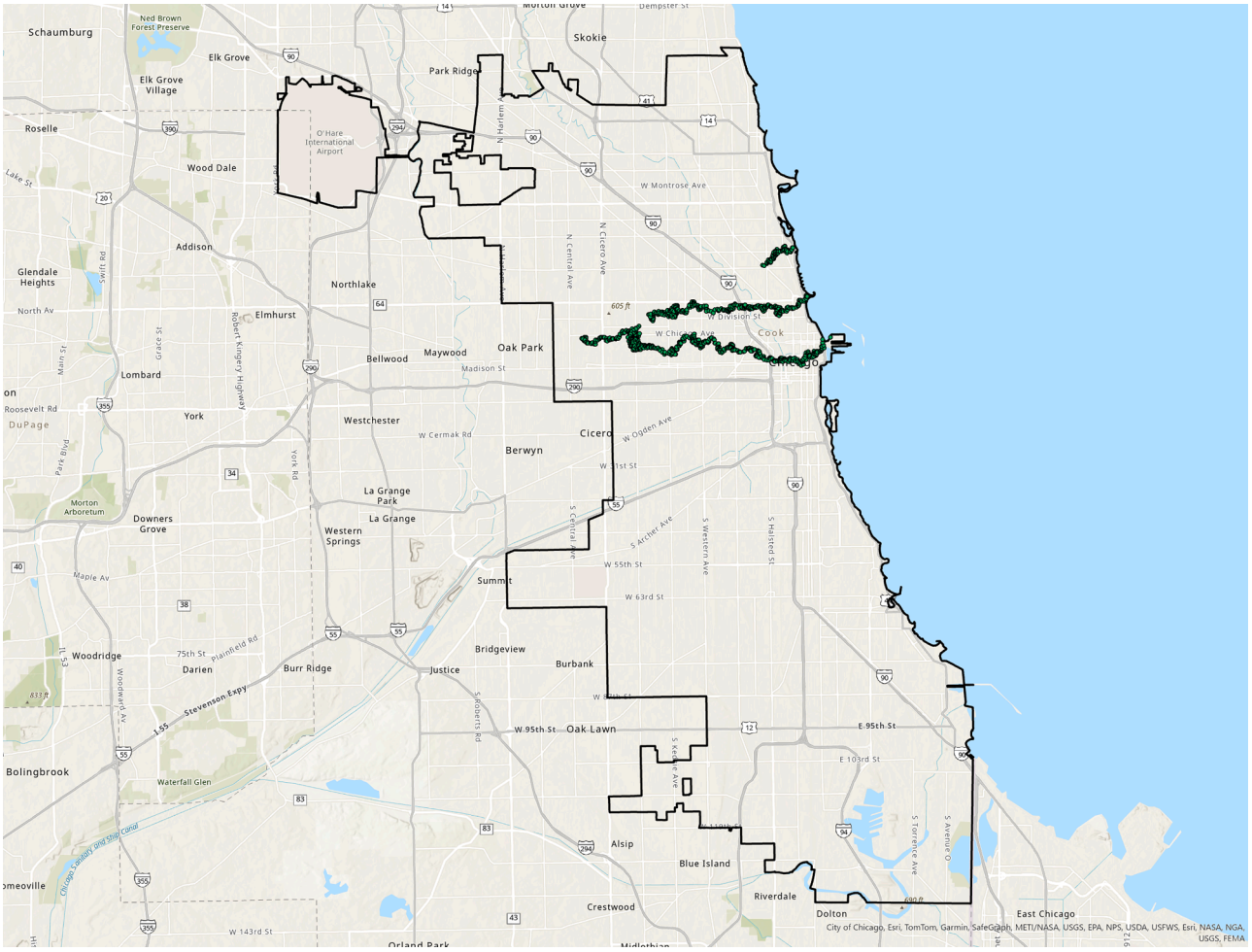


Fig. A.3. Fixed starting point and 1500 charging stations in Scenario 3.

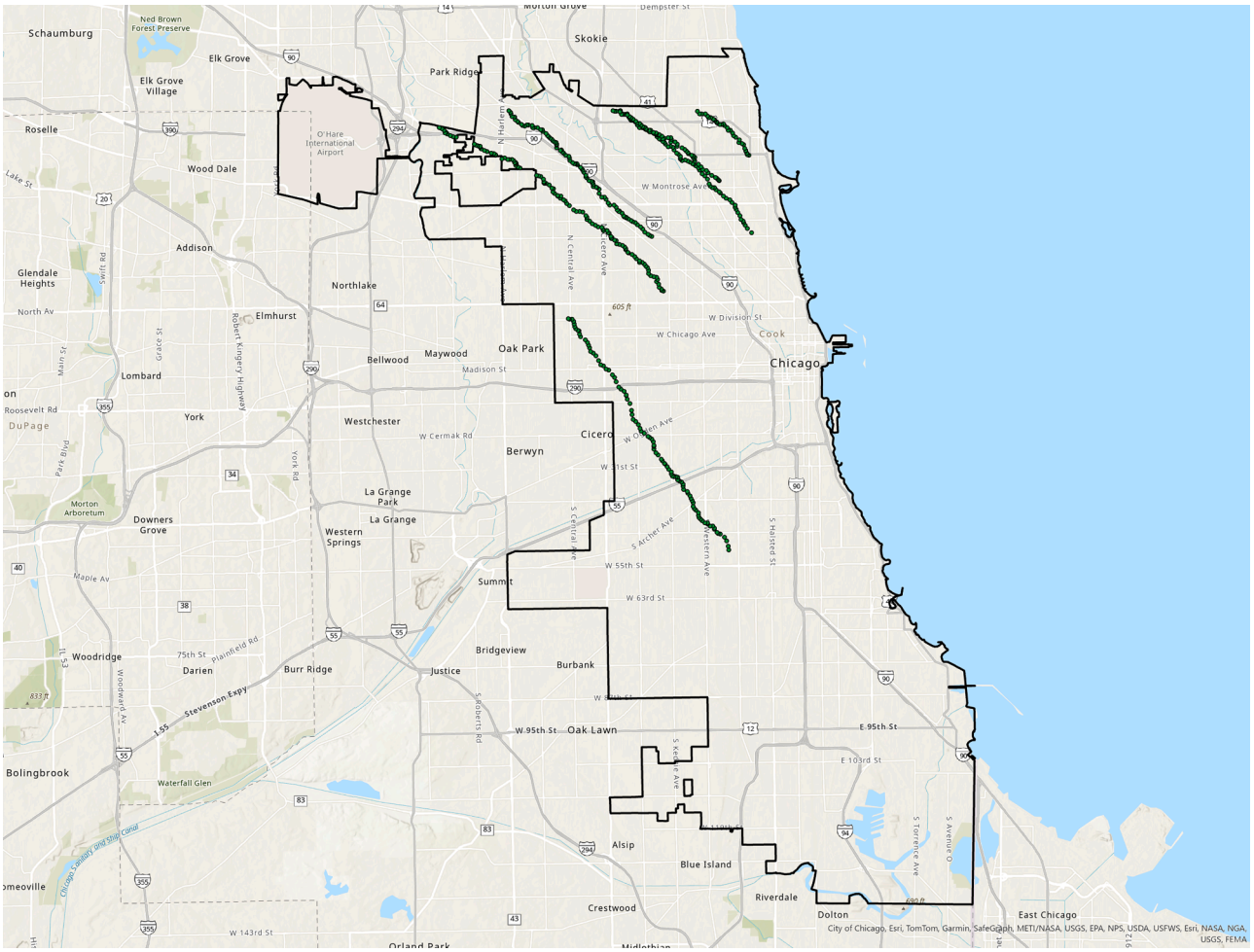


Fig. A.4. Random starting point and 500 charging stations in Scenario 4.

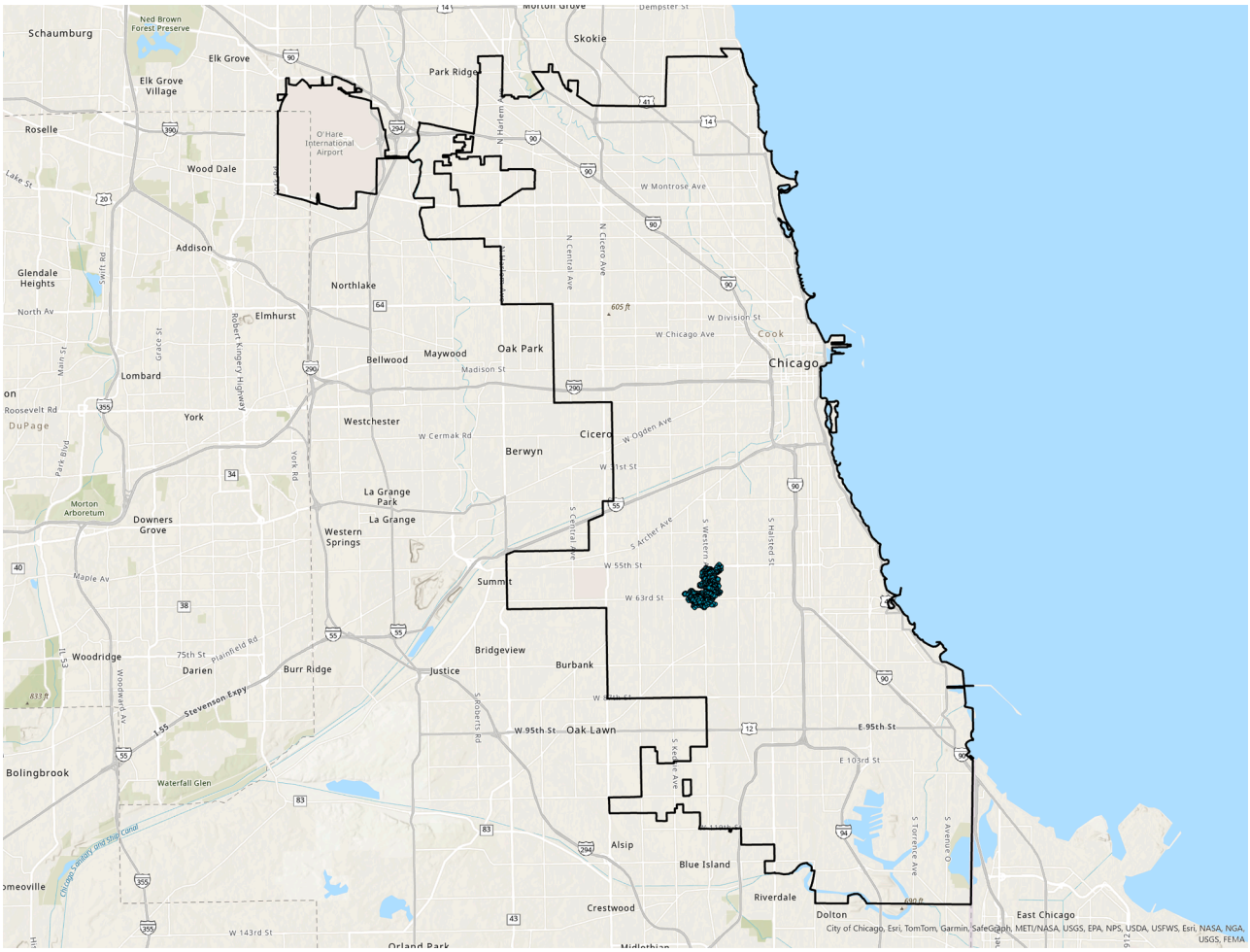


Fig. A.5. Random starting point and 1500 charging stations in Scenario 6.

References

Ali, A., Mahmoud, K., & Lehtonen, M. (2022). Optimal planning of inverter-based renewable energy sources towards autonomous microgrids accommodating electric vehicle charging stations. *IET Generation, Transmission & Distribution, 16*(2), 219–232. https://doi.org/10.1049/gtd2.12268

Arulkumaran, K., Deisenroth, M. P., Brundage, M., & Bharath, A. A. (2017). Deep reinforcement learning: A brief survey. *IEEE Signal Processing Magazine, 34*(6), 26–38. https://doi.org/10.1109/MSP.2017.2743240

A. Bui, P. Slowik, N. Lutsey, The International Council on Clean Transportation. (2020). Update on electric vehicle adoption across U.S. cities. Retrieved from chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://theicct.org/sites/default/files/publications/EV-cities-update-aug2020.pdf. Accessed May 11, 2024.

Chaudhari, K., Kandasamy, N. K., Krishnan, A., Ukil, A., & Gooi, H. B. (2018). Agent-based aggregated behavior modeling for electric vehicle charging load. *IEEE Transactions on Industrial Informatics, 15*(2), 856–868. https://doi.org/10.1109/TII.2018.2823321

Chicago Metropolitan Agency for Planning (CMAP). (2018). Land Use Data. Retrieved from https://www.cmap.illinois.gov/data/land-use. Accessed May 11, 2024.

Chris Gilligan, U.S. News & World Report. (2023). The 10 U.S. cities with the worst traffic. Retrieved from https://www.usnews.com/news/cities/articles/10-cities-with-the-worst-traffic-in-the-us. Accessed May 11, 2024.

City of Chicago. (2006). Average Daily Traffic Counts. Retrieved from https://www.chicago.gov/city/en/depts/cdot/dataset/averagedailytrafficcounts.html. Accessed May 11, 2024.

Deb, S., Tammi, K., Kalita, K., & Mahanta, P. (2018). Review of recent trends in charging infrastructure planning for electric vehicles. *Wiley Interdisciplinary Reviews: Energy and Environment, 7*(6), e306. https://doi.org/10.1002/wene.306

Erbaş, M., Kabak, M., Özceylan, E., & Çetinkaya, C. (2018). Optimal siting of electric vehicle charging stations: A GIS-based fuzzy multi-criteria decision analysis. *Energy, 163*, 1017–1031. https://doi.org/10.1016/j.energy.2018.08.140

Feng, J., Xu, S. X., & Li, M. (2021). A novel multi-criteria decision-making method for selecting the site of an electric-vehicle charging station from a sustainable perspective. *Sustainable Cities and Society, 65*, Article 102623. https://doi.org/10.1016/j.scs.2020.102623

García-Magariño, I., Palacios-Navarro, G., Lacuesta, R., & Lloret, J. (2018). ABSCEV: An agent-based simulation framework about smart transportation for reducing waiting times in charging electric vehicles. *Computer Networks, 138*, 119–135. https://doi.org/10.1016/j.comnet.2018.03.014

Google, Google Project Sunroof. (2019). Retrieved from https://sunroof.withgoogle.com/data-explorer/. Accessed May 11, 2024.

Guler, D., & Yomralioglu, T. (2020). Suitable location selection for the electric vehicle fast charging station with AHP and fuzzy AHP methods using GIS. *Annals of GIS, 26*(2), 169–189. https://doi.org/10.1080/19475683.2020.1737226

Heo, J., & Chang, S. (2024). Deep reinforcement learning for optimal planning of fast electric vehicle charging stations at a large scale. *Computing in Civil Engineering*, 680–688. https://doi.org/10.1061/9780784485248.082, 2023.

Hisoglu, S., Tuominen, A., & Huovila, A. (2023). An approach for selecting optimal locations for electric vehicle solar charging stations. *IET Smart Cities, 5*(2), 123–134. https://doi.org/10.1049/smc2.12058

Hodgson, M. J. (1990). A flow-capturing location-allocation model. *Geographical Analysis, 22*(3), 270–279. https://doi.org/10.1111/j.1538-4632.1990.tb00210.x

Huang, P., Ma, Z., Xiao, L., & Sun, Y. (2019). Geographic Information System-assisted optimal design of renewable powered electric vehicle charging stations in high-density cities. *Applied Energy, 255*, Article 113855. https://doi.org/10.1016/j.apenergy.2019.113855

Jordán, J., Palanca, J., Martí, P., & Julian, V. (2022). Electric vehicle charging stations emplacement using genetic algorithms and agent-based simulation. *Expert Systems with Applications, 197*, Article 116739. https://doi.org/10.1016/j.eswa.2022.116739

Kahraman, C., & Gündogdu, F. K. (2021). Decision making with spherical fuzzy sets. *Studies in Fuzziness and Soft Computing, 392*, 3–25. https://doi.org/10.1007/978-3-030-45461-6

Kaya, Ö., Tortum, A., Alemdar, K. D., & Çodur, M. Y. (2020). Site selection for EVCS in Istanbul by GIS and multi-criteria decision-making. *Transportation Research Part D: Transport and Environment, 80*, Article 102271. https://doi.org/10.1016/j.trd.2020.102271

Kuby, M., & Lim, S. (2005). The flow-refueling location problem for alternative-fuel vehicles. *Socio-Economic Planning Sciences, 39*(2), 125–145. https://doi.org/10.1016/j.seps.2004.03.001

Kuby, M., & Lim, S. (2007). Location of alternative-fuel stations using the flow-refueling location model and dispersion of candidate sites on arcs. *Networks and Spatial Economics, 7*, 129–152. https://doi.org/10.1007/s11067-006-9003-6

Lam, A. Y., Leung, Y. W., & Chu, X. (2014). Electric vehicle charging station placement: Formulation, complexity, and solutions. *IEEE Transactions on Smart Grid, 5*(6), 2846–2856. https://doi.org/10.1109/TSG.2014.2344684

Li, J., Liu, Z., & Wang, X. (2021). Public charging station location determination for electric ride-hailing vehicles based on an improved genetic algorithm. *Sustainable Cities and Society, 74*, Article 103181. https://doi.org/10.1016/j.scs.2021.103181

Li, C., Zhang, L., Ou, Z., Wang, Q., Zhou, D., & Ma, J. (2022). Robust model of electric vehicle charging station location considering renewable energy and storage equipment. *Energy, 238*, Article 121713. https://doi.org/10.1016/j.energy.2021.121713

Lillicrap, T.P., Hunt, J.J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., & Wierstra, D. (2015). Continuous control with deep reinforcement learning. arXiv preprint arXiv:1509.02971. https://doi.org/10.48550/arXiv.1509.02971.

Lim, S., & Kuby, M. (2010). Heuristic algorithms for siting alternative-fuel stations using the flow-refueling location model. *European Journal of Operational Research, 204*(1), 51–61. https://doi.org/10.1016/j.ejor.2009.09.032

Liu, Y., & Yu, R. (2018). Model-free optimal tracking control for discrete-time system with delays using reinforcement Q-learning. *Electronics Letters, 54*(12), 750–752. https://doi.org/10.1049/el.2017.3238

Liu, H. C., Yang, M., Zhou, M., & Tian, G. (2018). An integrated multi-criteria decision making approach to location planning of electric vehicle charging stations. *IEEE Transactions on Intelligent Transportation Systems, 20*(1), 362–373. https://doi.org/10.1109/TITS.2018.2815680

Liu, A., Zhao, Y., Meng, X., & Zhang, Y. (2020). A three-phase fuzzy multi-criteria decision model for charging station location of the sharing electric vehicle. *International Journal of Production Economics, 225*, Article 107572. https://doi.org/10.1016/j.ijpe.2019.107572

Liu, J., Sun, J., & Qi, X. (2023). Optimal placement of charging stations in road networks: A reinforcement learning approach with attention mechanism. *Applied Sciences, 13*(14), 8473. https://doi.org/10.3390/app13148473

McLaren, J., Miller, J., O'Shaughnessy, E., Wood, E., & Shapiro, E (2016). CO2 emissions associated with electric vehicle charging: The impact of electricity generation mix, charging infrastructure availability and vehicle type. *The Electricity Journal, 29*(5), 72–88. https://doi.org/10.1016/j.tej.2016.06.005

Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., & Riedmiller, M. (2013). Playing atari with deep reinforcement learning. arXiv preprint arXiv:1312.5602. https://doi.org/10.48550/arXiv.1312.5602.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidgeland, A., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., Kin, H., Kumaran, D., Wierstra, D., Legg, S., & Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature, 518*(7540), 529–533. https://doi.org/10.1038/nature14236

Mnih, V., Badia, A. P., Mirza, M., Graves, A., Lillicrap, T., Harley, T., Silver, D., & Kavukcuoglu, K. (2016). Asynchronous methods for deep reinforcement learning. In *International conference on machine learning* (pp. 1928–1937). PMLR.

Mozafar, M. R., Moradi, M. H., & Amini, M. H. (2017). A simultaneous approach for optimal allocation of renewable energy sources and electric vehicle charging stations in smart grids based on improved GA-PSO algorithm. *Sustainable Cities and Society, 32*, 627–637. https://doi.org/10.1016/j.scs.2017.05.007

Nandy, A., Ren, Z., Rathinam, S., & Choset, H. (2023). Heuristic Search for Path Finding with Refuelling. arXiv preprint arXiv:2309.10796. https://doi.org/10.48550/arXiv.2309.10796.

Nie, Y. M., Ghamami, M., Zockaie, A., & Xiao, F. (2016). Optimization of incentive polices for plug-in electric vehicles. *Transportation Research Part B: Methodological, 84*, 103–123. https://doi.org/10.1016/j.trb.2015.12.011

Pan, L., Yao, E., Yang, Y., & Zhang, R. (2020). A location model for electric vehicle (EV) public charging stations based on drivers' existing activities. *Sustainable Cities and Society, 59*, Article 102192. https://doi.org/10.1016/j.scs.2020.102192

Panah, P. G., Bornapour, S. M., Nosratabadi, S. M., & Guerrero, J. M. (2022). Hesitant fuzzy for conflicting criteria in multi-objective deployment of electric vehicle charging stations. *Sustainable Cities and Society, 85*, Article 104054. https://doi.org/10.1016/j.scs.2022.104054

Pareek, S., Sujil, A., Ratra, S., & Kumar, R. (2020). Electric vehicle charging station challenges and opportunities: A future perspective. In *2020 International Conference on Emerging Trends in Communication, Control and Computing (ICONC3)* (pp. 1–6). IEEE. https://doi.org/10.1109/ICONC345789.2020.9117473.

Petratos, A., Ting, A., Padmanabhan, S., Zhou, K., Hageman, D., Pisel, J.R., & Pyrcz, M.J. (2021). Optimal placement of public electric vehicle charging stations using deep reinforcement learning. arXiv preprint arXiv:2108.07772. https://doi.org/10.48550/arXiv.2108.07772.

Rane, N. L., Achari, A., Saha, A., Poddar, I., Rane, J., Pande, C. B., & Roy, R. (2023). An integrated GIS, MIF, and TOPSIS approach for appraising electric vehicle charging station suitability zones in Mumbai, India. *Sustainable Cities and Society, 97*, Article 104717. https://doi.org/10.1016/j.scs.2023.104717

Rani, P., & Mishra, A. R. (2021). Fermatean fuzzy Einstein aggregation operators-based MULTIMOORA method for electric vehicle charging station selection. *Expert Systems with Applications, 182*, Article 115267. https://doi.org/10.1016/j.eswa.2021.115267

Riemann, R., Wang, D. Z., & Busch, F. (2015). Optimal location of wireless charging facilities for electric vehicles: Flow-capturing location model with stochastic user equilibrium. *Transportation Research Part C: Emerging Technologies, 58*, 1–12. https://doi.org/10.1016/j.trc.2015.06.022

Rodriguez-Calvo, A., Cossent, R., & Frías, P. (2017). Integration of PV and EVs in unbalanced residential LV networks and implications for the smart grid and advanced metering infrastructure deployment. *International Journal of Electrical Power & Energy Systems, 91*, 121–134. https://doi.org/10.1016/j.ijepes.2017.03.008

Sanguesa, J. A., Torres-Sanz, V., Garrido, P., Martinez, F. J., & Marquez-Barja, J. M. (2021). A review on electric vehicles: Technologies and challenges. *Smart Cities, 4*(1), 372–404. https://doi.org/10.3390/smartcities4010022

Singh, H., Kavianipour, M., Soltanpour, A., Fakhrmoosavi, F., Ghamami, M., Zockaie, A., & Jackson, R. (2022). Macro analysis to estimate electric vehicles fast-charging infrastructure requirements in small urban areas. *Transportation Research Record, 2676*(11), 446–461. https://doi.org/10.1177/03611981221093625

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT Press.

Szepesvári, C. (2022). *Algorithms for reinforcement learning*. Springer Nature.

Tripathi, A. K., Agrawal, S., & Gupta, R. D. (2021). Comparison of GIS-based AHP and fuzzy AHP methods for hospital site selection: A case study for Prayagraj City, India. *GeoJournal*, 1–22. https://doi.org/10.1007/s10708-021-10445-y

U.S. Department of Renewable Energy. (2023). *Electric vehicle charging station locations, Tech*. Retrieved from https://afdc.energy.gov/stations#/find/nearest Accessed May 11, 2024.

Wang, C., Gao, Z., Yang, P., Wang, Z., & Li, Z. (2021). Electric vehicle charging facility planning based on flow demand—A case study. *Sustainability, 13*(9), 4952. https://doi.org/10.3390/su13094952

Wu, Y., Xie, C., Xu, C., & Li, F. (2017). A decision framework for electric vehicle charging station site selection for residential communities under an intuitionistic fuzzy environment: A case of Beijing. *Energies, 10*(9), 1270. https://doi.org/10.3390/en10091270

Yang, C., Komura, T., & Li, Z. (2017). Emergence of human-comparable balancing behaviours by deep reinforcement learning. In *2017 IEEE-RAS 17th international conference on humanoid robotics (Humanoids)* (pp. 372–377). IEEE. https://doi.org/10.1109/HUMANOIDS.2017.8246900.

Yang, Y., Zhang, Y., & Meng, X. (2020). A data-driven approach for optimizing the EV charging stations network. *IEEE Access : Practical Innovations, Open Solutions, 8*, 118572–118592. https://doi.org/10.1109/ACCESS.2020.3004715

Yu, Z., Wu, Z., Li, Q., & Bai, Q. (2022). A map matching-based method for electric vehicle charging station placement at directional road segment level. *Sustainable Cities and Society, 84*, Article 103987. https://doi.org/10.1016/j.scs.2022.103987

Zhang, C., Liu, Y., Wu, F., Tang, B., & Fan, W. (2020). Effective charging planning based on deep reinforcement learning for electric vehicles. *IEEE Transactions on Intelligent Transportation Systems, 22*(1), 542–554. https://doi.org/10.1109/TITS.2020.3002271

Zhao, H., & Li, N. (2016). Optimal siting of charging stations for electric vehicles based on fuzzy Delphi and hybrid multi-criteria decision making approaches from an extended sustainability perspective. *Energies, 9*(4), 270. https://doi.org/10.3390/en9040270

Zhao, Y., Guo, Y., Guo, Q., Zhang, H., & Sun, H. (2020). Deployment of the electric vehicle charging station considering existing competitors. *IEEE Transactions on Smart Grid, 11*(5), 4236–4248. https://doi.org/10.1109/TSG.2020.2991232