Spatial Equity Analysis of Public Electric Vehicle Charging Stations: A Case Study in Shanghai, China

Bale Chen[1], Jialin Liu[1]

[1]New York University Shanghai

December 21, 2022

Abstract

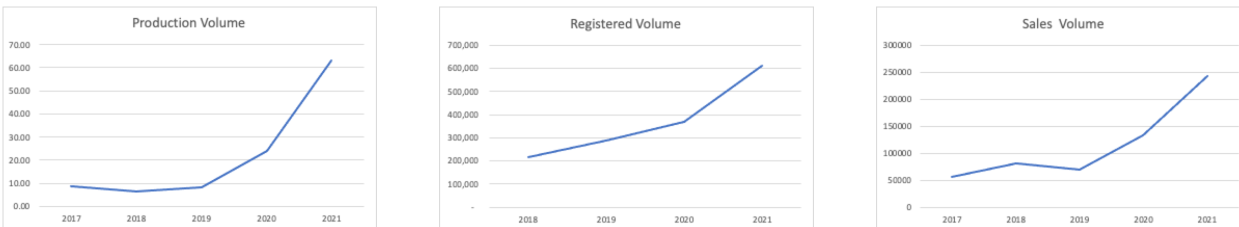
In recent years, electric vehicles (EVs) are increasingly regarded as a major solution to sustainable urban development, and the planning of charging infrastructure plays an essential role in supporting and promoting the EV popularization. Previous literature has mainly adopted optimization approaches to solve the transportation planning problem, but rarely have we seen a comprehensive spatial equity analysis on public EV charging stations (PEVCSs) with regard to major demographic, the built environment, and land use factors. Our research attempts to use a tailored version of Improved Gravity Model of Accessibility on a pixel-level and regression analysis to examine the spatial equity of PEVCSs in the context of Shanghai, China. With qualitative and quantitative analysis, this study finds that there exists a huge disparity in PEVCS access in urban and suburban areas. Regional patterns are also detected and analyzed on a district-level. Moreover, from regression analysis, male population is found to be positively associated with higher PEVCS accessibility while female population shows the opposite. Working age (20-50) population is surprisingly negatively correlated with PEVCS accessibility. Our methods are efficient and transferable in PEVCS spatial equity analysis with publicly available data and detailedly explained procedures.

Keywords: Public electric vehicle charging stations, spatial equity, accessibility, pixel-level analysis

1 Introduction

In the 11th Sustainable Development Goal, the United Nations highlighted the importance of reducing per capita adverse environment impact and providing universal access to inclusive, green and public spaces (2015). This global appeal has brought New Energy Vehicles under the spotlight. Electric vehicles (EVs), as the principal representative of the New Energy Vehicle family, are generally considered to be more sustainable than conventional gasoline-powered vehicles, as they produce zero tailpipe emissions and can be powered by renewable energy sources such as wind, hydropower, or solar energy. However, the sustainability of EVs depends not only on the vehicles themselves, but also on the infrastructure that supports them, including charging stations. The deployment of public EV charging stations (PEVCSs) can have a significant impact on the sustainability of transportation in a given geographic area. PEVCS is defined as publicly accessible charging stations that are different from home charging piles where individual EV users own. It provides charging service to everyone who possesses an EV. By providing equitable and reliable access to charging infrastructure, it can help to increase the adoption of EVs (Egnér & Trosvik, 2018; White et al., 2022), leading to reduced emissions and improved air quality.

Specifically in China, according to the New Energy Vehicle Industry Development Plan (2021 - 2035) proposed by the State Council, the nation has made huge contributions in structurally transforming the automobile industry and promoting sustainable development. However, it was also explicitly pointed out that urban infrastructure supporting the EV industry needs further perfection, and the charging station network is the primary focus to improve (2019). In the city of Shanghai, encouraging policies on purchasing EVs has been widely applied, which has caused a huge surge in both the supply and demand sides of the EV industry. The municipal government provides financial assistance to EV customers, offering free license plates specified for EVs, etc. These policies have led to a leap in sales, production, and registered volume of EVs in the recent 5 years (See Figure 1). Registered EVs in Shanghai, namely the EVs that are currently officially registered and in use, reached over 613000 in 2021, which is three times the number in 2018. With such a sharp increase in the number of EVs on the roads, determining whether the current charging infrastructure is sufficient is rather crucial to urban planning and urban life.



(a) Production Volume (b) Registered Volume (c) Sales Volume

Figure 1: Production, registered, and sales volume of EVs in Shanghai from 2017 to 2021 (Registered volume in 2017 is not available)

Sufficiency of charging infrastructure can be broken down into two aspects. One is an adequate quantity of PEVCSs, and the other is the equitable spatial distribution of such resources. Falchetta and Noussan have shown that a growing charging network doesn't mean a equitable access to EV charging facilities for a vast population (2021). Equity, different from equality, focuses on each individuals' unique needs and backgrounds, taking the unevenness in the past and present into consideration. Equality rather emphasizes neutrality and fairness which treat every one the same, with a aim to grant equal future opportunities (Minow, 2021). In the context of urban studies, spatial equity means that "there is an even distribution of services in relation to the needs, preferences and service standards of each resident" (Tsou et al., 2005). Spatial equity analysis thus can reflect how accessible the urban resources are to people with distinct demographic features and socioeconomic status and how they are distributed across different built environments and land use types.

We identified two main issues in the existing literature. Firstly, scholars mainly approaches the EV charging stations with an optimization perspective, where they build a mathematical model to optimize the current network layout with one or more objectives (Dong et al., 2019; Juvvala & Sarmah, 2021; J. Li et al., 2021; Xu et al., 2018; Zhang et al., 2019). With mathematical models, they focus on the future locationing the EV charging stations, while an analysis on current distribution is also important to urban planning. Mathematical models also fall short to model real-world factors since they usually assume EV users' charging or traveling behavior. For the existing literature on spatial equity analysis of PEVCSs, we observed that they use community-level (G. Li et al., 2022) or subdistrict-level analysis (Hsu & Fingerman, 2021) where they examine the spatial equity of PEVCSs to a whole community or subdistrict. However, considering a community or neighborhood as a receiver of public charging station service is less relevant since home charging is still more flavored way (Funke et al., 2019). In the literature, there's currently no pixel-level equity analysis that examines each general square of land in a city.

Given the increasing popularity of EVs and encouraging policies in Shanghai, studying the spatial equity of such public infrastructure is significantly relevant to urban planning. Therefore, in this study, we aim to understand how the accessibility of PEVCEs Shanghai correlates spatially with various demographic, built environment, and land use factors. Our main contributions are:

1. Developed a novel and easy-to-implement version of Improved Gravity Model (IGM) that is tailored to calculating the accessibility of PEVCSs, capturing the travel cost, PEVCS service capacity, and charging demand in a more integrated way.

2. Proposed a framework for the spatial equity analysis on a pixel level that can reflect the spatial equity of each general piece of land in a city with all the data sourced from publicly available platforms.

3. Qualitatively and quantatively analyzed the PEVCS accessibility in Shanghai from a perspective of spatial equity.

2 Previous Literature

2.1 Spatial Equity Analysis of Urban Infrastructure

In the subsection, we bring in a group of representive works in Table 1 that conducted spatial equity analysis on urban infrastructures to discuss what equity-related methods are implemented by previous researchers.

Authors & Year	Objective	Methods
(Chen et al., 2019)	Assessing accessibility-based service effectiveness of bus transits	Gini coefficient and Lorenz curve
(Cheng et al., 2020)	Examining geographical accessibility and spatial equity of multi-tier hospital care services	Gini coefficient and 2-step Floating Catchment Area
(Shi et al., 2022)	Assessing the accessibility of public hosiptals	IGM
(Shen et al., 2020)	Analyzing the spatial and social inequalities in urban sports facilities	Allocation index, cost distance analysis, and Geograpically weighted regression
(Chang et al., 2019)	Analyze the spatial inequalty of access to urban parks	Gravity model and linear regression analysis

Table 1: Studies related to spatial equity analysis

Studies used Gini coefficient as a way to capture the spatial equity, which can represent the overall degree of inequality of urban service provision across different places. Coupled with a measure that considers the demand of the population, the Gini coefficient can reflect the inequity. In Cheng et al.'s study, the 2-step Floating Catchment Area method incorporates the population density into the measurement, which allows the Gini coefficient to reflect the uneven distribution of the accessibility of medical care with respect to population density in an area. However, Chen et al. used a opportunity-based accessibility measure that can only reflect the service capacity of bus transit. In this case, using Gini index reflects how *unequal* the bus transit service are spatially distributed, regardless of the population density disparity across different places. The difference among multiple accessibility measures will be discussed in detail in Section 2.3. Allocation index used in Shen et al.'s paper can demonstrate how proportionate the urban facilities are distributed with regard to another feature (population density in their case), which is also a desirable indicator of spatial equity.

Multiple regression analysis methods are also adopted in studying spatial equity (Chang et al., 2019; Shen et al., 2020). Different from Gini coefficient and allocation index, regression analysis are generally able to study a one-to-many association. In other words, studying whether the accessibility of urban infrastructures are significantly associated with multiple factors is made possible in regression analysis, while allocation index and Gini coefficient generally can only reflect the disproportion between two variables. Thus, in this study, regression analysis is preferred since we want to reveal the relationship between accessibility of PEVCSs and multiple other factors.

2.2 Spatial Planning and Analysis of PEVCSs

In this subsection, we discuss the existing studies on PEVCS planning and analysis. As the main body of the PEVCS-related is related to location optimization, we look into its difference with spatial equity analysis. Firstly, spatial equity was rarely the objective of the optimization program. J. Li et al. and Juvvala and Sarmah focused on minimizing travel cost (2021, 2021); Anjos et al. considers the maximization of traffic flow (2020); Dong et al.'s model maximizes the PEVCS coverage area (2019). Using optimization models, mainly the locationing policies and charging strategies are tested with different objectives. In contrast, spatial equity analysis aims to find out whether spatial distribution of PEVCSs provides equitable access to different groups of people. Secondly, in those optimization approaches, present and past data is used to populate the model parameters to better simulate the real-world scenarios (J. Li et al., 2021). They tend not to use data for analytical purpose but rather a support for modeling, while in spatial analysis, real-world data is examined to generate insights about the current spatial layout of PEVCSs. In this sense, spatial equity analysis is focused on describing the present or the past, while optimization approach yields insights for future planning.

In the limited literature that studied the spatial equity of PEVCSs, G. Li et al. developed an accessibility measure based on cumulative opportunities approach, applied global and local Moran's I, and compared 10 major Chinese cities (2022). He found a significant inequity in most cities but with different regional characteristics. Inequity was also found in California across different racial and ethnic groups (Hsu & Fingerman, 2021). Similar results are also found by Roy and Law that area with higher minority population and lower socioeconomic status can lead to lower accessibility to EV charging resources (2022). These three studies all study the status quo of the PEVCS layout by comparing the regional difference. Temporal difference wasn't examined as time series analysis has a higher demand on data.

Another important fact is that all previous equity analysis on PEVCS are conducted on community level or subdistrict level, meaning that a community or a subdistrict is considered as a unit. Accessibility index is computed for each of the unit. For community level analysis by G. Li et al. (2022), they consider the people living in a community to access a nearby PEVCS. However, according to Funke et al. (2019), home charging is favored, meaning that people living in the neighborhood would tend to use their private charging piles if they have one. Thus, even if EV users don't have access to a nearby PEVCS, as long as they have a home charging pile, their charging needs are perfectly satisfied. This fact might make community-level analysis a less accurate method. Rather, we shall examine the accessibility of PEVCS from a more general location in the city. Subdistrict level analysis is overall satisfactory, but when computing the accessibility, it's difficult to consider a decay of accessibility with respect to distance. Subdistrict is too large to be considered a point from which we compute the distance to the nearby PEVCSs. Shi et al. attempted to use the subdistrict government office locations to compute the distance as a rough estimate, so that they are able to apply IGM (2022). Otherwise, only simple accessibility models without distance decay be applied (G. Li et al., 2022). Thus, in this study, we attempt to use pixel-level analysis, where we consider the accessibility of each pixel that is much smaller than a subdistrict.

2.3 Accessibility Measurement

As we analyze equity based on accessibility of PEVCSs, the measurement of accessibility is a key step in the loop. The main approaches to evaluate accessibility of a facility are listed and discussed below.

Naive Approaches One of the most rudimentary approach to measure accessibility is to see whether there is a facility within a threshold distance (Hsu & Fingerman, 2021; Smoyer-Tomic et al., 2008). This measure results in 0 or 1 as an indicator of such facility availability. Another approach is calculating the travel cost to the nearest facility (Sharkey et al., 2009; Shen et al., 2020), where travel costs can mainly be travel time or travel distance. This method is no longer a boolean indicator but outputs a continuous value for accessibility index.

Cumulative Opportunity Cumulative opportunity captures the number of opportunities available within a specified travel time/distance (Kelobonye et al., 2020). The general formulation is

$$A_i = \sum_j E_j c_{ij}$$

where A_i denotes the accessibility level of location i; E_j is the quantity of opportunities at location j; $c_{ij} = 1$ if the travel cost from location i to j is within a certain threshold, and $c_{ij} = 0$ otherwise. These notations will be used consistently throughout this paper.

As Kelobonye et al. also pointed out, the cumulative opportunity lacks consideration of the competition among the service receivers. In other words, a place can have sufficient exposure to a certain service and a large demand for this service at the same time, resulting in a less sufficient amount of such service per capita. To tackle this, G. Li et al. adopted the cumulative opportunity measure and added a demand term in the denominator to represent the negative association between accessibility and demand size. This modification led the model to be similar with the 2-Step Floating Catchment Area method.

2-Step Floating Catchment Area (2SFCA) 2SFCA method adopts a two step calculation of accessibility index. Firstly, it calculates the opportunity-to-demand ratio for each facility. Then, it sums up opportunity-to-demand ratio for a certain area of service receivers. The measurement can be formulated as

$$E_j = \frac{S_j}{\sum_{i \in \{i | c_{ij} \leq c_0\}} P_i}$$

then,

$$A_i = \sum_{j \in \{j | c_{ij} \leq c_0\}} E_j$$

where S_j denotes the capacity of facility j (i.e. the opportunity); P_i is the population density of location i; c_0 denotes the threshold distance that needs finetuning. This method is especially favored by medical care studies where they study the accessibility of medical care (Cheng et al., 2020; Kanuganti et al., 2016). With the population density term in the formulation, the 2SFCA method can model the accessibility of urban facilities in a more comprehensive and realistic manner.

Gravity Model Proposed by Walter G. Hansen, the gravity model utilizes a distance/time decay to model the decay of accessibility with travel cost (1959). Namely, for a certain facility, the people who are farther away from it would receive less service from the facility, while in 2SFCA model we see that people are assumed to receive the same amount of service as long as they are within the threshold distance. It's named gravity model because it follows the gravity law in Physics that the interaction between activities is directly proportional to their size and inversely proportional to the distance/cost of travelling between them (1959; 2012). The model can be written in our notation style as

$$A_i = \sum_{j=1}^{n} A_{ij} = \sum_{j=1}^{n} \frac{S_j}{c_{ij}^{\beta}}$$

Based on Hansen's formulation, Joseph and Bantock improved this model by introducing a competition factor into the model (Joseph & Bantock, 1982). In other words, the improved model considers the fact that availability of facilities not only depends on their capacity, but also on the people's demand level. The formulation is written below.

$$A_i = \sum_{j=1}^{n} A_{ij} = \sum_{j=1}^{n} \frac{S_j}{c_{ij}^{\beta} V_j}$$

where

$$V_j = \sum_{i=1}^{m} \frac{P_i}{c_{ij}^{\gamma}}$$

V_j is the competition level of facility j. β and γ are two parameters that governs the decay ratio of the travel cost's influence on accessibility. Note that the model considers the travel cost decay in computation of both competition factor and accessibility. This approach aligns with G. Li et al.'s modification to the cumulative opportunity model (2022). Adding a competition factor helps the model to better capture the interplay between the proximal service receivers and service supply. Taken together, in this study we adopted the IGM with a specific tailoring to our PEVCS data. It's the most all-rounded approach to our best knowledge to measure accessibility as it takes service capacity, travel cost decay, and competition factor all into account.

3 Methods

In this section, we present a PEVCS spatial equity analysis framework that incorporates the IGM, pixel-level analysis, and regression analysis (See Figure 2). We first gathered the data from publicly available platforms to guarantee the scalability of our framework. Then we performed necessary data processing techniques for pixel analysis, since pixel analysis generally requires the data to be in raster format. Meanwhile, we computed the accessibility index for each pixel using a tailored version of IGM. Using the resultant accessibility data together with demographic, built environment, and land use data, we conducted both linear and geographically weighted regression. Subsection 3.1 will explain the data collection and processing pipeline. 3.2 will be a detailed procedure of measuring pixel-level accessibility using Python and AMap API. POI-based land use categorization will be briefly covered in 3.3. Lastly in subsection 3.4, we will expound on the regression models.

3.1 Study area and Data

Our study area is Shanghai, one of the major cities in China, shown in Figure **??**. It's located on the east coast of China with a land area of 6340.50 squared kilometer and population of 24.88 million (Shanghai Municipal Bureau of Statistics, 2022). Note that we study 15 districts of Shanghai, excluding Chongming district due to data unavailability.

The PEVCS data is web scrapped from Lianlian Charging[1], the official municipal platform for public data collection and monitoring of charging facilities in Shanghai. The data was retrieved by an auto-crawling software called ScrapeStorm[2] on April 14, 2022. After scraping

[1]https://www.evchargeonline.com.cn/
[2]https://www.scrapestorm.com/

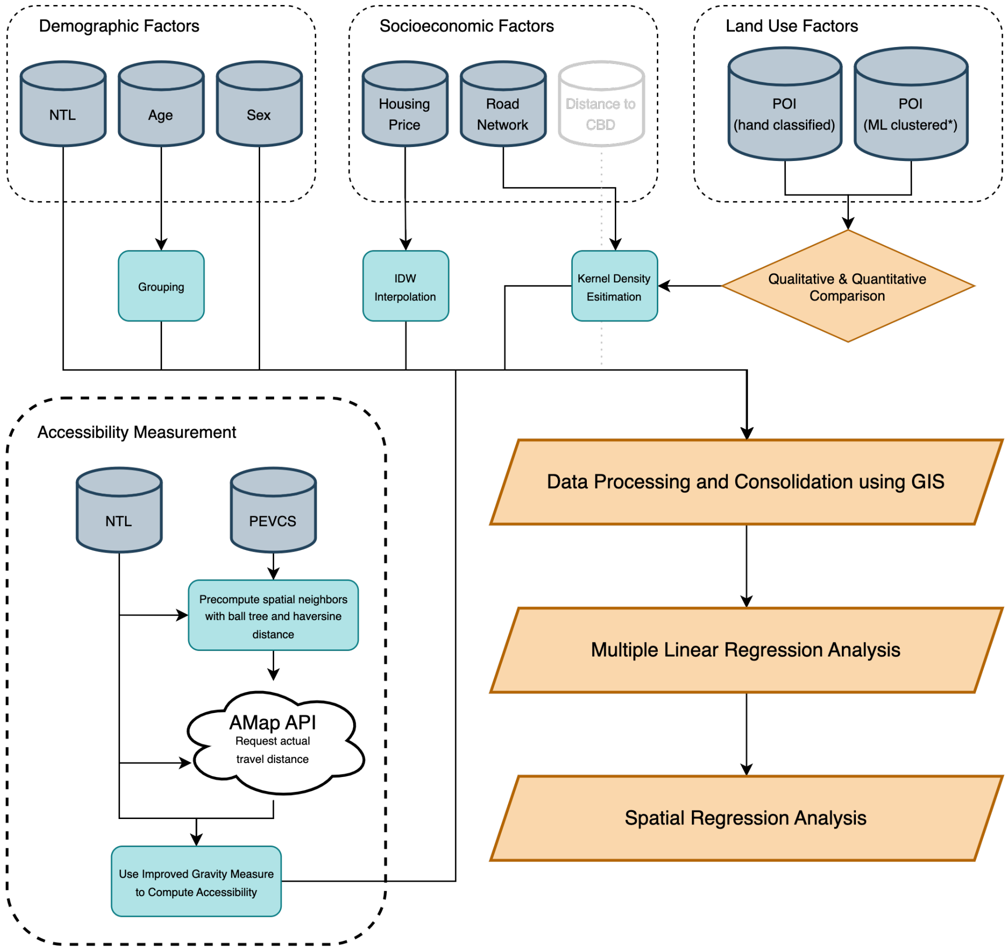


Figure 2: Research design flowchart

and cleaning the raw data, we geocoded the charging stations to using AMap API[3]. We ended up with 4477 datapoints of PEVCSs across Shanghai including information such as number of AC and DC charging piles (See Figure 3).

All other explanatory variables' data are presented in Table 2. All data are collected from open platforms that anyone can download freely. For road network, housing price, and POI data, they are originally vector data that is not eligible for pixel level analysis. To transfer their value to the pixel cells, we use two kinds of interpolation methods. For road network and POI data, we use kernel density estimation to rasterize the data points based

[3]https://lbs.amap.com/

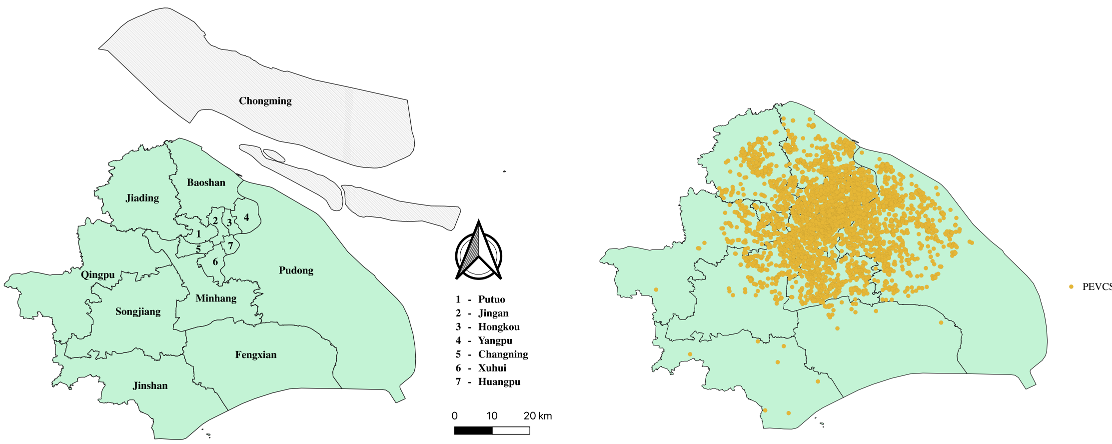


Figure 3: Study Area and PEVCS spatial distribution

Category	Data	Description	Source
Demographic	Night Time Light (NTL)	Night time light data that reflects the human activity and population density	NASA Earth(2015)
	Age	Number of people in each age group	WorldPop Hub(2020)
	Sex	Number of people of a certain sex	WorldPop Hub(2020)
Built Environment	Road Network	All roads in Shanghai	OpenStreetMap
	Housing price	Housing price of Shanghai neighborhoods	Fang.com[4]
Land Use	Points of Interest (POI)	POI point data obtained from Baidu Map API	Baidu API[5]

Table 2: Table of Data

on density (We convert road network into point data by taking the centroids of each road). As for housing price, we are not interested in the density of neighborhoods but rather the average housing price of each land parcel. We hence use inverse distance weighted (IDW) interpolation that assign a neighborhood's housing price value to its nearby pixels with a distance decay. The three demographic features were originally in raster format. Lastly, the raster data were aligned to the same resolution, 500m × 500m, and extracted to a dataframe using a fishnet. Therefore, each datapoint can be seen as a population point that has the following information: NTL value, number of people in each age group, number of male and female residents, kernel density of road network, kernel density of POI points of each land use type, housing price value based on IDW.

3.2 Measurement of Pixel-level Accessibility

Since in our spatial equity analysis, we mainly consider the equitable access to PEVCSs, it's essential to develop a suitable measurement of PEVCS accessibility. As we described in the literature section, the IGM (1982) were the most comprehensive measure that considers multiple aspects of accessibility including demand, capacity of facilities, and distance decay. In our case of PEVCS, we further developed this model to suit our regime. The tailored version can be formulated by

$$A_i = \sum_{j \in \{j | D_{ij} \leq d_0\}} \frac{\lambda N_{j,AC} + \mu N_{j,DC}}{D_{ij}^{\beta} V_j}$$

$$V_j = \sum_{i \in \{i | D_{ij} \leq d_0\}} \frac{P_i}{D_{ij}^{\gamma}}$$

where β and γ are the decay ratio, λ and μ are two coefficients for computing the capacity of each PEVCS. d_0 is the threshold driving distance. These parameters are determined by us. $N_{j,AC}$ and $N_{j,DC}$ means the number of AC or DC charging piles in PEVCS j. D_{ij} denotes the driving distance from population point i to j.

In practice, We set β and γ to be 1, which results in a linear decay. λ and μ are set to be 4 and 48 respectively because we consider the capacity of each PEVCS to be the number of EVs one can serve in a day. Based on the general charging cycle duration statistics (Hove & Sandalow, n.d.; Schroeder & Traber, 2012), we consider AC chargers can serve 4 EVs per day while DC charging piles can serve 48.

Another innovation in our measurement of accessibility lies in the algorithm that we use to compute the driving distance pairs and thus compute the accessibility (See Algorithm 1). We take advantage of the AMap API to compute the driving distance since it's much more accurate than traditional ArcGIS network analysis. The accuracy stems from its consideration of real road conditions and prediction based on real travel data from their application users.

Different from Shi et al. (Shi et al., 2022) who also applied a tailored version of IGM, we use the AMap to request for travel distance instead of travel time because for the same pair of locations, AMap API would generate different travel time at different request times. This is because AMap API takes factors like rush hours into account, which makes the time you send the request to the API a influencing factor. In contrast, travel distance is a more consistent regardless of at what time you use the API.

Moreover, almost all online map API has the limitation on request times per day. For AMap, one can only request it to compute distance for 3000 times, where in each request one can submit up to 100 pairs. With that being said, one can only compute distance for 300000 pairs of locations. However, with 25056 population points and 4477 PEVCS points, if we brutally compute distance for all population-PEVCS pairs, it would take us around a year to finish computing. Yet, most of the distance pairs are trivial since they are way larger than the threshold distance that we won't use in the gravity model Note that only pairs within the threshold distance would be considered in computing the accessibility. Therefore, we used a ball tree algorithm (Omohundro, 2009) that can compute the nearest neighbors based on haversine distance. The driving distance is larger than the haversine distance because haversine distance is the straight-line distance over the Earth sphere, while the actual driving route is rarely straight-line. If the haversine distance of a population-PEVCS pair is already larger than threshold, then this pair definitely exceeds the threshold distance. With this algorithmic design, computing all distance pairs is made possible within a day.

Algorithm 1 Algorithm for Computing Accessibility with Improve Gravity Model

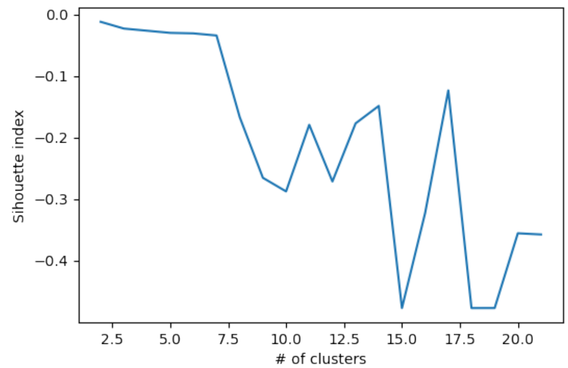
We have PEVCS data D_{PEVCS} and population points data D_{pop}
Initialize an empty set to store the generated distance pairs $\mathcal{M}$
Load the data D_{pop} into a Ball Tree $\mathcal{B}(\cdot,\cdot)$
for each PEVCS point p in D_{PEVCS} **do**
  $\mathcal{S}_{pop} \leftarrow \mathcal{B}(p, d_0)$(Generate a set of population points $\mathcal{S}_{pop}$ within the threshold distance d_0 to PEVCS point p by consulting the Ball Tree using haversine distance.)
  for each population point p' in $\mathcal{S}_{pop}$ **do**
    Request the AMap API to generate actual driving distance $d_{p,p'}$ between p and the population point p'.
    Append $(p, p', d_{p,p'})$ to the set $\mathcal{M}$.
  end for
end for

for each PEVCS point p in D_{PEVCS} **do**
  Find the set of population points within the driving distance threshold d_0 using $\mathcal{M}$ based on the aforementioned formula.
  Compute competition factor V for PEVCS point p
end for
for each population point p' in D_{pop} **do**
  Find the set of PEVCS points within the driving distance threshold d_0 using $\mathcal{M}$.
  Compute the accessibility index A for population point p' based on the aforementioned formula.
end for

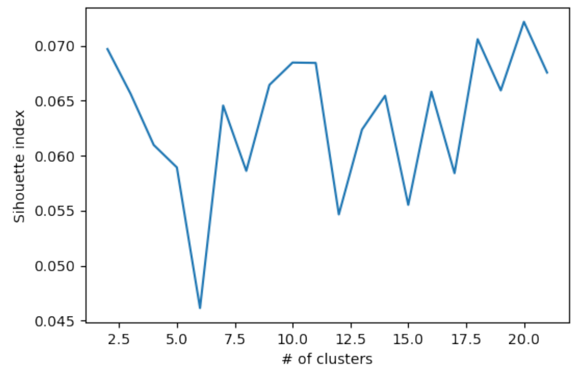
Eventually, we are able to generate a fishnet of points, each of which is assigned with an accessibility index. The index represents the general accessibility to PEVCS for the population within a $500m \times 500m$ area around a certain population point.

3.3 POI-based Land Use Clustering

This subsection will briefly cover the attempt to use embedding method from Natural Language Processing to cluster POI points for land use classification. Yao et al. (2017) and Yan et al. (2017) adopted the idea of word embedding from Google's Word2Vec (Mikolov et al., 2013). In Word2Vec, Mikolov et al. developed high-dimensional vector representation



(a) Silhouette score computed based on haversine distance



(b) Silhouette score computed based on l2 distance of representational vectors

Figure 4: Silhouette score of land use clusters generated by POI K-means clustering

for each word token using the nearby words, based on the notion that each word's meaning is determined by its context. Yan et al. and Yao et al. used the same concept, developing high dimensional vector representation for POI categories based on the spatially nearby POI points. For each POI category, a vector representation is trained. Then, K-means clustering algorithm can be used to unsupervisedly classify different POI categories into land use groups.

However, our implementation of such algorithm reports a bad silhouette index, which indicates a small distinction across different POI clusters and a less significant clustering pattern. From Figure 4, no matter how many clusters we use, the clustering algorithm is not able to generate a desirable silhouette score (silhouette score ranges from -1 to 1).

Thus, we ultimately used hand-classified land use where we manually agglomerate the different POI categories into commercial (e.g. malls, supermarkets, etc.), residential (e.g. neighborhoods, apartments, etc.), industrial (e.g. factories), institutional (e.g. governmental institutions, schools, etc.), transportation (e.g. bus transits, car parks, etc.), green space, and agricultural (e.g. farm, fisheries) land use.

3.4 Regression Models

Regression analysis was used to investigate the relationship between the accessibility of public EV charging stations and a set of explanatory variables. The relationship can thus be interpreted and analyzed from a perspective of spatial equity. Two types of regression were applied: linear regression (LR) and geographically weighted regression (GWR).

LR is a statistical method that allows us to model the linear relationship between a dependent variable and one or more independent variables. In this study, the dependent variable was the number of public EV charging stations, while the explanatory variables were night time light, population of different age groups, population of different sex, road network kernel density, housing price, and land use type represented by kernel density of POI. The goal of LR is to find the best fitting straight line that describes this relationship and detect the significance of each explanatory variable.

Geographically weighted regression (GWR) is a variant of LR that accounts for spatial autocorrelation, which refers to the tendency of nearby observations to be more similar to each other than to observations farther away. The model can be formulated in the following way:

$$y_i = \beta_0(s_i) + \sum_{j=1}^{p} \beta_j(s_i) x_{ij} + \epsilon_i$$

$\beta_0(s_i)$ is the intercept term at location i, and $\beta_j(s_i)$ is the coefficient of the jth independent variable at location i. The term s_i denotes the location i. The error term ϵ_i represents the deviation of the observed value y_i from the predicted value at location i.

The GWR model is similar to the LR model, except that the intercept and coefficients are allowed to vary spatially, rather than being estimated for the entire study area. This allows the GWR model to capture local variations in the relationships between the dependent and independent variables. We used GWR model on the same set of data as the LR, expecting that the GWR model can capture more spatiality of the data and provide some different insights on the dependent and independent variables.

Multi-scale Geographically Weighted Regression (Fotheringham et al., 2017) is an extension of the GWR model that allows for the analysis of relationships between variables at multiple scales, or distances from the center of each observation. This allows for the identification of spatial patterns that vary with distance from the center of each observation. However, due to its complexity brought by estimating coefficients on different scales and selecting what scales to use, the method is not feasible with our size of data in terms of computational intensity. Thus, we eventually used LR and GWR as the regression analysis models.

4 Results

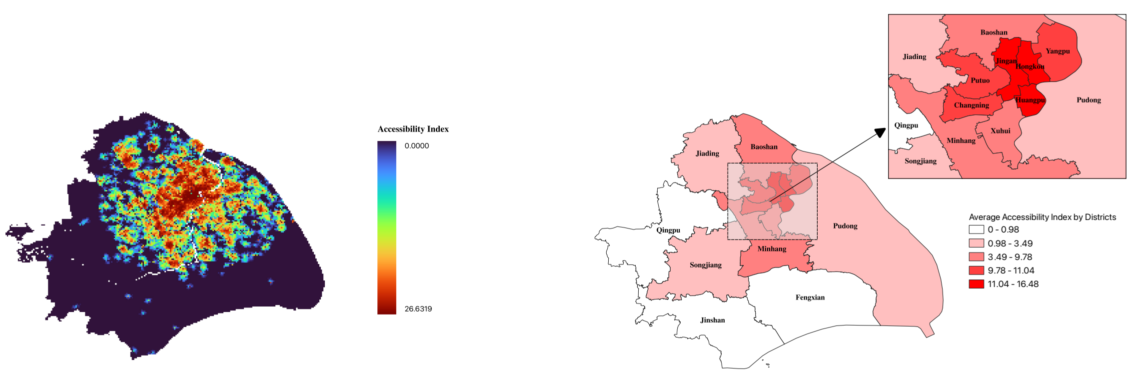
4.1 Spatial Accessibility Analysis

In this subsection we present the accessibility analysis from the IGM tailored for our dataset as we described in subsection 3.2. In Figure 5(a). We observed a noticeably higher accessibility of PEVCS in the city center and a lower accessibility in the suburban area. Comparing different districts (See Figure 5(b) and Table 3), we found that the central districts such as Huangpu, Jingan, Hongkou rank higher than non-central districts like Fengxian, Jinshan, Qingpu, etc. This is not a surprising statistic since we observe that PEVCSs are very sparsely distributed in the suburban area from Figure 3. Pudong district has both area closer to the city center and peripheral area has a relatively lower PEVCS accessibility in the rank.

We further look at the regional accessibility distribution of each district in Figure 6. We found that Jiading, Baoshan, Minhang, Songjiang and Pudong share the similar distribution pattern where there exist some high accessibility clusters while the other areas generally have little access to PEVCS. Although the high accessibility areas are mostly closer to the

District	Mean	Median	Standard Deviation
Huangpu	16.48	17.4	5.53
Jingan	16.03	17.1	6.45
Hongkou	13.03	12.59	6.46
Yangpu	10.54	10.15	7.34
Changning	10.38	10.9	7.16
Putuo	10.16	10.06	6.85
Xuhui	9.52	7.67	7.53
Minhang	5.94	4.11	6.11
Baoshan	4.06	1.5	5.67
Jiading	2.19	0.0	4.32
Pudong	2.63	0.0	4.62
Songjiang	1.06	0.0	2.67
Qingpu	0.66	0.0	2.1
Fengxian	0.12	0.0	0.94
Jinshan	0.00	0.00	0.35

Table 3: Accessibility index by district



(a) Accessibility of PEVCS overall

(b) Accessibility of PEVCS by district

Figure 5: Accessibility of PEVCS in Shanghai

city center, there still exist some high accessibility clusters that are not in close proximity to the center. Putuo, Xuhui, Changning, and Yangpu share the similar pattern where a certain part of the area has relatively lower accessibility level, while the other part has high PEVCS accessibility. No individual clusters are observed. For example, the area alongside the Huangpu River in Yangpu district and the southern part of Xuhui district has lower

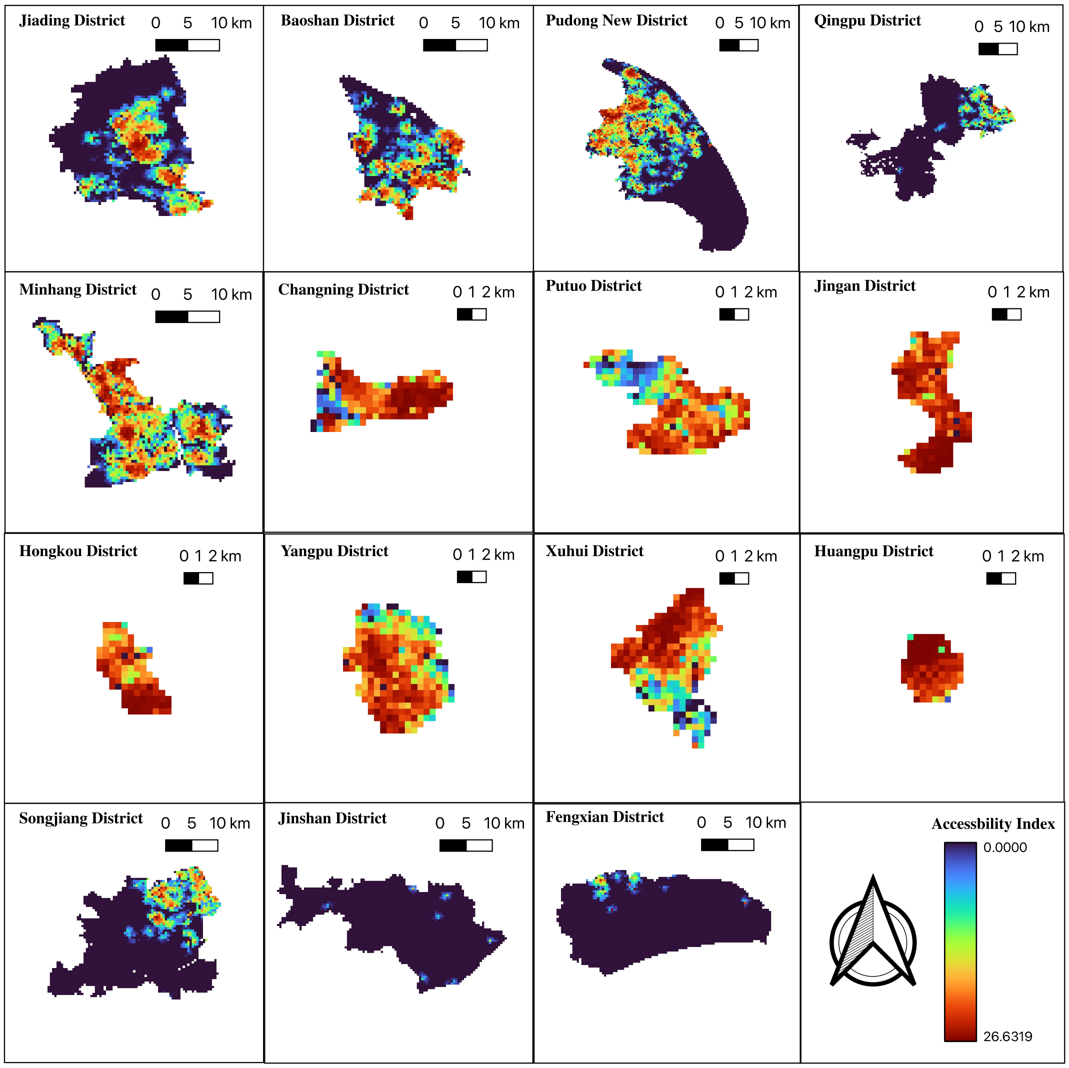


Figure 6: Accessibility pixel-level distribution of each district

access to PEVCSs. Jingan, Huangpu, Hongkou basicaly have high accessibility across the whole area, while Fengxian, Jiading, Qingpu demostrate the opposite.

4.2 Regression Results

Before conducting the regression analysis, we firstly compute the Pearson's correlation among all variables. Figure 7 shows a heatmap of the correlation coefficients. We observe that the male, female, and age group population are extremely correlated with each other. Their correlation coefficients are approaching 1, which indicates perfect or nearly perfect linear association. Thus, the original data has multicollinearity issue that can catastrophically degenerate the regression model. We hence combined the age group from 20 to 50, and removed the other age groups. We also tried to remove male or female population variable but it turns out that they don't cause too much multicollinearity and including both of them can significantly boost the regression model fitness.

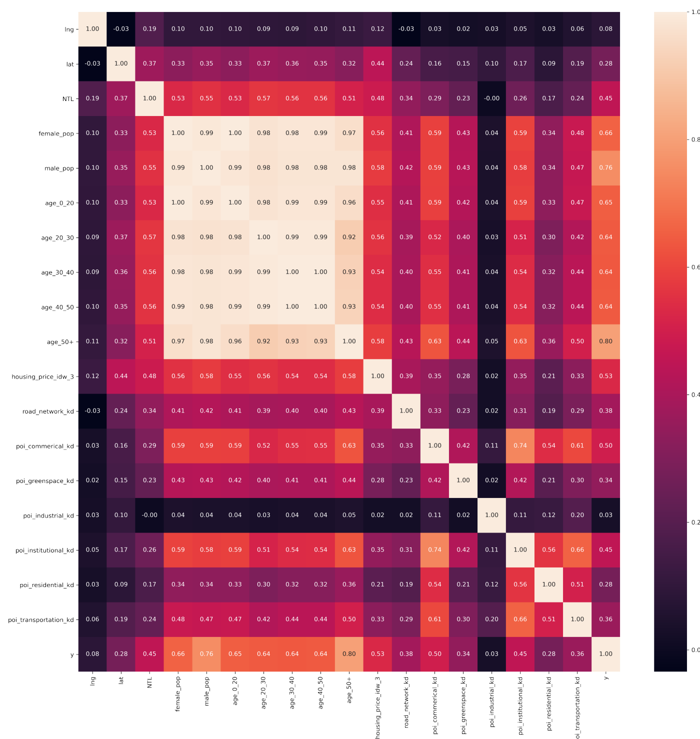


Figure 7: Pearson's correlation heatmap of all variables

The multiple linear regression results are presented in Table 4. Overall, the linear model without considering the spatial autocorrelation can already reach Adjusted $R^2 = 0.98$, which indicates a nearly perfect fit of the data. This result indicates that a LR model is already satisfactory to analyze the association between accessibility and various factors. In terms of statistical significance, we see that among all independent variables, only institutional land use and greenspace didn't demonstrate a significant association with accessibility. The other factors all show a strongly significant linear relationship with accessibility.

	Coeff. Est.	**Std. Err.**	**p-value**	**GWR** Mean Coeff. Est.	**GWR** Std. Err.
CONSTANT	0.000	0.001	1.000	0.009	0.074
NTL	0.005	0.001	0.000***	0.001	0.009
Female Population	-2.956	0.007	0.000***	-2.871	0.745
Male Population	5.131	0.007	0.000***	5.201	0.069
Age 20-50 Population	-1.508	0.006	0.000***	-1.599	0.599
Housing price	0.026	0.001	0.000***	0.010	0.026
Road Network	0.006	0.001	0.000***	0.003	0.015
POI commercial	0.014	0.002	0.000***	0.004	0.009
POI greenspace	-0.001	0.001	0.204	-0.001	0.004
POI industrial	-0.003	0.001	0.000***	0.001	0.003
POI institutional	-0.001	0.002	0.742	0.000	0.008
POI residential	0.011	0.001	0.000***	0.000	0.007
POI transportation	-0.012	0.001	0.000***	0.000	0.007

$N = 25012$,
Adj. $R^2 = .981$(LR)
Adj. $R^2 = .994$(GWR)

Table 4: Regression Results

As we standardized all variables before conducting the regression so that they are all on the same scale, the coefficient estimate shows the change in accessibility with 1 unit of change in the explanatory variable. For demographic variables, female/male population and Age 20-50 population demonstrates a fairly large absolute influence on the PEVCS accessibility. Meanwhile the NTL coefficient has a much lower value, which indicates that the population distribution doesn't have a strong magnitude of effect on the accessibility despite the statistical significance. These two observations affirms that sex structure and age group has a great scale of association with accessibility. Particularly, places with larger male population will have a significantly higher accessibility to PEVCS, while places with more female residents show the opposite. For adults aged 20-50, the PEVCS accessibility is negatively associated with the population.

Although the other factors all have lower magnitude of association with PEVCS accessibility, we can still infer the relative influence from the sign of the coefficient estimates. PEVCS accessibility is positively linked with housing price, road network, commercial and residential land use. Lower accessibility to PEVCS is reported in correspondence with higher industrial and transportation land use.

Next, GWR is performed on the same set of variables but also putting the spatial coordinates into play. The spatiality does improve the model fitness. However, since the LR is already sufficient, the raw improvement is trivial. Since GWR allows coefficient values to vary with spatial location, p-value is typically impossible to evaluate. The mean coefficient estimates are roughly in line with the LR coefficient estimates. Most of the association comes along with the age and sex structure, while built environment and land use factors contributing little to predicting accessibility level. Besides, the magnitude of built environment and land use factors' coefficients are even lower in GWR than in LR. The GWR thus further discriminates the factors that are strongly associated with accessibility with the ones with lower associations.

5 Discussion

This study probes into the spatial equity of PEVCS access by using the comprehensive IGM of accessibility and regression analysis. Whether the access is equitable is also analyzed with demographic, built environment, and land use factors to see how the PEVCS accessibility is changing in correspondence to those factors. In this section we interpret the analysis results from the spatial accessibility analysis and regression analysis. We also generate methodological insights as well as real-world implications.

In the spatial accessibility analysis, we tweaked the IGM to fit the PEVCS data and designed the efficient algorithm to realize the model in Python code. As a more sophisticated measure raises the bar for computation requirements, an resource-saving yet accurate algorithm is crucial. In our case, pixel-level analysis generally requires a high volume of data, and at the same time IGM computation complexity grows quadratically with data size. As we want to use Amap API for a more accurate calculation of driving distance, the API request limit is the main obstacle. Our ball tree implementation filters out unnecessary computation with ease and significantly increased efficiency. This method is generally applicable to distance-based measure of accessibility on a large dataset.

In Figure 5, the inequity in access to PEVCS between urban and suburban area is extremely evident, which is in line with the previous research on EV charging infrastructure (G. Li et al., 2022; Zhou et al., 2021). In districts such as Qingpu, Fengxian, and Jinshan, residents barely have any access to public EV charging infrastructure. Considering that our IGM takes population density into account, the disparity is rather significant and problematic. According to the Seventh National Census, population share in central districts (Huangpu, Yangpu, Jingan, Xuhui, Putuo, Hongkou) decreased by 3.4% while the suburban districts (all others except Pudong) grows by 2.5%. This counterurbanization phenomenon depicts that more people are living in the suburban districts, and the demand for PEVCS charging stations would also be expected to grow. Yet, the current distribution of PEVCS and accessibility level is definitely not sufficient. The centrality of PEVCS accessibility level doesn't coordinate with the population distribution, causing potential inequity in access.

With a more micro scope, we also found different regional distribution patterns of PEVCS accessibility. In some suburban districts, higher accessibility emerge in the form of a cluster that has a centroids. By consulting the map, these areas are generally areas with higher residential block density (e.g. Pujiang, Chuansha), newly developed area (e.g. Jiading new town, free trade zone), ,or places of attraction (e.g. Meilan Lake). Lower accessibility area were also observed in some central districts (e.g. Huajing in Xuhui District, Taopu in Putuo District). Area alongside Huangpu River also reports lower accessibility. One probable reason is that we compute driving distance between population points and PEVCS, and crossing the river would require a detour to some tunnels or bridges, so half of the area around is not within a close distance. This further validates the realisticness of our measurement of accessibility.

Quantitative results from regression analysis further reveal the inequity in age and sex structure. Higher male population area co-occur with places with higher PEVCS accessibility, while higher female population area are associated with lower PEVCS accessibility area. Sex structure is not typically included in previous PEVCS studies. This result implies that sex might also be an important factor to consider in order to promote an equitable access to the general public. Besides, population with working age (20-50), surprisingly, is negatively correlated with PEVCS accessibility, which contradicts the common sense that most drivers fall into this age group. One possible explanation is that more percentage elder people now live in the central area while more and more younger adults now tend to dwell in suburban area; floating population (Xie et al., 2016), floating population mainly migrant workers dominate the population in suburban area (He & Ning, 2015). Meanwhile, most PEVCS are located in the central area. Thus, this finding again calls for an extra attention in building public charging infrastructure in suburban and outskirt areas in Shanghai to accommodate the working population.

For built environment and land use factors, although we observe some statistically significant variables, but their magnitude of association with accessibility is too low. Not evident inequity are found in these two aspects. This observation might partially be attributed to the variable selection, since demographic factors are too strong predictors of accessibility.

Lastly, regular LR shows a desirable amount information about the relationship between multiple factors and PEVCS without considering the spatiality of the data. The LR model fits almost perfectly. GWR shows overall similar results in terms of coefficient estimates and a slightly higher adjusted R^2 value. The difference is that including spatiality in the regression model can further discriminate the variables with higher predictability on PEVCS accessibility, helping us to secure and validate the spatial inequity in age and sex structure.

6 Conclusion

As an important urban infrastructure, PEVCSs haven't been sufficiently studied in terms of spatial equity and accessibility. Yet, an equitable distribution could promote EV adoption and urban sustainability. Previous method framework are also not generally transferable to other studies in this area. To address these issues, we proposed the pixel-level accessibility analysis with a tailored version of IGM and a efficient algorithm. Regression analysis with LR and GWR are also conducted to quantitatively analyze the underlying factors that causes spatial inequity. This study found a significant inequity between urban and suburban area, as well as in age and sex structure. Also, from qualitative analysis, different regional accessibility patterns are also discovered.

With all publicly available data, our analysis framework is generally applicable to other study areas. Scholars and policymakers could use our method to evaluate spatial equity and accessibility of PEVCSs for research or policy-making aims with ease.

Our study is also subject to several limitations:

1. The PEVCS equity analysis is not well validated with actual EV driver population. In the measurement of accessibility and regression analysis, to analyze detect spatial equity, we should use the EV drivers' population instead of general population density. The findings can also be further analyzed with EV users' data. Due to data availability issues, we didn't put this into practice.

2. The regression analysis is not thorough enough. Built environment and land use factors are showing little contribution in the regression model. A step-wise regression or other

data processing or modeling regimes might generate more insights. The nearly perfect fit of linear model awaits more rigorous justification as well.

3. In qualitative analysis, a case study in some high/low accessibility clusters might explain the regional accessibility patterns more clearly.

4. POI-based land use clustering is not working well, but since it's not a major part of our study, we decided to use hand-classification instead. Yet, it could be interesting to see how unsupervised machine learning algorithm could classify land use differently from human based on POI points, and how it can aid spatial equity analysis.

5. It would be intellectually interesting if a time series analysis of accessibility change. Seeing how the spatial distribution of PEVCS has developed in recent years could potentially generate more insights in planning PEVCS in cities.

References

Anjos, M. F., Gendron, B., & Joyce-Moniz, M. (2020). Increasing electric vehicle adoption through the optimal deployment of fast-charging stations for local and long-distance travel. *European Journal of Operational Research*, *285*(1), 263–278. https://doi.org/https://doi.org/10.1016/j.ejor.2020.01.055

Bondarenko, M., Kerr, D., Sorichetta, A., Tatem, A., & WorldPop. (2020). Estimates of 2020 total number of people per grid square, adjusted to match the corresponding unpd 2020 estimates and broken down by gender and age groupings, produced using built-settlement growth model (bsgm) outputs. https://doi.org/10.5258/SOTON/WP00698

Chang, Z., Chen, J., Li, W., & Li, X. (2019). Public transportation and the spatial inequality of urban park accessibility: New evidence from hong kong. *Transportation Research Part D: Transport and Environment*, *76*, 111–122. https://doi.org/https://doi.org/10.1016/j.trd.2019.09.012

Chen, Y., Bouferguene, A., Shen, Y., & Al-Hussein, M. (2019). Assessing accessibility-based service effectiveness (absev) and social equity for urban bus transit: A sustainability perspective. *Sustainable Cities and Society*, *44*, 499–510. https://doi.org/https://doi.org/10.1016/j.scs.2018.10.003

Cheng, L., Yang, M., Vos, J. D., & Witlox, F. (2020). Examining geographical accessibility to multi-tier hospital care services for the elderly: A focus on spatial equity. *Journal of Transport Health*, *19*, 100926. https://doi.org/https://doi.org/10.1016/j.jth.2020.100926

Dong, G., Ma, J., Wei, R., & Haycox, J. (2019). Electric vehicle charging point placement optimisation by exploiting spatial statistics and maximal coverage location models.

Transportation Research Part D: Transport and Environment, *67*, 77–88. https://doi.org/10.1016/j.trd.2018.11.005

Egnér, F., & Trosvik, L. (2018). Electric vehicle adoption in Sweden and the impact of local policy instruments. *Energy Policy*, *121*, 584–596. https://doi.org/10.1016/j.enpol.2018.06.040

Falchetta, G., & Noussan, M. (2021). Electric vehicle charging network in europe: An accessibility and deployment trends analysis. *Transportation Research Part D: Transport and Environment*, *94*, 102813. https://doi.org/https://doi.org/10.1016/j.trd.2021.102813

Fotheringham, A. S., Yang, W., & Kang, W. (2017). Multiscale geographically weighted regression (mgwr). *Annals of the American Association of Geographers*, *107*(6), 1247–1265. https://doi.org/10.1080/24694452.2017.1352480

Funke, S., Sprei, F., Gnann, T., & Plötz, P. (2019). How much charging infrastructure do electric vehicles need? a review of the evidence and international comparison. *Transportation Research Part D: Transport and Environment*. https://doi.org/10.1016/j.trd.2019.10.024

This paper has some discussion on home charging verse public charging

Hansen, W. G. (1959). How Accessibility Shapes Land Use. *Journal of the American Institute of Planners*, *25*(2), 73–76. https://doi.org/10.1080/01944365908978307

He, D., & Ning, Y. (2015). Evolution of population structure and spatial distribution in shanghai since 2000. In R. Singh (Ed.), *Urban development challenges, risks and resilience in asian mega cities* (pp. 227–249). Springer Japan. https://doi.org/10.1007/978-4-431-55043-3_13

Hove, A., & Sandalow, D. (n.d.). ELECTRIC VEHICLE CHARGING IN CHINA AND THE UNITED STATES, 88.

Hsu, C.-W., & Fingerman, K. (2021). Public electric vehicle charger access disparities across race and income in California. *Transport Policy*, *100*, 59–67. https://doi.org/10.1016/j.tranpol.2020.10.003

Joseph, A. E., & Bantock, P. R. (1982). Measuring potential physical accessibility to general practitioners in rural areas: A method and case study. *Social Science & Medicine*, *16*(1), 85–90. https://doi.org/10.1016/0277-9536(82)90428-2

Juvvala, R., & Sarmah, S. P. (2021). Evaluation of policy options supporting electric vehicles in city logistics: A case study. *Sustainable Cities and Society*, *74*, 103209. https://doi.org/https://doi.org/10.1016/j.scs.2021.103209

Kanuganti, S., Sarkar, A. K., & Singh, A. P. (2016). Quantifying accessibility to health care using two-step floating catchment area method (2sfca): A case study in rajasthan. *Transportation Research Procedia*, *17*, 391–399. https://doi.org/https://doi.org/10.1016/j.trpro.2016.11.080

Kelobonye, K., Zhou, H., McCarney, G., & Xia, J. ((2020). Measuring the accessibility and spatial equity of urban services under competition using the cumulative opportunities measure. *Journal of Transport Geography*, *85*, 102706. https://doi.org/https://doi.org/10.1016/j.jtrangeo.2020.102706

Li, G., Luo, T., & Song, Y. (2022). Spatial equity analysis of urban public services for electric vehicle charging—Implications of Chinese cities. *Sustainable Cities and Society*, *76*, 103519. https://doi.org/10.1016/j.scs.2021.103519

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*, 103181. https://doi.org/10.1016/j.scs.2021.103181

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space [number: arXiv:1301.3781 arXiv:1301.3781 [cs]], (arXiv:1301.3781). http://arxiv.org/abs/1301.3781

Minow, M. (2021). Equality vs. equity. *American Journal of Law and Equality*, *1*, 167–193. https://doi.org/10.1162/ajle_a_00019

NASA Earth. (2015). Mod06$_l 2modis/terraclouds5 - minl2swath1kmand5km$. https://doi.org/10.5067/MODIS/MOD06_L2.006

Omohundro, S. M. (2009). Five balltree construction algorithms.

Páez, A., Scott, D. M., & Morency, C. (2012). Measuring accessibility: Positive and normative implementations of various accessibility indicators. *Journal of Transport Geography*, *25*, 141–153. https://doi.org/10.1016/j.jtrangeo.2012.03.016

Roy, A., & Law, M. (2022). Examining spatial disparities in electric vehicle charging station placements using machine learning. *Sustainable Cities and Society*, *83*, 103978. https://doi.org/10.1016/j.scs.2022.103978

Schroeder, A., & Traber, T. (2012). The economics of fast charging infrastructure for electric vehicles. *Energy Policy*, *43*, 136–144. https://doi.org/https://doi.org/10.1016/j.enpol.2011.12.041

Shanghai Municipal Bureau of Statistics. (2022). *Shanghai statistical yearbook 2021*. China Statistics Press.

Sharkey, J. R., Horel, S., Han, D., & Huber, J. C. (2009). Association between neighborhood need and spatial access to food stores and fast food restaurants in neighborhoods of colonias. *International Journal of Health Geographics*, *8*(1), 9. https://doi.org/10.1186/1476-072X-8-9

Shen, J., Cheng, J., Huang, W., & Zeng, F. (2020). An exploration of spatial and social inequalities of urban sports facilities in nanning city, china, 20.

Shi, Y., Yang, J., Keith, M., Song, K., Li, Y., & Guan, C. (2022). Spatial accessibility patterns to public hospitals in shanghai: An improved gravity model. *The Professional Geographer*, *74*(2), 265–289. https://doi.org/10.1080/00330124.2021.2000445

Smoyer-Tomic, K. E., Spence, J. C., Raine, K. D., Amrhein, C., Cameron, N., Yasenovskiy, V., Cutumisu, N., Hemphill, E., & Healy, J. (2008). The association between neighborhood socioeconomic status and exposure to supermarkets and fast food outlets. *Health Place*, *14*(4), 740–754. https://doi.org/https://doi.org/10.1016/j.healthplace.2007.12.001

State Council of People's Republic of China. (2019). New energy vehicle industry development plan (2021 - 2035).

Tsou, K.-W., Hung, Y.-T., & Chang, Y.-L. (2005). An accessibility-based integrated measure of relative spatial equity in urban public facilities. *Cities*, *22*(6), 424–435. https://doi.org/https://doi.org/10.1016/j.cities.2005.07.004

United Nations. (2015). Transforming our world: The 2030 agenda for sustainable development. https://sdgs.un.org/2030agenda

White, L. V., Carrel, A. L., Shi, W., & Sintov, N. D. (2022). Why are charging stations associated with electric vehicle adoption? Untangling effects in three United States metropolitan areas. *Energy Research & Social Science*, *89*, 102663. https://doi.org/10.1016/j.erss.2022.102663

Xie, B., Zhou, J., & Luo, X. (2016). Mapping spatial variation of population aging in china's mega cities. *Journal of Maps*, *12*(1), 181–192. https://doi.org/10.1080/17445647.2014.1000984

Xu, J., Zhong, L., Yao, L., & Wu, Z. (2018). An interval type-2 fuzzy analysis towards electric vehicle charging station allocation from a sustainable perspective. *Sustainable Cities and Society*, *40*, 335–351. https://doi.org/https://doi.org/10.1016/j.scs.2017.12.010

Yan, B., Janowicz, K., Mai, G., & Gao, S. (2017). From ITDL to Place2Vec: Reasoning About Place Type Similarity and Relatedness by Learning Embeddings From Augmented Spatial Contexts. *Proceedings of the 25th ACM SIGSPATIAL International*

Conference on Advances in Geographic Information Systems, 1–10. https://doi.org/10.1145/3139958.3140054

Yao, Y., Li, X., Liu, X., Liu, P., Liang, Z., Zhang, J., & Mai, K. (2017). Sensing spatial distribution of urban land use by integrating points-of-interest and Google Word2Vec model. *International Journal of Geographical Information Science*, *31*(4), 825–848. https://doi.org/10.1080/13658816.2016.1244608

Zhang, Y., Zhang, Q., Farnoosh, A., Chen, S., & Li, Y. (2019). Gis-based multi-objective particle swarm optimization of charging stations for electric vehicles. *Energy*, *169*, 844–853. https://doi.org/10.1016/j.energy.2018.12.062

Zhou, X., Han, P., Huang, J., & Yu, Z. (2021). The measurement method of spatiotemporal accessibility of electric vehicle charging stations in the dynamic time-dependent urban environment. *IOP Conference Series: Earth and Environmental Science*, *783*(1), 012078. https://doi.org/10.1088/1755-1315/783/1/012078