

## Research Article

# Analysis of Electric Vehicle Charging Behavior Patterns with Function Principal Component Analysis Approach

Chenxi Chen,<sup>1</sup> Yang Song,<sup>1</sup> Xianbiao Hu ,<sup>1</sup> and Ivan G. Guardiola <sup>2</sup>

<sup>1</sup>Department of Civil, Architectural and Environmental Engineering, Missouri University of Science and Technology, Rolla, MO 65409, USA

<sup>2</sup>School of Business Administration, Fort Lewis College, 1000 Rim Drive, Durango, CO 81301, USA

Correspondence should be addressed to Xianbiao Hu; [xbhu@mst.edu](mailto:xbhu@mst.edu)

Received 8 May 2020; Revised 9 November 2020; Accepted 11 November 2020; Published 23 November 2020

Academic Editor: Yuanyuan Zhang

Copyright © 2020 Chenxi Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This manuscript focused on analyzing electric vehicles' (EV) charging behavior patterns with a functional data analysis (FDA) approach, with the goal of providing theoretical support to the EV infrastructure planning and regulation, as well as the power grid load management. 5-year real-world charging log data from a total of 455 charging stations in Kansas City, Missouri, was used. The focuses were placed on analyzing the daily usage occupancy variability, daily energy consumption variability, and station-level usage variability. Compared with the traditional discrete-based analysis models, the proposed FDA modeling approach had unique advantages in preserving the smooth function behavior of the data, bringing more flexibility in the modeling process with little required assumptions or background knowledge on independent variables, as well as the capability of handling time series data with different lengths or sizes. In addition to the patterns revealed in the EV charging station's occupancy and energy consumption, the differences between EV driver's charging time and parking time were analyzed and called for the needs for parking regulation and enforcement. The different usage patterns observed at charging stations located on different land-use types were also analyzed.

## 1. Introduction

Electric vehicles (EVs) produce fewer emissions that contribute to climate change and smog than conventional vehicles and help the United States achieve a greater diversity of fuel choices available for transportation. The evolution of EVs has advanced from models best suited for commuting or traveling short distances to vehicles that can travel more than 200 or even 300 miles per charge.

Proper planning of the EV charging infrastructure and scientific determination of their locations are critical to promoting EV ownership and usage. Modeling efforts can be found in the literature, such as the electric vehicle infrastructure projection (EVI-Pro) model developed by the National Renewable Energy Lab to address the fundamental question of how much charging infrastructure is needed in the United States to support Plug-in-EVs (PEVs) [1]. The model generated a quantitative estimate for a US network of

nonresidential (public and workplace) EVSE that would be needed to support broader PEV adoption. He et al. studied how to optimally locate public charging stations on a road network, considering drivers' spontaneous adjustments and interactions of travel and recharging decisions [2]. A bilevel programming model with the consideration of EV's driving range was proposed in [3], with the upper level to optimize the position of charging stations so as to maximize the path flows that used the charging stations, while the user equilibrium of route choice with the EV's driving range constraint was formulated in the lower level. Other research on EV charging station locations can also be found in [4, 5] and many others.

Another approach to supporting the planning of charging infrastructure was to perform analysis of EV-related data, with the goal of identifying charging behavior patterns and inferring the scenarios of when and where people need to charge their vehicles. For example, the

driving data in Denmark was analyzed to extract the information of driving distances and driving time periods which were used to represent the driving requirements and the EV unavailability. The Danish National Transport Survey data were used to implement the driving data analysis [6]. The analysis of charge event data in Ireland for public charging infrastructure, including data from fast-charging infrastructure, and additionally a limited quantity of household data was performed in [7]. Sun et al. studied driver's charging timing decisions, in which a mixed logit model with unobserved heterogeneity is applied to panel data extracted from a two-year field trial on battery electric vehicle usage in Japan [8]. The analysis over the real-world dataset can also be found in [9, 10] and others.

This manuscript focused on performing analysis over the 5-year real-world charging event log data, from a total of 455 charging stations in Kansas City, Missouri (KCMO), with a functional data analysis (FDA) approach. The EV charging equipment recorded which vehicle was charged at which charging station, at what day and time. Such charging event log data contained many significant pieces of information for understanding EV charging patterns and user behavior. The goal of this research was to provide theoretical support to the EV infrastructure planning and regulation, as well as the power grid load management. We argue that compared with the existing research over the real-world charging event data, the proposed FDA modeling approach had many unique advantages over the prevailing discrete-based analysis models and led to some important insights that were difficult to model or discover with the other approaches.

Commonly, time series data (such as the EV charging log data used in this research) were treated as multivariate data because they were given as a finite discrete time series [11–13]. This usual multivariate approach completely ignored important information about the smooth functional behavior of the generating process that underpins the data [14]. For example, in our context, the vehicles' charging process was continuous and so was the time-dependent occupancy of a particular charging station. Additionally, in the previous research, performance measurements need to be defined by the researchers to extract useful information from the raw dataset, before any meaningful analysis can be performed. However, they were usually defined arbitrarily, based on the researcher's experience in the field. Instead of assuming a variety of explanatory variables, which was difficult or even impossible to enumerate and collect data for, FDA is much more flexible with little required assumptions or background knowledge on independent variables. Last not but least, time series data often has different time intervals or different lengths which are hard to deal with by other tools. In our context, some charging stations were more frequently used and might have thousands of charging records while others might only have a few hundred. It was thus impossible to apply principal component analysis (PCA) to the charging log dataset directly because of the dimension inconsistency.

The basic idea behind FDA is to express discrete observations arising from time series in the form of a function (i.e., to create functional data) that represents the entire

measured function as a single observation and then to draw the modeling and/or prediction information from a collection of functional data by applying statistical concepts from multivariate data analysis [15]. With this said, this manuscript firstly represented the EV charging dataset with a continuous functional form, then performed function principal component (FPC) analysis to identify the main contributing principal components (PC), and analyzed the dataset from different perspectives to understand EV owner's charging behavior patterns.

This research aimed to provide theoretical support to the EV infrastructure planning and regulation, as well as the power grid load management. To achieve such goals, the focuses were placed on three aspects. (1) The first aspect is the variability analysis of the daily usage patterns of all EV charging stations, in which the 24-hour occupancy of all charging stations in one day was treated as one continuous curve. Such analysis can provide insights and directly support the planning of new EV charging infrastructures. (2) The second aspect is the variability analysis of the daily energy consumption of all EV charging stations, in which the total energy consumption of all charging stations in one day was treated as one continuous curve. Such analysis was important from the power grid load management perspective. (3) At the station level, the usage pattern variabilities were analyzed, in which one station's usage over the entire observation period was treated as a continuous curve. This analysis revealed insights on the usage pattern differences at the station level and was combined with the land-use information for better EV charging infrastructure planning and management purposes.

The remaining part of this paper is organized as follows. The charging event log dataset used in this research is firstly presented in Section 2. Section 3 presents the analysis methodology, including the data smoothing, variable calculation, and the functional principal component analysis. The analysis results are shown and compared in Section 4. Section 5 concludes this research.

## 2. Data

This section presents the real-world charging event log data used in this research. The data was collected from 455 charging stations between January 2014 and November 2019 in Kansas City, Missouri (KCMO). The dataset included a total of 226,652 charging records from 4,921 users. Most of the stations were concentrated in the downtown area of KCMO. The spatial distribution of charging stations was shown in Figure 1, in which Figure 1(a) showed an overview and Figure 1(b) zoomed in to the downtown area.

In the collected dataset, each row contained the information of a charging event and had a total of 30 columns/attributes. Table 1 showed the sample data from the dataset, in which only the most critical and relevant information was displayed. The complete dataset included information of the following three categories:

- (1) Charging station information: including a unique station ID, station name, address and zip code where

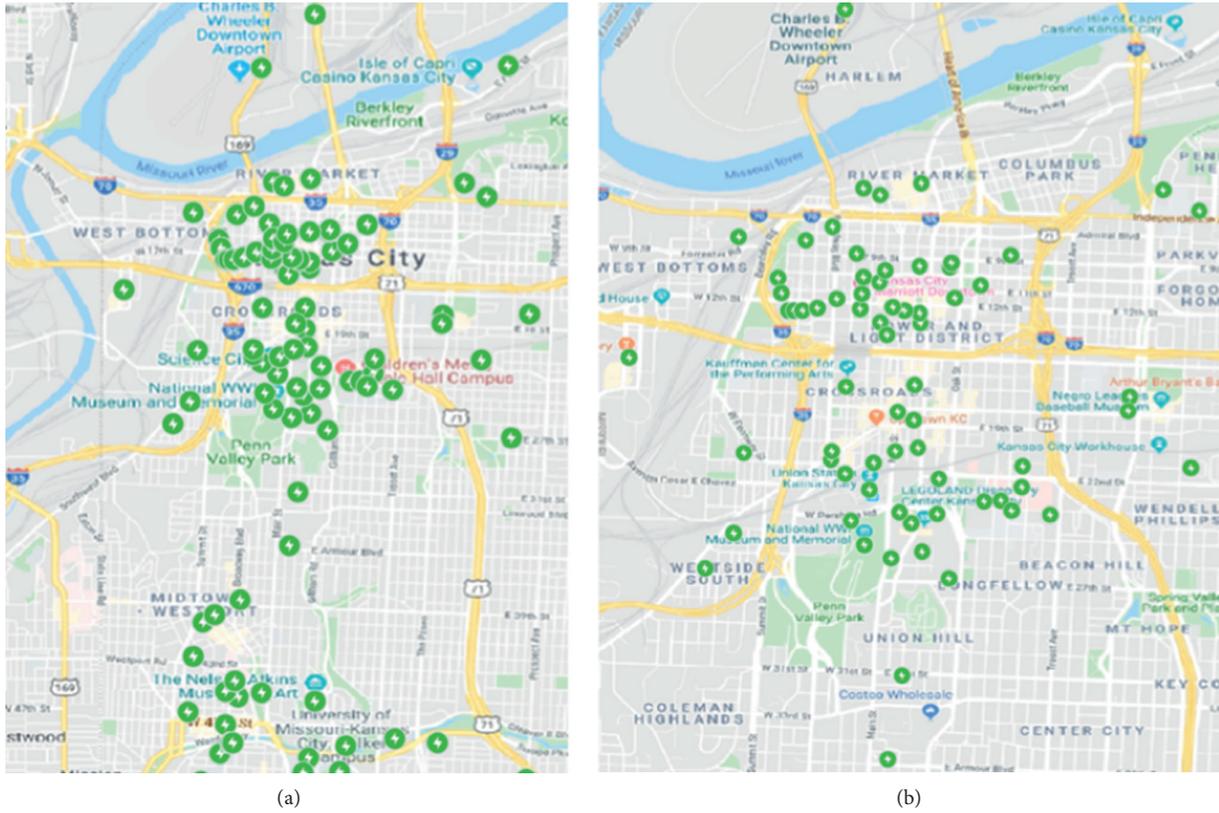


FIGURE 1: Locations of the charging station in KCMO. (a) Overview. (b) Downtown.

TABLE 1: Sample Charging Event Log data.

Station name	Start date	End date	Time duration (hh: mm:ss)	Charging time (hh: mm:ss)	Energy (kWh)	Latitude	Longitude	User ID
KCPL/@JE DUNN	9/18/2018	9/18/2018	1:23:15	1:23:02	5.934	39.1011	94.5763	756927
PG125 C	14:27	15:50						
KCPL/@WOLF PG-129A	9/18/2018	9/18/2018	8:42:04	2:15:18	6.766	39.0999	94.5791	1533491
KCPL/@LOOSE PRK-121A	9/18/2018	9/18/2018	0:00:30	0:00:00	0	39.0346	94.5947	598567

the station was located at, MAC address, latitude and longitude of the station, and type of the charging ports which included level 1, level 2, and DC fast charge

- (2) Electric vehicle attribute: including a unique ID of the electric vehicle and zip code where this electric was registered in (which is usually the zip code of the driver’s home)
- (3) Charging event data: including the start date and time of the charging event, end date and time of the charging event, charging time which is equal to the end time minus start time, total duration which included not only the time spent on charging but also the time spent on parking afterward, start state of charge (SOC), end SOC, energy charged, Greenhouse Gas (GHG) saving, and information on how was the charging event ended (e.g., terminated by

customer or server). Duration is the total time that a station is occupied, which is one of the most significant properties we are interested

### 3. Methodology

This section presents the analysis methodology used in this manuscript, including a brief overview of the function data analysis approach, charging pattern definition, and functional principal component analysis.

3.1. FDA Method Overview. To process “curve-like” data that are continuous in nature, such as the time-dependent charging station usage rates of this manuscript, one advanced and popular method is functional data analysis [16]. Apart from the commonly seen multivariate data analysis approaches, the proposed FDA approach considered EV

charging usage as a function of time; thus, all the EV charging events that were sampled in different scales, from different charging stations built at different time periods, and used with different frequencies with different data sizes, were all modeled uniformly by functions. In other words, under the functional data analysis approach, each charging pattern to be defined in Section 3.2 was treated as one functional data. By applying basis expansion techniques such as B-spline expansion denoted in (4) [17, 18], each charging pattern can be modeled and expressed in a functional form.

$$y(t) = \sum_{j=1}^M \beta_j \varphi_j(t), \quad (1)$$

where  $y(t)$  is the original function,  $\varphi_j(t)$  is  $j^{\text{th}}$  basis functions, and  $\beta_j$  is the coefficient of the corresponding basis function.

With such a data analysis approach, all charging patterns which were sampled in different scales can be uniformly expressed in the same functional form. Additional benefits of such an approach also included the reduction of unnecessary noise in raw data by basis expansion smoothing. Based on this model, all the information from the raw data can be projected to  $M$  basis coefficients  $\beta_j$ . Obtaining the basis coefficients can be done through an ordinary least square (OLS) regression. This process is also known as B-spline smoothing. Section 3.3 will further describe the basis expansion and modeling process.

**3.2. Charging Pattern Definition.** This subsection defines the three charging patterns to be analyzed, corresponding to the three analyses performed in the numeric analysis section.

**3.2.1. Daily Usage Occupancy.** Daily usage occupancy was defined to measure the 24-hour time-dependent usage occupancy within a single day, by aggregating the charging events at all charging stations.

A binary variable  $u_{j,d}(t)$  was firstly defined to denote the usage condition of a charging station  $j$  in hour  $t$  at day  $d$ . If the charging station was used for at least once,  $u_{j,d}(t) = 1$ , else 0.

$$u_{j,d}(t) = \begin{cases} 1, & \text{the station was in use,} \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

Next, the 24-hour time-dependent occupancy for each day can be calculated by aggregating all charging station's usage, so that in the end, one curve was generated to represent the daily usage occupancy of each day.

$$u_d(t) = \frac{1}{J} \sum_j u_{j,d}(t), \quad (3)$$

where  $u_d(t)$  means the average occupancy of time  $t$  at day  $d$ , and  $J$  means the total number of stations on day  $d$ . Note  $J$  is day-dependent, so  $J = J(d)$ .

**3.2.2. Daily Energy Consumption.** As shown in Table 1, the energy consumption  $e_{i,d}$  associated with each charging event  $i$  at day  $d$  was recorded and thus was directly available. First,  $e_{i,d}$  was proportionally assigned to each hour, so that

$$e'_{i,d}(t) = \frac{e_{i,d}}{l_{i,d}} * l_{i,d}(t), \quad (4)$$

in which  $e'_{i,d}(t)$  was the proportion of energy consumption  $e_{i,d}$  in hour  $t$ ,  $l_{i,d}$  was the duration of charging event  $i$ , and  $l_{i,d}(t)$  was the proportion of  $l_{i,d}$  in hour  $t$ .

Next, the 24-hour time-dependent energy consumption for each day can be calculated by doing aggregation over all charging stations, so that in the end, one curve was generated to represent the daily energy consumption of each day.

$$e_d(t) = \frac{1}{J} \sum_i e'_{i,d}(t). \quad (5)$$

**3.2.3. Station-Level Occupancy.** Similar to the daily usage occupancy calculation, to analyze the difference between stations, aggregations can be performed over the days. For each station  $j$ , its aggregated occupancy at time  $t$  was denoted as  $u_j(t)$  and calculated as follows.

$$u_j(t) = \frac{1}{T} \sum_d u_{j,d}(t), \quad (6)$$

where  $u_{j,d}(t)$  was calculated from (2), and  $T$  denoted the total number of days in the analyzed time period. In the end, one curve was generated to represent the aggregated usage occupancy of each charging station.

**3.3. Data Smoothing.** This subsection focuses on how to represent the charging patterns defined above as curves. Since  $u_d(t)$ ,  $e_d(t)$ , and  $u_j(t)$  are all time-dependent, they can be represented by  $(t, u_d(t), e_d(t), \text{ and } u_j(t))$ . Based on B-spline expansion, these discrete points can be modeled by a continuous function:

$$\begin{aligned} u_d(t) &= \sum_{j=1}^M \alpha_{i,j} \varphi_j(t) = [\alpha_{i,1} \ \alpha_{i,2} \ \dots \ \alpha_{i,M}] \begin{bmatrix} \varphi_1(t) \\ \varphi_2(t) \\ \dots \\ \varphi_M(t) \end{bmatrix} = \alpha_i^T \phi(t), \\ e_d(t) &= \sum_{j=1}^M \beta_{i,j} \varphi_j(t) = [\beta_{i,1} \ \beta_{i,2} \ \dots \ \beta_{i,M}] \begin{bmatrix} \varphi_1(t) \\ \varphi_2(t) \\ \dots \\ \varphi_M(t) \end{bmatrix} = \beta_i^T \phi(t), \\ u_j(t) &= \sum_{j=1}^M \gamma_{i,j} \varphi_j(t) = [\gamma_{i,1} \ \gamma_{i,2} \ \dots \ \gamma_{i,M}] \begin{bmatrix} \varphi_1(t) \\ \varphi_2(t) \\ \dots \\ \varphi_M(t) \end{bmatrix} = \gamma_i^T \phi(t). \end{aligned} \quad (7)$$

An example of B-spline expansion was depicted in Figure 2, where a smoothed function (solid black curve) was represented as a summation of B-spline basis functions (dashed black curves) to model the raw daily usage occupancy data (red diamond). The heights of these basis

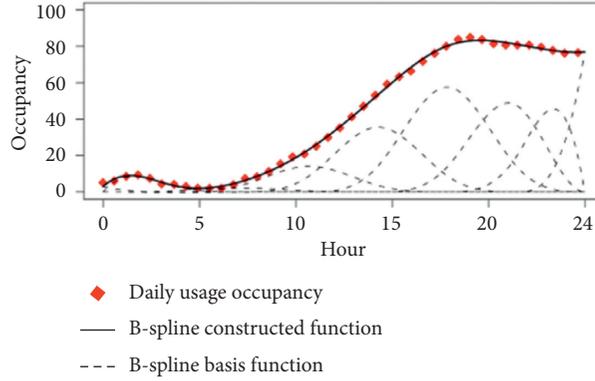


FIGURE 2: Illustration of B-Spline expansion.

functions were determined by the basis coefficients  $\alpha_i^T$ ,  $\beta_i^T$ , and  $\gamma_i^T$ . Such basis expansion method was advantageous in terms of transferring a high volume of data points into several basis functions' coefficients without losing the original pattern [11].

To obtain the basis coefficients  $\alpha_i^T$ ,  $\beta_i^T$ , and  $\gamma_i^T$ , the least square regression model was constructed as follows.  $u_d(t)$  was used as an example to avoid repetition, but the method presented hereinafter was directly applicable to  $e_d(t)$  and  $u_j(t)$  as well.

$$u\_SSE_i = \sum_{k=1}^{K_i} [u_d(t) - \alpha_i^T \varphi(t_{i,k})]^2. \quad (8)$$

To simplify the notations for the least square model, some matrix-formed data were introduced as follows:

$$u_d(t) = \begin{bmatrix} u_{di}(1) \\ u_{di}(2) \\ \dots \\ u_{di}(k) \end{bmatrix}, \quad (9)$$

$$\Phi_i = \begin{bmatrix} \varphi_1(t_{i,1}) & \varphi_2(t_{i,1}) & \dots & \varphi_M(t_{i,1}) \\ \varphi_1(t_{i,2}) & \varphi_2(t_{i,2}) & \dots & \varphi_M(t_{i,2}) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_1(t_{i,K_i}) & \varphi_2(t_{i,K_i}) & \dots & \varphi_M(t_{i,K_i}) \end{bmatrix},$$

where  $u_{di}(k)$  was a vector that contained raw data points in day  $i$ .  $\Phi_i$  was a  $K_i \times M$  matrix; each column was the basis function value at all the time points. By reconstructing these usage occupancy data, the least square model can be rewritten in a simple quadratic form:

$$u\_SSE_i = (u_{di} - \Phi_i \alpha_i)^T (u_{di} - \Phi_i \alpha_i). \quad (10)$$

Thus, the basis coefficients for day  $i$  can be estimated by

$$\hat{\alpha}_i = (\Phi_i^T \Phi_i)^{-1} \Phi_i^T u_{di}. \quad (11)$$

Through the B-spline model and least square regression, all three charging patterns defined above were converted into the basis coefficients. The functions can be obtained by  $u_d(t) = \hat{\alpha}_i^T \varphi(t)$ ;  $e_d(t) = \hat{\beta}_i^T \varphi(t)$ , and  $u_j(t) = \hat{\gamma}_i^T \varphi(t)$ .

**3.4. Functional Principal Component Analysis.** After data smoothing, functional PCA was enabled as a powerful tool of the FDA approach to explore the curve's underlying features. In multivariate data analysis, PCA was commonly used to convert a large number of variables to some comprehensive variables that are much less in quantity but account for the highest variability. The mathematical solution of this problem was similar to finding the eigenvalue and the new variables were the functional principal components (FPCs).

In the FDA approach, the analyzed function contained information of a set of specific variables at enormous time points in a time interval. As a result, the work was confronted with the curse of dimensionality if the time was seen as the independent variable in the functional case. Consequently, the functional PCA method can be applied for the purpose of dimension reduction. In [19, 20], FPCA was employed as a data dimensionality reduction technique in the modeling of traffic flow patterns, which inhibit similar functional characteristics observed in EV charging. The approach was similar to the multivariate case. The dependent variable  $x_i(s)$  ( $s \in T$ ) was relative to  $x_{ij}$  in multivariate case.

$$f_i = \int \beta(s) x_i(s) ds = \int \beta x_i, \quad (12)$$

where  $\beta(s)$  was the weight value and  $\beta^k(s)$  denoted the weight function of  $k^{\text{th}}$  principal component. The variance function can be represented as  $\text{Var}(f) = (1/N) \sum_{i=1}^N (\int \beta x_i)^2$ .

Let  $f^k$  denote the  $k^{\text{th}}$  principal component, where  $k = 1, 2, \dots, K$ , so the relationship is  $\text{Var}(f^1) > \text{Var}(f^2) > \dots > \text{Var}(f^K)$ .

To calculate the first principal component, we just need to solve the following optimization problem:

$$\max \frac{1}{N} \sum_{i=1}^N \left( \int \beta^1 x_i \right)^2$$

s.t.

$$\int [\beta^1(s)]^2 ds = \int [\beta^1]^2 = \|\beta^1\|^2 = 1, \quad (13)$$

and the  $k^{\text{th}}$  principal component can be calculated by the following optimization problem:

$$\begin{aligned} & \max \frac{1}{N} \sum_{i=1}^N \left( \int \beta^k x_i \right)^2 \\ & \text{s.t.} \\ & \int [\beta^k]^2 = 1; \int \beta^k \beta^{k-m} = 0, \quad m = 1, \dots, k-1. \end{aligned} \quad (14)$$

The covariance of  $x(s)$  and  $x(t)$  can be calculated by

$$v(s, t) = \frac{1}{N-1} \sum_{i=1}^N x_i(s)x_i(t). \quad (15)$$

The weight function of functional principal components  $\beta(s)$  is needed to satisfy the following secular equation:

$$\int v(s, t)\beta(t)dt = \lambda\beta(s), \quad (16)$$

where  $\lambda$  was the eigenvalue and  $\lambda_k / \sum_{k=1}^{N-1} \lambda_k$  meant the proportion of variability which the  $k^{\text{th}}$  principal component accounted for. The left side of (6) was an integral transform  $V$  of the weight function  $\beta(s)$  with the kernel of the transform  $v$  defined by

$$V\beta(s) = \int v(s, t)\beta(t)dt. \quad (17)$$

The covariance operator was denoted by  $V$ . Therefore, (17) can be expressed as

$$V\beta(s) = \lambda\beta(s). \quad (18)$$

Equation (1) can be calculated through several methods, and we can calculate the FPC score  $f_i$  through (12).

#### 4. Numeric Analysis

In this section, the numeric analysis was performed with the goal of understanding the EV owner's charging behavior patterns. The focuses were placed on three aspects: (1) variability analysis of the daily usage patterns of all EV charging stations, (2) variability analysis of the daily energy consumption of all EV charging stations, and (3) at the station level, the usage pattern differences analyzed.

**4.1. Daily Usage Pattern Variability Analysis.** To analyze the time-dependent usage pattern variabilities, the time-dependent occupancy of each day was calculated by aggregating all charging stations, so that in each year, a total of 365 curves were obtained, with each curve representing the occupancy of a day. Function PCA was then applied to extract the FPC from the dataset. It was observed that FPC1 accounted for 94% of the variance, and FPC2 accounted for 3%. When combined together, they reflected 97% of the data's variability and were kept for further analysis.

Figure 3 showed a way to look at the two FPCs and how they supported the unique analysis that FDA enabled. X-axis represented the time (0–24 hours in a day), and Y-axis represented the percentage of charging stations that were occupied at that time. The blue curve in both subfigures (a for FPC1 and b for FPC2) stood for the mean occupancy of all charging stations, while the green and red curves stood for the functions adding and subtracting one functional principal component. For example, in Figure 3(a), the green curve was generated by adding one FPC1 to the mean function represented by the blue curve, and the red curve was generated by subtracting one FPC1 to the mean function.

The first principal component focused on daytime between 7 am and 5 pm, which corresponded to the time that public charging stations were busiest in the day, especially workdays. Therefore, the first FPC essentially distinguished between working days and nonworking days. This observation was directly supported by Figure 4, in which almost all weekdays (blue dots) were located on the right-hand side of the plot, indicating a higher FPC1 score (X-axis), while almost all weekend days (red dots) were located on the left-hand side of the plot with lower FPC1 score. A few exceptions were identified in the plot and turned out to be the holidays, such as Labor Day and Independence Day, so these were nonworking days as well.

The second FPC accounted for only 3% variability and mainly captured the variance in the evening time from midnight to 6 am and again from 6 pm to midnight. The days with higher usage after 6 pm and before 6 am and with slightly less or average usage in the daytime would receive higher scores. However, due to the dominance of FPC1, the effect of FPC2 was rather limited.

Figure 5 presented the monthly and yearly charging usage patterns. The X-axis was the score of FPC1 and the Y-axis was that of FPC2. Figure 5(a) showed the monthly pattern with the colors standing for 12 months, respectively. No clear monthly pattern was observed.

Figure 5(b) showed the yearly pattern with the colors standing for years from 2014 to 2019. Dots in 2014 were almost invisible due to the low data size and overlap with pink color. The observation led to a clear pattern that as time went by, the scores of both FPC1 and FPC2 increased significantly. That meant that for the days with a higher FPC1 score, the occupancy continued to increase at an ever-increasing speed, while for the days with higher FPC2 scores, its morning and evening usage also increased significantly. This interpretation was in line with the rapid increase of EV ownership in Kansas City at a 78% year-over-year growth rate [21] and emphasized the needs for more charging infrastructures in the region.

Figure 6(a) shows the clustering result of the data. To make sure that similar data sizes are studied, data from 2016 to 2018 are selected for clustering. Compared with Figure 6(b), the result indicates that the data points in 2015

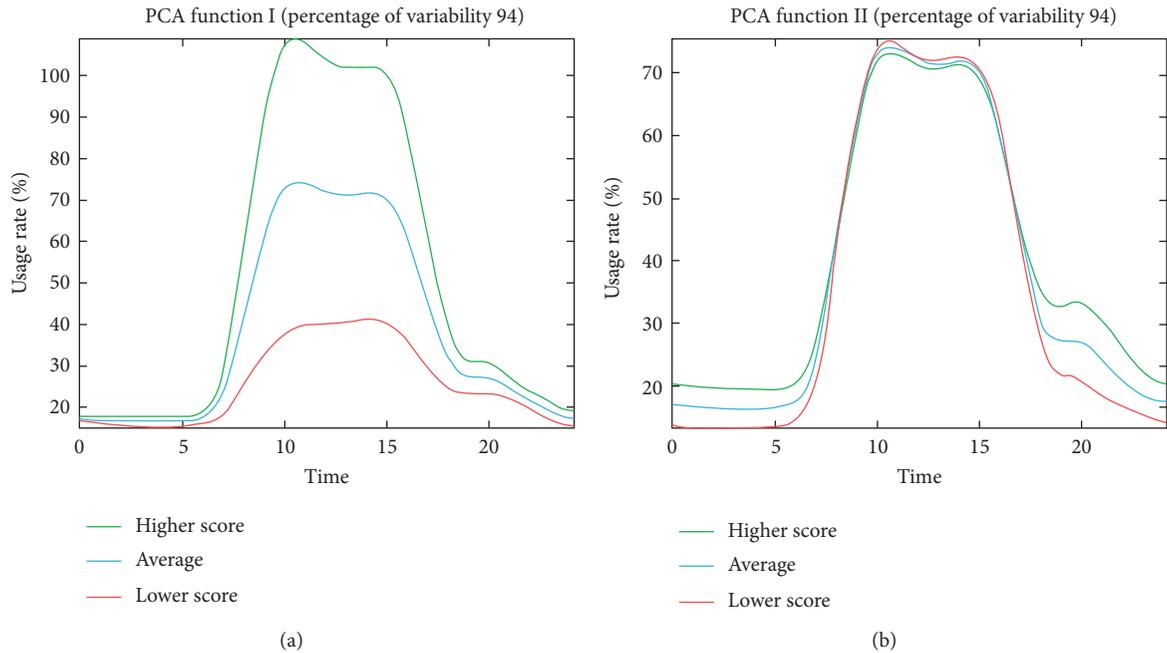


FIGURE 3: Effect of the first 2 FPCs on daily occupancy. (a) Left: the first FPC. (b) Right: the second FPC.

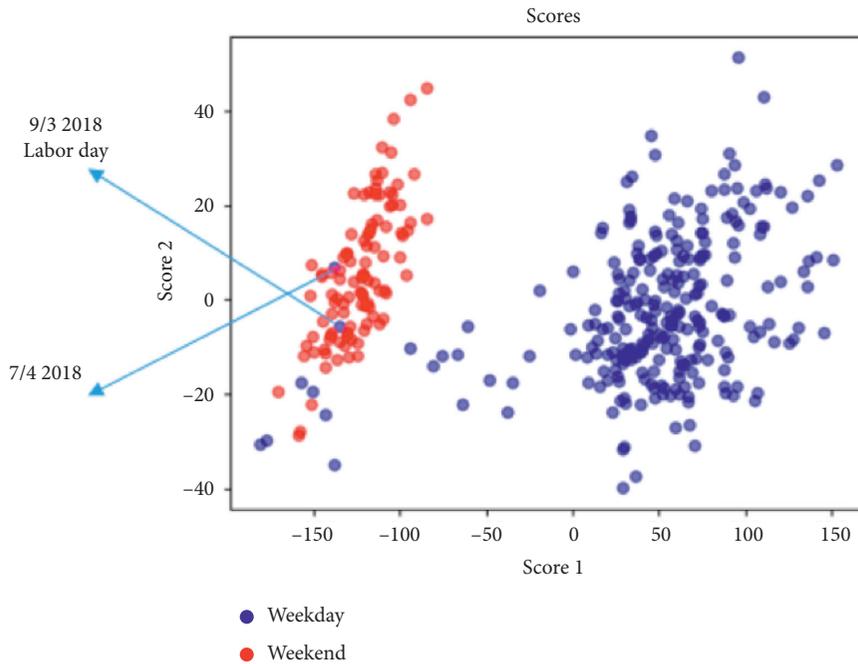


FIGURE 4: FPC scores of daily charging data.

have a lower FPC1 score and FPC 2 scores and are obviously separated from the other data points. However, the difference between 2017 and 2018 is less significant, which means that they have a similar occupancy pattern.

**4.2. Daily Energy Consumption Variability Analysis.** This section aimed to analyze the energy consumption variability caused by EV charging, which had a significant impact on the power grid and was helpful for grid load

management. Similar to the analysis above in Section 4.1, the time-dependent energy consumption of each day was calculated by aggregating all charging stations, so that in each year a total of 365 curves were obtained, with each curve representing the energy consumption of a day. Function PCA was then applied to extract the FPC from the dataset. It was observed that FPC1 accounted for 81% of the variance, and FPC2 accounted for 5%. So when combined together, they reflected 86% of the data's variability and

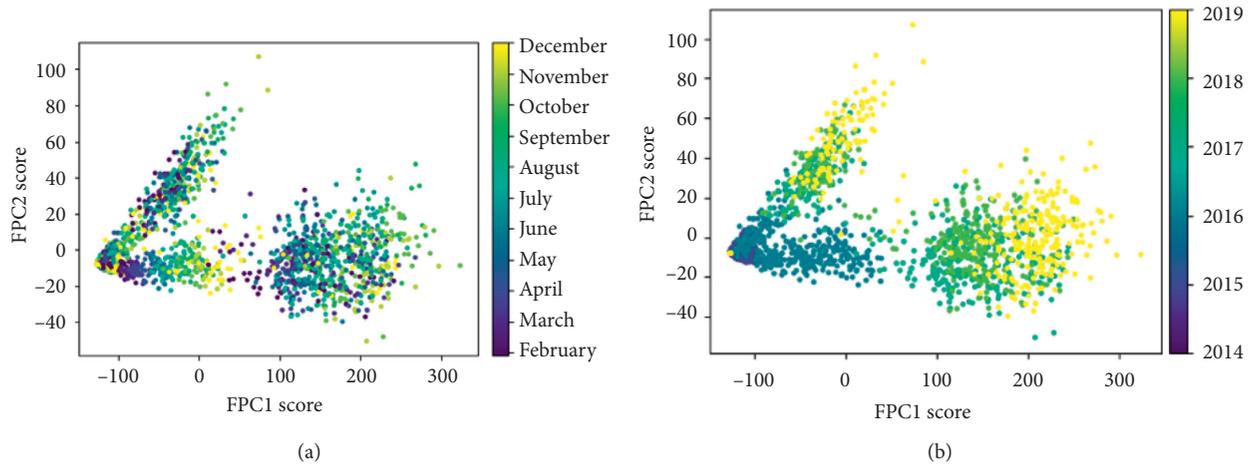


FIGURE 5: FPC scores for (a) monthly and (b) yearly pattern.

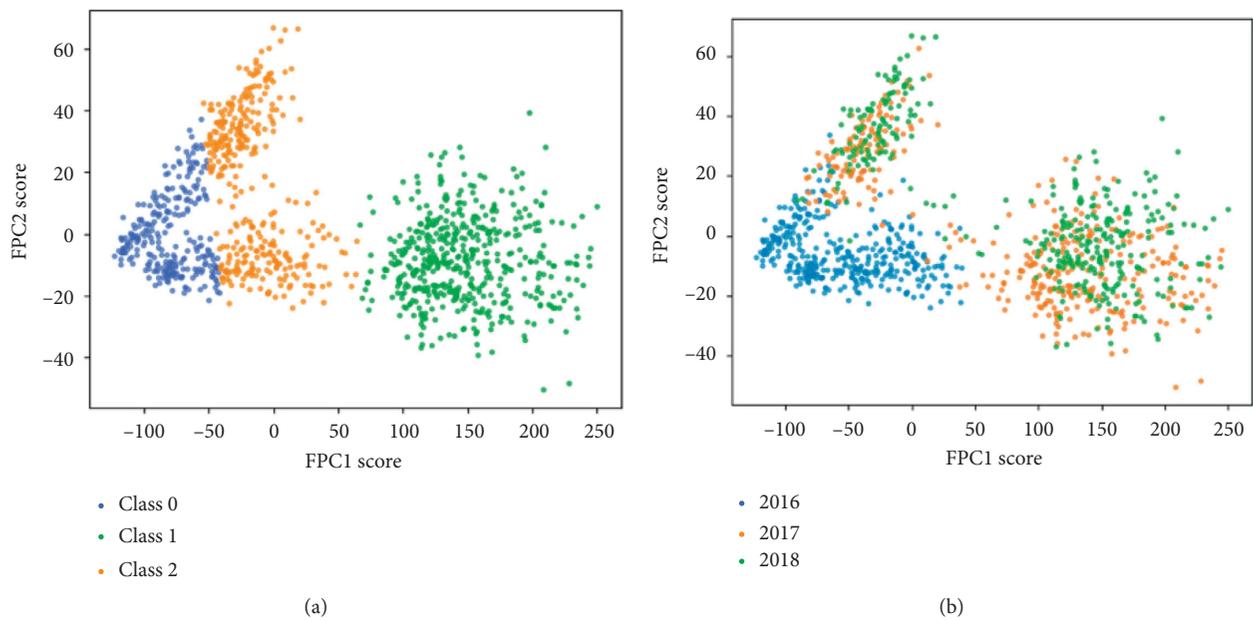


FIGURE 6: (a) FPC scores clustering (b) FPC scores during 2016–2018.

were kept for further analysis. The results were shown in Figure 6.

FPCA analysis of energy consumption resulted in some very different patterns when compared with daily occupancy. FPC1 mostly captured the variance of energy consumption during the morning peak between 7 am and 11 am. During this time range, the green curve increased dramatically, representing the days with a higher FPC1 score, and the required energy to charge EVs in the morning would be higher. On the other hand, FPC2 mostly captured the variance of energy consumption during the evening peak between 4 pm and 9 pm. In other words, if a day was observed to have a higher FPC2 score, its impacts on the power grid in the evening hours would be significantly increased.

A comparison between Figures 3 and 6 led to some interesting conclusions. While Figure 3 indicated that

from an occupancy perspective, the peak hour during the day started from as early as 7 am and did not end until 5 pm, Figure 6 suggested that the impact on the power grid became low after 11 am. This suggested that some vehicles did not leave the charging stations after they were fully charged, under which circumstances, the charging stations continued to be occupied (and thus unavailable to the other EV drivers), but from a power grid perspective, they did not require any energy. To validate such interpretation, the team went on to compare the charging event duration and the time EV actually spent on charging. The finding was as follows: while 40% of EVs left the charging stations within 1 minute after they are charged, the remaining 60% of EVs continued to park at the charging stations for various durations, and among them, two-thirds (or 40% of the entire population) would

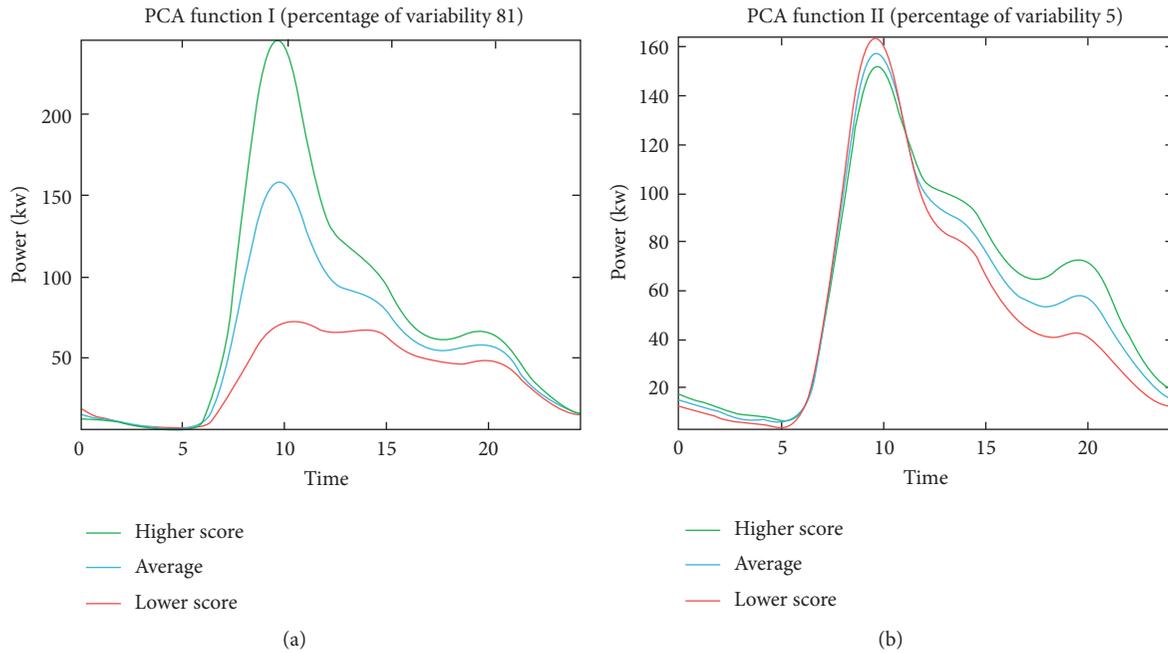


FIGURE 7: Effect of the first 2 FPCs on daily energy consumption. (a) Left: first FPC. (b) Right: second FPC.

even occupy the charging stations for at least an hour. While the discrepancy between EV owner's daily activity and the time needed for charging was understandable, the longer-than-reasonable parking behavior effectively reduced the availability of charging stations to the other EV drivers and, in our view, called for the need for parking regulation and enforcement.

**4.3. Station Occupancy Variability Analysis.** Different from the above analysis performed from the daily perspective, this analysis in this section examined the occupancy at the station level. So, each curve represented one charging station's 24-hour occupancy rates with all days aggregated, and a total of 455 curves (representing a total of 455 charging stations) were derived. Function PCA was then applied to extract the FPC from the dataset. It was observed that FPC1 accounted for 85% of the variance, and FPC2 accounted for 8%. So, when combined together, they reflected 93% of the data's variability and were kept for further analysis. The results were shown in Figure 7.

In the morning before 6 am and in the evening after 5 pm, stations with higher FPC1 scores were utilized more often than average, while in the daytime, their utilization rates were lower. On the other hand, stations with higher FPC2 values were utilized more often than average in the first half of a day (before noon) but were used less often in the second half of a day (afternoon).

An intuitive guess was these patterns might be attributed to the differences in the land-use patterns. As such, all 455 charging stations were mapped to five categories of land-use types: (1) recreational, which was

meant to be used for the enjoyment of the people who used it, such as arts center and theater; (2) commercial, which was designated for businesses, warehouses, shops, and any other infrastructures related to commerce, such as plazas, hotels, and hospitals; (3) transport, which was built for the structures that help people get from one destination to the other, such as airport; (4) industrial such as the plant and industrial parks; and (5) residential, such as apartments and condominiums. Then, the scores of FPC1 and FPC2 were plotted in Figure 8, in which Figure 8(a) had all land-use types together, while Figures 8(b)~8(f) stood for commercial, residential, industrial, transportation, and recreational.

No clear patterns can be found in Figure 9(a), in which charging stations of all land-use types were plotted together. However, when they were separated, some conclusions can be drawn. (1) For charging stations that were built on residential (Figure 9(c)), transport (Figure 9(e), and recreational (Figure 9(f)) areas, the majority of the dots in those subfigures had positive FPC1 scores. In other words, charging stations in these three categories shared a common pattern that they were used more often in the evening than in the daytime. Considering the nature of activities happening at these locations, this interpretation was consistent with what was observed in the real life. (2) FPC1 values for commercial (Figure 9(b)) and industrial (Figure 9(d)) areas were mixed and thus inconclusive to identify clear patterns. (3) Charging stations in the recreational area, in general, had negative FPC2 scores, meaning that they were used more in the second half of the day than before noon. Again, this seemed to be in line with our understanding of human behavior patterns.

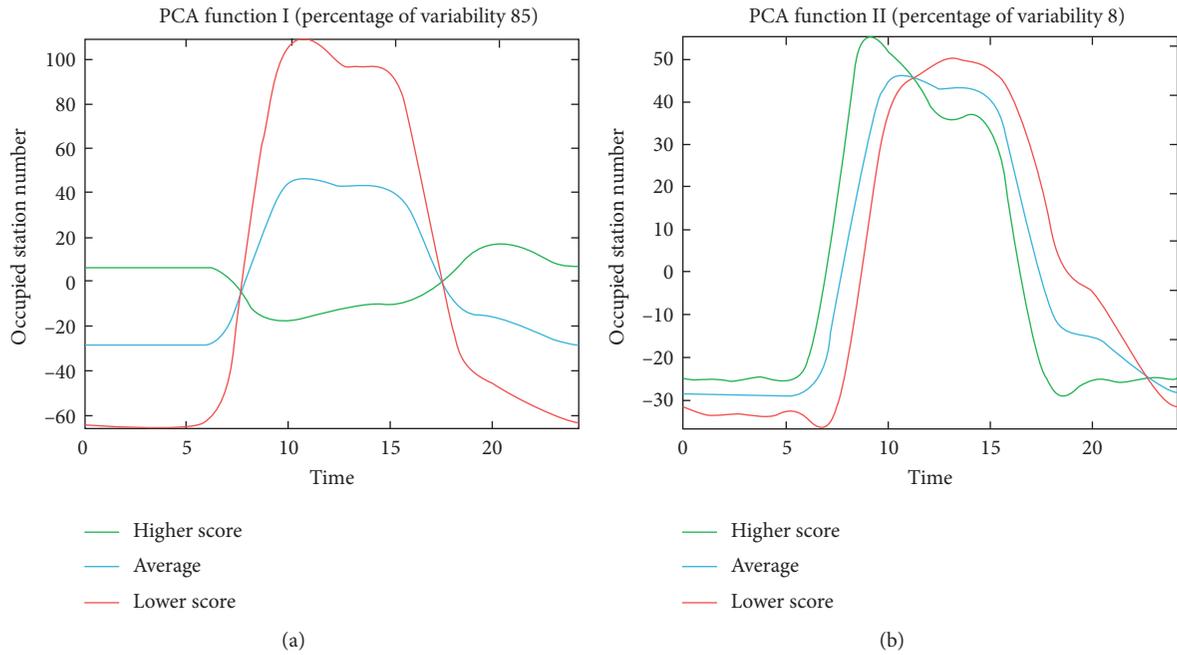


FIGURE 8: Effect of the first 2 FPCs on station occupancy. (a) Left: the first FPC. (b) Right: the second FPC.

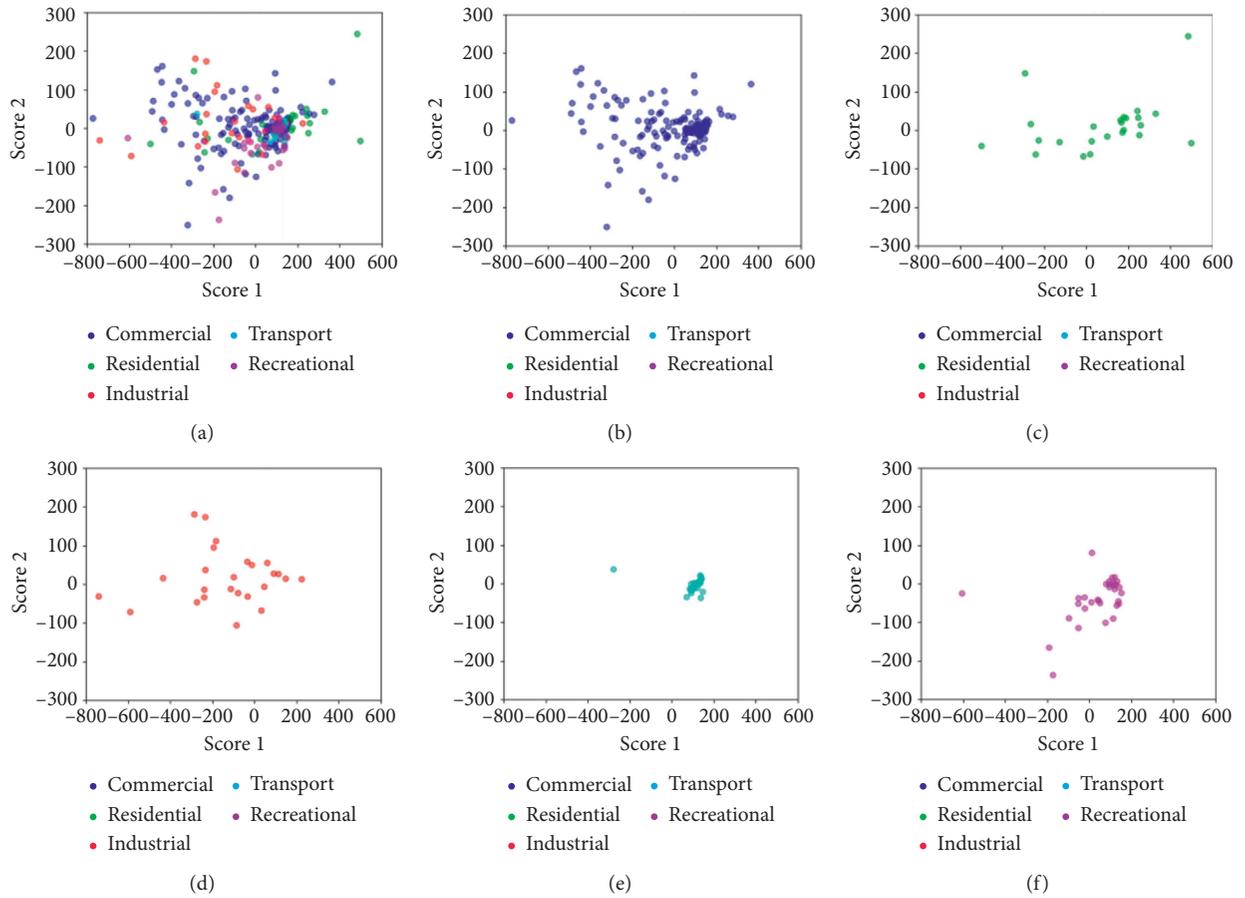


FIGURE 9: FPC scores of stations with different POI. (a) Scores. (b) Commercial. (c) Residential. (d) Industrial. (e) Transport. (f) Recreational.

## 5. Conclusions

In this manuscript, the focus was placed on analyzing the electric vehicle's usage behavior pattern with a functional data analysis approach, specifically, based on functional principal component analysis. Compared with the traditional discrete-based analysis models, the proposed FDA modeling approach had unique advantages in preserving the smooth function behavior of the data, bringing more flexibility in the modeling process with little required assumptions or background knowledge on independent variables, as well as the capability of handling time series data with different lengths or sizes. 5-year real-world charging event log data from a total of 455 charging stations in Kansas City, Missouri (KCMO), was used. The daily usage variability, daily energy consumption variability, and station-level usage variability were analyzed, with the goal of providing theoretical support to the EV infrastructure planning and regulation, as well as the power grid load management. In addition to the patterns revealed in the EV charging station's occupancy and energy consumption, the differences between EV driver's charging time and parking time were analyzed and called for the needs for parking regulation and enforcement. The different usage patterns associated with charging stations of different land-use types were also analyzed.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Disclosure

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This material is based on the work supported by the U.S. Department of Energy, Office of Energy Efficiency and

Renewable Energy's (EERE) Vehicle Technologies Office under the Award Number DE-EE008474. The authors are also thankful for the support from the Metropolitan Energy Center (MEC), the City of Kansas City Missouri (KCMO), Lilypad, Mid-America Regional Council (MARC), and Evergy (formerly Kansas City Power and Light Company (KCP & L)).

## References

- [1] E. W. Wood, C. L. Rames, and M. Muratori, *New EVSE Analytical Tools/Models: Electric Vehicle Infrastructure Projection Tool (EVI-Pro)*, National Renewable Energy Lab.(NREL), Golden, CO, USA, 2018.
- [2] F. He, Y. Yin, and J. Zhou, "Deploying public charging stations for electric vehicles on urban road networks," *Transportation Research Part C: Emerging Technologies*, vol. 60, pp. 227–240, 2015.
- [3] J. He, H. Yang, T.-Q. Tang, and H.-J. Huang, "An optimal charging station location model with the consideration of electric vehicle's driving range," *Transportation Research Part C: Emerging Technologies*, vol. 86, pp. 641–654, 2018.
- [4] O. Arslan and O. E. Karaşan, "A Benders decomposition approach for the charging station location problem with plug-in hybrid electric vehicles," *Transportation Research Part B: Methodological*, vol. 93, pp. 670–695, 2016.
- [5] Z.-H. Zhu, Z.-Y. Gao, J.-F. Zheng, and H.-M. Du, "Charging station location problem of plug-in electric vehicles," *Journal of Transport Geography*, vol. 52, pp. 11–22, 2016.
- [6] Q. Wu, A. H. Nielsen, J. Ostergaard et al., "Driving pattern analysis for electric vehicle (ev) grid integration study," in *Proceedings of the 2010 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT Europe)*, Gothenburg, Europe, October 2010.
- [7] P. Morrissey, P. Weldon, and M. O'Mahony, "Future standard and fast charging infrastructure planning: an analysis of electric vehicle charging behaviour," *Energy Policy*, vol. 89, pp. 257–270, 2016.
- [8] X.-H. Sun, T. Yamamoto, and T. Morikawa, "Charge timing choice behavior of battery electric vehicle users," *Transportation Research Part D: Transport and Environment*, vol. 37, pp. 97–107, 2015.
- [9] T. D. Chen, K. M. Kockelman, and M. Khan, "Locating electric vehicle charging stations: parking-based assignment method for seattle," *Transportation Research Record*, vol. 2385, no. 1, pp. 28–36, Washington, USA, 2013.
- [10] F. Jabeen, D. Olaru, B. Smith, T. Braunl, and S. Speidel, "Electric vehicle battery charging behaviour: findings from a driver survey," in *Proceedings of the Australasian Transport Research Forum*, Brisbane, Australia, October 2013.
- [11] X. Hu, Y. Yuan, X. Zhu, H. Yang, and K. Xie, "Behavioral responses to pre-planned road capacity reduction based on smartphone GPS trajectory data: a functional data analysis approach," *Journal of Intelligent Transportation Systems*, vol. 23, no. 2, pp. 133–143, 2019.
- [12] Z. Li, R. Kluger, X. Hu, Y.-J. Wu, and X. Zhu, "Reconstructing vehicle trajectories to support travel time estimation," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2672, no. 42, pp. 148–158, 2018.
- [13] X. Zhu, X. Hu, and Y.-C. Chiu, "Design of driving behavior pattern measurements using smartphone global positioning system data," *International Journal of Transportation Science and Technology*, vol. 2, no. 4, pp. 269–288, 2013.

- [14] P. J. Green and B. W. Silverman, *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*, CRC Press, Boca Raton, FL, USA, 1993.
- [15] S. Ullah and C. F. Finch, "Applications of functional data analysis: a systematic review," *BMC Medical Research Methodology*, vol. 13, no. 1, p. 43, 2013.
- [16] J. Ramsay and B. W. Silverman, *Functional Data Analysis*, Springer, Berlin, Germany, 2005.
- [17] X. Hu, X. Zhu, Y.-C. Chiu, and Q. Tang, "Will information and incentive affect traveler's day-to-day departure time decisions? - an empirical study of decision making evolution process," *International Journal of Sustainable Transportation*, vol. 14, no. 6, pp. 403–412, 2019.
- [18] X. Hu, X. Zhu, Y. Yuan, A. Rehan, Y.-C. Chiu, and M. Zmud, "A multi-resolution approach to investigate the impacts of pre-planned road capacity reduction based on smartphone gps trajectory data - a case study of mopac expressway in Austin, Texas," in *Proceedings of the Transportation Research Board 96th Annual Meeting*, Washington DC, USA, January 2017.
- [19] I. G. Guardiola, T. Leon, and F. Mallor, "A functional approach to monitor and recognize patterns of daily traffic profiles," *Transportation Research Part B: Methodological*, vol. 65, pp. 119–136, 2014.
- [20] I. M. Wagner-Muns, I. G. Guardiola, V. Samaranyake, and W. I. Kayani, "A functional data analysis approach to traffic volume forecasting," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 3, pp. 878–888, 2017.
- [21] Clean Charge Network, "Kansas city zooms ahead in electric vehicle growth," 2017, <https://cleanchargenetwork.com/kansas-city-zooms-ahead-in-electric-vehicle-growth/>.