*Article*

# Modeling the Charging Behaviors for Electric Vehicles Based on Ternary Symmetric Kernel Density Estimation

**Lixing Chen** [1,*]**, Xueliang Huang** [2,*] **and Hong Zhang** [1]

[1] School of Electrical &Information Engineering, Jiangsu University of Technology, Changzhou 213001, China; hazh0216@jsut.edu.cn

[2] School of Electrical Engineering, Southeast University, Nanjing 210096, China

[*] Correspondence: chenlixing@jsut.edu.cn (L.C.); xlhuang@seu.edu.cn (X.H.); Tel.: +136-1611-8650 (L.C.)

check for updates

**Abstract:** The accurate modeling of the charging behaviors for electric vehicles (EVs) is the basis for the charging load modeling, the charging impact on the power grid, orderly charging strategy, and planning of charging facilities. Therefore, an accurate joint modeling approach of the arrival time, the staying time, and the charging capacity for the EVs charging behaviors in the work area based on ternary symmetric kernel density estimation (KDE) is proposed in accordance with the actual data. First and foremost, a data transformation model is established by considering the boundary bias of the symmetric KDE in order to carry out normal transformation on distribution to be estimated from all kinds of dimensions to the utmost extent. Then, a ternary symmetric KDE model and an optimum bandwidth model are established to estimate the transformed data. Moreover, an estimation evaluation model is also built to transform simulated data that are generated on a certain scale with the Monte Carlo method by means of inverse transformation, so that the fitting level of the ternary symmetric KDE model can be estimated. According to simulation results, a higher fitting level can be achieved by the ternary symmetric KDE method proposed in this paper, in comparison to the joint estimation method based on the edge KDE and the ternary t-Copula function. Moreover, data transformation can effectively eliminate the boundary effect of symmetric KDE.

**Keywords:** electric vehicles; charging behaviors; kernel density estimation; optimum bandwidth

## 1. Introduction

Nowadays, a considerable part of the total emissions is due to contribution of road traffic [1,2]. For this reason, the development of the new technologies (as electric vehicles (EVs) and bicycles for sustainable mobility) is necessary to reduce the emissions. Nevertheless, a huge amount of disorderly EVs charging behaviors could pose a wide range of challenges to the safe, economic, and efficient operation of the charging network (CN) with the large-scale development of EVs. As for a power grid (PG), these challenges originate from the uncertain distribution of the EVs charging load [3–5]. In terms of EVs, these challenges stem from the randomness of EVs driving and charging behaviors [6]. As a result, problems such as uneven EVs distribution in the CN, congestion in some charging stations (CSs), and long waiting time for charging, and waste of charging resources are caused [7]. The above problems raised in the operation of CN will become a "bottleneck" to the EVs development, which should be solved without delay.

Fortunately, the distribution of EVs charging load is also characterized by flexible load with a potential for regulation although it is uncertain. The potential can be found in two aspects. On the one hand, the CN (as shown in Figure 1) integrates the attributes of a PG, a transportation network

(TN), and an information network (IN), which allows EV users to select a CS base on bidirectional information interaction with the help of the navigation system and the CS administration center, so as to achieve the regulation of EVs charging load between CSs [8,9]. On the other hand, the advancement of battery technology and charging facility technology, especially the emergence of multi-gun charging piles can provide the necessary conditions for regulating EVs charging load in CSs [10]. Therefore, the study of regulating EVs charging load distribution appears to be particularly important and will create opportunities for the safe, economic, and efficient operation of the CN.
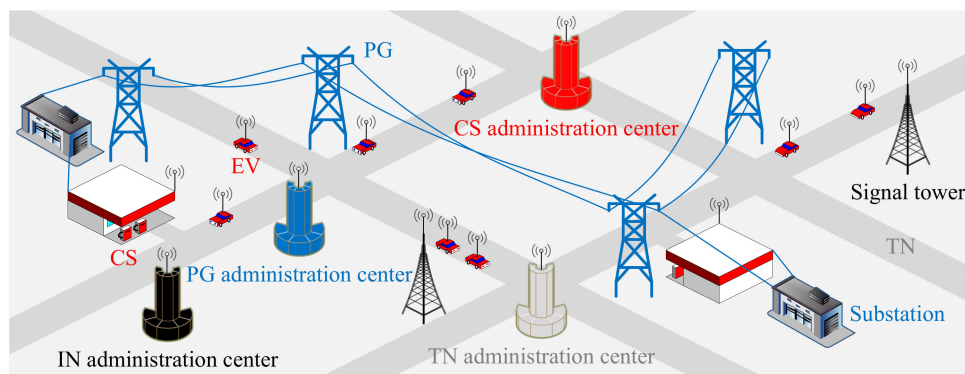


**Figure 1.** Diagram of the charging network (CN).

Modeling of EVs charging behaviors, charging load modeling, the charging impact on the PG, orderly charging strategy, and planning of charging facilities have been focuses of domestic and foreign studies concerning the regulation of EVs charging load distribution, and EVs charging behaviors should be modelled accurately for the latter four research contents. It can be observed that the modeling study of EVs charging behaviors is the key to the regulation of EVs charging load distribution. However, with low popularization of EVs at the beginning, it was difficult to obtain the actual driving data. Considering that drivers' driving times and distances are constant, it is common to study the modeling of the EVs charging behaviors with fuel vehicle data. From the perspective of modeling data, these research data are derived from the national household travel survey (NHTS) data of America [11], daily traffic flow data of Jiangsu Province of China [10], and daily traffic flow data of South Korea [12]. Characteristic parameters of fuel vehicles should be converted into characteristic parameters of EVs charging via certain assumptions when the above data are applied in the actual study. For instance, it must be assumed that the daily driving mileage is positively correlated with the charging time [13], that the end time of the last trip is equivalent to the initial charging time [14], that the initial charging time is independent of the charging time [15], and that the battery of each EV is fully charged [16]. These assumptions might lead to large deviations between Monte Carlo simulation results and actual conditions.

As EVs have witnessed rapid growth in recent years, a higher accuracy can be gained from the modeling of charging behaviors with the actual EVs data. In that case, studies on the modeling of charging behaviors with the actual EVs operating data have been rising gradually. In terms of modeling data, these research data are derived from charging data [3,4] of electric taxis in Shenzhen, EVs charging data [17] in Kanagawa, EV parking and charging data [18,19] of CSs in Netherlands, EV charging data [20] of CSs in UK, and EVs charging data [21] of CSs in Nanjing, etc. In terms of modeling objects, these objects primarily consist of the arrival time (or initial charging time), the staying time (that contains waiting time and charging time), and the charging capacity. Additionally, the charging capacity modeling can be divided into three parts: the starting state of charge (SOC) modeling, the ending SOC modeling, and the battery pack capacity modeling. In terms of modeling approaches, these approaches are divided into two main categories: parametric estimation (PE) and non-parametric estimation (NPE). When EVs charging behaviors are modelled with PE, the variable to be estimated conforming to a certain known probability distribution should be presupposed before estimating parameters of the assumed

probability distribution with EVs charging behaviors data for testing. Common PE methods are composed of Gaussian distribution [6], mixed Gaussian distribution [4], mixed Beta distribution [19], and Weibull distribution [22] etc. The premise of using the PE approach is to presuppose that the probability distribution is approaching the actual situation. Otherwise, the established EVs charging behaviors model might be significantly deviated from the actual situation. When EVs charging behaviors are modelled with NPE, presupposition is not required. Instead, the probability distribution of the variable to be estimated is gained directly under the drive of EVs operating data. In this way, deviations caused by wrong selection of PE models can be effectively addressed.

At present, there are two approaches to EVs charging behaviors modeling by NPE. One uses the unitary kernel density estimation (KDE) or the unitary self-adaptive KDE [23,24] for modeling with taking no account of the correlation between variables, the other uses the estimation associating the edge distribution function (common empirical distribution function or unitary KDE [25]) with the Copula function [21] for modeling with considering the correlation between variables. In terms of modeling accuracy, the second approach has higher modeling accuracy than the first one. Nevertheless, in the second approach, errors can be found in the estimation of each edge distribution and the overall relationship of describing ternary variables with the correlation. Consequently, the deviation of the joint density estimation generated by the above errors is large, causing a large deviation between the simulation results and the actual results. With this as the cutting point, this paper introduces an accurate joint modeling approach of the arrival time, the staying time, and the charging capacity for the EVs charging behaviors in the work area based on ternary symmetric KDE in accordance with the actual data. The rest of the paper is organized as follows: a ternary symmetric KDE model of EVs charging behaviors is built in Section 2. In Section 3, we solve the model with using KDE software package and toolbox of MATLAB. In Section 4, the ternary symmetric KDE method, and the joint estimation method based on the edge KDE and the ternary Copula function are investigated, and the results are analyzed from the simulations. Finally, Section 5 concludes this paper.

## 2. Ternary Symmetric KDE Model of EVs Charging Behaviors

### 2.1. Data Transformation Model

There is a boundary bias in the symmetric KDE, namely, a large deviation will arise in the boundary estimation [26]. In order to eliminate the boundary bias, Box-Cox transformation was selected in this paper to transform EVs data in the work area from all dimensions. Its normal transformation and its corresponding inverse transformation are expressed as Equations (1) and (2), respectively [27].

$$\Omega'_i = \begin{cases} \frac{\Omega_i^{\lambda}-1}{\lambda} & \lambda \neq 0 \\ \log(\Omega_i) & \lambda = 0 \end{cases}, \quad i = 1,2,\cdots,n \ , \tag{1}$$

$$\Omega_i = \begin{cases} (\lambda\Omega'_i+1)^{\frac{1}{\lambda}} & \lambda \neq 0 \\ \exp(\Omega'_i) & \lambda = 0 \end{cases}, \quad i = 1,2,\cdots,n \ , \tag{2}$$

where $\log(\cdot)$ is the logarithmic function; $\exp(\cdot)$ is the exponential function; $n$ is the number of EVs samples; $\lambda$ is a parameter to be estimated. In this paper, the estimated parameter $\lambda$ can be solved with the maximum likelihood method [27] and the corresponding logarithmic likelihood function is represented by Equation (3).

$$L(\lambda) = -\frac{n}{2}\log\left[\text{var}\left(\frac{\Omega^{\lambda}-1}{\lambda}\right)\right] + n(\lambda-1)\text{mean}[\log(\Omega)], \tag{3}$$

where $\text{var}(\cdot)$ is the variance function; $\text{mean}(\cdot)$ is the mean value function; $\Omega$ is the transformed data of EVs.

The parameter $\lambda$ value is estimated as

$$\hat{\lambda} = \arg \max_{\lambda \in R}[L(\lambda)], \tag{4}$$

where $\max(\cdot)$ is the maximum function.

### 2.2. Ternary Symmetric KDE Model

The modeling object in this paper is ternary variables that involve the overall relationship among the EVs arrival time, the staying time, and the charging capacity. In order to effectively describe its overall characteristics, a ternary symmetric KDE model is established as follows.

Suppose that $(X_{11}, X_{12}, X_{13}), (X_{21}, X_{22}, X_{23}), \cdots, (X_{n1}, X_{n2}, X_{n3})$ are $n$ three-dimensional (3D) samples of the EVs charging behaviors, where $(X_{n1}, X_{n2}, X_{n3})$ is the nth 3D sample, and the subscript n of $X_{nd}(d = 1, 2, 3)$ refers to the number of the samples; $K(\cdot)$ is a given Borel measurable function on $R^3$; $h_{n1} > 0$, $h_{n2} > 0$, $h_{n3} > 0$ are constants related to $n$, and when $n$ is infinite, $h_{n1}$, $h_{n2}$, $h_{n3}$ tend to 0.

$$\hat{f}_K(x, y, z) = \frac{1}{n h_{n1} h_{n2} h_{n3}} \sum_{i=1}^{n} K\left(\frac{X_{i1} - x}{h_{n1}}, \frac{X_{i2} - y}{h_{n2}}, \frac{X_{i3} - z}{h_{n3}}\right), \tag{5}$$

Equation (5) is called the ternary KDE of $f(x, y, z)$ [23,24]; $h_{n1}$, $h_{n2}$, $h_{n3}$ are bandwidths; $(x, y, z)$ are valued by dividing the cube space that is composed of the maximum values $(X_{max1}, X_{max2}, X_{max3})$ and the minimum values $(X_{min1}, X_{min2}, X_{min3})$ of three dimensions in the given sample into $m^3$ segments. Sequentially, $(x, y, z)$ are taken as $x_i = X_{min1} + (X_{max1} - X_{min1}) \times i/m, y_j = X_{min2} + (X_{max2} - X_{min2}) \times j/m, z_k = X_{min3} + (X_{max3} - X_{min3}) \times k/m$; $K(\cdot)$ is a kernel function, which can be expressed by the product kernel function [28].

$$K(\cdot) = \prod_{d=1}^{3} K_d(\cdot), \tag{6}$$

where $K_d(\cdot)$ is unitary kernel function, which exerts no influence on the overall relationship of variables. In order to ensure rational estimation, a symmetric unimodal probability density function centered on 0 is usually selected. Additionally, the same pattern is generally adopted [28].

Uniform kernel, Gaussian kernel, Epanechnikov kernel, and quartic kernel are common kernel functions [23]. In this paper, the Gaussian kernel is selected as the kernel function of the ternary symmetric KDE model, as expressed in Equation (7).

$$K_d(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right), d = 1, 2, 3, \tag{7}$$

The ternary symmetric KDE model for EVs arrival time, the staying time, and the charging capacity is established as Equation (8).

$$\hat{f}_K(x, y, z) = \frac{1}{n(\sqrt{2\pi})^3 \prod\limits_{d=1}^{3} h_{nd}} \sum_{i=1}^{n} \exp\left\{-\left(\frac{X_{i1} - x}{\sqrt{2}h_{n1}}\right)^2 - \left(\frac{X_{i2} - y}{\sqrt{2}h_{n2}}\right)^2 - \left(\frac{X_{i3} - z}{\sqrt{2}h_{n3}}\right)^2\right\}, \tag{8}$$

### 2.3. Optimum Bandwidth Model

Selecting an appropriate bandwidth is extremely important for the actual use of ternary symmetric KDE. When a small bandwidth is selected, the randomness effect will be prominent, resulting in irregular shapes of estimation results. When large bandwidth is selected, the averaging effect will be prominent, possibly causing the structural characteristics of data [29] to be masked and the over-slick estimation result. Therefore, the least square cross-validation was adopted in this paper, so that

bandwidths can be directly and automatically generated by data [23] to ensure the appropriateness of the bandwidth. The integral squared error (ISE) between the ternary symmetric KDE and the true density is defined as

$$\text{ISE}\big(\hat{f}_K\big) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \big(\hat{f}_K(x,y,z) - f(x,y,z)\big)^2 dxdydz, \tag{9}$$

where the ISE reaching the minimum bandwidth is called the optimal bandwidth. To make it simple, solving the optimal bandwidth of Equation (9) can be equivalent to solving the optimal bandwidth of Equation (10).

$$R\big(\hat{f}_K\big) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{f}_K^2(x,y,z) dxdydz - 2\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{f}_K(x,y,z) f(x,y,z) dxdydz, \tag{10}$$

In Equation (10), $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{f}_K(x,y,z) f(x,y,z) dxdydz = E\big[\hat{f}_K(x,y,z)\big]$, so $n^{-1} \sum_{i=1}^{n} \hat{f}_K^{(-i)}(X_{i1}, X_{i2}, X_{i3})$ is an unbiased estimation, in which

$$\hat{f}_K^{(-i)}\big(X_{i1}, X_{i2}, X_{i3}\big) = \frac{1}{(n-1)\prod\limits_{d=1}^{3} h_{nd}} \sum_{j=1,i\neq j}^{n} K\bigg(\frac{X_{j1}-X_{i1}}{h_{n1}}, \frac{X_{j2}-X_{i2}}{h_{n2}}, \frac{X_{j3}-X_{i3}}{h_{n3}}\bigg), \tag{11}$$

By calculation

$$n^{-1}\sum_{i=1}^{n} \hat{f}_K^{(-i)}(X_{i1}, X_{i2}, X_{i3}) = \frac{1}{n(n-1)\prod\limits_{d=1}^{3} h_{nd}}\left\{\sum_{i=1}^{n}\sum_{j=1}^{n} K\bigg(\frac{X_{j1}-X_{i1}}{h_{n1}}, \frac{X_{j2}-X_{i2}}{h_{n2}}, \frac{X_{j3}-X_{i3}}{h_{n2}}\bigg) - nK(0,0,0)\right\}, \tag{12}$$

Taking Equations (6), (11) and (12) into account, the second term in Equation (10) is

$$\frac{-2}{n(n-1)\prod\limits_{d=1}^{3} h_{nd}}\left\{\sum_{i=1}^{n}\sum_{j=1}^{n} \exp\bigg\{\sum_{d=1}^{3} \log\bigg[K\bigg(\frac{X_{jd}-X_{id}}{h_{nd}}\bigg)\bigg]\bigg\} - nK^3(0)\right\}, \tag{13}$$

Additionally, the first term in Equation (10) can be calculated as

$$\frac{1}{n^2 \prod\limits_{d=1}^{3} h_{nd}}\sum_{i=1}^{n}\sum_{j=1}^{n} \exp\bigg\{\sum_{d=1}^{3} \log\bigg[K^*\bigg(\frac{X_{jd}-X_{id}}{h_{nd}}\bigg)\bigg]\bigg\}, \tag{14}$$

When the Gaussian kernel is selected as the kernel function, the $K^*(\cdot)$ can be expressed as

$$K^*(u) = \frac{1}{\sqrt{4\pi}} \exp\bigg(-\frac{u^2}{4}\bigg), \tag{15}$$

When $n$ is large, $n-1$ in Equation (13) can be replaced with $n$. In this way, a final scaling function can be obtained by substituting Equations (13) and (14) into Equation (10).

$$M(h_{n1}, h_{n2}, h_{n3}) = \frac{1}{n^2 \prod\limits_{d=1}^{3} h_{nd}}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\{\exp(\sum_{d=1}^{3} \log(K^*(\frac{X_{jd}-X_{id}}{h_{nd}}))) - 2\exp(\sum_{d=1}^{3} \log(K(\frac{X_{jd}-X_{id}}{h_{nd}}))) + 2nK^3(0)\right\}, \tag{16}$$

The optimum bandwidth model obtained through verification using the least squares method can be established as

$$\left(h_{\text{opt1}}, h_{\text{opt2}}, h_{\text{opt3}}\right) = \min_{h_{n1}>0, h_{n2}>0, h_{n3}>0} M(h_{n1}, h_{n2}, h_{n3}), \tag{17}$$

### 2.4. Estimation Evaluation Model

A ternary symmetric KDE model is constructed on the basis of the actual EVs operating data in the work area. Moreover, the Monte Carlo method [30] is utilized to generate the simulation data of the required scale. Then, simulation data and original data are presented in a 3D right-angle plane using a 3D scatter density map. Three binary frequency histograms are selected to describe the overall characteristics of the simulated data and the original data, so as to quantitatively evaluate the fitting level of the estimated model. Meanwhile, corresponding frequency matrices can be also obtained. In order to evaluate the difference of the frequency matrix, the similarity between the matrices A and B is defined as

$$r = \frac{\sum_{i=1}^{n} \sum_{j=1}^{m} \left(A_{ij} - \overline{A}\right)\left(B_{ij} - \overline{B}\right)}{\sqrt{\left(\sum_{i=1}^{n} \sum_{j=1}^{m} \left(A_{ij} - \overline{A}\right)^2\right)\left(\sum_{i=1}^{n} \sum_{j=1}^{m} \left(B_{ij} - \overline{B}\right)^2\right)}}, \tag{18}$$

where $\overline{A} = \text{mean2}(A)$ and $\overline{B} = \text{mean2}(B)$; $\text{mean2}(\cdot)$ is the mean function of all elements in a matrix.

After obtaining the similarity of frequency matrix, the estimation model is defined to fit the evaluation indexes concerning the advantages and disadvantages of original data. It is composed of Equations (19) and (20).

$$\mu_r = \frac{1}{n} \sum_{i}^{n} r_i, \tag{19}$$

$$\sigma_r = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (r_i - \mu_r)^2}, \tag{20}$$

where $\mu_r$ and $\sigma_r$ are the mean and the standard deviation of the matrix similarities corresponding to the generated data and the original data after $n$ times of simulations. The greater $\mu_r$ indicates the higher the fitting level of original data of the model; while the smaller $\sigma_r$ indicates the higher the stability of the model fitting the original data.

## 3. Algorithm Flowchart

The data transformation model, the ternary symmetric KDE model, the optimal bandwidth model, and the estimation evaluation model are established in this paper, of which, the data transformation model and the optimal bandwidth model are unconstrained optimization models. Functions in Toolbox are used in MATLAB to solve the unconstrained optimization model and the estimation evaluation model. In addition, the KDE software package [31] is also used to model the ternary symmetric KDE. The specific process is presented as follows:

Step 1:　Extract effective data records of the EVs arrival time, stay time, and charging capacity in the work area to conduct normal transformations on it with Equation (1) sequentially and estimate the transformed parameters with the Equations (3) and (4) sequentially before solving the parameter estimation model with the genetic algorithm (GA). Moreover, the data transformation model is solved by using the GA function in Toolbox in MATLAB to obtain the estimated value of the transformation parameter;

Step 2:　Substitute the original data and the transformed data into Equations (16) and (17), respectively, and solve the optimal bandwidth model through using the Fmincon function in Toolbox in MATLAB to obtain corresponding optimal bandwidths;

Step 3: After reading the original and transformed data and the corresponding optimal bandwidth, model the ternary symmetric KDE on the original data and the transformed data distributions through calling the KDE software package in MATLAB. At the same time, estimation modeling is conducted on the original data and transformed data, respectively, by means of using the Ksdensity function and the Copulafit function in Toolbox in MATLAB;

Step 4: Implement the Monte Carlo random sampling through calling the Sample function in the KDE software package in MATLAB by setting the number of simulations and the sampling size, and reading the results of the ternary symmetric KDE. Meanwhile, the Monte Carlo random sampling can be implemented through calling the Copularnd function in Toolbox in MATLAB. It should be noted that the random sampling of the transformed data estimation should be inversely transformed in Equation (2);

Step 5: Read the frequency matrix of the binary frequency histogram of the original data and the simulated data, and solve the frequency matrix similarity between the original data and the simulated data through calling the Corr2 function in Toolbox in MATLAB. Besides, results of the ternary symmetric KDE are compared with that of the joint estimation based on the edge KDE and the Copula function to verify the effectiveness of the ternary symmetric KDE model.

## 4. Numerical Simulation

### 4.1. EVs Data

More than 17,000 data records from September 1, 2016 to December 31, 2017 were extracted and used in this paper from a EVs charging service company platform in Nanjing, each of which contains the tab and the name of the CS, the arrival time, the departure time, the charging capacity, the charging amount, the order type, the payment method, the order number and the order status, etc. This data is a crucial support for the analysis of EVs charging behaviors. Four sets of data records, including the name of the CS (it is used for dividing the charging area), the arrival time, the staying time (the difference between the departure time and the arrival time), and the charging capacity were extracted for the study. Apart from the abnormal data record that should be eliminated as required, there was more than 14,000 effective data records.

The randomness is large since the data volume per day per station is small among the effective data. In order to investigate the charging behaviors law of EVs, effective data records are aggregated from space and time. First of all, every CS is classified into the work area, the residential area, the commercial area, and other areas through searching the name of the CS from the map at the spatial dimension, so as to spatially aggregate the effective data records. Next, the EVs arrival times of multiple days are aggregated into one day at the time dimension, so as to carry out time aggregation on the effective data records.

After time-space aggregation, data records in the work area, the residential area, the commercial area, and other areas are 11,740, 681, 1042, and 792, respectively. As can be seen from data, the work area has the largest number of data records (up to 82%), while few records can be found in the residential area, the commercial area, and other areas (accounting for less than 7.3%). In that case, in order to guarantee the rational modeling, the combined distribution of the EVs arrival time (AT), the staying time (ST), and the charging capacity (CC) were analyzed with the EVs operating data in the work area as the original data (as shown in Figure 2) for concluding the charging behaviors law of EV users, which can provide the basis for studying EVs charging load modeling, the impact of charging on the PG, orderly charging strategies, and planning of charging facilities.
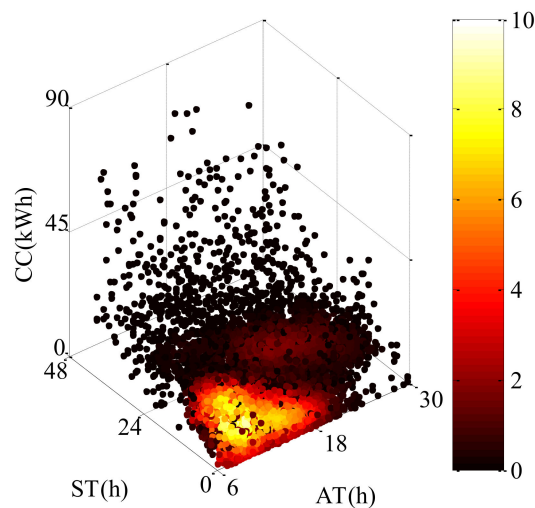
**Figure 2.** Original data for electric vehicles (EVs) in the work area.

*4.2. Results and Analysis*

In order to verify the effectiveness of the method proposed in this paper, the joint estimation method based on the edge KDE and the Copula function [21,25,28] was compared, defined, and simulated in the following four cases:

Case 1: Ternary symmetric KDE was conducted on the original data, and simulation data were generated with the Monte Carlo method;

Case 2: After conducting normal transformation on the original data, ternary symmetric KDE was performed on the transformed data. Then, the simulated data that were generated by the Monte Carlo method were inversely transformed as the final simulation data;

Case 3: Joint estimation based on the edge KDE and the Copula function was conducted on the original data, and simulation data were generated with the Monte Carlo method;

Case 4: After conducting normal transformation on the original data, joint estimation based on the edge KDE and the Copula function was conducted on the original data. Then, the simulated data that were generated by the Monte Carlo method were inversely transformed as the final simulation data.

Furthermore, the simulation samples number can be set to 11,740, and the simulation times with Monte Carlo method can be set to 10,000. The experimental platform was built in MATLAB (R2014a 64-bit), based on a Hewlett Packard personal computer (CPU: Intel Core i7-8700 3.2 GHz, RAM: 8 GB, OS: Windows 10 64-bit).

(1)    Data Transformation Analysis

In this paper, transformed parameters of the AT, the ST, and the CC were estimated with the maximum likelihood estimation method. Estimated results of various parameters are shown in Figure 3. When the parameter estimation results are taken as −0.5575, 0.2170, and 0.3992, corresponding maximum likelihood functions are maximized. Based on this, the data transformation model was utilized to perform normal transformations on the AT, the ST, and the CC in the original data, respectively. The scatter density map (yellow means high density) of the transformed data (without units) is shown in Figure 4. A total of 74.3% of the scattered points (with yellow) gather around the center area in Figure 4, and 87.3% of the scattered points (with non-yellow) in Figure 2 are relatively dispersed. Therefore, most of the scattered points in the 3D space gather around the center area after the normal transformation of the original data.
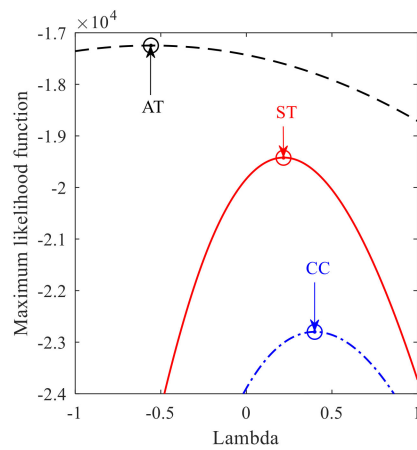
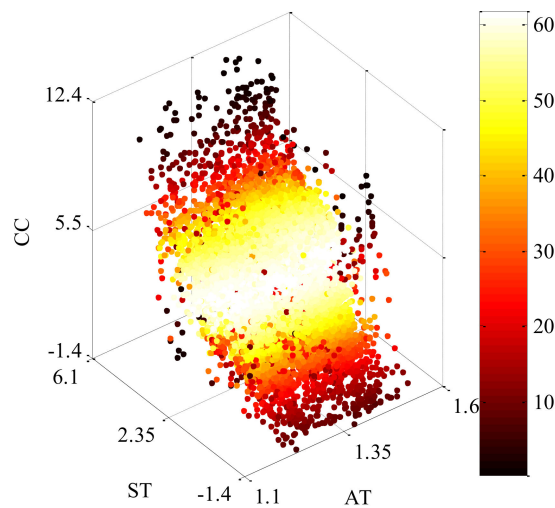**Figure 3.** Estimated results of various parameters.



**Figure 4.** Transformed data for EVs.

In general, the skewness can be used to detect whether the data conforms to a normal distribution. If so, it is equal to 0 [25]. In order to better illustrate the effect of data transformation, the frequency histograms of the original data and the transformed data (without units) in all dimensions are given in Figure 5. Additionally, the corresponding skewness results were calculated as (0.7994, 2.1144, 1.8073), and (0.0830, −0.0062, 0.0192). Therefore, the distributions of the transformed data in all dimensions are closer to the normal distributions than those of the original data, and a better normal transformation effect can be observed in the ST and the CC data.
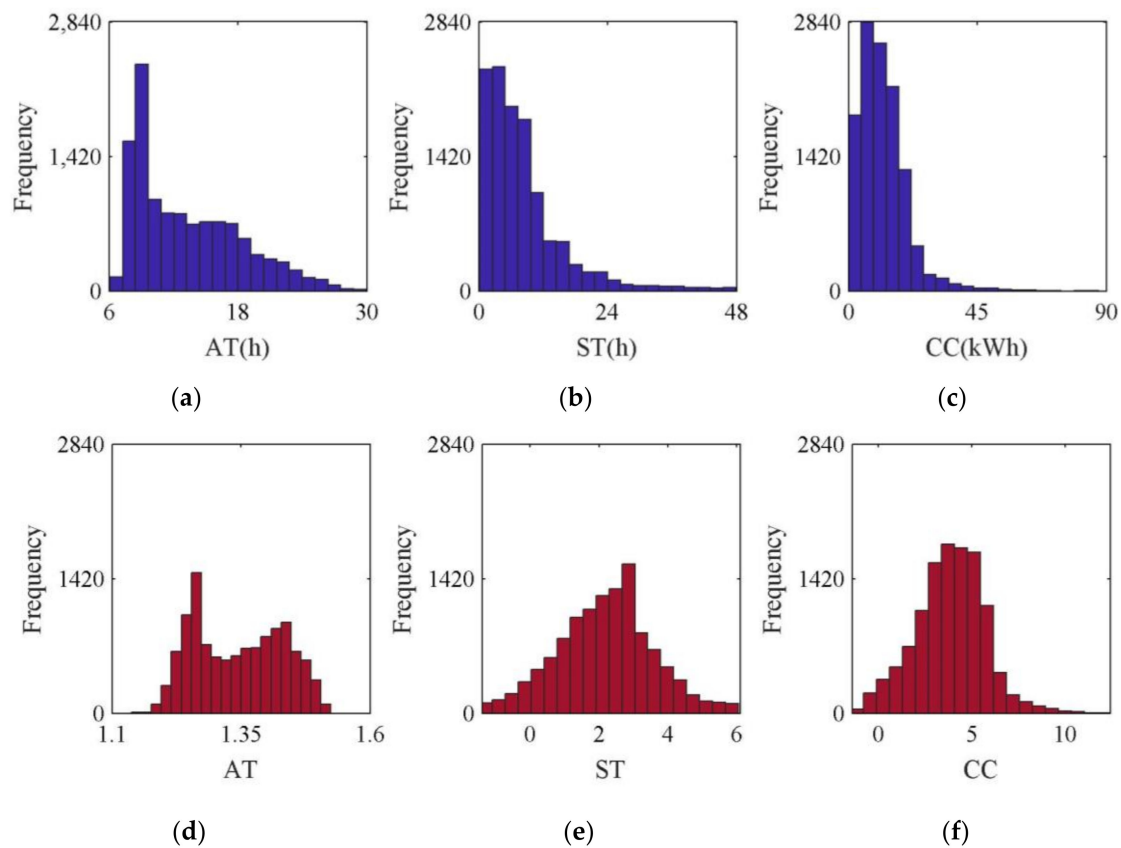
**Figure 5.** The frequency histograms of (**a**) arrival time (AT) original data, (**b**) staying time (ST) original data, (**c**) charging capacity (CC) original data, (**d**) AT transformed data, (**e**) ST transformed data, (**f**) CC transformed data.

(2)　Estimation Model Analysis

The optimal bandwidths of the original data and the transformed data were calculated with the least square cross-validation method, as shown in Table 1. When the optimal bandwidths $\left(h_{opt1}, h_{opt2}, h_{opt3}\right)$ are taken as (0.5865, 0.0733, 0.2027) and (0.0069, 0.0456, 0.0908), respectively, the corresponding scaling function can be minimized.

**Table 1.** Optimal bandwidths of the original data and the transformed data.

| Items | $\left(h_{opt1}, h_{opt2}, h_{opt3}\right)$ | **Min(M)** |
|---|---|---|
| Original data | (0.5865, 0.0733, 0.2027) | −0.0012 |
| Transformed data | (0.0069, 0.0456, 0.0908) | −0.5371 |

Based on this, estimation results can be obtained by conducting the ternary symmetric KDE on the original data and the transformed data. Estimation results that are four-dimensional data are projected on three coordinate planes for output. The ternary symmetric KDE results of the original data and transformed data in each coordinate plane are shown in Figure 6.
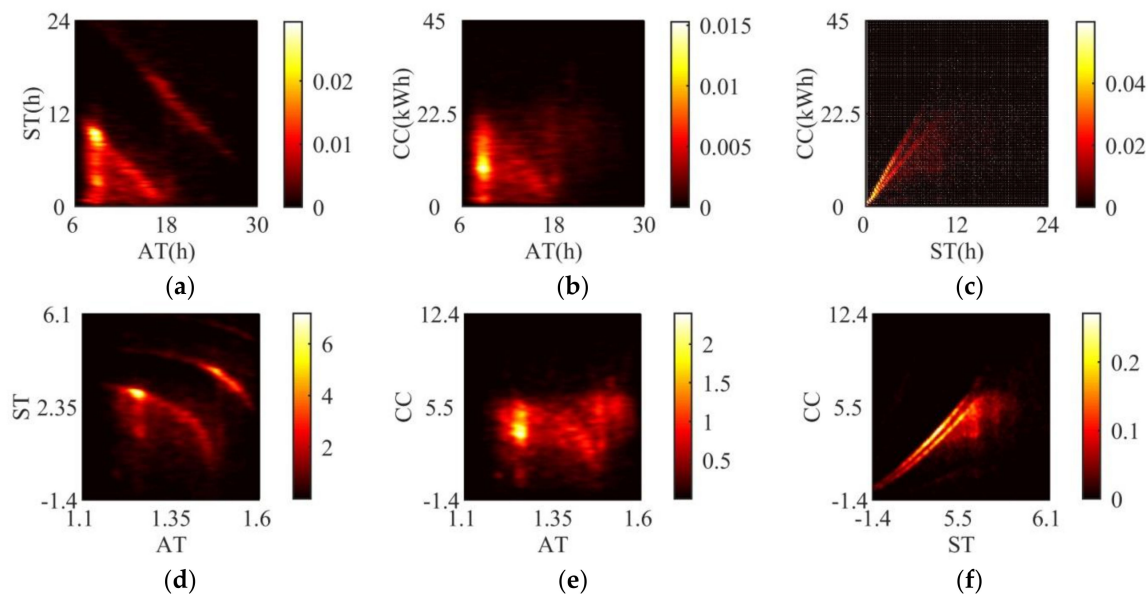
**Figure 6.** Ternary symmetric kernel density estimation (KDE) results of (**a**) AT-ST original data, (**b**) AT-CC original data, (**c**) ST-CC original data, (**d**) AT-ST transformed data, (**e**) AT-CC transformed data, (**f**) ST-CC transformed data.

Moreover, the edge KDE should be given to the to-be-compared joint estimation model based on the edge KDE and the Copula function. The edge KDEs of the original data and the transformed data are shown in Figure 7. On this basis, an appropriate ternary Copula function was selected. The ternary normal Copula function (TNCF) and the ternary t-Copula function (TTCF) are commonly ternary Copula functions [25]. The linear correlation parameter matrices of the Copula functions of the raw data and the transformed data are shown in Table 2.
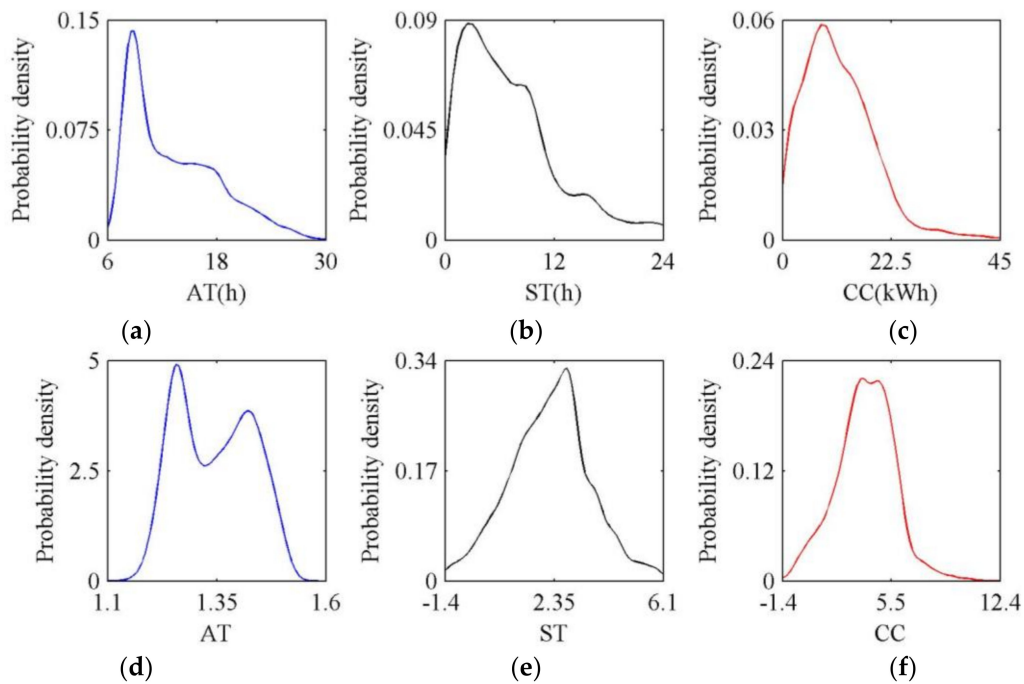


**Figure 7.** The edge KDEs of the (**a**) AT original data, (**b**) ST original data, (**c**) CC original data, (**d**) AT transformed data, (**e**) ST transformed data, (**f**) CC transformed data.

**Table 2.** Linear correlation parameter matrices of the Copula functions of the original data and the transformed data.

| Items | Original Data | | | Transformed Data | | |
|:---:|:---:|:---:|:---:|:---:|:---:|:---:|
| TNCF | 1.000 | 0.1718 | 0.1520 | 1.000 | 0.1561 | 0.1402 |
| | 0.1718 | 1.000 | 0.6557 | 0.1561 | 1.000 | 0.6581 |
| | 0.1520 | 0.6557 | 1.000 | 0.1402 | 0.6581 | 1.000 |
| TTCF | 1.000 | 0.1600 | 0.1338 | 1.000 | 0.1520 | 0.1318 |
| | 0.1600 | 1.000 | 0.7314 | 0.1520 | 1.000 | 0.7276 |
| | 0.1338 | 0.7314 | 1.000 | 0.1318 | 0.7276 | 1.000 |

As shown in Table 2, the original data are characterized by a good linear correlation in comparison to transformed data. After that, the squared Euclidean distance between the Copula function and the empirical Copula function [25] were calculated to evaluate their strengths and weaknesses. Regarding original data, corresponding Euclidean distance calculation results are 3.4119 and 3.2634, respectively; concerning transformed data, corresponding Euclidean distance calculation results are 3.2667 and 3.1489, respectively, which show that original data and transformed data can be better fitted in the TTCF model. In this paper, the joint estimation model based on the edge KDE and the TTCF was chosen as the comparison object.

(3)    Evaluation Analysis for Estimation Model

A total of 10,000 simulation times were conducted to generate simulated data samples for evaluating the fitting level of the ternary symmetric KDE model and the joint estimation model based on the edge KDE and the TTCF. To make it simple, a simulation sample was selected. To select the simulated data (whose structure is more similar to that of the original data), the axis ranges of the scatter density maps were set to $AT \in [6, 30], ST \in [0, 48], CC \in [0, 90]$, and the radii of the benchmark spheres of the scatter density maps were set to 3.4066 according to $\sqrt{[\text{range}(AT)/30]^2 + [\text{range}(ST)/30]^2 + [\text{range}(CC)/30]^2}$, where $\text{range}(\cdot)$ is the function that is the difference between the maximum value and the minimum value of the axis. The scatter density maps of the simulated data in four cases are shown in Figure 8. The overall 3D structures of the original data (as shown in Figure 2) and simulated data in case 1 and case 2 are approaching. In comparison, the overall 3D structures of the simulated data in case 3 and case 4 are significantly different. In conclusion, the joint estimation model based on the edge KDE and the TTCF might lead to a large deviation between the simulation results and the actual results, while the ternary symmetric KDE is more suitable for describing the overall relationship among the AT, the ST, and the CC, presenting a high fitting level.
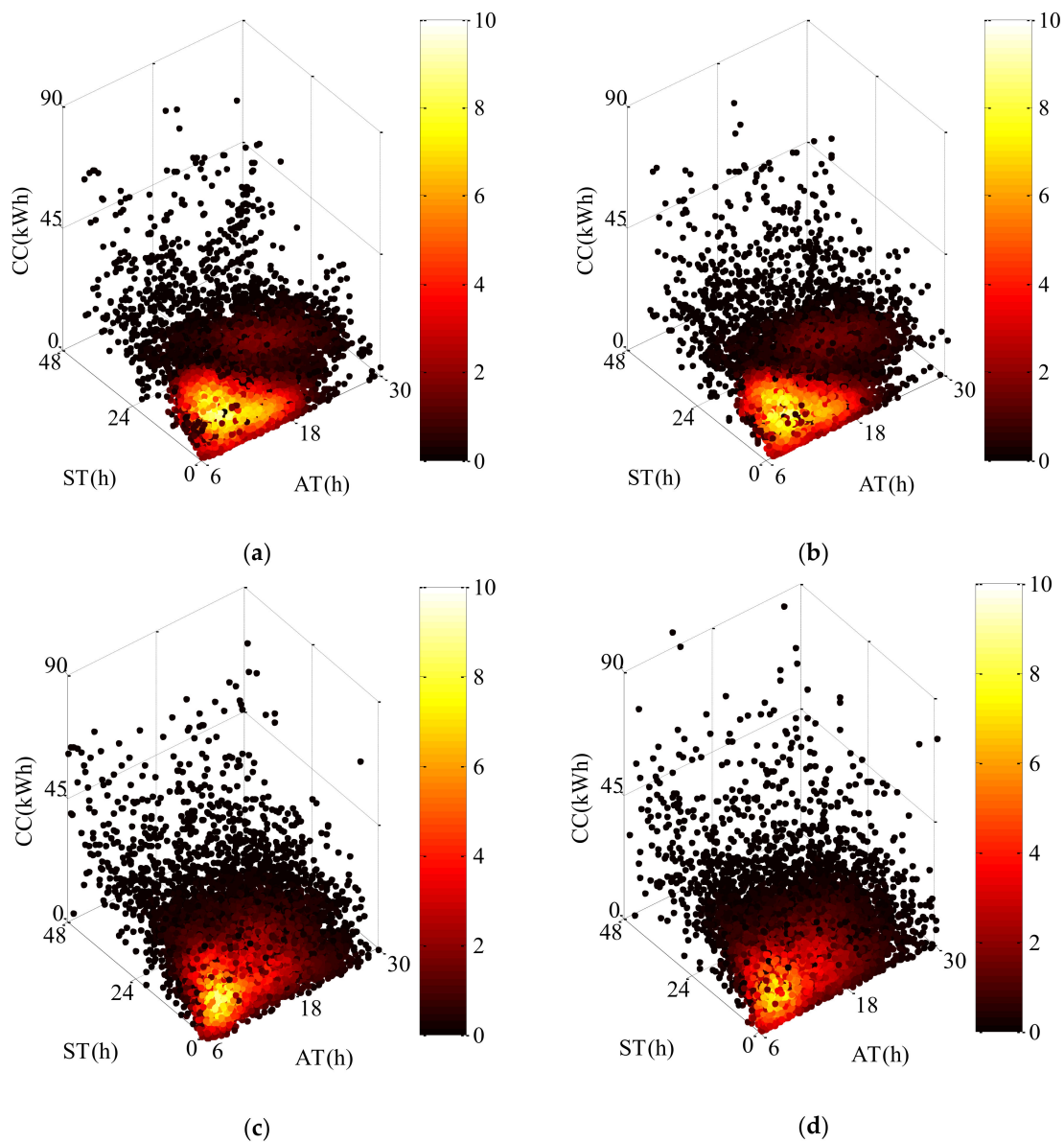
**Figure 8.** Scatter density maps of the simulated data in (**a**) case 1, (**b**) case 2, (**c**) case 3, (**d**) case 4.

As shown in Figure 9, the frequency histograms of the original data and the frequency histograms (for describing four-dimensional data) of the simulated data in two cases in different coordinate planes are presented to compare the accuracy levels of the ternary symmetric KDE in case 1 and case 2. The frequency matrix similarities between the simulated data and the original data in each coordinate plane in two cases can be shown in Table 3. The similarity between the frequency matrix of the simulated data and the original data in each coordinate plane in the case 2 is higher than that in the case 1. Meanwhile, the corresponding frequency histogram has a high goodness of fit. Hence, the transformed data can eliminate the boundary bias of the symmetric KDE in a more effective way in comparison to the original data, improving the fitting level of the ternary symmetric KDE model.
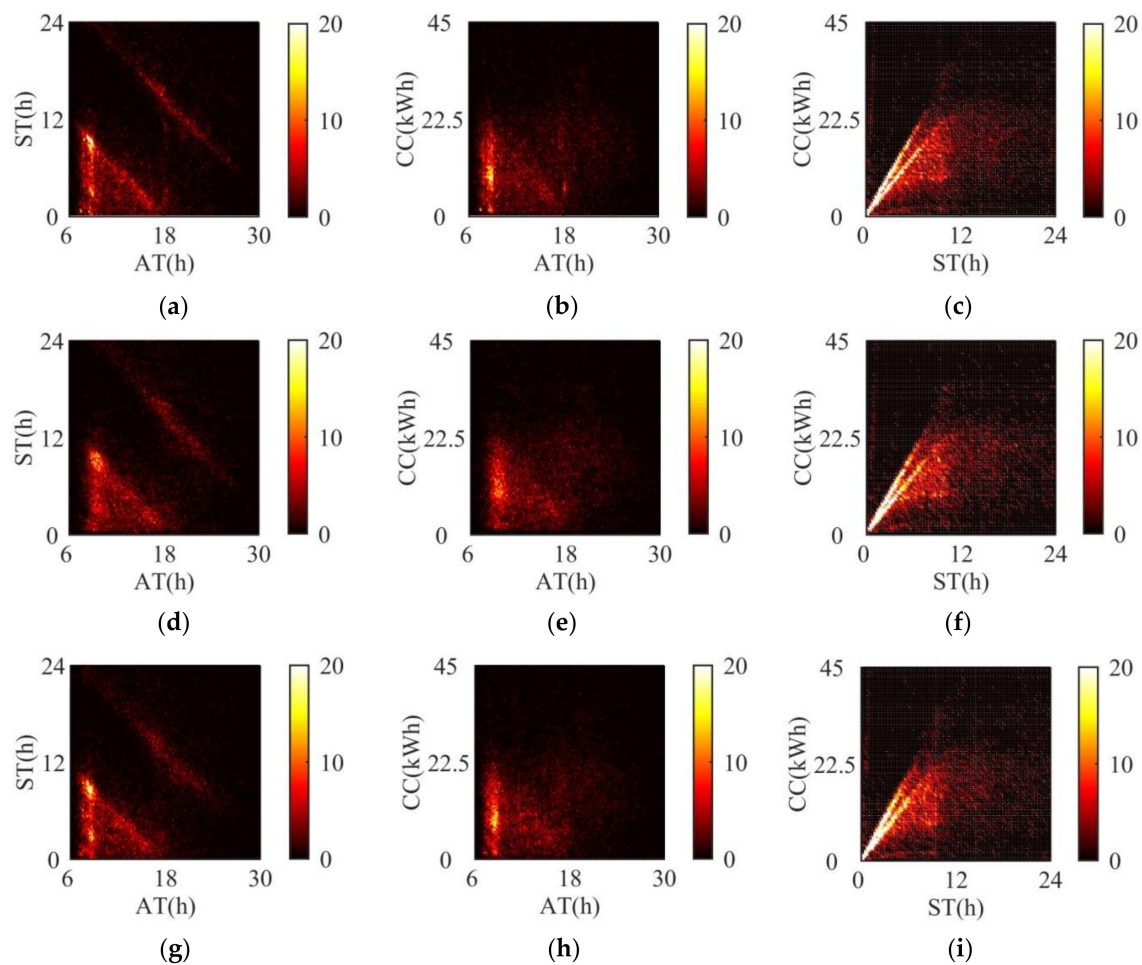
**Figure 9.** Frequency histograms of (**a**) AT-ST original data, (**b**) AT-CC original data, (**c**) ST-CC original data, (**d**) AT-ST simulated data in case 1, (**e**) AT-CC simulated data in case 1, (**f**) ST-CC simulated data in case 1, (**g**) AT-ST simulated data in case 2, (**h**) AT-CC simulated data in case 2, (**i**) ST-CC simulated data in case 2.

**Table 3.** The frequency matrix similarities between the simulated data and the original data in each coordinate plane in two cases.

| Items | Case 1 | Case 2 |
|-------|--------|--------|
| AT-ST | 0.6805 | 0.8001 |
| AT-CC | 0.7102 | 0.8751 |
| ST-CC | 0.9530 | 0.9630 |

The mean and the standard deviation of the frequency matrix similarities between the simulated data and the original data in two cases under 10,000 Monte Carlo simulations are shown in Table 4. Compared with case 1, the simulated data and the original data in each coordinate plane have a higher mean value of frequency matrix similarities and a lower standard deviation in case 2, that is, the ternary symmetric KDE model has a high fitting level and stability under transformed data.

**Table 4.** The mean and the standard deviation of the frequency matrix similarities between the simulated data and the original data in two cases.

|  | Mean | | Standard Deviation | |
|---|---|---|---|---|
|  | Case 1 | Case 2 | Case 1 | Case 2 |
| AT-ST | 0.7499 | 0.8256 | 0.0650 | 0.0461 |
| AT-CC | 0.7592 | 0.8612 | 0.0597 | 0.0321 |
| ST-CC | 0.9075 | 0.9130 | 0.0864 | 0.0815 |

## 5. Conclusions

An accurate joint modeling approach of the arrival time, the staying time, and the charging capacity for the EVs charging behaviors in the work area based on ternary symmetric KDE was proposed in accordance with the actual data. It can provide the basis for studying EVs charging load modeling, the impact of charging on the PG, orderly charging strategies, and planning of charging facilities. The main contributions of this paper can be summarized as follows:

- Effective data records are aggregated from space and time. On the one hand, every CS is classified into the work area, the residential area, the commercial area, and other areas through searching the name of the CS from the map at the spatial dimension. On the other hand, EVs arrival times of multiple days are aggregated into one day at the time dimension.
- A ternary symmetric KDE model of EVs charging behaviors is established. A higher fitting level can be achieved by the proposed ternary symmetric KDE method, in comparison to the joint estimation method based on the edge KDE and the TTCF.

The main conclusions of this paper can be summarized as follows: (1) Through the Box-Cox transformation, the original data can be transformed to make its distribution characteristic more normal. (2) The deviation between the simulation results and the actual results can be effectively reduced by the ternary symmetric KDE, in comparison to the joint estimation model based on the edge KDE and the Copula function. (3) With the Box-Cox transformation, the ternary symmetric KDE model has a high fitting level and stability, and its boundary effect can be effectively eliminated.

Simulated data samples of the arrival time, staying time, and charging capacity with the overall relationship can be generated on a large-scale with the Monte Carlo method, and the charging behaviors modeling results. The simulated data samples can be directly used for modeling the charging load of a certain scale EVs in work areas, with the actual power data of the charging pile that needs to be given separately since it is not stored in the platform. In the future, the charging behaviors modeling results can be modified with the continual accumulation of the actual running data of EVs in other areas, and the modified results can also be used for modeling the charging load of a certain scale EVs in other areas.

It is worth mentioning that the original data drives the charging behaviors modeling to obtain the modeling results that would otherwise be unavailable. Moreover, the waiting time has not been added to the modeling in this paper due to the charging characteristics of EVs in work area. However, the subsequent research will focus on the EVs charging behaviors modeling in highway fast-charging stations where the waiting time is an important (and a new compared to refueling) factor in case of EVs.

**Author Contributions:** L.C. established a ternary symmetric KDE model for EVs charging behaviors and designed the experiments; L.C. and H.Z. contributed analysis tools; L.C. wrote the paper. This work was performed under the advisement and regular feedback of X.H. All authors have read and agree to the published version of the manuscript.

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

## References

1.  Abagnale, C.; Cardone, M.; Iodice, P.; Strano, S.; Terzo, M.; Vorraro, G. Power requirements and environmental impact of a pedelec. A case study based on real-life applications. *Environ. Impact Assess. Rev.* **2015**, *53*, 1–7. [CrossRef]
2.  Iodice, P.; Senatore, A. Industrial and Urban Sources in Campania, Italy: The Air Pollution Emission Inventory. *Energy Environ.* **2015**, *26*, 1305–1317. [CrossRef]
3.  Yang, T.Y.; Xu, X.X.; Guo, Q.L.; Zhang, L.; Sun, H.B. EV charging behavior analysis and modelling based on mobile crowdsensing data. *IET Gener. Transm. Distrib.* **2017**, *11*, 1683–1691. [CrossRef]
4.  Rao, R.; Cai, H.; Xu, M. Modeling electric taxis' charging behavior using real-world data. *Int. J. Sustain. Transp.* **2018**, *12*, 452–460. [CrossRef]
5.  Shao, Y.C.; Mu, Y.F.; Yu, X.D.; Dong, X.H.; Jia, H.J.; Wu, J.Z.; Zeng, Y. A spatial-temporal charging load forecast and impact analysis method for distribution network using EVs-traffic-distribution model. *Proc. CSEE* **2017**, *37*, 5207–5219.
6.  Su, J.; Lie, T.; Zamora, R. Modelling of large-scale electric vehicles charging demand: A New Zealand case study. *Electr. Power Syst. Res.* **2019**, *167*, 171–182. [CrossRef]
7.  Chen, L.; Chen, Z.; Huang, X.; Jin, L. A Study on Price-Based Charging Strategy for Electric Vehicles on Expressways. *Energies* **2016**, *9*, 385. [CrossRef]
8.  Luo, Y.; Zhu, T.; Wan, S.; Zhang, S.; Li, K. Optimal charging scheduling for large-scale EV (electric vehicle) deployment based on the interaction of the smart-grid and intelligent-transport systems. *Energy* **2016**, *97*, 359–368. [CrossRef]
9.  Wei, W.; Wu, L.; Wang, J.; Mei, S. Network Equilibrium of Coupled Transportation and Power Distribution Systems. *IEEE Trans. Smart Grid* **2018**, *9*, 6764–6779. [CrossRef]
10. Chen, L.; Huang, X.; Chen, Z.; Jin, L. Study of a New Quick-Charging Strategy for Electric Vehicles in Highway Charging Stations. *Energies* **2016**, *9*, 744. [CrossRef]
11. U.S. Department of Transportation. Federal Highway Administration. National Household Travel Survey. Available online: https://nhts.ornl.gov/ (accessed on 19 December 2019).
12. Arias, M.B.; Bae, S. Electric vehicle charging demand forecasting model based on big data technologies. *Appl. Energy* **2016**, *183*, 327–339. [CrossRef]
13. Lojowska, A.; Kurowicka, D.; Papaefthymiou, G.; Van Der Sluis, L. Stochastic Modeling of Power Demand Due to EVs Using Copula. *IEEE Trans. Power Syst.* **2012**, *27*, 1960–1968. [CrossRef]
14. Wang, H.L.; Zhang, Y.J.; Mao, H.P. Charging load forecasting method based on instantaneous charging probability for electric vehicles. *Electr. Power Autom. Equip.* **2019**, *39*, 207–213.
15. Wang, D.; Gao, J.; Li, P.; Wang, B.; Zhang, C.; Saxena, S. Modeling of plug-in electric vehicle travel patterns and charging load based on trip chain generation. *J. Power Sources* **2017**, *359*, 468–479. [CrossRef]
16. Yi, T.; Zhang, C.; Lin, T.; Liu, J. Research on the spatial-temporal distribution of electric vehicle charging load demand: A case study in China. *J. Clean. Prod.* **2020**, *242*, 118457. [CrossRef]
17. Sun, X.-H.; Yamamoto, T.; Morikawa, T. Fast-charging station choice behavior among battery electric vehicle users. *Transp. Res. Part D Transp. Environ.* **2016**, *46*, 26–39. [CrossRef]
18. Sadeghianpourhamami, N.; Refa, N.; Strobbe, M.; Develder, C. Quantive analysis of electric vehicle flexibility: A data-driven approach. *Int. J. Electr. Power Energy Syst.* **2018**, *95*, 451–462. [CrossRef]
19. Flammini, M.G.; Prettico, G.; Julea, A.; Fulli, G.; Mazza, A.; Chicco, G. Statistical characterisation of the real transaction data gathered from electric vehicle charging stations. *Electr. Power Syst. Res.* **2019**, *166*, 136–150. [CrossRef]
20. Xydas, E.; Marmaras, C.; Cipcigan, L.M.; Jenkins, N.; Carroll, S.; Barker, M. A data-driven approach for characterising the charging demand of electric vehicles: A UK case study. *Appl. Energy* **2016**, *162*, 763–771. [CrossRef]
21. Chen, Z.; Zhang, Z.; Zhao, J.; Wu, B.; Huang, X. An Analysis of the Charging Characteristics of Electric Vehicles Based on Measured Data and Its Application. *IEEE Access* **2018**, *6*, 24475–24487. [CrossRef]
22. Zhao, H.; Yan, X.; Ren, H. Quantifying flexibility of residential electric vehicle charging loads using non-intrusive load extracting algorithm in demand response. *Sustain. Cities Soc.* **2019**, *50*, 101664. [CrossRef]
23. Xue, L. *Introduction to Modern Nonparametric Statistics*; Science Press: Beijing, China, 2015.

24. Silverman, B.W. *Density Estimation for Statistics and Data Analysis*; Springer Science and Business Media LLC: Berlin, Germany, 1986.
25. Xie, Z. *MATLAB Statistic Analysis and Application: 40 Cases Analysis*; Beihang University Press: Beijing, China, 2015.
26. Miao, P.B.; Yu, J.; Liu, G.P.; Liang, M.; Li, W.Y.; Ren, Z.Y. Electric public bus load model based on improved kernel density estimation and latin hypercube sampling. *Trans. China Electrotech. Soc.* **2016**, *31*, 187–193.
27. Box, G.E.P.; Cox, D.R. An Analysis of Transformations Revisited, Rebutted. *J. Am. Stat. Assoc.* **1982**, *77*, 209–210. [CrossRef]
28. Wang, R.; Thakur, C.S.; Hamilton, T.J.; Tapson, J.; Van Schaik, A. A neuromorphic hardware architecture using the Neural Engineering Framework for pattern recognition. *IEEE Trans. Biomed. Circuits Syst.* **2015**, *11*, 574–584. [CrossRef] [PubMed]
29. Yang, N.; Huang, Y.; Ye, X.W.; Li, H.S.; Li, S.Y.; Dong, B.T. Modeling of output correlation of multiple wind farms based on adaptive multivariable nonparametric kernel density estimation. *Proc. CSEE* **2018**, *38*, 3805–3812.
30. Rubinstein, R.Y.; Kroese, D.P. *Simulation and the Monte Carlo Method*; John Wiley & Sons: Hoboken, NJ, USA, 2011.
31. Kernel Density Estimation Toolbox. Available online: https://www.ics.uci.edu/~{}ihler/code/kde.html (accessed on 19 December 2019).