# Exploring the Attractiveness of Residential Areas for Human Activities Based on Shared E-Bike Trajectory Data

**Xiaoqian Cheng** [1], **Weibing Du** [1,2,*], **Chengming Li** [1,3], **Leiku Yang** [1] and **Linjuan Xu** [4]

[1] School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo 454000, China; cxq@hpu.edu.cn (X.C.); cmli@casm.ac.cn (C.L.); yanglk@hpu.edu.cn (L.Y.)

[2] State Key Laboratory of Desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi 830011, China

[3] Chinese Academy of Surveying and Mapping, Beijing 100830, China

[4] Key Laboratory of Sediment, Ministry of Water Resources, Yellow River Institute of Hydraulic Research, Yellow River Conservancy Commission, Zhengzhou 450003, China; xulinjuan@hky.yrcc.gov.cn

\* Correspondence: dwb@hpu.edu.cn

**Abstract:** Human activities generate diverse and sophisticated functional areas and may impact the existing planning of functional areas. Understanding the relationship between human activities and functional areas is key to identifying the real-time urban functional areas based on trajectories. Few previous studies have analyzed the interactive information on humans and regions for functional area identification. The relationship between human activities and residential areas is the most representative for urban functional areas because residential areas cover a wide range and are closely connected with human life. The aim of this paper is to propose the CARA (Commuting Activity and Residential Area) model to quantify the correlation between human activities and urban residential areas. In this model, human activities are represented by hot spots extracted by the Gaussian Mixture Model algorithm while residential areas are represented by POI (point of interest) data. The model shows that human activities and residential areas present a logarithmic relationship. The CARA model is further assessed by retrieving urban residential areas in Tengzhou City from shared e-bike trajectories. Compared with the actual map, the accuracy reaches 83.3%, thus demonstrating the model's reliability and feasibility. This study provides a new method for functional areas identification based on trajectory data, which is helpful for formulating the urban people-oriented policies.

**Keywords:** shared e-bike trajectory; commuting activity; urban residential area; machine learning; urban hot spots

## 1. Introduction

Human activities not only generate diverse and sophisticated functional areas, but may also change the existing planning of functional areas [1]. Understanding the relationship between human activities and urban regions is key to functional area identification [2]. In the big data era, various trajectory data can be easily obtained because of the popularity of location-aware devices and smart sensors in a city. These trajectory data not only convey the underlying information on people and cities, but also imply the interactive information of people and the urban environment [3,4]. Therefore, trajectory data provide a new opportunity for deeply mining the inner relationship between human activities and the surrounding environment. Exploring the inner relationship and identifying functional areas based on trajectory data are helpful for city planners and administrators to comprehend urban dynamics and evaluate urban environments timely and rapidly.

At present, few studies have identified functional areas based on trajectories by considering the inner relationship between human activities and functional areas. Reades et al. explored the relationship between human activities and business areas based on mobile phone data, in which human activities are described by mobile phone usage [5]. They also successfully identified different mobile usage patterns between business and residential areas. Calabrese et al. used an eigen decomposition method to identify different patterns between human activities and the urban environment, and then combined the results with a clustering method to classify campus physical environments based on Wi-Fi network data [6]. Based on taxi pick-ups and drop-offs, Liu et al. built temporal activity variations of each parcel, and then clustered the parcels with temporal activity variations to identify "source-sink" areas of Shanghai [7]. The temporal activity variations of different functional areas were built based on taxi or bicycle trajectories, and the Support Vector Machine (SVM) method was then used to identify functional areas in Hangzhou City [8,9]. It is difficult to obtain the standard temporal activity variation of each functional area, which need prior knowledge or affluent high-precision samples training [10,11]. Furthermore, the temporal activity variation of each functional area is easily subjected to the region size and population density.

Shared e-bikes, as a new emerging green travel mode, have been widely adopted by local citizens, especially in the second- and third-tier cities. Compared with taxies, shared e-bikes can freely travel along large roads and narrow alleys and have better accessibility especially in the areas with large traffic volume. Compared with bicycles, shared e-bikes can travel a longer distance with a higher speed due to the electric power driven [12]. Most of the literature related to e-bikes focus on travel behavior analysis [13,14], travel mode choice [15–17], and environmental protection [18], aiming to promote e-bikes usage. Few contemporary research has explored the relationship between human activities and the social function of regions and identified functional areas based on e-bike trajectory.

Among urban functional areas, residential areas cover a wide range and are closely connected with human living life [19]. The relationship between human activities and residential areas is the most representative. Commuting represents an important component of daily travel, and it refers to the journey between home and the workplace [20]. Citizens usually leave "home" and travel to the "workplace" in the morning and return "home" from the "workplace" in the evening. Exploring the relationship between human commuting activities and residential areas is helpful for identifying urban residential areas. Delineating residential areas using e-bike trajectories is difficult because the relationship between human activities and the social function of residential areas is ambiguous. In this paper, we attempt to quantify for the first time the correlation between human activities and residential areas based on e-bike trajectories, and then retrieve residential areas from human activities to determine the spatial distribution. The objectives are to (1) quantify the inner relationship between commuting activities and residential areas; and (2) delineate the residential areas by shared e-bike trajectory. This study provides a new perspective for functional area identification based on trajectory data and will thus be helpful for formulating urban people-oriented policies.

The remainder of this paper is organized as follows. In Section 2, a review of the relevant literature on the relationship between human activities and urban regions is presented. In Section 3, the study area and the dataset are briefly introduced, and then the workflow and the core technology are illustrated. In Section 4, the CARA (Commuting Activity and Residential Area) model is built to quantify the correlation between human activities and residential areas, and then the residential areas in Tengzhou City are retrieved from e-bike trajectory data. In Section 5, the experimental results are evaluated and discussed, and in Section 6, the conclusions are presented.

## 2. Related Works

The study presented here is focused on "identifying residential areas" and "determining the relationships between human activities and functional areas." The related works on these two aspects are summarized.

Residential areas can be determined by identifying urban functional areas. Conventional urban functional area identification is largely based on the land use status and questionnaire surveys [21], which are time-consuming, labor-intensive, and cannot reflect the city structure in real time [22]. Some researchers also used POI (point of interest) data to identify functional areas [23,24]. This method ignores the inner correlation between human activities factors and urban functional areas; thus, the identified results cannot reflect the influence of human activities on the dynamic changes of region social function [25].

Big data provides an opportunity to identify functional areas by data-driven methods. Accompanied with the rapid development of LBS (location-based service) technology, various trajectory data can be easily obtained, such as taxi trajectories, bicycle trajectories, e-bike trajectories, and mobile phone data. Trajectory data include the geographic locations of objects that change over time [26]; thus, these data are a fine-scale representation of the spatiotemporal footprint of human activities. Thus, scholars have combined trajectory data and other data sources (POI, social media data, satellite image) to improve the identification results [27–29]. For instance, to fully exploit taxi trajectory and POI data, a topic-based inference model is applied to discover urban functional areas [1]. In the model, trajectory data reflect human movements from one place to another while the POIs located in a region reflect the function of that region. This method implicitly considers spatial interaction information rather than the influence of region social function on human mobility.

The collective dynamic of a whole city has a strong spatiotemporal pattern, which suggests that the population of each region in a city varies with time due to different human activities. These population variations can be represented by human activities identified via trajectory data, such as picking up a bicycle or making a phone call [2]. This temporal variation shows the interactive information between human activities and locations with certain social functions. Generally, regions with different social functions have different temporal activity variations. For instance, residential areas are characterized by large travel volume, although the centralized travel time is limited to two peaks periods of commuting. However, railway stations have a large traffic volume all day except late at night. Obviously, temporal activity variations are the interpreted indicators of social functions and are thus analogous to the reflection spectrum curve of objects in remote sensing images [11]. Therefore, temporal activity variation can be assessed by advanced machine learning algorithms, such as SVM (Support Vector Machine), EM (Expectation Maximization), deep learning, or random forest [8,9,30], and are used to identify urban functional areas through trajectory data [31,32]. The temporal activity variation-based method of functional area identification considers the interactive information between human activities and functional areas, although the interactive information is qualitatively described. However, the standard temporal activity variation pattern of each functional area is difficult to obtain; thus, the method is limited in applicability and robustness. In addition, the temporal activity variation pattern of each functional area is obtained based on prior knowledge or training a high number of precise samples [10].

Human activities and urban regions are closely intertwined [33]. Understanding the interactive information between human activities and functional areas is essential for functional area identification and land use analysis [23,34,35]. The relationship between human activities and functional areas has always been a hot topic in the urban transportation and urban planning domain field, and plays an important role in public policymaking [36]. Many scholars have focused on the relationship between travel behavior and land use, and they found that land use patterns determine the need to travel, and travel behaviors also affect the land use pattern and lead to improved urban development [20]. Additionally, the influence of the urban built environment on some special travel behaviors was explored. For instance, Boarnet and Crane built a model to study the relationship between the urban built environment and nonwork travel behavior, and found significant linkages between land use and travel speed and distance in San Diego [33]. Handy et al. studied the influence of the built environment on commuting behavior and found that the influence was not only statistically significant but also practically important [37]. These studies were mainly carried out using travel survey data, which are time-consuming to collect and cannot describe human movements in real time. With the popularity

of LBS technology, the relationship between travel behavior and the built environment was studied based on trajectory data. Ahmadreza et al. built a linear mixed model to explore the influence of the factors that affect bicycle flows (e.g., weather, temporal, bicycle infrastructure, land use, and built environment) based on bicycle trajectory data [38]. Wafic et al. studied the bike sharing demand under the impacts of the built environment and weather condition in order to optimize the bicycle share system [39]. In these studies, the associated models mainly represent the comprehensive influence of the external environmental factors on human travel behavior and cannot be used to infer region social function from trajectory data.

Previous studies on "functional area identification" and the "relationships between human activities and functional areas" have various shortcomings. First, most of the studies on functional area identification do not consider the interactions between human activities and regional social functions, which limits the accuracy and real-time performance of functional area identification. Second, the majority of the relationships between human activities and functional areas focus on how human travel behavior is affected by the external environment, which cannot be used to inversely identify functional regions. Third, most study data are concentrated on mobile phone data, taxi and bicycle trajectory data, and survey data, while few studies have focused on shared e-bike trajectory data. Thus, this paper is conducted based on shared e-bike trajectory data in Tengzhou City and aims to provide insights on the relationship between human activities and residential areas and guidance for residential area delineation.

## 3. Materials and Methods

### 3.1. Study Area and Dataset

Shared electric bikes (e-bikes for short) are an emerging green travel mode. Compared with bicycles, e-bikes have distinct advantages on long trips, during periods of poor air quality or weather, and in areas with challenging topography [40]. The popularity of shared e-bikes is beneficial to the sustainable development of urban transportation, regional socioeconomics, and the environment. With the use of an electronic fence, shared e-bikes can be picked up or returned freely within the station-free bike sharing system using a smart phone. Therefore, they are widely used in the downtown areas of Tengzhou City.

Tengzhou City, a prefecture-level city, is located in southwestern Shandong Province in eastern China. The geographic coordinates of the city are 116°49′–117°24′ N and 34°50′–35°17′ E. Tengzhou is one of the most beautiful ecotourism cities in China, and the city includes four blocks (located in the central area) and seventeen towns (located in the surrounding area) (seen in Figure 1).

Human activity data were obtained from the real-time captured GPS trajectory points from May to July 2018 in the urban area of Tengzhou City, Shangdong Province. Each e-bike sends GPS information to a specified internet address every minute through an integrated GPS communication module. The trajectory points of shared e-bikes can be obtained by a high-frequency timer system based on an HTTP (hyper-text transfer protocol) data acquisition interface provided by the shared e-bike operator. The acquired trajectory data set contains 2,128,290 GPS points covering the downtown areas (Longquan block, Jinghe block, Shannan block, and Beixin block) of Tengzhou City, which involves 960 e-bikes. Each of the raw GPS points includes the e-bike ID (Station ID), data acquisition time (timestamp), geographic location (latitude and longitude coordinates), and predicted mileage (anticipated mileage).
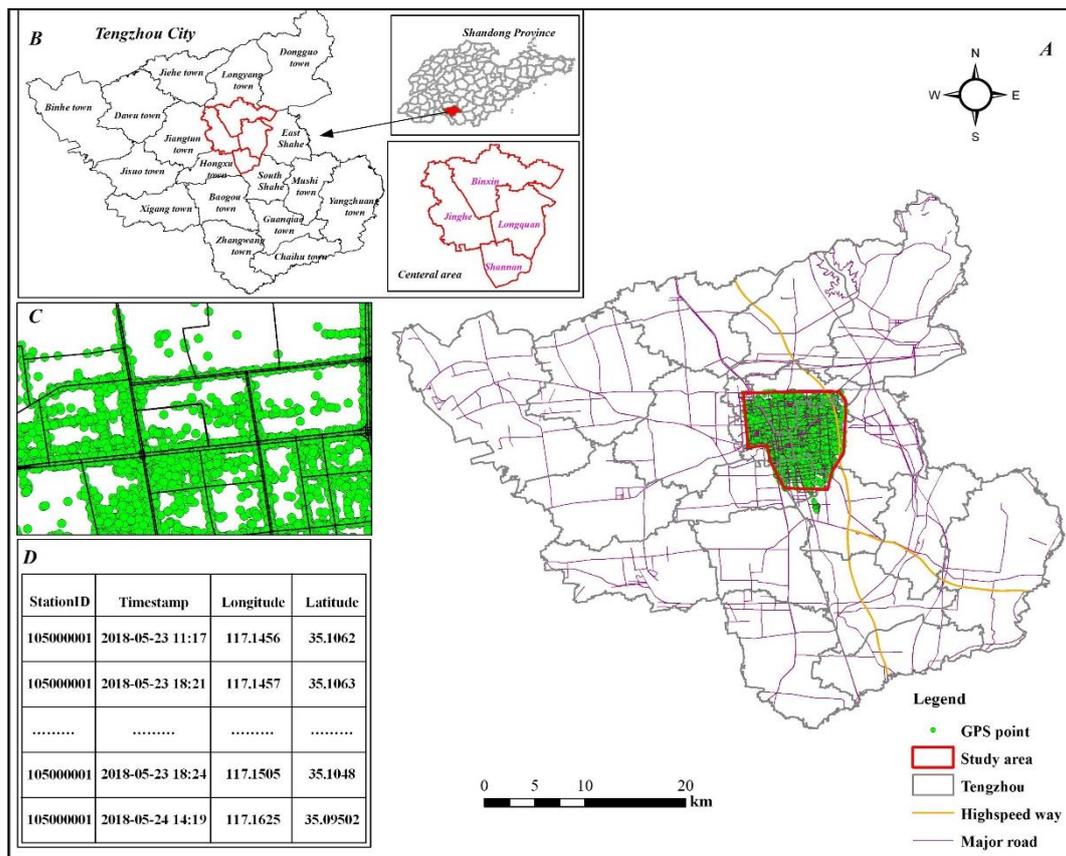
**Figure 1.** Study area and trajectory data. Part **A** shows the study area in Tengzhou, and the green dots indicate the trajectory data for the shared e-bikes. Part **B** shows the location of Tengzhou City in Shandong Province. Part **C** shows the spatial pattern of the trajectory data in Tengzhou. Part **D** illustrates the attributes of raw trajectory GPS points.

## 3.2. Methodology

Commuting behavior links workplace areas and residential areas and is one of the important activities in human daily life. This paper proposes a model to quantify the inner correlation between human commuting activities and residential zones by trajectory data. The raw trajectories are cleaned to obtain OD (Origin-Destination) points, which imply that certain activities are carried out at certain times in a region. Based on these OD points, temporal mobility patterns are analyzed to form the e-bike trajectory data related to residential areas. Then, urban hot spots are detected by the GMM (Gaussian Mixture Model), a self-adapting machine leaning algorithm, in which the Bayesian information criterion (BIC) is used to determine the key parameter, that is, the number of the urban hot spots. POI data are the point data used to describe physical entities in the real world. The spatial distribution reflects the function of a region. Combining POI data and urban hot spots, the CARA model is constructed quantitatively to mine their inner correlation. The CARA model vividly depicts the correlation between human commuting activities and residential zones. Finally, based on the CARA model, the residential zone and boundary in Tengzhou City are re-delineated through e-bike trajectory data to validate the feasibility of the model. The overall workflow is illustrated in Figure 2.
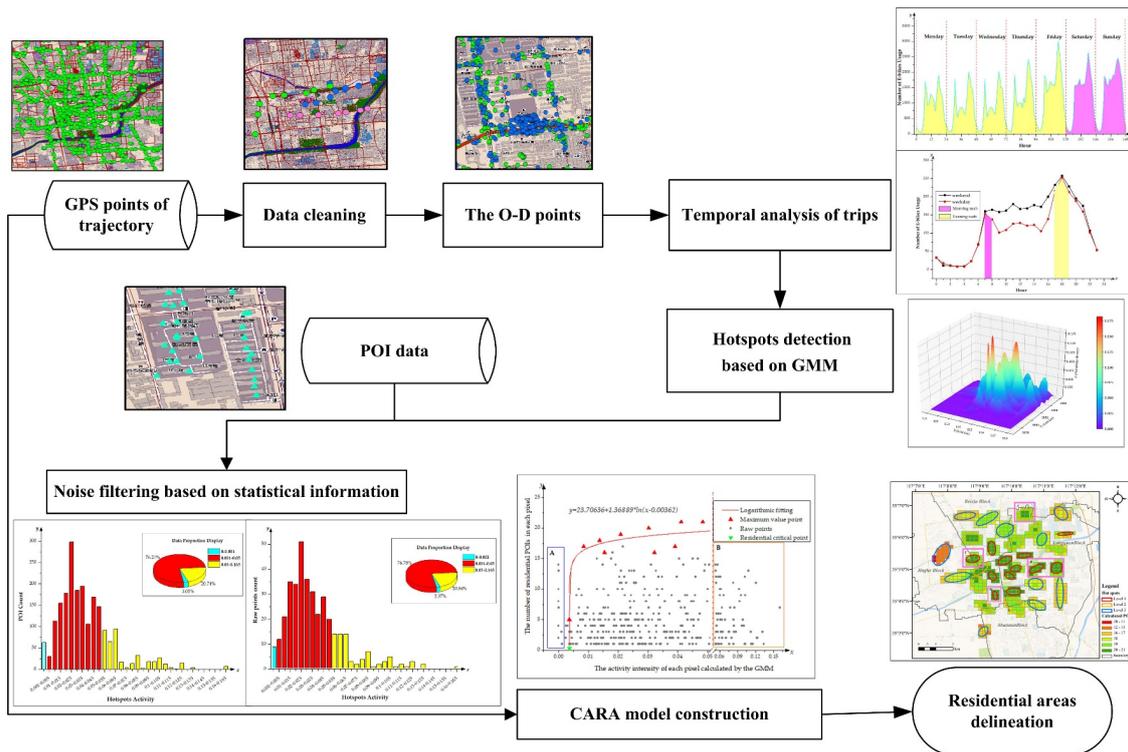
**Figure 2.** The overall workflow of the study.

### 3.2.1. Gaussian Mixture Model

In the study, hot spot detection is a core technology. To accommodate the nonuniformity of trajectory data, a self-adopting machine learning algorithm GMM (Gaussian mixture model) is adopted. The GMM is a classical machine learning algorithm that is most widely used for feature recognition, data classification, and image segmentation [41,42]. The GMM assumes that all the data are generated from a superposition of a finite number of Gaussian distributions with some unknown parameters. In fact, almost any density of data can be approximated at an arbitrary accuracy, if sufficient Gaussian components are used and the parameter values (mean, weight, and covariance values) of each Gaussian component in the linear combination are adjusted. In this paper, a GMM with two variables is applied to GPS trajectory data to detect activity hot spots.

Let $x = \{x_i; i = 1, 2, \cdots, n\}$ and $y = \{y_i; i = 1, 2, \cdots, n\}$ denote the GPS trajectory data sets, where $(x_i, y_i)$ are the longitude and latitude of an arbitrary GPS point, respectively, $i$ is the index of the GPS points, and $n$ is the total number of trajectory points. To cluster a trajectory data set of n points into $K$ labels, the GMM assumes that each observation $x_i$ is considered independent of the label $\Omega_k(k = 1, 2, \cdots, K)$, where $\Omega_k$ is a sample set of all the trajectory points with label $k$. The corresponding probability density equation can be expressed as follows:

$$p(x_i, y_i | \Pi, \Theta) = \sum_{k=1}^{K} \pi_k \cdot \Phi(x_i, y_i | \theta_k) \tag{1}$$

where $\Pi = \{\pi_k | k = 1, 2, \cdots, K\}$ is the set of weights for each component and satisfies the constraints $0 \le \pi_k \le 1$, $\sum_{k=1}^{K} \pi_k = 1$, and $\Phi(x_i, y_i | \theta_k)$ is a Gaussian component of the mixture model with a Gaussian distribution, and it is parameterized by the symbol $\theta_k$. The symbol $\theta_k$ includes the mean vector $\mu_k \left(\mu_k \in R^2\right)$ and covariance matrix $\Sigma_k \left(\Sigma_k \in R^{2\times2}\right)$ of the $K$th component, which can be efficiently

estimated with the expectation maximization (EM) algorithm. $\Phi(x_i, y_i | \theta_k)$ describes the Gaussian distribution of the trajectory data, which can be expressed as follows:

$$\Phi(x_i, y_i | \theta_k) = \Phi(x_i, y_i | \mu_{xk}, \mu_{yk}, \Sigma_k) = \frac{1}{(2\pi) \cdot |\Sigma_k|^{1/2}} \exp\left(-\frac{1}{2} \begin{bmatrix} x_i - \mu_{xk} \\ y_i - \mu_{yk} \end{bmatrix}^T \Sigma_k^{-1} \begin{bmatrix} x_i - \mu_{xk} \\ y_i - \mu_{yk} \end{bmatrix}\right) \tag{2}$$

$\Theta$ is the parameter set for all the components, and $\Theta = \{\theta_k | k = 1, 2, \cdots, K\}$. As the observation $(x_i, y_i)$ is the independent variable being modeled, the joint conditional density of the trajectory data set $(x, y) = \{x_i, y_i; i = 1, 2, \cdots, n\}$ can be modeled as follows:

$$p(x_i, y_i | \Pi, \Theta) = \prod_{i=1}^{n} \sum_{k=1}^{K} \pi_k \Phi(x_i, y_i | \theta_k) \tag{3}$$

The estimation of model parameters $(\pi, \mu, \Sigma)$ is key to constructing a well-suited GMM from random data. In this paper, the EM algorithm is used to estimate parameters with a predefined number K of components until the iterative process converges [43].

### 3.2.2. Bayesian Information Criterion

The number of components in the GMM needs to be defined in advance. A good model can balance the relationship between complexity and descriptive ability. The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are frequently used to find a good model for determining the number of components K in a bivariate GMM. Compared with the AIC, the BIC has advantages in modeling convergence for large data sets [44]. Therefore, the BIC is adopted in this paper to determine the number of components in the GMM.

The BIC, which was first proposed by Schwartz [45], uses a Bayesian framework for model selection to establish a model with the maximum posterior probability or maximum likelihood under certain conditions. This approach has been widely used in audio recognition and image processing and is characterized by high efficiency, high accuracy, and strong robustness [46]. The BIC can be used to determine the number of components in a GMM and is usually used in combination with the expectation maximization algorithm. The mathematical expression of the BIC is as follows:

$$BIC = k\ln(n) - 2\ln(L(\theta; x, y)) \tag{4}$$

where the positive integer $k$ denotes the number of components in the GMM, $n$ denotes the size of the sample data set, and $L(\theta; x, y)$ is the likelihood function of the GMM. As $k$ increases from 1 to a large number, the best GMM is obtained when the BIC value is minimized. The equation is written in the following form.

$$k = \min_{k}(BIC(k)) \tag{5}$$

## 4. Results

### 4.1. Temporal Mobility Pattern of Shared E-Bikes

As the raw GPS trajectory points are disordered, the raw trajectory data are cleaned first by the MCHLC (multirule-constrained homomorphic linear clustering) algorithm [47]. After data cleaning, 143,910 trips were obtained. The OD pairs of each trip were used to analyze citizens' travel behaviors. Figure 3A presents the aggregated hourly rhythm of shared e-bike usage. The variation in hourly rhythm indicates that shared e-bike travel is related to daily behaviors on different days of the week. There is a similar variational pattern from Monday to Friday and a distinct pattern on weekends.
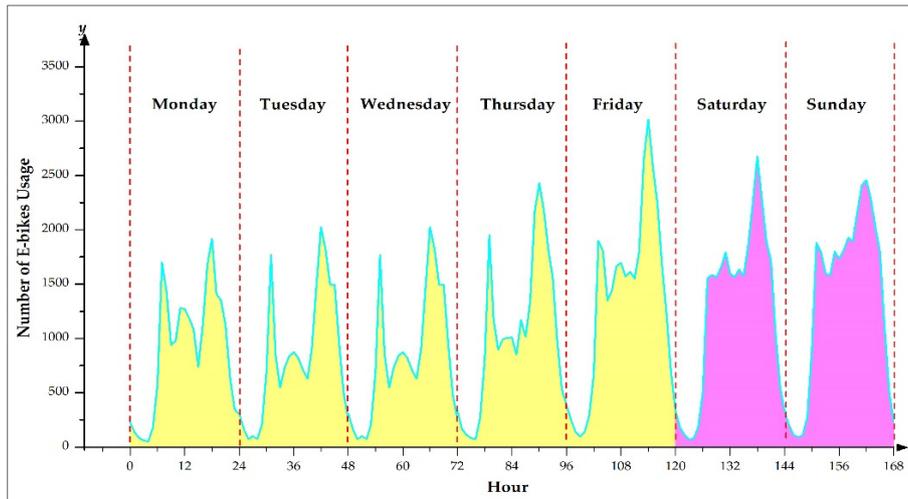
**(A)**
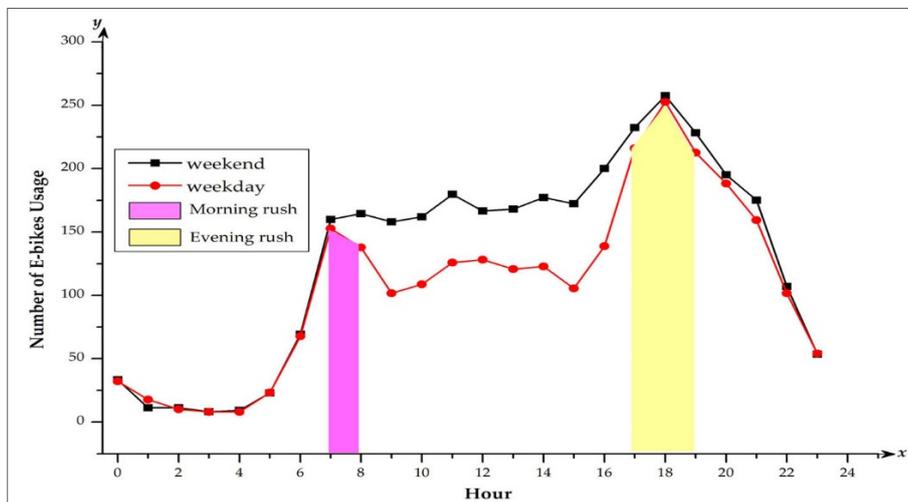


**(B)**

**Figure 3.** Hourly distribution of e-bikes daily usage. (**A**) Hourly distribution of e-bike daily usage from Monday to Sunday. (**B**) Hourly distribution of e-bike average daily usage on weekdays and weekends.

On weekdays, there are usually three peak periods in a day, and they occur in the morning, evening, and at noon. Among the three peaks, the evening peak is highest, morning peak second, and midday peak lowest. Dong, 2018 noted a similar result that there were three obvious peaks (morning peak, midday peak, and evening peak) of e-bike usage in a day in Zhongshan City [17]. It is also a characteristic of the travel behavior of small- and medium-sized cities in China. Although a similar pattern is observed on weekdays, the total usage on Friday (especially after the evening rush hours) approaches that on the weekend. Compared with the weekdays, an obvious evening peak only occurs on weekends. E-bike usage fluctuates slightly from 7:00 a.m. to 3:00 p.m., and a small peak occurs at noon (Figure 3B).

There are obvious discrepancies in the travel patterns on weekdays and weekends. The travel pattern on weekdays suggests that the morning and evening rush periods for e-bikes in the central area of Tengzhou City are 7:00–8:00 a.m. and 17:00–19:00 p.m., when people are highly mobile to commute. Therefore, the trips during rush hours on weekdays are mainly home-to-work journeys. Therefore, the O-D point of trips during rush hours on weekdays is strongly associated with residential and workplace areas, which supports a subsequent investigation of the correlation between commuting activities and residential areas.

### 4.2. Hot Spot Detection Based on the Gaussian Mixture Model

The travel pattern of shared e-bikes on weekdays shows that trips during morning and evening rushes are concentrated mainly on job–home journeys. This finding suggests that the origin point of trips during morning rush hour and the destination point of trips during evening rush hour are highly correlated with residential areas. To measure the correlation between commuting activity and residential area, the origin points of trips during morning rush hour on weekdays are selected.

The GMM is adopted to detect activity hot spots because the multiple components in the model have a nonuniform distribution. To detect activity hot spots with the GMM, the number of Gaussian components $K$ is first determined based on the BIC. To obtain the best GMM, a series of BIC values are calculated for values of $K$ varying from 1 to 100. Figure 4 shows that the best model is composed of 32 Gaussian components, that is, the value of $K$ corresponding to the minimum BIC value is 32.
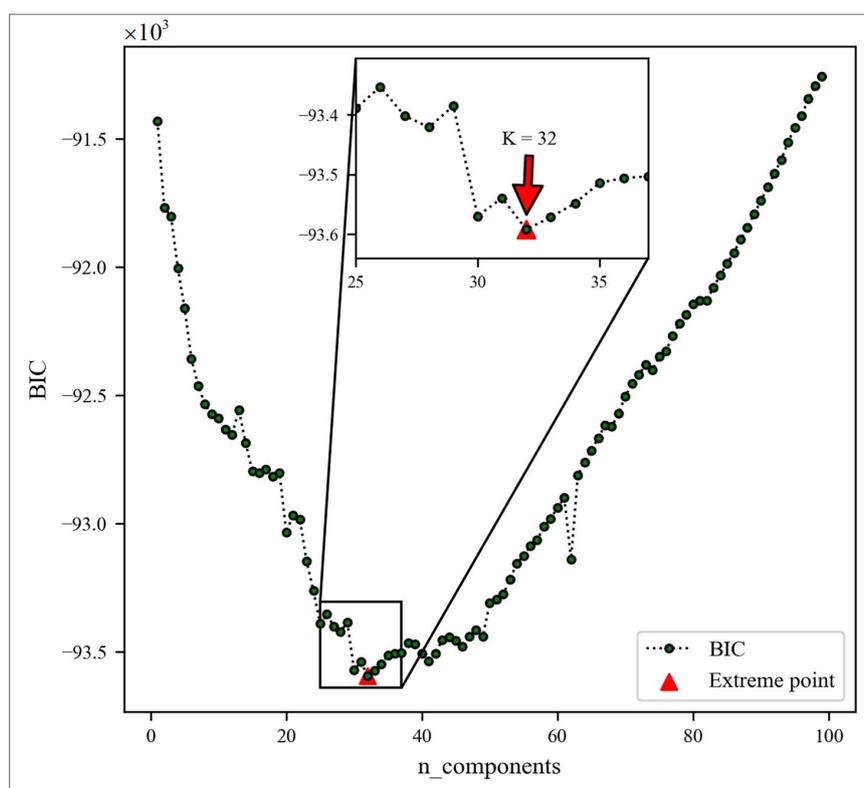


**Figure 4.** Variation curve between the components and the Bayesian information criterion (BIC) value. The *x*-axis is the number of Gaussian components, labeled n_components. The *y*-axis is the BIC value corresponding to different components.

The distribution of origin points (green dots in Part A in Figure 5) in the morning rush hour period is simulated by the GMM with the best number of components at $K = 32$. The points with nonuniform densities can be well depicted by the GMM with different component distributions (see Part B in Figure 5). The nonuniform density distribution implies that the activity degrees at different locations vary with the uneven population distribution. The probability of the GMM denotes the activity intensity, and the shape of each component reflects the dispersion degree, center position, and maximum activity intensity of each activity hot spot. Thirty-two Gaussian components represent the regions in which activities are locally concentrated during morning rush hours, and the peak value of each component reflects the degree of activity intensity in the local region. All the activity hot spots in the study area are divided into three levels based on the average index (*Ave* = 0.067) and average deviation index (*Ave_Dev* = 0.037) of the peak value distribution. Level 1 denotes the activity intensity of hot spots larger than the average index, indicating a large traffic flow for shared e-bikes in this

region. Level 2 denotes an activity intensity value between the average index and average deviation index values, indicating a typical traffic flow in the region. Level 3 denotes an activity intensity lower than the average deviation index value, indicating that few people choose shared e-bikes in the region.
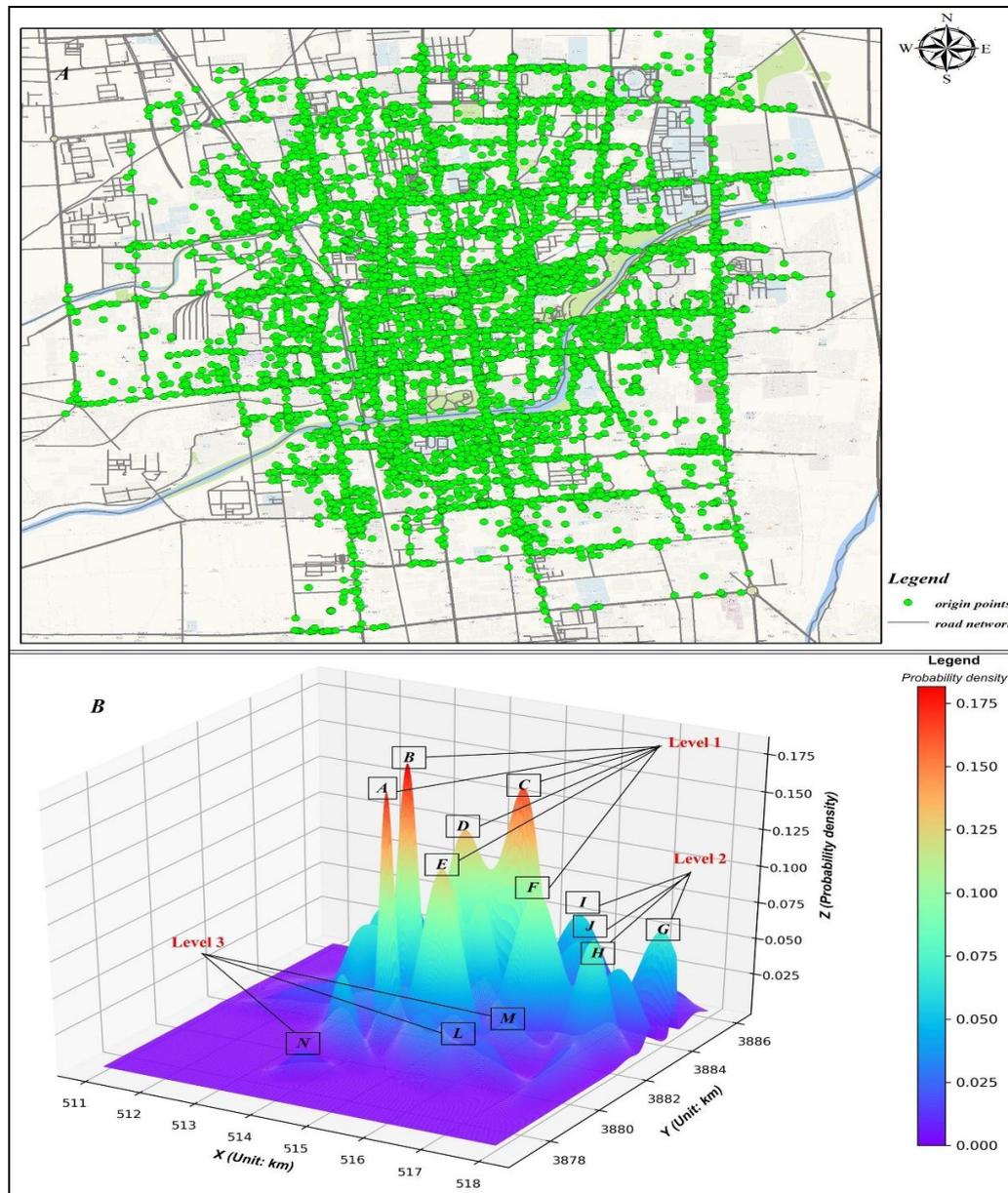


**Figure 5.** Distribution of the origin points during morning rush hours. Part **A** shows the discrete distribution of the origin points of trips in the morning rush hour period. Part **B** is the simulation results of the Gaussian mixture model (GMM). *A* to *N* denote Gaussian components with different shapes, reflecting the nonuniform point density.

The standard deviation indicator was applied to delimit the urban hotspot scope [48]. Notably, Borruso applied three standard deviation units to delimit the CBDs (central business districts) of two midsize urban areas in northeastern Italy [49]. To further examine the hotspot "cores," one standard deviation and two standard deviations were computed for the hotspot centrality. The result suggests that one standard deviation unit is the best indicator for hotspot scope delineation in our study. The hotspot "core" information, as an expression of human activity, is used to construct the model of the relationship between human activity and functional zone.

The activity hot spots at different levels are displayed in Figure 6 and the distribution of activity hot spots at different levels was analyzed with the Standard Error Ellipse function in ArcGIS software. The Level 1 hot spots are mainly distributed at the junction of the Beixin, Jinghe, and Longquan blocks. The flatness value of the error ellipse is 0.35, and the direction angle is 67.87°. The direction angle of the error ellipse indicates that the residents are more active in the east–west direction than in the north–south direction, which is consistent with the distribution of residential areas in the city center (residential areas are scattered along the main east–west roads, including Jiefang Road, Jinghe Road, and Xueyuan Road). Compared with the Level 1 hot spots, the Level 2 hot spots display a more distinct trend, with a flatness value of 0.41 and direction angle of 25.75° for the error ellipse. The Level 2 hot spots are mainly located in the north and northeast parts of the city, especially in the emerging developing regions located in the northern part of the Longquan block. The Level 3 hot spots are located near the periphery of the city. It is noteworthy that the outer periphery of the city is characterized by small rural villages according to the actual Gaode Map. The low activity intensity levels imply that the usage of shared e-bikes is less common in rural regions than in urban regions.
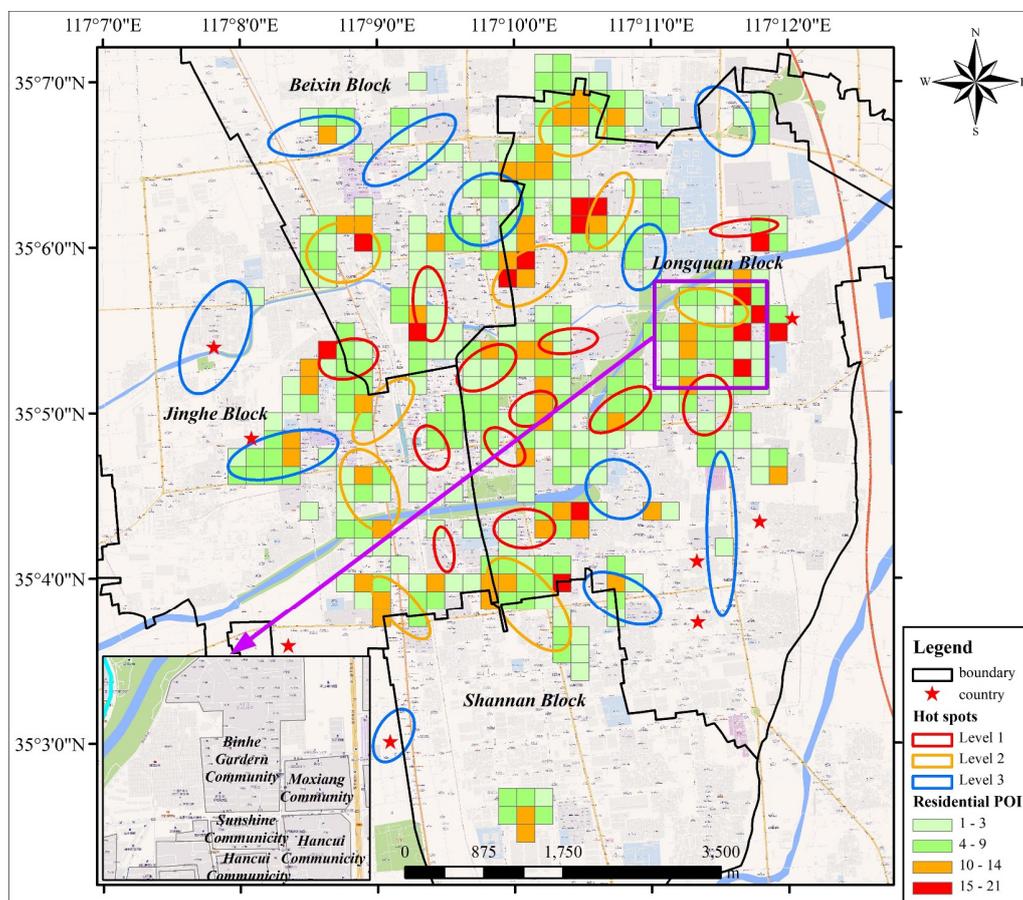


**Figure 6.** Distribution of the points of interest (POIs) data and hot spots at different levels.

### 4.3. CARA Model Construction

To verify the aforementioned hypothesis that the hot spots of origin points during morning rush hours are located in residential zones, POIs obtained from the Baidu Map API (http://lbsyun.baidu.com/) using the Python language are introduced and categorized in accordance with Baidu's internal POI standards (http://lbsyun.baidu.com/index.php?title=lbscloud/poitags). The POI data collected for this study include multiple categories, such as residential zones, shopping malls, transportation facilities, educational facilities, financial institutions, tourist attractions, medical facilities, restaurants, corporate facilities, and community services. These POIs are point features that describe entities including

geolocation coordinates (latitude and longitude coordinates), names, and classes; thus, they not only reflect the basic activities of urban residents (e.g., living, working, commuting, and recreation) but also the function of a region [50]. In total, 16,626 data points were acquired and 2573 POIs were located in residential areas. A coordinate transformation was performed due to the encryption form of the POI data.

The POI data for residential zones were aggregated to an appropriate spatial analysis unit to reveal the hot spot distribution. The spatial analysis unit size ranged from 100 to 500 m with an interval of 50 m. The optimum unit size was found to be 200 m, which is within the range suggested by some urban geographers (200–300 m is suitable for fine-spatial-resolution data in urban centers) [51,52]. The density distribution of residential POIs in the study area is displayed with a regular grid cell (200 m × 200 m) in Figure 6. The density distribution of POIs indicates that residential zones in Tengzhou City are mainly located in the city center at the junction of the four jurisdictional blocks, and these zones extend to the north and east. Considering the distribution of residential POIs, the distribution of hot spots with different intensities verifies the speculation that activities during morning rush hours are concentrated in residential areas.

We further model the activity hot spots (human activities) and the POIs (region function) to investigate the relationship between these factors and verify the feasibility of residential area identification based on the trajectories of shared e-bikes. As for the statistical distribution of POIs, the activity intensity of each spatial analysis unit can be calculated by the GMM. The spatial units with nonzero POI distributions in residential areas are selected for modeling, and these areas are represented by gray dots in Figure 7.
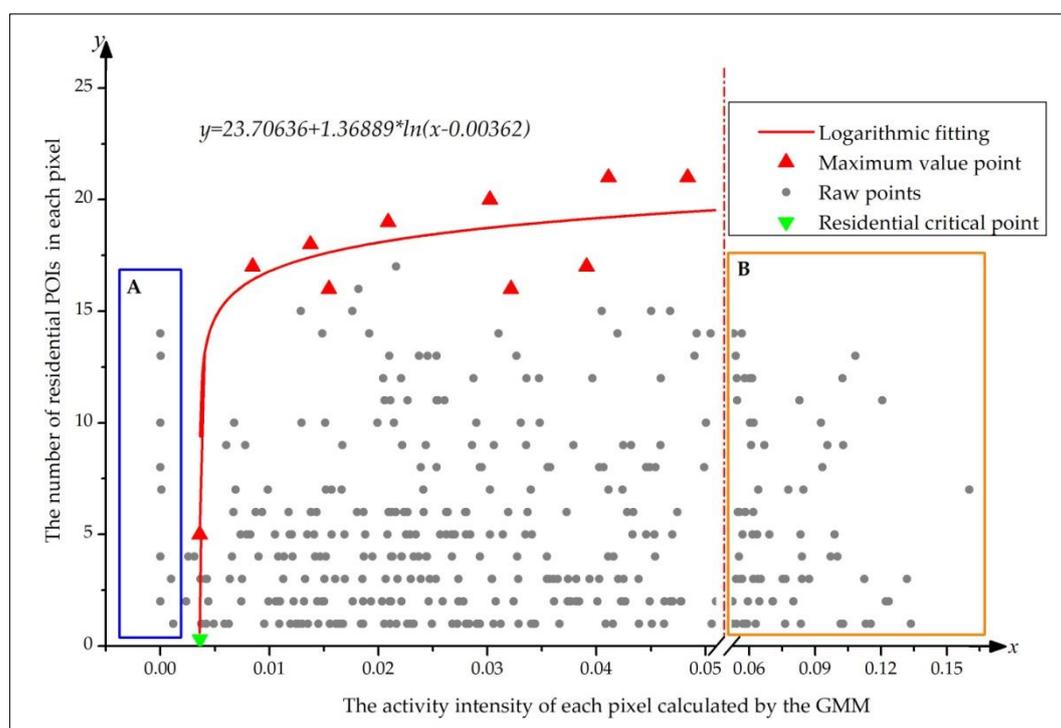


**Figure 7.** Quantitative analysis of POI and activity intensity data. Part **A** shows the pixels with a low activity intensity (close to zero) and large number of POIs. Part **B** highlights the pixels with a high activity intensity and few POIs.

Before modeling, noise points should be filtered and removed to reduce the influence of noise on modeling. The distributions of POIs and spatial grid units corresponding to different activity intensity values are statistically analyzed in Figure 8. The points in Part A of Figure 7 denote the spatial units with activity intensities close to zero (the value lies in the range of 0–0.001), as denoted by the cyan bars in Figure 8. These spatial units are uncommon, accounting for only 2.37% of all units, and the

POIs of these units account for only 3.05% of all POIs. Such spatial grid units are mainly located in the periphery of the city and characterized by low activity intensities and high POI values, suggesting that residents in these areas seldom select e-bikes for travel. Additionally, this finding indicates that shared e-bike users are only a limited portion of the population, which is a common issue for trajectory data. The points in Part B of Figure 7 denote the spatial grid units with activity intensity values in the range of 0.005–0.165, as denoted by the yellow bars in Figure 8. The statistical results indicate that the activity intensity values of these spatial units span large and reach 69.7%; however, these grid units account for only 20.84% of all grid units, and only 20.74% of POIs are in these grid units. Such spatial grid units are mainly located in the downtown area of Tengzhou City, which is characterized by high intensity values but low POI values, indicating the mixed functionality of the downtown area. Therefore, the points in Parts A and B in Figure 7 cannot be concretely used to identify residential areas and are excluded from modeling.
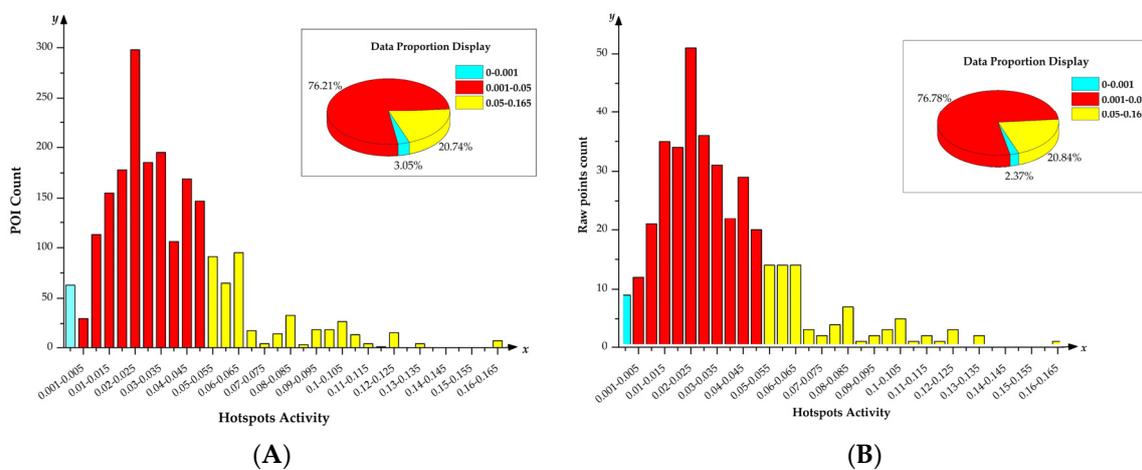


**Figure 8.** Statistical results for hotspot activities in different interval values. (**A**) Statistical results of POI count for different hotspot activity values. (**B**) Statistical results of the raw pixel count for different hotspot activity values. Note: The pie chart shows the proportions of POI data and raw pixels corresponding to hotspot activity value in three different ranges.

The points with activity intensity values between 0.001 and 0.005 were used for modeling. Figure 7 shows that the activity intensities of the trajectory and POI distributions exhibit a many-to-one functional relationship because of the uncertainty in human activities. Thus, the activity intensity observations need to be normalized. To explore the maximum impact of the activity intensity on residential identification, a maximization process method similar to the MVC (maximum value composite) approach of remote sensing is adopted. That is, according to the activity intensity value, the spatial grid units were divided into many groups with an interval of 0.005. Then, the spatial grid unit with the maximum intensity value in each interval was selected to model, as denoted by the red triangles in Figure 7. The CARA model is expressed by a logarithmic function.

$$\mathbf{RA}_{cnt} = 23.7036 + 1.36889 * ln(\mathbf{CA} - 0.00362) \qquad (6)$$

where **CA** is the intensity value of the commuting activity of each spatial grid unit, which can be obtained from the hot spot distribution after GMM processing, and $\mathbf{RA}_{cnt}$ is the number of residential POIs in the corresponding grid.

The R-squared value is 0.876, indicating that the logarithmic model can effectively reflect the correlation between human activity and residential functional areas with high accuracy. It is noteworthy that the boundary of the residential area can also be delineated based on this specific logarithmic model. If the activity intensity value of a spatial unit is lower than 0.0036 (the green triangle in Figure 7), then the probability of the grid unit being classified as a residential area is negligible.

Based on the CARA model, the residential areas of Tengzhou City can be re-delineated, as shown in Figure 9. The result is consistent with the urban planning of Tengzhou City and further corroborates the aforementioned speculation that citizens start their daily activities in residential areas during the morning rush hour period. In addition, the rural villages located at the periphery of the city can be identified, even if the density of the POI data is low.
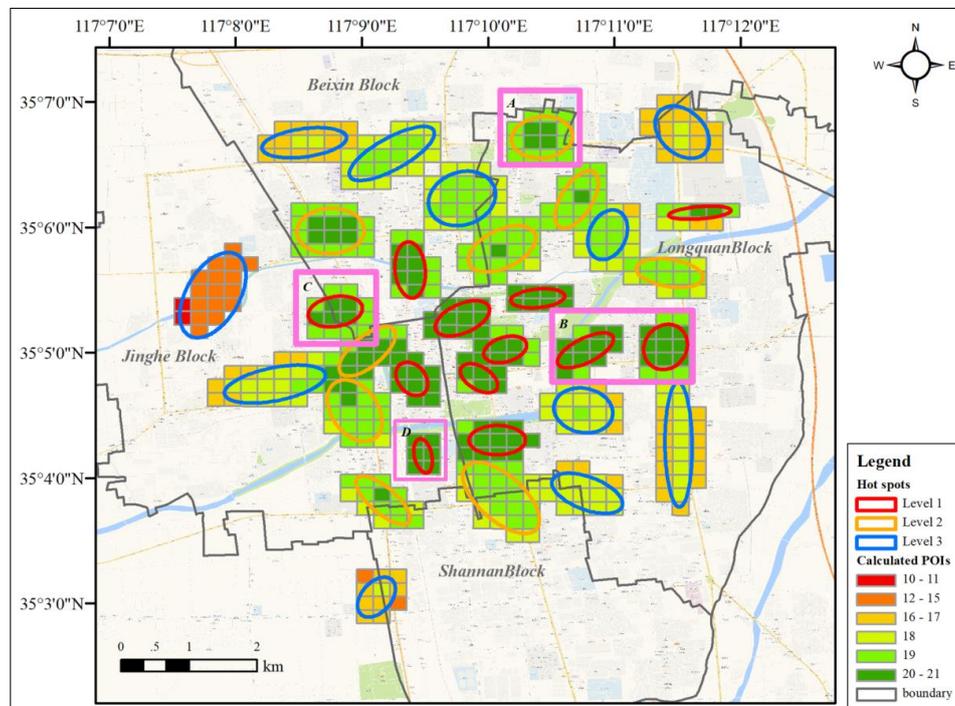


**Figure 9.** Delineated results of residential areas based on the Commuting Activity and Residential Area (CARA) model.

## 5. Discussion

### 5.1. Delineated Result of Urban Residential Areas and Evaluation

By a comparison with Baidu Map information, each grid function can be obtained through the geographic coordinate inverse calculation module. As a grid unit may cover two different land use parcels, a score is used to evaluate the relevance of each grid to residential area. The score of each grid is calculated by the area index $A_{index}$. The symbol $Area_{insection}$ denotes the intersection area between the grid unit of the delineated results and the actual residential area. The symbol $Area_{grid}$ denotes the area of each grid unit. The symbol $A_{index}$ denotes the relevance of the grid to the residential areas. If $A_{index}$ is less than 0.5, then the grid is considered independent of the residential areas; otherwise, the grid is regarded as part of a residential area. A score for each grid can be obtained from the following criteria.

$$A_{index} = Area_{insection} / Area_{grid} \tag{7}$$

$$Score_i = \begin{cases} 1 & A_{index_i} \geq 0.5, i = 1, 2, \cdots, n \\ 0 & A_{index_i} < 0.5, i = 1, 2, \cdots, n \end{cases} \tag{8}$$

$$Precision = \left( \sum_{i=1}^{n} Score_i / n \right) \times 100\% \tag{9}$$

where $n$ is the number of grid units in the identified result. The precision is the relevance of the delineated result to the residential areas and calculated by the score of each grid unit.

Compared with the actual map, four typical regions in Figure 9 are selected to further clarify the reliability of the results. These regions are located to the north, east, west, and south of the downtown area of the city. Part A lies in the emerging urban area in the northern part of the city; this area includes three large communities: Apple Garden, Tongsheng Garden, and Champion House. The identification result in Part A includes 15 grid units, twelve of which are highly associated with the aforementioned communities. The precision index reaches 80%.

Part B is a cluster of modern communities located in the eastern part of the city, including the Voyage First International Community, Hancui Garden, Moxiang Holy House, and Huilong Harmony Garden. Residents in these communities prefer to travel by shared e-bikes, thus forming high-intensity hot spots. All the grid units are associated with residential areas.

Part C is another community cluster located west of the downtown area. Here, communities with different scales are built next to each other, and the largest are Jinkongfu Community, Western Song Community, and Geological Home. These communities are not along the major road in the area (Parallel Avenue), although the trajectories are mainly along this road. Thus, the precision index is low at only 78.6%.

Unlike Part C, Part D is a community cluster located along Datong Avenue, and it is composed of the different districts of Venice Community. Hence, the results obtained with the trajectory along the road are highly consistent with the actual scenario.

Based on the CARA model, the delineated residential areas from trajectory data encompass 480 grid units, including the villages at the periphery of the city. According to the actual map and the score index, the precision of the delineation of residential areas reaches 83.3%, which suggests that the proposed model is feasible.

## 5.2. Influencing Factors for the CARA Model

Commuting behaviors usually occur during the morning and evening rush hours, suggesting that the attractiveness of residential areas and workplaces to residents have distinct time characteristics. Therefore, the time characteristic is a factor that needs to be considered when building the CARA model. To further investigate the influence of the time factor on the CARA model, two additional data sets are used and their relevance for residential areas is compared. One is the set of destination points of trajectories in evening rush hours on weekdays; and the other is the mixture set of original points of trajectories in evening rush hours and destination points of trajectories in evening rush hours on weekdays. Notably, the weekdays for evening rush hours do not include Friday, because the usage of shared e-bikes during the Friday evening peak is close to that on weekends. The hot spots results for the two data sets are shown in Figure 10, in which Part A is the result for the evening rush data, and Part B is the result based on the morning–evening rush data.

Compared with the result for the morning peak period, similar distributions of different-level hot spots are observed in the evening peak and morning–evening peak periods. However, the number of hot spots with high activity intensity increases significantly with the increasing usage of shared e-bikes. When the usage of shared e-bikes increases, the number of hot spots (especially hot spots with high activity intensity) increases significantly (Table 1). This finding indicates that the hot spots are related to the volume of trajectory data.

In Figure 10, some hot spots are not associated with residential areas, suggesting that residents in Tengzhou ride shared e-bikes to different destinations in the evening peak period. To verify this finding, the functions of hot spots are investigated based on actual maps. As shown in Figure 11, compared with the Baidu Map information, although most hot spots in the evening peak period are associated with residential areas (the green circles), some hot spots are associated with shopping malls (the red circles) and entertainment venues (the pink circles). This finding suggests that residents with available time enjoy riding shared e-bikes for relaxation; in addition, based on the morning–evening peak trajectory data, as shared e-bike usage in the evening peak period is dominant (the trajectory volume in the evening peak period is approximately 1.5 times that in the morning peak period), the hot spot results are more similar to those in the evening peak period.
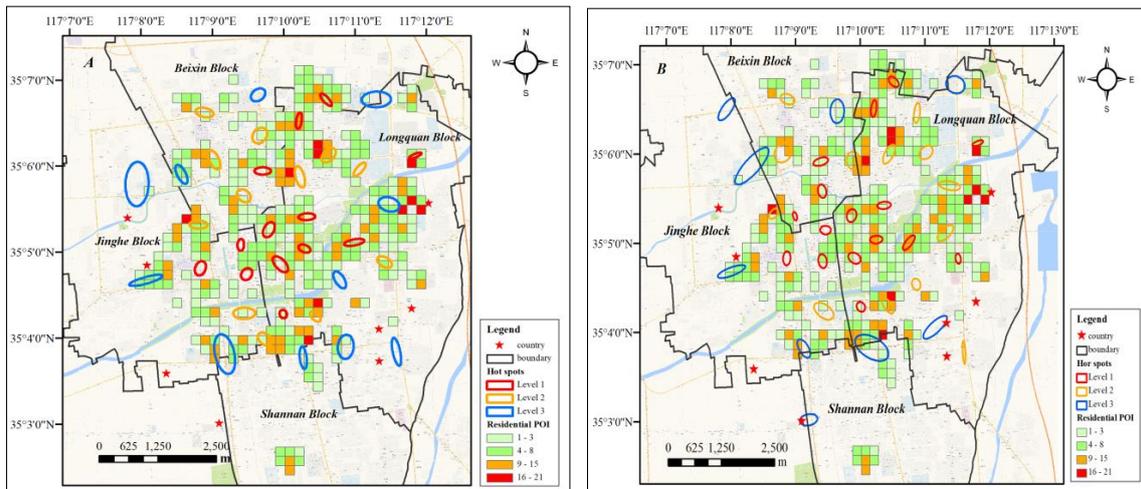
**Figure 10.** Hot spot results for different data sets. (**A**) Detected result of hot spots for the data set in the evening rush period. (**B**) Detected result of hot spots for the data set in the morning–evening rush period.
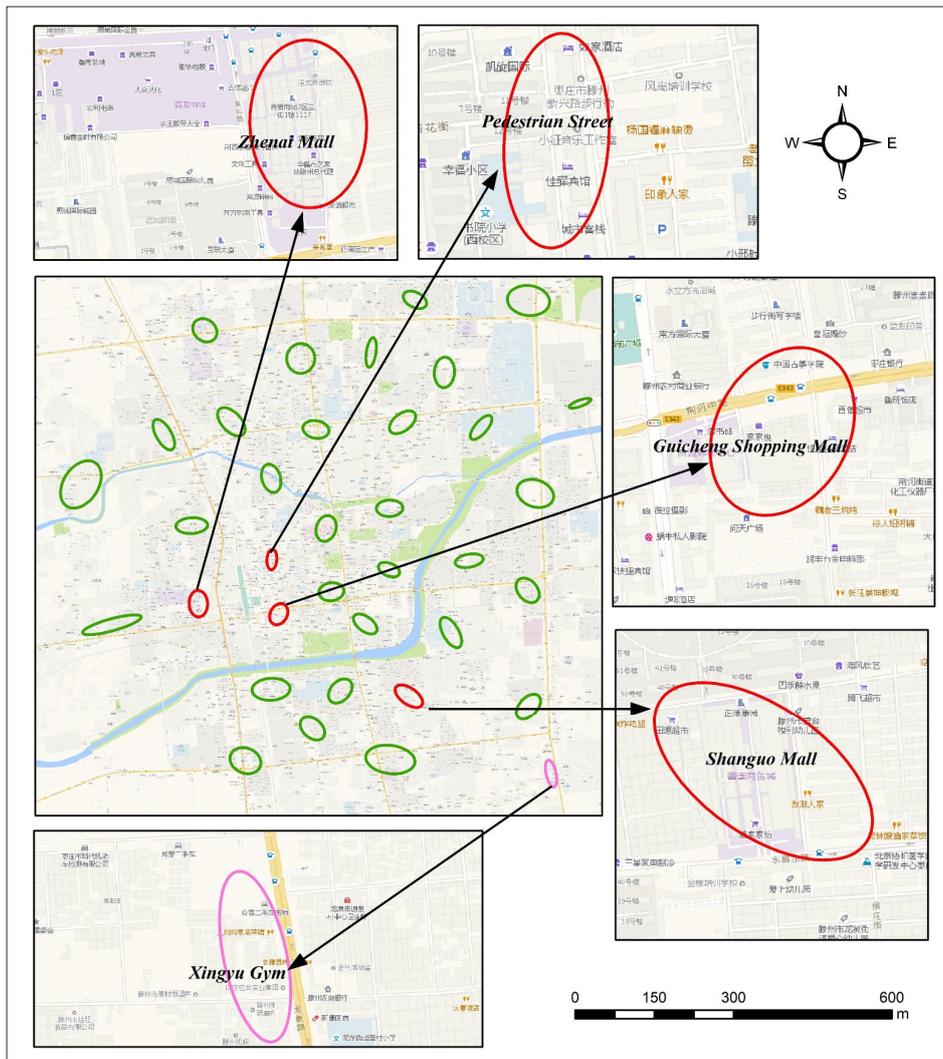


**Figure 11.** Diverse locations of hot spots in the evening peak period. The green circles indicate the hot spots associated with residential areas, the red circles indicate the hot spots associated with shopping malls, and the prink circle indicate the hot spots associated with entertainment.

We further investigate the functions of hot spots through the geographical coordinate inverse calculation module of Baidu Map. Here, the function types of hot spots include residential areas, shopping malls, entertainment venues, and villages. The statistical results in Table 1 show that 90.6% of the hot spots in the morning peak period are associated with residential areas, while only 77.8% of residential areas are identified as hot spots during the evening peak period. The result suggests that the attractiveness of residential area to human activities is strongest during the morning rush hours. The diversity of human travel during the evening peak period interferes with the construction of the CARA model. Therefore, compared with the data volume, the data time characteristic has a greater influence on model construction, because it may result in the heterogeneity of human activities.

**Table 1.** Statistics of hot spot types for different data sets.

| Hot spots | Morning | Evening | Morning and Evening |
|---|---|---|---|
| Residential areas | 29 | 28 | 30 |
| Shopping malls | 0 | 4 | 3 |
| Entertainment venues | 0 | 1 | 1 |
| Villages | 3 | 3 | 4 |
| Residential relevance (%) | 90.6 | 77.8 | 78.9 |

Note: Morning, Evening, and Morning and Evening denote the trajectory data for the morning peak, evening peak, and combined morning–evening peak periods, respectively.

## 6. Conclusions

Understanding the relationship between human activities and functional areas is key to identifying urban functional areas. Trajectories imply rich interactive information of human-region, which is the basis for urban functional area identification. In this paper, the CARA model is proposed to quantify the correlation between commuting activities and residential areas. The residential areas are retrieved from the shared e-bike trajectory data to verify the quantitative model. The main conclusions are as follows.

(1) The GMM combined with BIC can accurately detect hot spots, including weak hot spots, when the density of trajectory data is heterogeneous. Based on the "activity degree" indicator, the hot spots can be classified into different levels.

(2) The CARA model is built to quantify the correlation between human commuting activity and residential area, and the $R^2$ coefficient reaches 0.876. Based on the model, residential areas can be delineated with high precision (83.3%). The result validates the feasibility and reliability of the CARA model.

(3) Although human activities have distinct temporal characteristics, different human activities may occur at the same time. The diversity of human activities results in interference when modeling the correlation between certain human activities and functional areas. The attractiveness of residential area to human activities is strongest during the morning rush hours on weekdays. The origins of e-bike trajectory data during morning rush period are most strongly correlated with urban residential areas.

(4) Compared with the amount of data, the data time characteristic can affect the homogeneity of human activities and has a more obvious influence on the model of correlation between human activities and functional areas.

Other functional areas, such as workplace and entertainment areas, can also be obtained from the shared e-bike trajectory based on the concept provided in this study. In our study, each functional region is regarded as a region with only one social function; thus, the mixed functional areas identification cannot be identified. Developing a method of identifying mixed functional areas is the focus of our future work.

## References

1. Yuan, J.; Zheng, Y.; Xie, X. Discovering Regions of Different Functions in a City Using Human Mobility and POIs. In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, Beijing, China, 12–16 August 2012.
2. Liu, X.; Kang, C.; Gong, L.; Liu, Y. Incorporating spatial interaction patterns in classifying and understanding urban land use. *Int. J. Geogr. Inf. Sci.* **2016**, *30*, 334–350. [CrossRef]
3. Pan, G.; Qi, G.; Zhang, W.; Li, S.; Wu, Z.; Yang, L.T. Trace analysis and mining for smart cities:Issues, methods, and application. *IEEE Commun. Mag.* **2013**, *51*, 120–126. [CrossRef]
4. Mou, N.; Zhang, H.; Chen, J.; Zhang, L.; Dai, H. A Review on the Application Research of Trajectory Data Mining in Urban Cities. *J. Geo-Inf. Sci.* **2015**, *17*, 1136–1142. [CrossRef]
5. Reades, J.; Calabrese, F.; Ratti, C. Eigenplaces: Analysing cities using the space-time structure of the mobile phone network. *Environ. Plan. B Plan. Des.* **2009**, *36*, 824–836. [CrossRef]
6. Calabrese, F.; Reades, J.; Reades, J. Eigenplaces: Segmenting Space through Digital Signatures. *IEEE Pervasive Comput.* **2010**, *9*, 78–84. [CrossRef]
7. Liu, Y.; Wang, F.; Xiao, Y.; Gao, S. Urban land uses and traffic 'source-sink areas': Evidence from GPS-enabled taxi data in Shanghai. *Landsc. Urban Plan.* **2012**, *106*, 73–87. [CrossRef]
8. Pan, G.; Qi, G.; Wu, Z.; Zhang, D.; Li, S. Land-Use Classification Using Taxi GPS Trace. *IEEE Trans. Intell. Transp. Syst.* **2013**, *14*, 113–123. [CrossRef]
9. Xu, H.; Ying, J. Recognizing Social Function of Urban Regions by Using Data of Public Bicycle Systems. *Chin. J. Electron.* **2019**, *28*, 13–20. [CrossRef]
10. Liu, Y.; Gao, S.; Gong, L.; Zhi, Y. Social Sensing:A New Approach to Understanding Our Socioeconomic Environments. *Ann. Assoc. Am. Geogr.* **2015**, *105*, 512–530. [CrossRef]
11. Zhou, T.; Liu, X.; Qian, Z.; Chen, H.; Tao, F. Automatic Identification of the Social Functions of Areas of Interest (AOIs) Using the Standard Hour-Day-Spectrum Approach. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 7. [CrossRef]
12. Inagaki, T.; Mimura, Y.; Ando, R. An Analysis on Excursion Characteristics of Electric Assist Bicycles by Travel Behavioral Comparison Based on Trajectory Data. In Proceedings of the 2012 15th International IEEE Conference on Intelligent Transportation Systems, ITSC 2012, Anchorage, AK, USA, 16–19 September 2012; pp. 433–437.
13. Plazier, P.A.; Weitkamp, G.; van den Berg, A.E. "Cycling was never so easy!" An analysis of e-bike commuters' motives, travel behaviour and experiences using GPS-tracking and interviews. *J. Transp. Geogr.* **2017**, *65*, 25–34. [CrossRef]
14. Lopez, A.J.; Astegiano, P.; Gautama, S.; Ochoa, D.; Tampère, C.M.J.; Beckx, C. Unveiling E-Bike Potential for Commuting Trips from GPS Traces. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 190. [CrossRef]
15. Campbell, A.A.; Cherry, C.R.; Ryerson, M.S.; Yang, X. Factors influencing the choice of shared bicycles and shared electric bikes in Beijing. *Transp. Res. Part C Emerg. Technol.* **2016**, *67*, 399–414. [CrossRef]

16. Edge, S.; Dean, J.; Cuomo, M.; Keshav, S. Exploring e-bikes as a mode of sustainable transport: A temporal qualitative study of the perspectives of a sample of novice riders in a Canadian city. *Can. Geogr.* **2018**, *62*, 384–397. [CrossRef]

17. Meixuan, D. Factors Affecting E-bike Mode Choice in a Medium-sized Chinese City. *Am. J. Transp. Logist.* **2018**, *1*, 1–19. [CrossRef]

18. Cherry, C.R.; Weinert, J.X.; Xinmiao, Y. Comparative environmental impacts of electric bikes in China. *Transp. Res. Part D* **2009**, *14*, 281–290. [CrossRef]

19. Huang, H.; Li, Q.; Zhang, Y. Urban Residential Land Suitability Analysis Combining Remote Sensing and Social Sensing Data: A Case Study in Beijing, China. *Sustainability* **2019**, *11*, 2255. [CrossRef]

20. Antipova, A.; Wang, F.; Wilmot, C. Urban land uses, socio-demographic attributes and commuting: A multilevel modeling approach. *Appl. Geogr.* **2011**, *31*, 1010–1018. [CrossRef]

21. Yue, Y.; Lan, T.; Yeh, A.G.O.; Li, Q.-Q. Zooming into individuals to understand the collective: A review of trajectory-based travel behaviour studies. *Travel Behav. Soc.* **2014**, *1*, 69–78. [CrossRef]

22. Maat, K.; Wee, B.V.; Stead, D. Land use and travel behaviour: Expected effects from the perspective of utility theory and activity-based theories. *Environ. Plan. B Plan. Des.* **2005**, *32*, 33–46. [CrossRef]

23. Andrade, R.; Alves, A.; Bento, C. POI Mining for Land Use Classification: A Case Study. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 493. [CrossRef]

24. Jiang, S.; Alves, A.; Rodrigues, F.; Ferreira, J., Jr.; Pereira, F.C. Mining point-of-interest data from social networks for urban land classification and disaggregation. *Comput. Environ. Urban Syst.* **2015**, *53*, 36–46. [CrossRef]

25. Zhang, H.; Wang, R.; Chen, B.; Hou, Y.; Qu, D. Dynamic Identification of Urban Functional Areas and Visual Analysis of Time-varying Patterns Based on Trajectory Data and POIs. *J. Comput.-Aided Des. Comput. Graph.* **2018**, *30*, 1728–1740. [CrossRef]

26. Spaccapietra, S.; Parent, C.; Damiani, M.L.; Macedo, J.A.D.; Porto, F.; Vangenot, C. A conceptual view on trajectories. *Data Knowl. Eng.* **2008**, *65*, 126–146. [CrossRef]

27. Mazimpaka, J.; Timpf, S. Exploring the Potential of Combining Taxi GPS and Flickr Data for Discovering Functional Regions. In *AGILE 2015*; Springer International Publishing: Cham, Switzerland, 2015. [CrossRef]

28. Qian, Z.; Liu, X.; Tao, F.; Zhou, T. Identification of Urban Functional Areas by Coupling Satellite Images and Taxi GPS Trajectories. *Remote Sens.* **2020**, *12*, 2449. [CrossRef]

29. Yuan, N.J.; Zheng, Y.; Xie, X.; Wang, Y.; Zheng, K.; Xiong, H. Discovering Urban Functional Zones Using Latent Activity Trajectories. *IEEE Trans. Knowl. Data Eng.* **2015**, *27*, 712–725. [CrossRef]

30. Ma, Y.; Liu, S.; Xue, G.; Gong, D. Soft Sensor with Deep Learning for Functional Region Detection in Urban Environments. *Senors* **2020**, *20*, 3348. [CrossRef]

31. Gao, Q.; Fu, J.; Yu, Y.; Tang, X. Identification of urban regions'functions in Chengdu, China, based on vehicle trajectory data. *PLoS ONE* **2019**, *14*, 1–17. [CrossRef]

32. Pei, T.; Sobolevsky, S.; Shaw, S.-L.; Zhou, C. A new insight into land use classification based on aggregated mobile phone data. *Int. J. Geogr. Inf. Sci.* **2014**, *28*, 1988–2007. [CrossRef]

33. Boarnet, M.; Crane, R. The influence of land use on travel behavior: Specification and estimation strategies. *Transp. Res. Part A Policy Pract.* **2001**, *35*, 823–845. [CrossRef]

34. Gao, S.; Janowicz, K.; Couclelis, H. Extracting urban functional regions from points of interest and human activities on location-based social networks. *Trans. GIS* **2017**, *21*, 446–467. [CrossRef]

35. Terroso-Saenz, F.; Muñoz, A. Land use discovery based on Volunteer Geographic Information classification. *Expert Syst. Appl.* **2020**, *140*, 112892–112906. [CrossRef]

36. Hong, J.; Shen, Q.; Zhang, L. How do built-environment factors affect travel behavior? A spatial analysis at different geographic scales. *Transportation* **2014**, *41*, 419–440. [CrossRef]

37. Handy, S.; Cao, X.; Mokhtarian, P.L. Self-Selection in the Relationship between the Built Environment and Walking. *J. Am. Plan. Assoc.* **2006**, *72*, 55–74. [CrossRef]

38. Faghih-Imani, A.; Eluru, N.; El-Geneidy, A.M.; Rabbat, M.; Haq, U. How land-use and urban form impact bicycle flows: Evidence from the bicycle-sharing system (BIXI) in Montreal. *J. Transp. Geogr.* **2014**, *41*, 306–314. [CrossRef]

39. El-Ass, W.; Mahmoud, M.S.; Habib, K.N. Effects of built environment and weather on bike sharing demand: A station level analysis of commercial bike sharing in Toronto. *Transportation* **2015**, *44*, 589–613. [CrossRef]

40. Fishman, E.; Wei, H. Bikeshare: A Review of Recent Literature. *Urban Transp. China* **2016**, *36*, 92–113. [CrossRef]

41. Ji, Z.; Huang, Y.; Xia, Y.; Zheng, Y. A robust modified Gaussian mixture model with rough set for image segmentation. *Neurocomputing* **2017**, *266*, 550–565. [CrossRef]

42. Shi, X.; Li, Y.; Zhao, Q. Flexible Hierarchical Gaussian Mixture Model for High-Resolution Remote Sensing Image Segmentation. *Remote Sens.* **2020**, *12*, 1219. [CrossRef]

43. Huang, Z.; Chau, K. A new image thresholding method based on Gaussian mixture model. *Appl. Math. Comput.* **2008**, *205*, 899–907. [CrossRef]

44. Mehrjou, A.; Hosseini, R.; Araabi, B.N. Improved Bayesian information criterion for mixture model selection. *Pattern Recognit. Lett.* **2016**, *69*, 22–27. [CrossRef]

45. Schwarz, G.E. Estimating the Dimension of a Model. *Ann. Stat.* **1978**, *6*, 461–464. [CrossRef]

46. Chen, G.; Zhang, Y.; Liang, D. A novel method for image segmentation based on the BIC. *J. Liaoning Tech. Univ. (Nat. Sci.)* **2016**, *35*, 1359–1362. [CrossRef]

47. Cheng, X.; Li, C.; Du, W.; Shen, J.; Dai, Z. Trip Extraction of Shared Electric Bikes Based on Multi-Rule-Constrained Homomorphic Linear Clustering Algorithm. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 526. [CrossRef]

48. Chainey, S.; Reid, S.; Stuart, N. When is a hotspot a hotspot? A procedure for creating statistically robust hotspot maps of crime. In *Socio-Economic Applications of Geographic Information Science*; Kidner, D., Higgs, G., White, S., Eds.; Taylor and Francis: London, UK, 2002; pp. 22–36.

49. Borruso, G.; Porceddu, A. A Tale of Two Cities: Density Analysis of CBD on Two Midsize Urban Areas in Northeastern Italy. In *Geocomputation and Urban Planning*; Murgante, B., Borruso, G., Lapucci, A., Eds.; Springer: Berlin/Heidelberg, Germany, 2009; pp. 37–56. [CrossRef]

50. He, Q.; He, W.; Song, Y.; Wu, J.; Yin, C.; Mou, Y. The impact of urban growth patterns on urban vitality in newly built-up areas based on an association rules analysis using geographical 'big data'. *Land Use Policy* **2018**, *78*, 726–738. [CrossRef]

51. Xia, Z.; Li, H.; Chen, Y.; Liao, W. Identify and Delimitate Urban Hotspot Areas Using a Network-Based Spatiotemporal Field Clustering Method. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 344. [CrossRef]

52. Maantay, J.A.; Maroko, A.R.; Herrmann, C. Mapping Population Distribution in the Urban Environment: The Cadastral-based Expert Dasymetric System (CEDS). *Cartogr. Geogr. Inf. Sci.* **2007**, *34*, 77–102. [CrossRef]