remote sensing

MDPI

Article

Sensing Travel Source–Sink Spatiotemporal Ranges Using Dockless Bicycle Trajectory via Density-Based Adaptive Clustering

Yan Shi [1], Da Wang [1], Xiaolong Wang [1], Bingrong Chen [1], Chen Ding [1] and Shijuan Gao [1,2,*]

[1] Department of Geo-Informatics, Central South University, Changsha 410006, China; csu_shiy@csu.edu.cn (Y.S.); 215001023@csu.edu.cn (D.W.); 8211180227@csu.edu.cn (X.W.); brchen@csu.edu.cn (B.C.); jagndc@csu.edu.cn (C.D.)

[2] Information and Network Center, Central South University, Changsha 410006, China

* Correspondence: gaoshijuan@csu.edu.cn

Abstract: The travel source–sink phenomenon is a typical urban traffic anomaly that reflects the imbalanced dissipation and aggregation of human mobility activities. It is useful for pertinently balancing urban facilities and optimizing urban structures to accurately sense the spatiotemporal ranges of travel source–sinks, such as for public transportation station optimization, sharing resource configurations, or stampede precautions among moving crowds. Unlike remote sensing using visual features, it is challenging to sense imbalanced and arbitrarily shaped source–sink areas using human mobility trajectories. This paper proposes a density-based adaptive clustering method to identify the spatiotemporal ranges of travel source–sink patterns. Firstly, a spatiotemporal field is utilized to construct a stable neighborhood of origin and destination points. Then, binary spatiotemporal statistical hypothesis tests are proposed to identify the source and sink core points. Finally, a density-based expansion strategy is employed to detect the spatial areas and temporal durations of sources and sinks. The experiments conducted using bicycle trajectory data in Shanghai show that the proposed method can accurately extract significantly imbalanced dissipation and aggregation events. The travel source–sink patterns detected by the proposed method have practical reference, meaning that they can provide useful insights into the redistribution of bike-sharing and station resources.

Keywords: urban remote sensing; urban human mobility; spatiotemporal cluster; travel source–sink; sharing resource sustainability

check for updates

Citation: Shi, Y.; Wang, D.; Wang, X.; Chen, B.; Ding, C.; Gao, S. Sensing Travel Source–Sink Spatiotemporal Ranges Using Dockless Bicycle Trajectory via Density-Based Adaptive Clustering. *Remote Sens.* **2023**, *15*, 3874. https://doi.org/10.3390/rs15153874

Academic Editor: Yuji Murayama

Received: 25 June 2023
Revised: 29 July 2023
Accepted: 2 August 2023
Published: 4 August 2023

1. Introduction

To date, population patterns and facilities have become more concentrated and complex and need to be considered in terms of sustainable urbanization. Unfortunately, the imbalanced relationship between human activities and facility resources can lead to persistent congestion and high energy consumption, which brings enormous challenges to sustainability in highly urbanized development [1,2]. With the revival of the sharing concept, the public has realized that public transportation, such as buses, taxis, and shared bicycles, has an important role in saving transportation resources and relieving intensive transportation [3,4]. Empirical studies have shown that public transportation effectively reduces urban traffic congestion and resource waste [5,6]. The basis for ensuring the sustainability of public transportation is the dynamic balance of public transportation and resource recycling [7].

As a highly variable urban subsystem, transportation is more susceptible to imbalanced human–land relationships [8,9]. A large number of anomalous dissipations or aggregations may emerge among human individuals commuting through different urban areas. This significantly imbalanced phenomenon in the number of people entering and

exiting in a short period is called a travel source–sink [10], which may occur due to the attraction of major activities or sudden disasters. The travel source and sink of taxis represent areas where human departure and arrival preferences are different, which is conducive to the timely adjustment of taxis' supply to improve the travel efficiency in demand-excess areas, and reduce taxi resource waste in supply excess areas [11]. The consequences of shared bicycles' travel source–sink patterns are even worse, because they lack one-to-one drivers to dispatch each bicycle to urgently requisite areas in real time. Therefore, a large number of flexible, dockless shared bicycles often appear to show an anomalous scarcity or accumulation phenomena in important travel areas, such as residences, workplaces, or transit hubs [12]. Travel source–sink areas and periods of sharing transportation resources can clearly indicate the target extents that need to be reallocated by external manual forces, providing accurate references and insights into unusual events for maintaining urban traffic health. It is important to optimize the sustainable circulation of limited transportation resources, which can further promote the sustainable urban development goals of cities and communities given current global concerns.

Encouragingly, urban remote sensing is providing increasing theoretical and empirical evidence for addressing urban issues, such as traffic systems, medical health, and green spaces [13–15]. Plentiful image remote sensing technologies have effectively supported the large-scale detection of urban facility distribution [16]. However, cities do not entail only the coverage distribution of buildings, impervious surfaces, parks, and other facilities on the land, but also complex human activities among these urban facilities [17,18]. Correspondingly, the sensing of human activity phenomena is an emerging exploration in urban remote sensing [19,20]. Floating car trajectory is considered to be generalized remote sensing that senses human activities and maps human–land relationships with the help of positioning satellites, which has shown an important potential in urban structure analyses [18], vitality monitoring [21], functional area identification [22], and other applications. However, existing studies have mostly treated travel source–sink detection as a spatial count problem of geographic units with a specified shape [10,23], which makes it hard to distinguish shape-flexible travel source–sink patterns from dynamic, balanced hotspots. In fact, a travel source–sink area can be regarded as a dense aggregation area of a single variate, whose number is plenarily dominant relative to another variate. Thus, as shown in Figure 1, travel source–sink ranges can be detected through a bivariate cluster with a significantly biased single variate. This paper proposes a density-based adaptive bivariate clustering algorithm to extract the spatiotemporal extents of travel source–sink areas. The contributions of this paper are as follows:

(1) Proposing a density-based bivariate clustering method to identify travel source–sink ranges. Compared to pre-setting the boundaries using a geographic unit strategy, density-based clustering could avoid the potential neglect of arbitrarily shaped travel source–sink ranges.
(2) Improving an adaptive source–sink clustering estimation approach to solve the uncertainty of hyper-parameters. This paper employs statistical significance testing, which is more executable for cases without reliable a priori knowledge of a dataset.
(3) Developing a spatiotemporal joint strategy to overcome the separate extraction of source–sink periods and areas. Snapshot time and partition space can easily lead to bias in the estimation of consecutive spatiotemporal events in the real world.

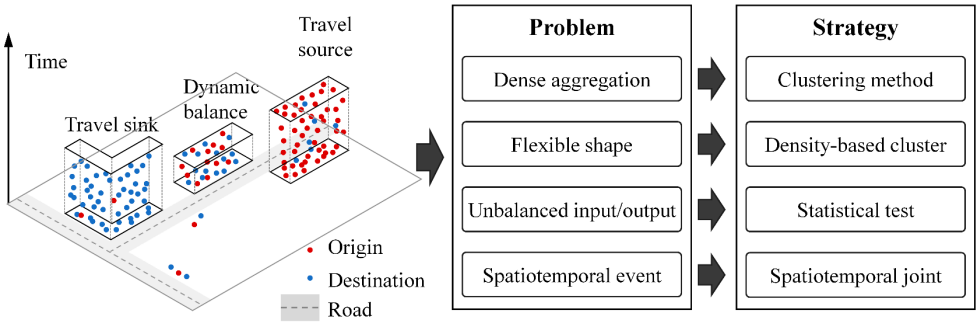


Figure 1. The structure of the problem and strategy for travel source–sink detection.

Section 2 reviews the related work about travel source–sink detection. The study area and the data utilized are presented in Section 3. Section 4 describes the research framework and the proposed method. Section 5 reveals the experimental results of the dockless bicycle source–sink spatiotemporal extents based on global and typical scenes. The discussion and conclusion are reported in Section 6 to summarize the interesting findings and highlight future works.

2. Related Work

A large amount of available floating car trajectory data has been utilized to extract human travel origin and destination (OD) and to measure the source–sink level of a selected area [10]. According to different spatial extraction strategies, the existing travel source–sink detection methods can be divided into unit feature measurement-based and point event clustering-based methods.

Unit feature measurement-based methods first use time slices and road networks or spatial grids to divide a study area into geographic basic units. Based on this, a statistical index of these geographic basic units is constructed to measure the imbalanced degree between inflow and outflow. Liu et al. [10] evenly divided a study area into 1 km grids to calculate the time series of the difference between the inflow and outflow of each grid, and then the k-means algorithm [24] was used to classify the spatial grids into six typical source–sink patterns. Gao et al. [25] focused on the imbalanced distribution of bicycle-sharing supply and demand around subway stations. Based on the outflow and inflow, a ratio index of grids was constructed to cluster the source–sink areas into five different patterns. In addition to spatial grids, administrative districts [17], functional areas [22], and blocks [9] are also commonly used as spatial units for measuring human mobility. For example, Fang et al. [23] used mobile phone signal data received by towers to calculate cumulative net flow, further revealing the temporal stability of the sources and sinks of different signal towers. Dividing geographic units is a common preprocessing approach before sensing geographical phenomena [26]. However, as shown in Figure 2, the pre-set boundaries of geographic basic units are prone to rigidly dividing continuous intensive source–sink areas, resulting in a serious underestimation of the source or sink level. At the same time, how to determine the size of geographic basic units and the threshold for identifying a source–sink area is a supervise-hard problem, which brings obstacles to the generalizability and interpretability of unit feature measurement-based methods.

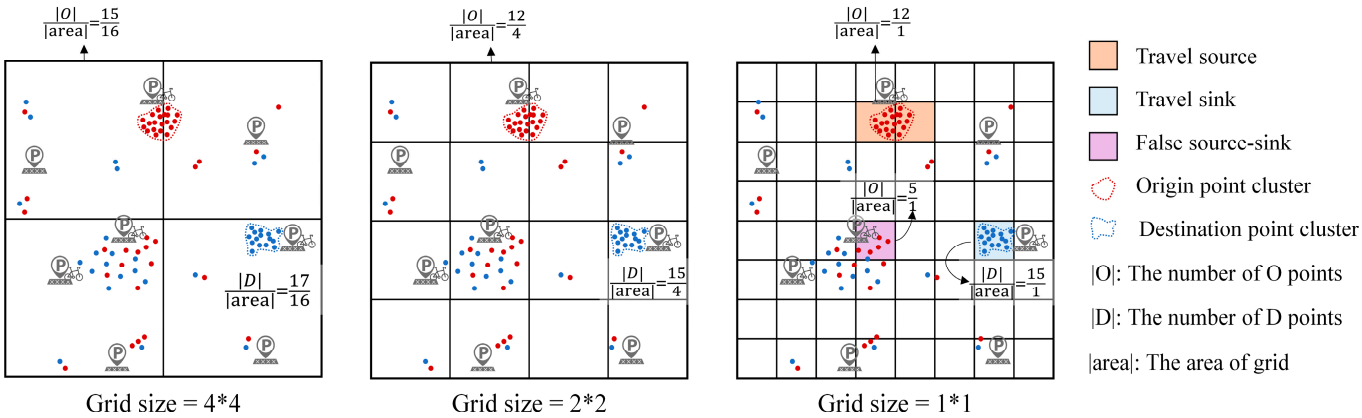


Figure 2. Geographic basic unit boundaries lead to source–sink underestimation.

Point event clustering-based methods avoid the difficulties of spatiotemporal division by aggregating OD point positions to form continuous dense areas or periods with significantly imbalanced inflow and outflow. Existing univariate trajectory point event clustering has been widely used in many scenarios, such as travel hotspot extraction [27], accident-prone area assessment [28], and traffic congestion detection [29]. Clustering-based source–sink detection methods belong to binary variable clustering tasks and can be used to detect symbiotic or imbalanced relationships between two variables. When the travel flows from different sources are treated as binary variables for clustering, an imbalanced flow relationship indicates significant competition between two types of flows, while a similar flow relationship indicates that they are in a coordinated, symbiotic relationship [30,31]. A large number of spatial statistical clustering methods have been utilized to aggregate different types of flow aggregation [30–33], while other methods are still rare. Similarly, performing clustering with O points and D points as binary variables can detect imbalanced relationships between inflow and outflow, which is useful for discovering potential interest areas with scarce or abundant transportation resources. Liu et al. [34] constructed a spatial scan statistic in a road network and used a Monte Carlo simulation to evaluate the significance of the source–sink area aggregation via multi-directional optimization operations, eventually extracting the volcano and black hole patterns formed by the imbalanced distribution of motor vehicle OD points in the urban road network space. However, the computational cost of a Monte Carlo simulation and multi-directional optimization operations is too high for computers [31]. Density-based clustering algorithms may be a possible choice, because density calculation and core point extension can be rapidly performed with the help of spatial indexing techniques. However, several pre-set critical parameters may unconsciously affect the clustering detection results of density-based clustering. Although there are interactive parameter recommendation schemes, this is still a manual process and is limited in terms of estimating the migration at different study areas [35]. Pei et al. [36] extended the DBSCAN algorithm [37], which is applicable to two types of point data, and found three kinds of taxi OD density relationships using a spatial point database.

Existing point event clustering-based methods [29–34,36] provide encouraging references for identifying OD distribution imbalances; however, there are still several challenges in quickly and adaptively identifying travel source–sink areas with any shape in a complicated urban space. Firstly, spatial statistical-based clustering methods [30,33] do not perform well in detecting arbitrarily shaped OD clustering areas, such as linear, T-shaped, or cross-shaped areas, which always appear in a traffic network space [31]. In addition, how these methods can efficiently process large spatial databases has not been properly addressed. Secondly, density-based clustering methods can overcome the problem of identifying clusters of any shape by treating trajectory points as dense-level test objects, but how to avoid multiple sensitive parameters is still a challenge, especially when analysts lack clear prior knowledge about the OD point datasets being used. Finally, the source–sink

temporal feature is ignored by existing point event clustering-based methods, which project OD points over a long period onto a spatial surface [29,36]. Utilizing spatial clustering to detect source–sink areas will result in OD points in the same source or sink area, which will not be continuous and tight in timestamps. Only by accurately indicating the life period and space location of source–sink areas can the optimization of anomalous imbalanced traffic phenomena be effectively supported.

To address the above challenges, this paper proposes a novel adaptive bivariate clustering algorithm to reveal the spatiotemporal ranges of travel source–sink areas, indicating the imbalanced aggregation between the origins and destinations implied in individual travel trajectories. Specifically, the density-based clustering algorithm is developed to solve the challenge of arbitrarily shaped clusters, a statistical test is proposed to overcome the prior parameter uncertainty when identifying source–sink patterns, and a spatiotemporal joint clustering strategy is developed to solve the spatiotemporal characteristics of traffic events.

3. Study Area and Dataset

Shanghai is one of the four municipalities directly under the Central Government of China and is a mega city in eastern China. The urbanization rate in Shanghai is close to 90%. According to the Shanghai Statistical Yearbook [38], the population of Shanghai in 2021 was about 24.89 million, while its administrative area is only 6340 km^2, with a population density as high as 3926 people per square kilometer. Limited land resources and construction costs restrict the expansion of transportation resources. Thus, the dynamic balance of sharing transportation resources is an important issue to ensure the high-quality development and sustainability of the city. Figure 3 shows the study area selected in this paper, which is a core area of Shanghai. Although the study area is a plain terrain, the river network is developed, which leads to the urban infrastructure and transportation construction being intensive and complex.

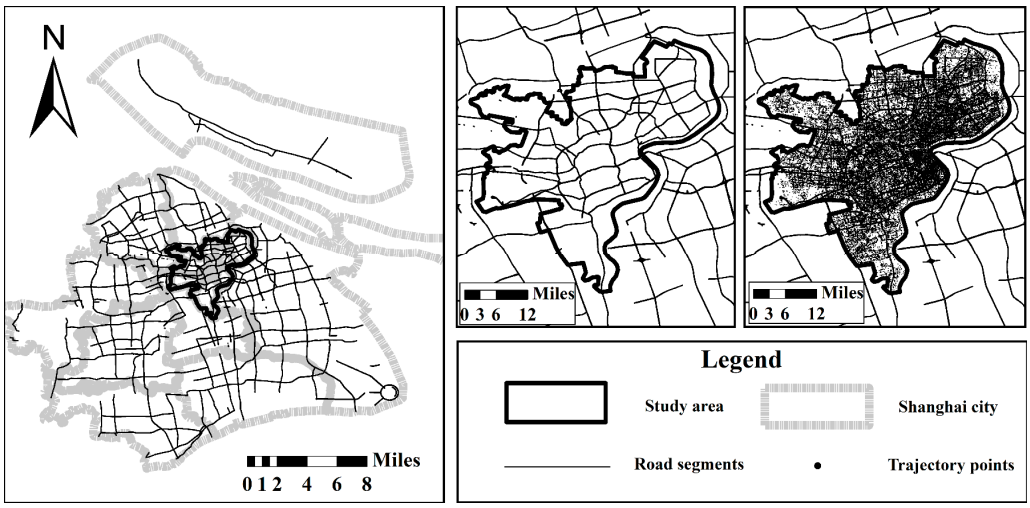


Figure 3. Spatial distribution of the study area and practical bicycle trajectory points.

Sharing bicycles are a new type of public transportation that have emerged in recent years. With the advantages of flexibility, a low cost, and sustainability, sharing bicycles constitute an important part of the urban transportation system in China. Shared bicycles are mostly distributed in areas with interconnected multi-transportation areas, such as residential areas, shopping malls, and subway stations, complementing the traditional public transportation system. As a key transportation to serve human travel for the "last mile", sharing bicycles are of great significance to governmental intelligent transportation construction and international sustainability goals. The free and flexible employment of shared bicycles greatly meets the various dynamic travel needs of urban residents for commuting, entertainment, shopping, and other purposes. Unfortunately, sharing bicycles

can easily lead to imbalanced source–sink phenomena, i.e., waste accumulation or resource scarcity in specific areas, especially during peak hours. This study collected Shanghai sharing bicycle trajectory data from 2 to 8 September 2018, where more than 11 million trajectory points were generated from over 250,000 shared bicycles across 24 h. The data on these sharing bicycle trajectory points contained bicycle ID, geographical location, timestamp, and bicycle status (parking or riding). Specifically, the trajectory data from 7:00 to 9:00 on 2 September are shown in the right panel of Figure 3.

Due to the positioning uncertainty of the spatial positioning equipment and parking lock system, bicycle trajectory data preprocessing was performed. First, the trajectory data with the same ID were sorted in an ascending order according to timestamp, and each travel was distinguished according to bicycle status. The first and last points in each complete travel were extracted as the O point and D point to form an OD pair. Then, we deleted OD pairs with a travel time of less than 1 min or more than 1 h, which is rare in bike rides. Finally, we deleted OD pairs outside the study area, in water systems, or in buildings.

4. Methods

This paper proposes a density-based bivariate clustering method to sense dense areas with a significant imbalanced distribution of two type points, which addresses the limitations of prior parameter sensitivity and spatiotemporal feature separation. The proposed framework is shown in Figure 4. The first step is optimal spatiotemporal neighborhood construction. A Gaussian kernel density function is employed to represent the local spatiotemporal point density field, whose stability is further measured to detect the optimal spatiotemporal radius. The second step involves the adaptive identification of spatiotemporal core entities. A spatiotemporal random distribution and density-based statistical test method is constructed to adaptively identify the dense core point for each spatiotemporal cluster. The final step involves travel source–sink range detection. A bivariate statistic test method is constructed to adaptively detect the source or sink core neighborhood where the OD point distribution is significantly imbalanced. A spatiotemporal neighbor expansion strategy is designed to extract the source and sink areas with arbitrary shapes and durations, where the source areas refer to regions where the O points are significantly denser than the D points, and the sink areas are regions where the D points are significantly denser than the O points.

4.1. Optimal Spatiotemporal Neighborhood Construction

A crucial aspect of the density-based clustering approach is to check the density of geographic entities by defining a spatial neighborhood radius, *eps*, and a minimum number of elements, *minpts*, within a neighborhood. Previous density-based clustering algorithms have usually determined these two key parameters via a preset way [29,36,37]. However, setting *eps* without careful thought often results in unpredictable and worse clustering results [39]. The adaptive generation of a spatial neighborhood radius based on geographic entity distribution features has achieved credible application [29], whereas the adaptive construction of a stable spatiotemporal neighborhood based on spatiotemporal entities' distribution has not gotten enough attention.

Figure 5a shows the spatial neighborhood diagram generated using the spatial density-based clustering algorithm. It can be seen that too large a neighborhood radius will merge small clusters into a too-large cluster, and too small a neighborhood radius usually will not appear to have enough neighborhood entities and generate too much false noise. This paper develops a spatiotemporal kernel density estimation to measure the intensity field of neighborhood entities in the case of various neighbor radii, and then the instability is utilized to measure the uncertainty of the neighborhoods of all entities and further determine the optimal spatiotemporal neighborhood size.

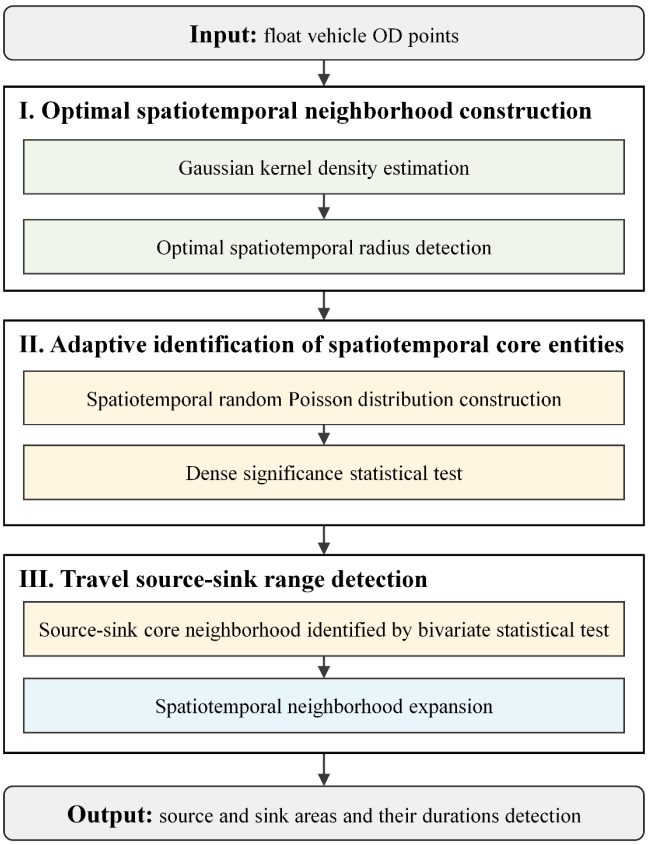


Figure 4. The framework of the proposed method.

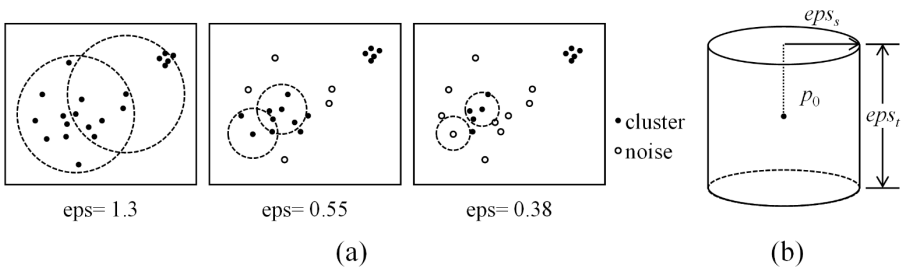


Figure 5. The (**a**) spatial (dotted line circle) and (**b**) spatiotemporal neighborhood (dotted line column) of geographic entities generated using the density-based clustering method.

Given any point p_0 in a real geographic entity database ***GED***, its spatiotemporal kernel density [40], $STKD(p_0)$, is estimated as follows:

$$STKD(p_0) = \frac{1}{Nh_s^2 h_t} \sum_{j=1}^{N} K_s\left(\frac{ds(p_0,\ p_j)}{h_s}\right) * K_t\left(\frac{dt(p_0, p_j)}{h_t}\right) \tag{1}$$

where p_j denotes any other geographic entities in the ***GED*** and N denotes the number of whole entities in the ***GED***; ds and dt represent the spatial and temporal distance between p_0 and p_j, respectively; Ks and K_t denote the spatial and temporal kernel density functions, respectively, where planar and one-dimensional Gaussian kernel functions are utilized in this paper; and h_s and h_t represent the spatial and temporal bandwidths, i.e., the selected spatial radius eps_s and temporal radius eps_t. Entities beyond the h_s and h_t of the spatiotemporal kernel density bandwidths are constrained without superimposing density intensities and creating neighbor relationships.

For any specified eps_s and eps_t, the corresponding intensity field $IF(eps_s, eps_t)$ of geographic entities can be calculated through spatiotemporal kernel density estimation. Further, the instability of $INS(eps_s, eps_t)$ can be measured as follows [41]:

$$INS(eps_s, eps_t) = -\sum_{i=1}^{N} \frac{STKD(p_i)}{\sum_{k=1}^{N} STKD(p_k)} \log\left(\frac{STKD(p_i)}{\sum_{k=1}^{N} STKD(p_k)}\right) \tag{2}$$

The smaller the $INS(eps_s, eps_t)$, the greater the asymmetry among $STKD(p_i)$, and the easier it is to extract dense neighborhoods [27]. Thus, this paper selects the $(eps_{s\_opt}, eps_{t\_opt})$ that minimizes the $INS(eps_s, eps_t)$ as the optimal spatial and temporal neighbor radius:

$$\left(eps_{s\_opt}, eps_{t\_opt}\right) = argmin(INS(eps_s, eps_t)) \tag{3}$$

Further, this method can generate a spatiotemporal neighborhood *STN* for each point entity as a cylinder enclosed by $(eps_{s\_opt}, eps_{t\_opt})$. In practice, it is suggested to use a candidate temporal radius $ht \in (0,\ 60\ \text{min}]$ with a step size of 5 min, and a candidate spatial radius $hs \in (0,\ 300\ \text{m}]$ with a step size of 20 m, which ensures that the approximate optimal parameter combination $(eps_{s\_opt}, eps_{t\_opt})$ can be obtained within the executable time.

4.2. Adaptive Identification of Spatiotemporal Core Entities

It is hard to cause large-scale source–sink disasters with sporadic origins and destinations for sharing transport resources, such as dockless bikes. The travel source–sink pattern of shared bicycles has a dense distribution of OD points, which reflects a large number of bicycles being parked or being ridden away in the real world. Thus, based on the construction of the optimal spatiotemporal neighborhood *STN*, the next important step is to estimate whether the space–time neighborhood is dense or sparse.

Previous research has shown that clusters exhibit entity aggregation relative to random distributions [34]. In the case of random distribution without clusters, the probability of a certain number of k entities appearing in a spatial neighborhood can be considered as conforming to the Poisson distribution [29], which is widely used to describe the number of random events occurring per spatial or temporal unit [36]. In a spatiotemporal coupled three-dimensional random space, the theoretical probability of k entities appearing in a spatiotemporal neighborhood of any entity $STN(p_i)$ is the probability density function of the Poisson distribution:

$$\begin{gathered} p(|STN(p_i)| = k) = e^{-\rho} \frac{\rho^k}{k!} \\ \rho = N * \frac{V(STN(p_i))}{V(ST)} \\ V(STN(p_i)) = \pi * eps_{s\_opt}^2 * eps_{t\_opt} \end{gathered} \tag{4}$$

where N denotes the total number of entities; $V(\cdot)$ represents the volume calculation operation; and *ST* represents the potential spatiotemporal range that entities may appear within, which is regarded as a spatiotemporal body composed of a potential spatial region and potential period in the case of an absence of periodic traffic control, such as no parking at nighttime. The potential spatial region is generated through the union of all the OD point buffers, whose radius is set according to the positioning error of the floating vehicle. The potential period is selected as a whole day due to the fact that shared bicycles can be freely borrowed and returned 24 h a day.

Given any entity p_i and its spatiotemporal neighborhood $STN(p_i)$ in the ***GED***, the probability that the number of entities in any other $STN(p_j)$ in the case of a hypothetical random distribution is greater than that in $STN(p_i)$, is calculated as follows [29]:

$$\begin{aligned} P(core(p_i)) &= P\left(\left|STN\left(p_j\right)\right| \geq |STN(p_i)|\right) = 1 - P\left(\left|STN\left(p_j\right)\right| < |STN(p_i)|\right) \\ &= 1 - \sum_{k=0}^{|STN(p_i)|-1} p\left(\left|STN\left(p_j\right)\right| = k\right) = 1 - \sum_{k=0}^{|STN(p_i)|-1} e^{-\rho} \frac{\rho^k}{k!} \end{aligned} \tag{5}$$

where $|STN(p_i)|$ and $|STN(p_j)|$ denote the number of entities in $STN(p_i)$ and $STN(p_i)$, respectively. If $P\left(\left|STN\left(p_j\right)\right| \geq \left|STN(p_i)\right|\right)$ is less than a given statistically significant level α (e.g., 0.05), then $STN(p_i)$ is more likely to be a spatiotemporally dense neighborhood than $STDN(p_i)$, the probability of which cannot be exceeded by a random neighborhood, and p_i can be regarded as a core spatiotemporal entity $c\_\ p_i$.

4.3. Travel Source–Sink Range Detection

The travel source–sink pattern of shared bicycles not only shows a dense aggregation of dockless shared bikes, but it also has a significant imbalance in terms of the number of origins and destinations, which leads to scarcity or waste rather than a dynamic balance of bicycle resources. For the OD points in the ***GED***, there are only two states between an origin and a destination. In a case that lacks origin or destination point aggregation, the point entities in any selected region randomly belong to O or D, according to the probability of binomial distribution. Given any spatiotemporally dense neighborhood $STDN(p_h)$ of a core point c_p_h in the ***GED***, the probability of k origin points appearing in the $STDN(p_h)$ follows a binomial distribution according to the hypothesis of balanced distribution:

$$p(|STDN(p_h, O)| = k) = C^k_{|STDN(p_h)|} * \sigma^k * (1-\sigma)^{|STDN(p_h)|-k} \ s.t.\ 0 \leq k \leq |STDN(p_h)| \tag{6}$$

where $|STDN(p_h, O)|$ and $|STDN(p_h)|$ denote the number of origin points and number of whole points in $STDN(p_h)$, respectively, and σ represents the ratio of origin points in the ***GED***, e.g., 0.5. On this basis, the probability of $STDN(p_i)$ being a source region can be expressed as the significance level of aggregating the origin points as follows:

$$\begin{aligned} P(sour(p_i)) &= p(|STDN(p_h, O)| \leq |STDN(p_i, O)|) = \sum_{k=0}^{|STDN(p_i,O)|} p(|STDN(p_h, O)| = k) \\ &= \sum_{k=0}^{|STDN(p_i,O)|} C^k_{|STDN(p_h)|} * \sigma^k * (1-\sigma)^{|STDN(p_h)|-k} \ s.t.\ 0 \leq k \leq |STDN(p_i,\ O)| \end{aligned} \tag{7}$$

According to the significance level α used in the previous section, if $P(sour(p_i))$ is larger than the significance level $1 - \alpha$, p_i and $STDN(p_i)$ can be labeled as a spatiotemporal source core point $oc\_\ p_i$. and a source neighborhood $OSTDN(p_i)$, respectively. Similarly, if $P(sink(p_i))$ is less than a given statistical significance level α, p_i and $STDN(p_i)$ can be regarded as a spatiotemporal sink core point $dc\_\ p_i$ and a sink neighborhood $DSTDN(p_i)$, respectively.

The density-based expansion strategy is employed to extract the travel source–sink pattern. The detailed steps are as follows:

(1) Select any source core point $oc\_\ p_i$ as the seed point, and define the other origin points p_j within the $OSTDN(p_i)$ as dense reachable points from $oc\_\ p_i$. On this basis, aggregate all the dense reachable points from $oc\_\ p_i$ into a source pattern $OP1$.
(2) For any other source core point p_k in $OP1$, keep the aggregation operation in (1) to expand $OP1$ until all the source core points in $OP1$ have been visited.
(3) Reselect the other source core points that have not been visited in the ***GED*** as new seed points, and repeat the above process (1)~(2) to generate a new OP, until there are no visited source core points within the ***GDE***.
(4) Finally, the spatiotemporal extents of the travel source patterns are merged through the whole entities in $OP = \{OP_1, OP_2, \ldots\}$.

Similarly, with any sink core point $dc\_\ p_i$ as the seed point, the other destination points within the $DSTDN(p_i)$ are defined as dense reachable points. All dense reachable points are iteratively aggregated into a sink pattern $DP1$. Based on the above spatial expansion strategy, all the sink patterns extracted from the ***GED*** can be represented as $DP = \{DP_1, DP_2,$

. . .}. The GIS- or RS-based visualization platform can intuitively exhibit and interactively analyze the spatiotemporal range variations in travel source–sink patterns.

5. Experimental Results

5.1. Experimental Comparisons and Quantitative Evaluation

Before applying the proposed method to the real bicycle trajectory data, this section constructs quantitative evaluation indexes according to the characteristics of source–sink patterns to evaluate the performance of the proposed method. Considering that travel sources and sinks are dense aggregation areas where the number of single variates is plenarily dominant relative to another, the clustering evaluation indexes and extra imbalanced metrics are employed to estimate the experimental results. Specifically, the Davies–Bouldin (DB) index [42] and Calinski–Harabaz (CH) index [43] for clustering evaluation are utilized to assess the cluster quality:

$$\begin{gathered} DB = \frac{1}{k}\sum_{i=1}^{k} \max_{i \neq j}\left(\frac{avg\left(C_i\right)+avg\left(C_j\right)}{d_{cen}\left(C_i,C_j\right)}\right) \\ avg(C_i) = \frac{2}{n_i*(n_i-1)}\sum_{1<p<q<n_i} dis\left(x_p, x_q\right) \\ d_{cen}\left(C_i, C_j\right) = dis\left(u_i, u_j\right) \end{gathered} \tag{8}$$

$$CH = \left(\frac{\sum_{i=1}^{k} n_i * |u_i - u|^2}{k-1}\right) / \left(\frac{\sum_{i=1}^{k}\sum_{j=1}^{ni}\left|x_j - u_i\right|^2}{N-k}\right) \tag{9}$$

where, k and n_i represent the number of clusters and number of points in the ith cluster C_i, respectively; x_p and x_q denote the points; u_i and u_j denote the centers of cluster C_i and C_j; $dis(\cdot)$; and N represents the distance between two points and the number of all the points in the ***GED***. It is easy to understand that a smaller DB indicates a better clustering performance, and a larger CH is better. The DB index prefers intra-cluster closeness, and the CH index focuses on inter-cluster discrimination [44]. A comprehensive comparison of several assessment metrics helps to more objectively cognize the clustering performance. Meanwhile, the Net Flow Ratio (NFR) [45] for OD imbalanced metrics is utilized to evaluate the travel source–sink level:

$$NFR = \frac{1}{k}\sum_{i=1}^{k}\left(\frac{|O_i|-|D_i|}{|O_i|+|D_i|}\right) \tag{10}$$

where $|O_i|$ and $|D_i|$ denote the number of origin points and destination points of cluster C_i, respectively, and k represents the number of clusters. A larger NFR indicates that the number of origin points is greater than the number of destination points; conversely, a smaller NFR indicates that the number of origin points is less than the number of destination points. Therefore, a greater NFR is better for the source and a smaller NFR is better for the sink. The upper and lower bounds of NFR are $[-1, 1]$.

Considering that the proposed method belongs to the density-based spatiotemporal clustering approach, the ST-DBSCAN algorithm [46] is used as the comparison for the clustering level assessment. The ST-DBSCAN algorithm is considered to be an excellent algorithm based on density, which has been widely applied in spatiotemporal database clustering to deal with flexibly shaped aggregations. It is known that the ST-DBSCAN algorithm needs to specify the priori parameters. We followed the heuristic strategy for parameter recommendation in the original paper to ensure the effectiveness of the algorithm. Meanwhile, considering that the proposed method aims to detect the source–sink area of OD imbalance, Pei's method [36] is used to compare the source–sink detection level evaluation. Pei's method is a typical approach to detecting sources and sinks in spatial databases, which has been verified as effective in simulated and trajectory datasets. Pei's method requires specifying two parameters, which follow the parameter values recommended in the original paper.

Table 1 shows the quantitative evaluation results of the detected travel sources and sinks by the proposed method, ST-DBSCAN algorithm, and Pei's method. Since the ST-DBSCAN algorithm is not designed for source–sink detection, this paper clusters origin points and destination points separately as source and sink pattern results. As shown in Table 1, the proposed method exhibits the smallest DB index for sink detection and a similar DB index for source detection, in contrast with the other methods. Furthermore, the greater CH index similar to the ST-DBSCAN algorithm indicates that the proposed adaptive spatiotemporal clustering method can easily detect dense clusters in spatiotemporal databases. In fact, the clustering evaluation of the ST-DBSCAN algorithm is slightly better than that of the proposed method. This can be acceptable because the ST-DBSCAN algorithm does not need to consider the imbalanced distribution of OD points, which results in aggregating more dense clusters. It is can be seen from the local view of Figure 6a,b that the source areas and sink areas detected by ST-DBSCAN show the same distribution. Meanwhile, the NFR closer to 0 of the ST-DBSCAN algorithm indicates that the univariate clustering algorithm finds it easy to detect dynamic balanced areas, but hard to extract sources and sinks with imbalanced OD points. On the other hand, the NFR index of the proposed method and Pei's method are comparable, indicating that both can detect source and sink areas. Figure 6c,d show the significant ability to distinguish source–sink and dynamic balanced areas of the proposed method and Pei's method. However, Pei's method tends to extract more fragmented clustering patterns. Meanwhile, Pei's method is only applicable to static spatial databases, and it cannot detect the duration of the source or sink. The proposed method enhances the source–sink detection performance in spatiotemporal databases, which is vital, because OD points are labeled with nonnegligible timestamps in real life [29].

Table 1. Quantitative evaluations of the experimental results by the proposed method, Pei's method and ST-DBSCAN.

Method	Type	DB	CH	NFR
The proposed method	Source	1.59	21.36	0.81
	Sink	1.30	18.70	−0.79
ST-DBSCAN algorithm	Origin point cluster	1.56	23.90	0.28
	Destination point cluster	1.46	20.37	−0.27
Pei's method	Source	1.60	12.97	0.77
	Sink	1.34	9.52	−0.64

Table 1 exhibits that the source–sinks detected by the proposed method perform a credible CH and DB index and the maximum NFR proves the clustering effectiveness and source–sink detection ability.

5.2. Spatiotemporal Distribution of Bicycle Travel Sources and Sinks

Figure 7a shows the spatial distribution of the bicycle travel sources and sinks detected by the proposed method during the mornings (07:00–09:00) and evenings (17:00–19:00) on weekends and workdays, where the travel source patterns are labeled by a bright orange color and the travel sink patterns are drawn in a blue color. Meanwhile, Figure 7b exhibits the bicycle travel source and sink levels at the zone level aggregated by 1 km $\times$ 1 km grids in the study area to intuitively visualize the heterogeneity of the spatial distribution. The diverse colors of the cells represent the area differences between the sources and sinks in the same grid, where a positive value denotes that the source area is greater than the sink area; otherwise, the source area is smaller than the sink area.

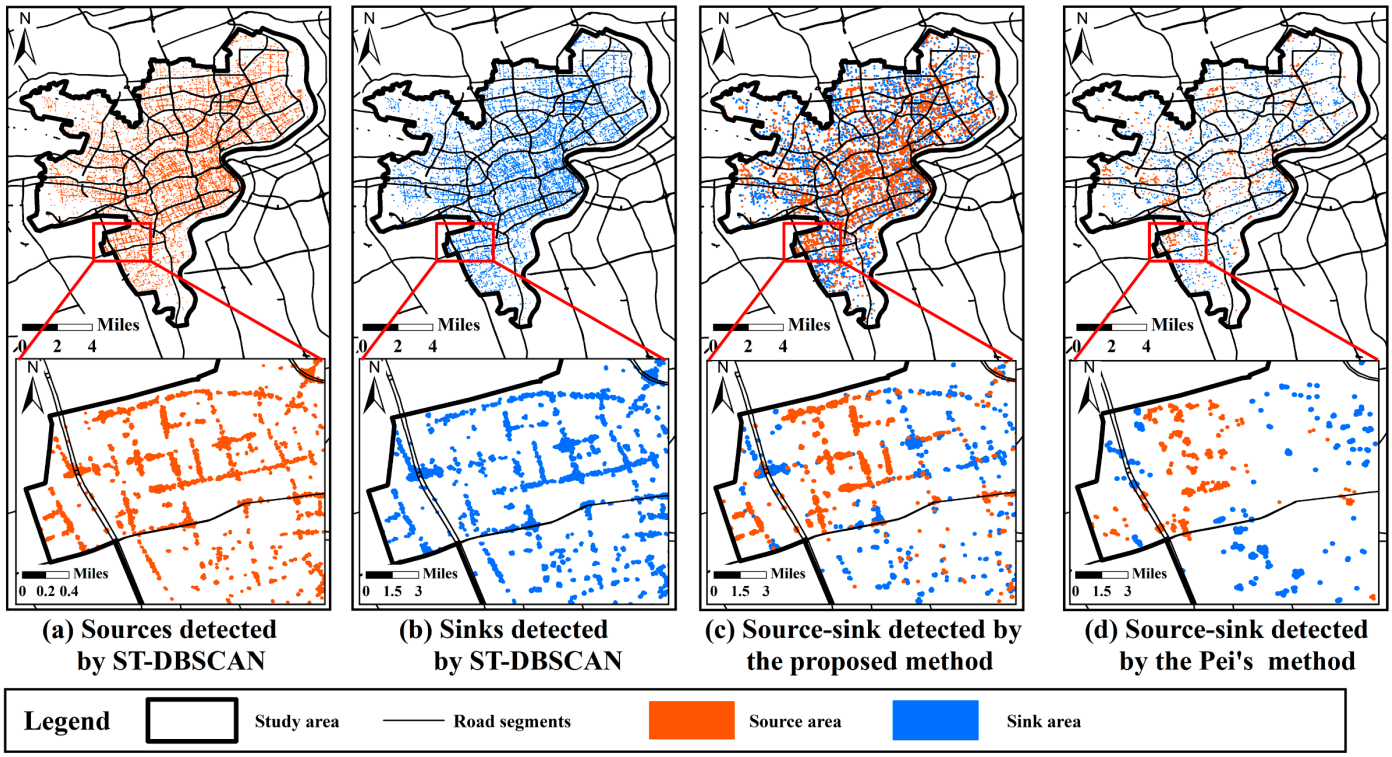
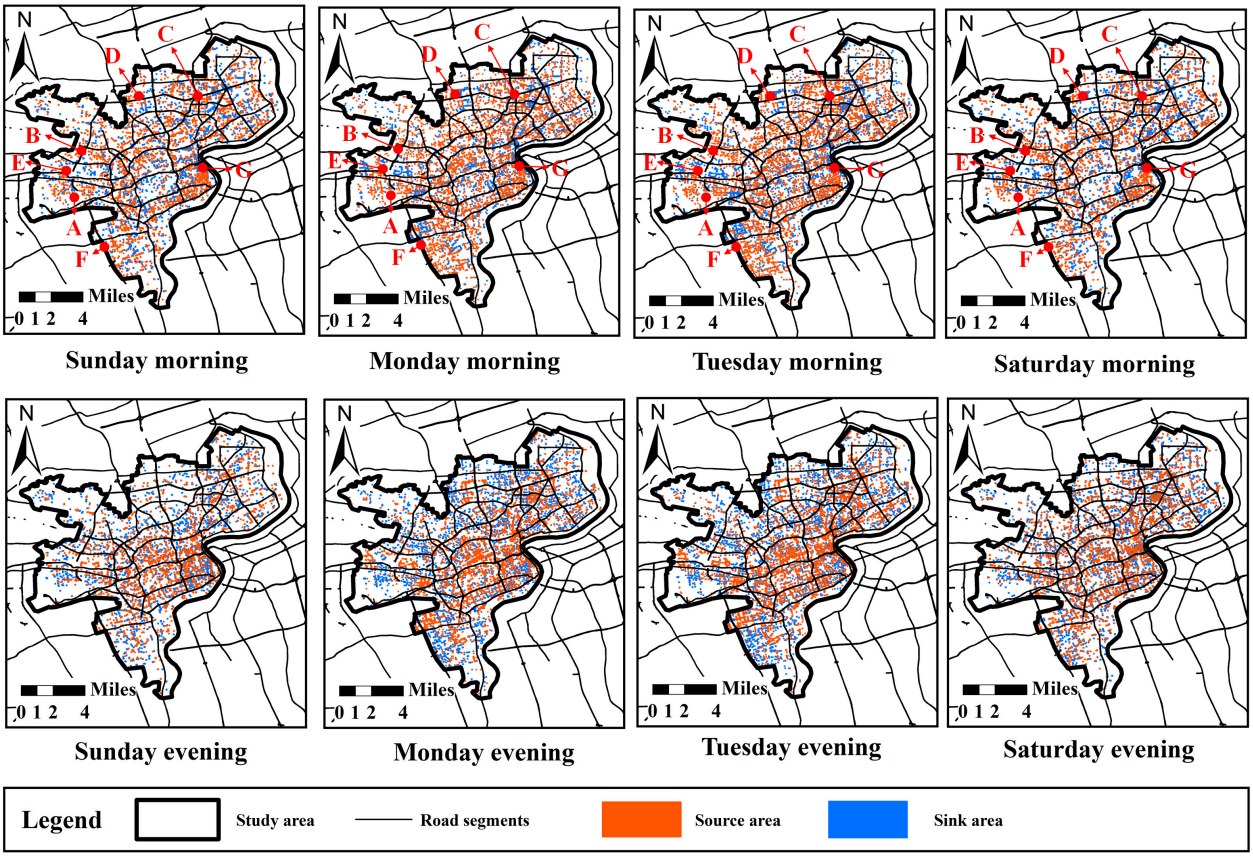


Figure 6. (**a**) Bicycle travel sources and (**b**) sinks detected by the ST-DBSCAN algorithm; bicycle travel source–sink detected by (**c**) the proposed method and (**d**) Pei's method during 17:00–19:00 on 3 September.

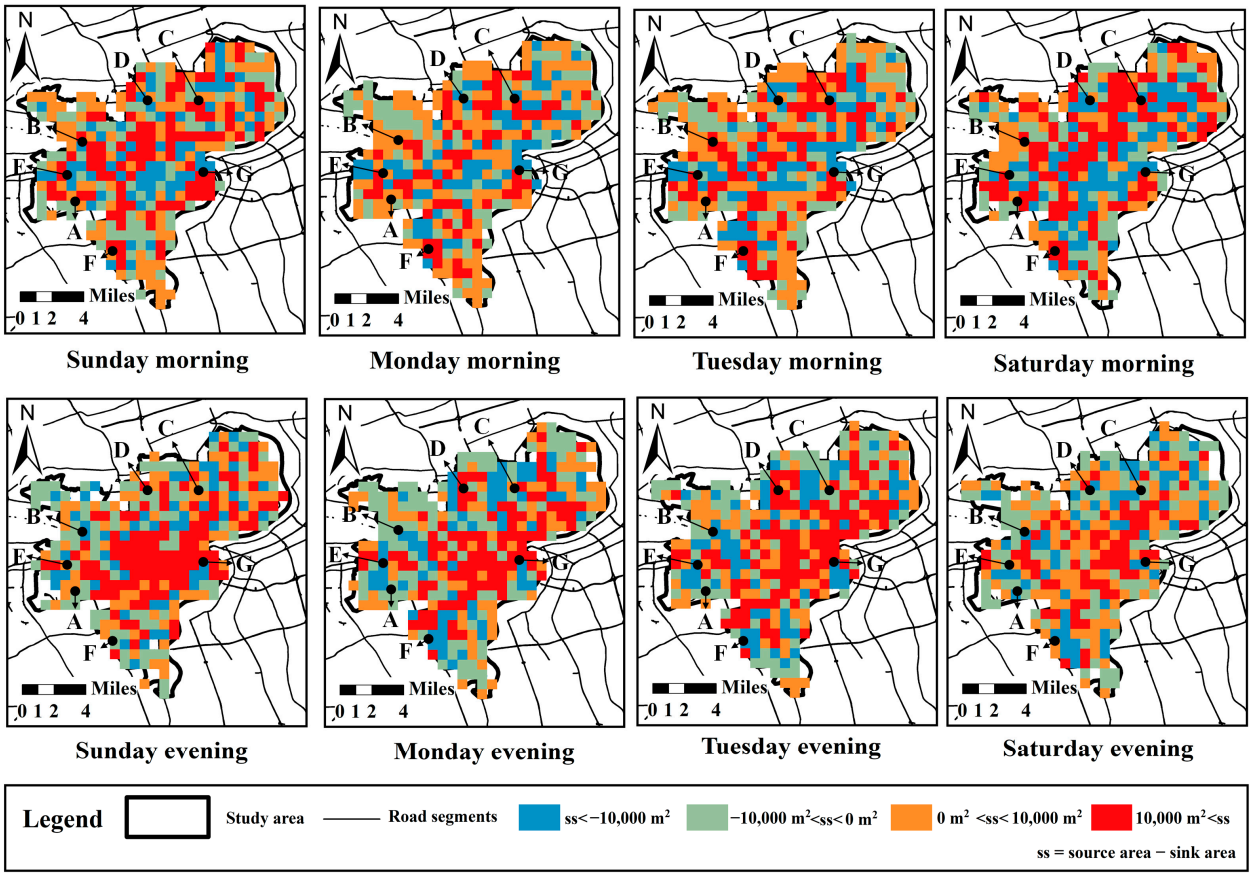
From the perspective of intraday, the spatial distribution of the travel source–sink areas in the morning and the distribution in the afternoon are almost symmetrical. The source patterns are widely distributed, but the sink patterns are more concentrated in the morning, which represents an obvious tendency towards activity aggregation. On the contrary, the source patterns show a concentrated distribution, but the sink patterns express a widespread distribution in the evening, indicating higher preferences for the dispersed activity of urban residents and tourists. These interesting distributions are more significant in Figure 7b. In detail, a source pattern in the morning mostly becomes a sink pattern in the afternoon, and a sink pattern in the morning mostly transfers to a source pattern in the afternoon, such as in areas A, B, and C. Areas A, B, and C are surrounded by a large number of residential communities, schools, zoos, and parks. This may be related to the commute of residents from communities and students from middle schools by bicycle for "the last mile" or "the first mile", which is more common on workdays. Meanwhile, parks and other relaxation places attract plentiful residents and tourists on weekends, thereby forming obvious source and sink patterns. However, due to the diversity of travel purposes on weekends, there are exceptions to the symmetry laws of these travel source–sink patterns in the morning and evening. Residential communities can form a workday-like source pattern on weekends, but areas that only have workplaces and lack commercial centers and leisure places hardly form a large number of sink patterns on weekends in a similar way to workdays, such as areas D, E, and F. This subtle variation can be finely captured, as shown in Figure 7a. Specifically, areas D, E, and F are composed of a large number of high-tech parks, media enterprises, information service companies, and other high-density and low-area intensive workplaces, which have little imbalanced human flow to form sink patterns on weekends.

From the perspective of interday, the source–sink patterns on workdays and weekends show different characteristics. Firstly, the number of source–sink areas on weekends is less than that on workdays, especially because the imbalance in bicycle use on weekend nights is sharply reduced. The source–sink patterns on weekends are more concentrated in the urban center of Shanghai, which reflects the vitality of the human activities of a large city in the evening. Furthermore, the sources and sinks on weekends show a more mixed distribution than those on workdays, which may be due to bicycles becoming the main way of transportation between dense scenic spots and commercial centers. It is worth noting that the Bund (area G), a famous scenic spot in Shanghai, shows a more significant sink pattern on weekends than on weekdays in Figure 7a, indicating its strong travel attraction and potential crowd-gathering risk.



(**a**)

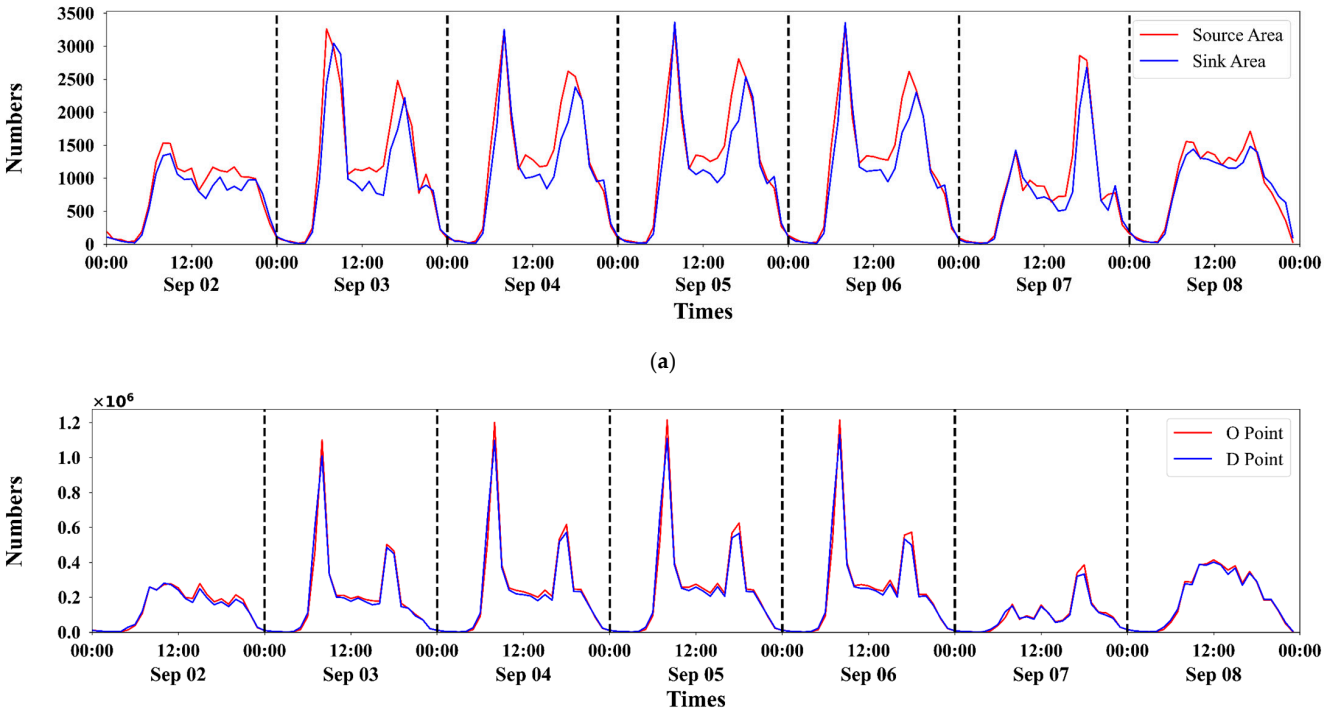
Figure 7. *Cont.*



(**b**)

Figure 7. Spatial distribution of (**a**) bicycle travel sources and sinks detected by the proposed method and (**b**) their grid-level statistical distribution during 07:00–09:00 and 17:00–19:00 on Sunday (2 September), Monday (3 September), Tuesday (4 September), and Saturday (8 September).

Figure 8 presents the temporal variation in the number of travel sources and sinks detected by the proposed method. The number of sources and sinks on workdays shows a bimodal structure, while the occurrence of sources and sinks on weekends remains stable during the daytime. The peaks on workdays occur around 8:30 a.m. and 5:30 p.m., which may denote the regular commuting patterns of residents and students. The intensity of the morning source–sink peaks is greater than that in the afternoon, while the duration of the afternoon source–sink peaks is longer than that in the morning, reflecting different paces of life. The exception is that the morning source–sink peak of 7 September was not more significant than others. When combining the data shown in Figure 8b with a historical weather check, it can be seen that thundershowers and a low visibility might have led to a lower number of bicycle trips on the morning of 7 September compared to other workdays. Another interesting phenomenon is that the number of sources was usually always greater than the number of sinks, especially at noontime and in the afternoon, which presents an imbalance in bicycle travel source–sink events, where multiple bicycle travel source areas usually converged into fewer sink areas. The variations in the number of O points and D points show the same curves, while the number of source and sink areas formed by these significantly imbalanced O points and D points are different, which reflects the necessity of urban event sensing using trajectory data, such as vehicle observation.



(**a**)

(**b**)

Figure 8. Temporal variation in the number of (**a**) bicycle travel sources and sinks detected by the proposed method, and (**b**) bicycle travel O and D points.

5.3. Typical Bicycle Travel Source and Sink Scenarios

Figure 9 shows the spatiotemporal variation in the bicycle travel sources and sinks near the Caohejing High-Tech Park, which is a typical technology-intensive workplace, aggregating a large number of technology development firms, financial services, information consulting services, and services from other industries. It is obvious that this high-tech park presented significant sinks in the morning of workdays, and the surrounding subway stations and residential communities are the related sources, which are accounted for by hordes of workers going to work by bicycle, thereby forming these mismatched origins and destinations. This pattern is reversed on workday afternoons. On workdays, the nearby Caohejing High-Tech Station, Guilin Road Station, and Hongcao Road Station presented sources in the morning and sinks in the evening, and the farther Hongmei Road Station and Guilin Park Station showed slighter travel sources and sinks. On weekends, the Caohejing High-Tech Park hardly exhibited severe bicycle travel source–sink patterns. However, the surrounding communities still presented distinct bicycle travel sources in the morning and sinks in the evening, as residents have travel demands other than work on the weekends. Through this detailed analysis of a local source–sink scenario, it can be seen that the proposed method can effectively detect the arbitrary shapes of source–sink regions, such as line shapes, T shapes, or cross shapes along the streets. This is important in urban traffic pattern sensing, since humans are constrained when traveling in a multifarious road network space.

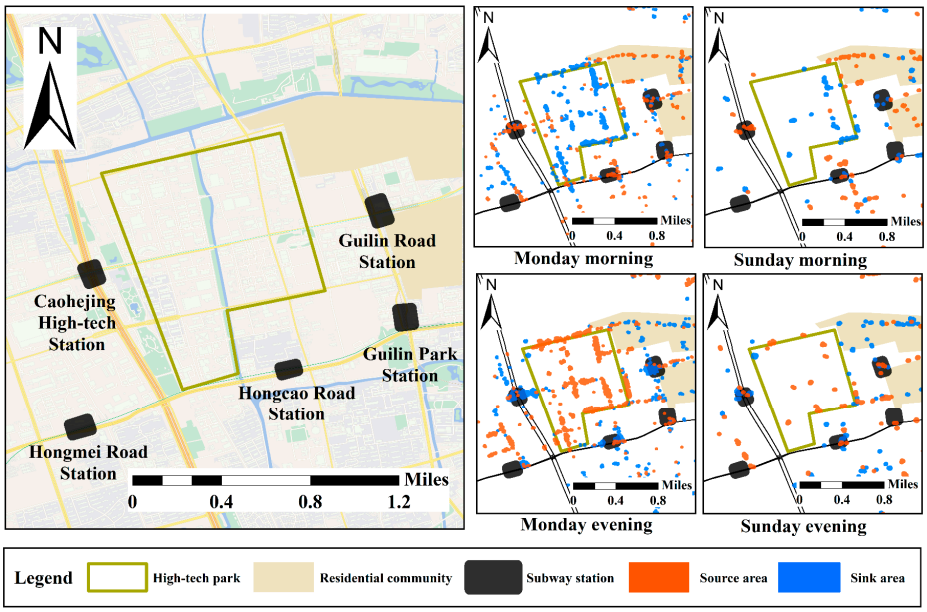


Figure 9. Spatial distribution of bicycle travel sources and sinks at Caohejing High-tech Park.

Figure 10 shows the variation in the bicycle travel source–sink patterns of the selected Caohejing High-Tech Subway Station (CH Subway Station for short). The light red color and lavender color bars indicate the dominant source and sink types within the CH Subway Station, respectively. Regardless of whether it was on workdays or weekends, the CH Subway Station showed a source pattern in the morning and a sink pattern in the afternoon. The difference is that the bicycle travel source–sink pattern on weekends was more scattered and showed instability, and that on workdays was more fixed and intense. The blue lines represent the net flow, and it is clear that the sink intensity of the CH Subway Station was greater than the source intensity on workdays. If these imbalanced areas with an arbitrary shape of shared bicycle resources are not detected in time, this can easily cause future challenges in terms of bicycle accumulation and waste in the long term.

Not all regions had stronger bicycle travel source–sink patterns on workdays than on weekends. Figure 11 shows another typical travel source–sink pattern in a commercial district and high-grade residential community. The green box is Yu Garden, a famous commercial center and scenic spot in Shanghai. It is obvious that there were no unbearable bicycle travel source–sink patterns at Yu Garden on workdays, while on weekends, the large number of sources and sinks, with a wide range, reflects the significant attraction of this commercial and entertainment area to residents and tourists on weekends. Thus, the shared bicycle resources in such areas are more imbalanced on weekends, which should be paid more attention to. Expectedly, the Yu Garden Subway Station acts as an aggregation of bicycle travel sources and sinks. Surprisingly, the residential communities around Yu Garden were not sources or sinks, showing that residents rarely ride shared bikes for transportation, which is different from the residential communities around the Caohejing High-Tech Park.

Figure 12 shows the temporal variation in the bicycle travel source and sink patterns of Yu Garden, which receives a large number of travelers, and where dockless shared bicycles take the role of "the last mile" and "the first mile" transportation. From the perspective of intraday, Yu Garden showed travel sink patterns in the morning and travel source patterns in the afternoon, which is in accordance with its commercial characteristics. From the perspective of interday, the bicycle travel source–sink patterns were more persistent and intense on weekends than on workdays. Meanwhile, the number of O points and D points on workdays reveals the same variation, indicating that Yu Garden is a bicycle dynamic area with a balanced influx and outflux on workdays.

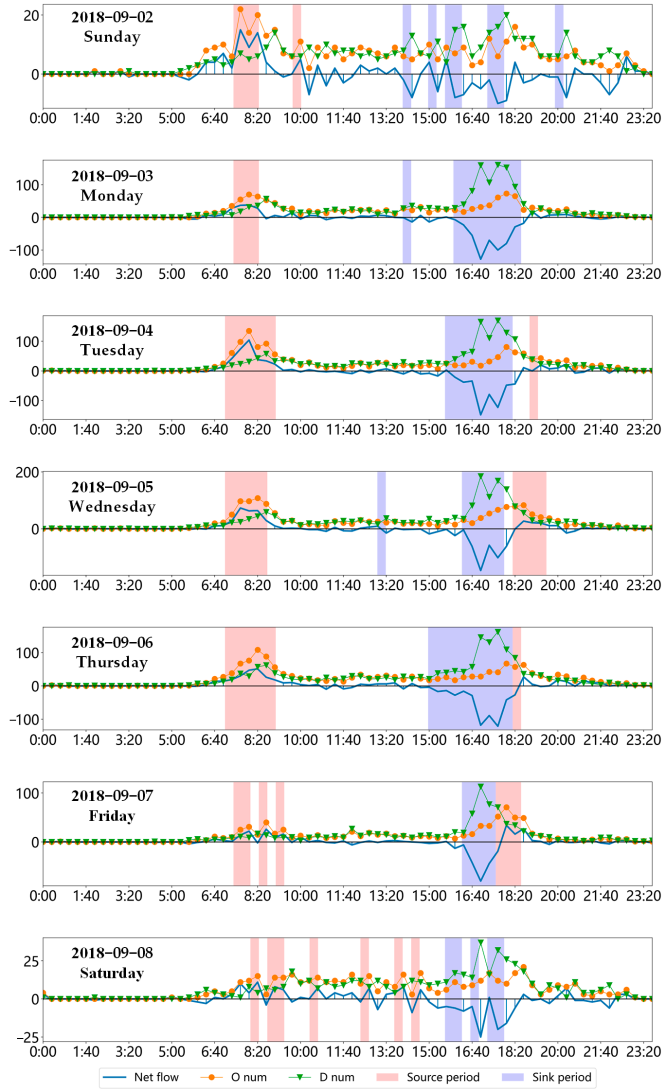


Figure 10. Temporal variation in bicycle travel sources and sinks at Caohejing High-Tech Subway Station.

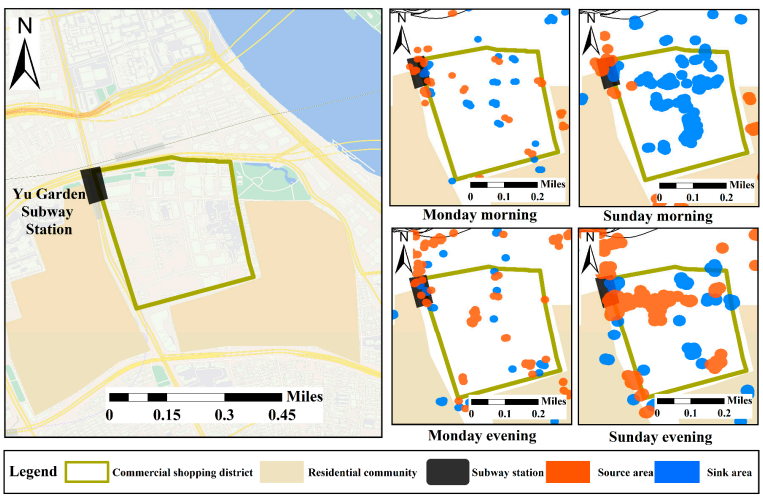


Figure 11. Spatial distribution of bicycle travel sources and sinks at the Yu Garden commercial district.

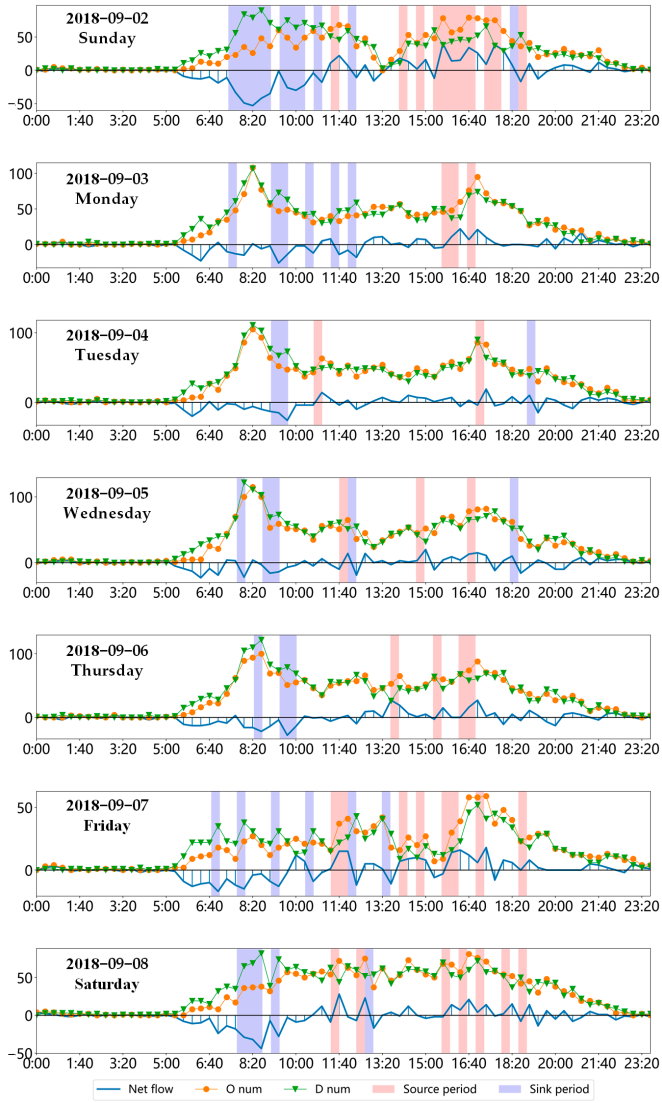


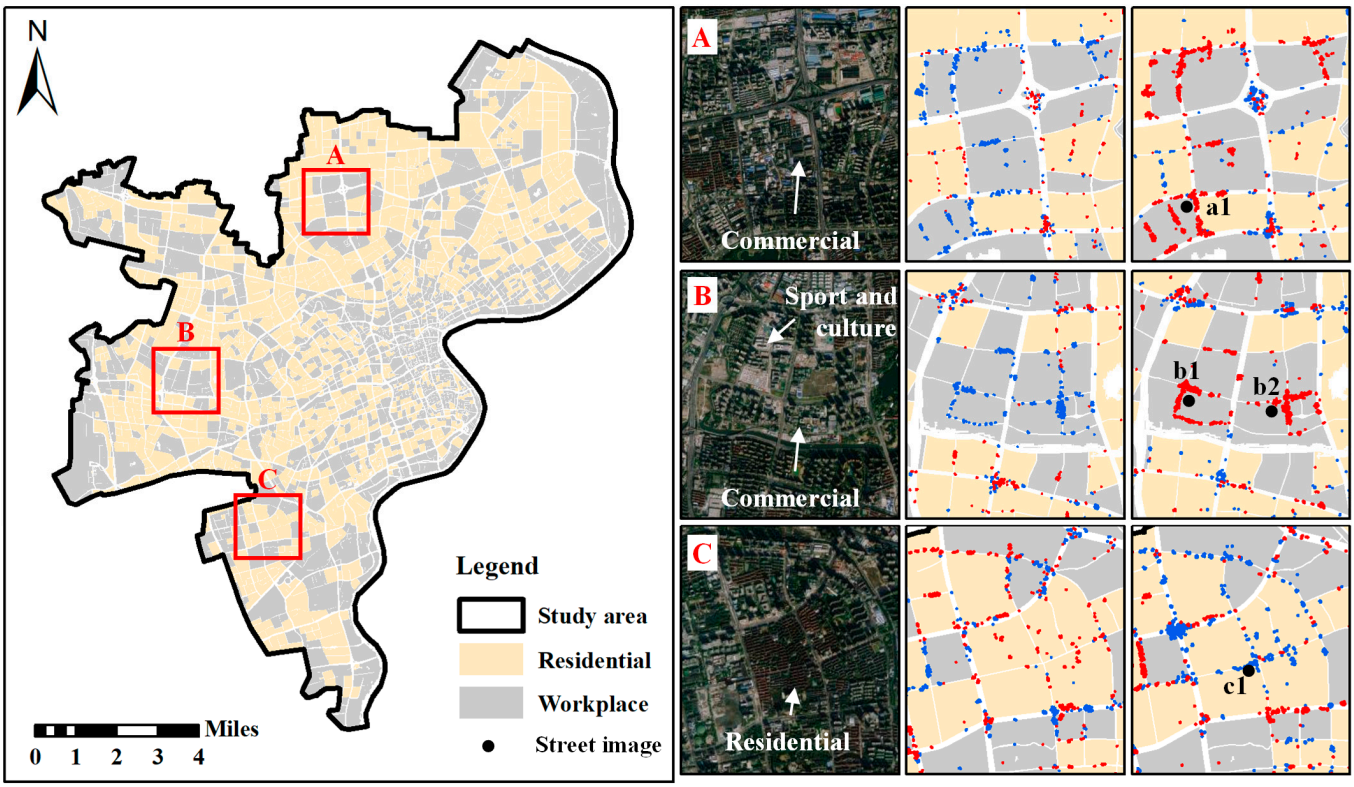
Figure 12. Temporal variation in bicycle travel sources and sinks at Yu Garden.

6. Discussion and Conclusions

6.1. Discussion

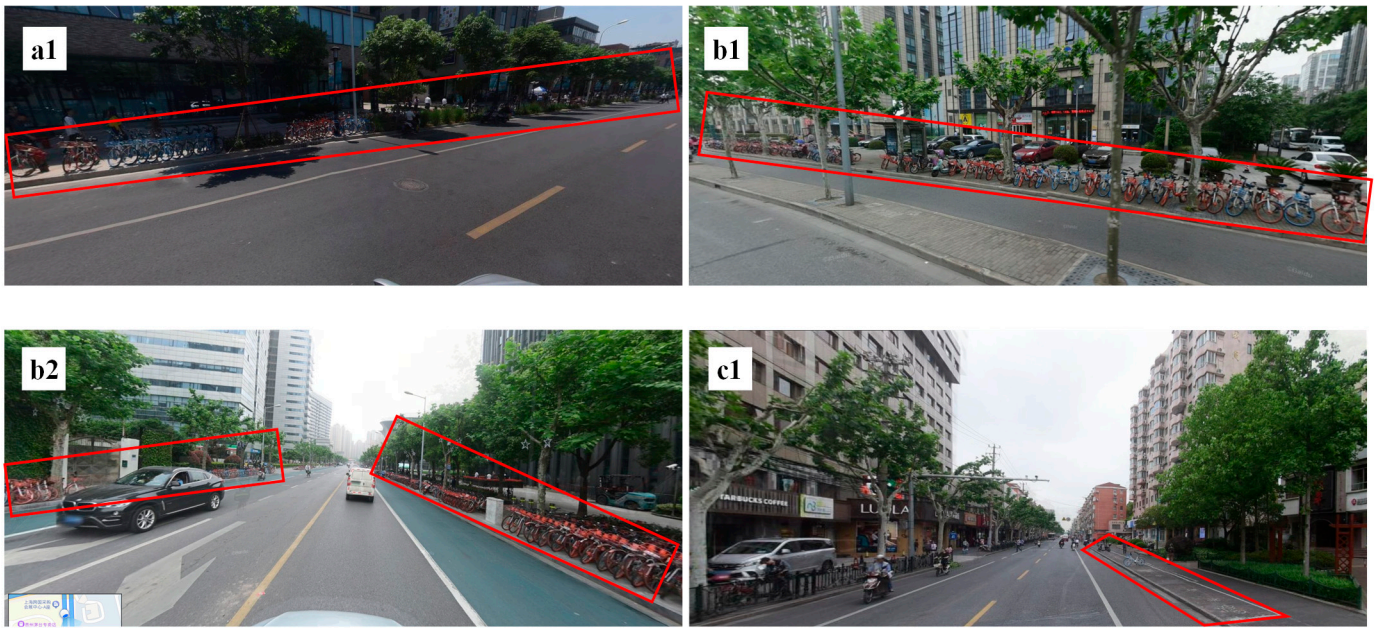
With the development of remote sensing data using images, videos, and geographical locations, combining novel data-mining approaches is an important methodological approach in digital earth sensing research. Since location sensing data do not have a regular grid arrangement like image sensing data, sensing human mobility patterns emerging from individual trajectories within urban areas is a new challenge in urban remote sensing. In this paper, we design an adaptive clustering method to extract the spatiotemporal ranges of travel source–sink patterns with imbalanced resource inflows and outflows, which can serve as a sustainable resource for sensing traffic flow anomalies such as those of shared bicycles. The proposed method addresses two challenges. On the one hand, the process based on point entity clustering alleviates the modifiable areal unit problem [47]. The experiments show that the proposed method can effectively identify arbitrarily shaped spatial clusters and micromesh periods. On the other hand, the proposed method develops spatiotemporal statistical tests for source and sink events, which do not require users to blindly set crucial parameters. Thus, the proposed method can be easily applied in the identification of shared resource imbalance, detection of traffic anomalies, and division of traffic service boundaries.

The bicycle travel source–sinks detected by the proposed method can be combined with multisource remote sensing data to analyze human mobility patterns and potential risks. As shown in Figure 13, we overlaid the bicycle travel source–sink patterns detected by the proposed method with urban land use [48] identified from remote sensing images to assess the source–sink patterns in different urban functional areas. The left panel of Figure 13a reveals the spatial distribution of the residential areas and workplaces of the study area. The right panels of Figure 13a present the satellite images and travel source–sink areas of the labeled regions (i.e., A, B, and C) in the left panel of Figure 13a, as detected by the proposed method. Figure 13b shows the street view images of the labeled representative points (i.e., a1, b1, b2, and c1) in the right panels of Figure 13a. It is well known that travel source–sink patterns may appear around residential areas and workplaces on workdays. Areas A and B are typical commercial places for finance, internet technology, e-commerce, media, and so on, which present high buildings and a hard ground. The street view images from Baidu Map show that these two areas contain a high intensity of sharing bikes and a large number of workers. Thus, the managers of shared bicycles should supervise and optimize imbalanced source–sink patterns in real time to ensure the sustainable circulation of bicycle resources in these areas. Area C denotes a middle-to-low-grade residential zone with a large area and is surrounded by commercial areas and traffic stations, where bikes provide the main transportation mode to go to work and to travel to other places. The bicycle travel sinks are stronger than the travel sources in Area C, indicating that Area C has a high risk of bicycle resource imbalance if there is no timely intervention to allocate the resource. We inspected the street view images in Area C and found that it has plenty of parking zones, but not enough shared bikes, which is a caution for shared bicycle resources' sustainability and fairness. Urban roads are easily covered by trees and street view images are static; thus, we recommend the use of other urban sensing data, such as location data, to extract urban events.



(**a**)

Figure 13. *Cont.*



(**b**)

Figure 13. The distribution of (**a**) bicycle travel source–sink patterns on 3 September (Monday) and urban land use [48] extracted by multisource remote sensing, and (**b**) the matched street view images from Baidu Map.

Using the spatiotemporal ranges of the bicycle travel source–sink patterns detected by the proposed method can provide useful insights into urban vitality sensing and the development of balanced measures of resource supply and demand. Generally speaking, a travel source represents the dissipation of a large number of traffic outflows, while a travel sink denotes the convergence of a large number of traffic inflows, and other areas without sources or sinks present a blank or a dynamic balance of inflows and outflows. For dynamically balanced areas (such as transit stations), regulators do not need to take extra efforts to ensure resource sustainability. For persistent travel source events, more attention should be paid to fairness challenges due to resource scarcity, and for persistent travel sink events, resource waste issues should be of more concern.

6.2. Conclusions

This paper proposed a density-based adaptive clustering method to sense traffic anomalies due to travel sources and sinks by utilizing bicycle trajectory data. The proposed method first constructed an optimal spatiotemporal neighborhood based on the OD points obtained from the field instability measure using kernel density estimation. Second, the spatiotemporal core entities with a high density were adaptively detected by using statistical hypothesis tests for random distribution, according to a given significance level. On this basis, the source and sink core entities with imbalanced influxes and outfluxes were identified using statistical hypothesis tests for binary uniform distribution, according to a given significance level. Finally, the spatial areas and temporal durations of the travel source and sink events were extracted using a spatiotemporal expansion strategy.

This paper detected the emergence of travel source–sink events within travel OD entities. Analyzing the driving relationship among urban structure, social economy, and travel source–sink areas based on causal inference is an important area of research to solve issues related to imbalances in inflow and outflow in the future. Further, combining the trajectories of different trips in urban areas can comprehensively determine the variation in source–sink patterns due to urban human mobility activities, which provides useful insights for urban resource intensification and the development of policy decisions.

Author Contributions: Conceptualization, Y.S. and D.W.; methodology, D.W.; software, X.W.; validation, B.C., C.D. and S.G.; formal analysis, D.W.; investigation, Y.S.; resources, Y.S.; data curation, B.C. and C.D.; writing—original draft preparation, D.W.; writing—review and editing, Y.S.; visualization, X.W.; supervision, S.G.; project administration, Y.S.; funding acquisition, Y.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key R&D Program of China (No. 2021YFB3900904), the National Nature Science Foundation of China (No. 42071452), the Hunan Provincial Natural Science Foundation of China (No. 2022JJ20059), and the Central South University Innovation-Driven Research Program (No. 2023CXQD013).

Data Availability Statement: The shared bicycle trajectory data used in this paper were purchased from companies with confidentiality agreements, and are not publicly applicable.

Acknowledgments: We thank the Editors and reviewers for their support and comments.

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

References

1. Wang, M.; Debbage, N. Urban Morphology and Traffic Congestion: Longitudinal Evidence from US Cities. *Comput. Environ. Urban Syst.* **2021**, *89*, 101676. [CrossRef]
2. Cao, C.; Zhen, F.; Huang, X. How Does Perceived Neighborhood Environment Affect Commuting Mode Choice and Commuting CO2 Emissions? An Empirical Study of Nanjing, China. *Int. J. Environ. Res. Public Health* **2022**, *19*, 7649. [CrossRef]
3. Coretti Sanchez, N.; Martinez, I.; Alonso Pastor, L.; Larson, K. On the Simulation of Shared Autonomous Micro-Mobility. *Commun. Transp. Res.* **2022**, *2*, 100065. [CrossRef]
4. Gao, K.; Yang, Y.; Gil, J.; Qu, X. Data-Driven Interpretation on Interactive and Nonlinear Effects of the Correlated Built Environment on Shared Mobility. *J. Transp. Geogr.* **2023**, *110*, 103604. [CrossRef]
5. Nieuwenhuijsen, M.J. Urban and Transport Planning Pathways to Carbon Neutral, Liveable and Healthy Cities; A Review of the Current Evidence. *Environ. Int.* **2020**, *140*, 105661. [CrossRef]
6. Zhao, X.; Ke, Y.; Zuo, J.; Xiong, W.; Wu, P. Evaluation of Sustainable Transport Research in 2000-2019. *J. Clean Prod.* **2020**, *256*, 120404. [CrossRef]
7. Bardaka, E.; Hajibabai, L.; Singh, M.P.; Murukannaiah, P.K. Reimagining Ride Sharing: Efficient, Equitable, Sustainable Public Microtransit. *IEEE Internet Comput.* **2020**, *24*, 38–44. [CrossRef]
8. Wang, Y.; Tong, D.; Li, W.; Liu, Y. Optimizing the Spatial Relocation of Hospitals to Reduce Urban Traffic Congestion: A Case Study of Beijing. *Trans. GIS* **2019**, *23*, 365–386. [CrossRef]
9. Liu, B.; Deng, M.; Yang, J.; Shi, Y.; Huang, J.; Li, C.; Qiu, B. Detecting Anomalous Spatial Interaction Patterns by Maximizing Urban Population Carrying Capacity. *Comput. Environ. Urban Syst.* **2021**, *87*, 101616. [CrossRef]
10. 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]
11. Liu, Z.; Li, J.; Wu, K. Context-Aware Taxi Dispatching at City-Scale Using Deep Reinforcement Learning. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 1996–2009. [CrossRef]
12. Jia, R.; Chamoun, R.; Wallenbring, A.; Advand, M.; Yu, S.; Liu, Y.; Gao, K.; Jia, R.; Chamoun, R.; Wallenbring, A.; et al. A Spatio-Temporal Deep Learning Model for Short-Term Bike-Sharing Demand Prediction. *Electron. Res. Arch.* **2023**, *31*, 1031–1047. [CrossRef]
13. Cai, Y.; Zhang, M.; Lin, H. Estimating the Urban Fractional Vegetation Cover Using an Object-Based Mixture Analysis Method and Sentinel-2 MSI Imagery. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2020**, *13*, 341–350. [CrossRef]
14. Guo, Y.; Wu, C.; Du, B.; Zhang, L. Density Map-Based Vehicle Counting in Remote Sensing Images with Limited Resolution. *ISPRS-J. Photogramm. Remote Sens.* **2022**, *189*, 201–217. [CrossRef]
15. Chen, J.; Dai, X.; Guo, Y.; Zhu, J.; Mei, X.; Deng, M.; Sun, G. Urban Built Environment Assessment Based on Scene Understanding of High-Resolution Remote Sensing Imagery. *Remote Sens.* **2023**, *15*, 1436. [CrossRef]
16. Yao, S.; Li, L.; Cheng, G.; Zhang, B. Analyzing Long-Term High-Rise Building Areas Changes Using Deep Learning and Multisource Satellite Images. *Remote Sens.* **2023**, *15*, 2427. [CrossRef]
17. Liu, X.; Gong, L.; Gong, Y.; Liu, Y. Revealing Travel Patterns and City Structure with Taxi Trip Data. *J. Transp. Geogr.* **2015**, *43*, 78–90. [CrossRef]
18. Xu, J.; Liu, J.; Xu, Y.; Lv, Y.; Pei, T.; Du, Y.; Zhou, C. Identification of Spatial and Functional Interactions in Beijing Based on Trajectory Data. *Appl. Geogr.* **2022**, *145*, 102744. [CrossRef]
19. Liu, Y.; Liu, X.; Gao, S.; Gong, L.; Kang, C.; Zhi, Y.; Chi, G.; Shi, L. Social Sensing: A New Approach to Understanding Our Socioeconomic Environments. *Ann. Assoc. Am. Geogr.* **2015**, *105*, 512–530. [CrossRef]
20. Liu, Q.; Huan, W.; Deng, M. A Method with Adaptive Graphs to Constrain Multi-View Subspace Clustering of Geospatial Big Data from Multiple Sources. *Remote Sens.* **2022**, *14*, 4394. [CrossRef]

21. Zhang, Z.; Zhang, Y.; He, T.; Xiao, R. Urban Vitality and Its Influencing Factors: Comparative Analysis Based on Taxi Trajectory Data. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2022**, *15*, 5102–5114. [CrossRef]
22. Pan, G.; Qi, G.; Wu, Z.; Zhang, D.; Li, S. Land-Use Classification Using Taxi GPS Traces. *IEEE Trans. Intell. Transp. Syst.* **2013**, *14*, 113–123. [CrossRef]
23. Fang, Z.; Yang, X.; Xu, Y.; Shaw, S.-L.; Yin, L. Spatiotemporal Model for Assessing the Stability of Urban Human Convergence and Divergence Patterns. *Int. J. Geogr. Inf. Sci.* **2017**, *31*, 2119–2141. [CrossRef]
24. Macqueen, J. Some Methods for Classification and Analysis of Multivariate Observations. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Oakland, CA, USA, 27 December–7 January 1967; pp. 281–297.
25. Gao, Y.; Song, C.; Guo, S.; Pei, T. Spatial-temporal Characteristics and Influencing Factors of Source and Sink of Dockless Sharing Bicycles Connected to Subway Stations. *J. Geo-Inf. Sci.* **2021**, *23*, 155–170. [CrossRef]
26. Openshaw, S. *The Modifiable Areal Unit Problem. Concepts and Techniques in Modern Geography*; GeoBooks: Norwich, UK, 1984; p. 38.
27. Zhao, P.; Qin, K.; Ye, X.; Wang, Y.; Chen, Y. A Trajectory Clustering Approach Based on Decision Graph and Data Field for Detecting Hotspots. *Int. J. Geogr. Inf. Sci.* **2017**, *31*, 1101–1127. [CrossRef]
28. Steenberghen, T.; Dufays, T.; Thomas, I.; Flahaut, B. Intra-Urban Location and Clustering of Road Accidents Using GIS: A Belgian Example. *Int. J. Geogr. Inf. Sci.* **2004**, *18*, 169–181. [CrossRef]
29. Shi, Y.; Wang, D.; Tang, J.; Deng, M.; Liu, H.; Liu, B. Detecting Spatiotemporal Extents of Traffic Congestion: A Density-Based Moving Object Clustering Approach. *Int. J. Geogr. Inf. Sci.* **2021**, *35*, 1449–1473. [CrossRef]
30. Tao, R.; Thill, J.-C. Flow Cross K-Function: A Bivariate Flow Analytical Method. *Int. J. Geogr. Inf. Sci.* **2019**, *33*, 2055–2071. [CrossRef]
31. Liu, W.; Liu, Q.; Yang, J.; Deng, M. A Network-Constrained Clustering Method for Bivariate Origin-Destination Movement Data. *Int. J. Geogr. Inf. Sci.* **2023**, *37*, 767–787. [CrossRef]
32. Liu, Q.; Yang, J.; Deng, M.; Liu, W.; Xu, R. BiFlowAMOEBA for the Identification of Arbitrarily Shaped Clusters in Bivariate Flow Data. *Int. J. Geogr. Inf. Sci.* **2022**, *36*, 1784–1808. [CrossRef]
33. Gao, Y.; Li, T.; Wang, S.; Jeong, M.-H.; Soltani, K. A Multidimensional Spatial Scan Statistics Approach to Movement Pattern Comparison. *Int. J. Geogr. Inf. Sci.* **2018**, *32*, 1304–1325. [CrossRef]
34. Liu, Q.; Wu, Z.; Deng, M.; Liu, W.; Liu, Y. Network-Constrained Bivariate Clustering Method for Detecting Urban Black Holes and Volcanoes. *Int. J. Geogr. Inf. Sci.* **2020**, *34*, 1903–1929. [CrossRef]
35. Song, C.; Pei, T.; Ma, T.; Du, Y.; Shu, H.; Guo, S.; Fan, Z. Detecting Arbitrarily Shaped Clusters in Origin-Destination Flows Using Ant Colony Optimization. *Int. J. Geogr. Inf. Sci.* **2019**, *33*, 134–154. [CrossRef]
36. Pei, T.; Wang, W.; Zhang, H.; Ma, T.; Du, Y.; Zhou, C. Density-Based Clustering for Data Containing Two Types of Points. *Int. J. Geogr. Inf. Sci.* **2015**, *29*, 175–193. [CrossRef]
37. Ester, M.; Kriegel, H.P.; Sander, J.; Xiaowei, X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. *KDD* **1996**, *96*, 226–231.
38. Shanghai Municipal Bureau of Statistics; National Bureau of Statistics Shanghai Survey Team. *Shanghai Statistical Yearbook 2022*; China Statistics Press: Beijing, China, 2023.
39. Deng, M.; Liu, Q.; Cheng, T.; Shi, Y. An Adaptive Spatial Clustering Algorithm Based on Delaunay Triangulation. *Comput. Environ. Urban Syst.* **2011**, *35*, 320–332. [CrossRef]
40. Li, Q.; Racine, J.S. *Nonparametric Econometrics: Theory and Practice*; Princeton University Press: Princeton, NJ, USA, 2006.
41. Shannon, C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 623–656. [CrossRef]
42. Davies, D.L.; Bouldin, D.W. A Cluster Separation Measure. *IEEE Trans. Pattern Anal. Mach. Intell.* **1979**, *PAMI-1*, 224–227. [CrossRef]
43. Caliński, T.; JA, H. A Dendrite Method for Cluster Analysis. *Commun. Stat.—Theory Methods* **1974**, *3*, 1–27. [CrossRef]
44. Arbelaitz, O.; Gurrutxaga, I.; Muguerza, J.; Pérez, J.M.; Perona, I. An Extensive Comparative Study of Cluster Validity Indices. *Pattern Recognit.* **2013**, *46*, 243–256. [CrossRef]
45. Guo, D.; Zhu, X.; Jin, H.; Gao, P.; Andris, C. Discovering Spatial Patterns in Origin-Destination Mobility Data. *Trans. GIS* **2012**, *16*, 411–429. [CrossRef]
46. Birant, D.; Kut, A. ST-DBSCAN: An Algorithm for Clustering Spatial–Temporal Data. *Data Knowl. Eng.* **2007**, *60*, 208–221. [CrossRef]
47. Openshow, S.; Taylor, P.J. A Million or so Correlation Coefficients: Three Experiments on the Modifiable Areal Unit Problem. In *Statistical Applications in the Spatial Sciences*; Wrigley, N., Ed.; Pion: London, UK, 1979; pp. 127–144.
48. Gong, P.; Chen, B.; Li, X.; Liu, H.; Wang, J.; Bai, Y.; Chen, J.; Chen, X.; Fang, L.; Feng, S.; et al. Mapping Essential Urban Land Use Categories in China (EULUC-China): Preliminary Results for 2018. *Sci. Bull.* **2020**, *3*, 182–187. [CrossRef] [PubMed]