Article

Understanding Spatiotemporal Patterns of Human Convergence and Divergence Using Mobile Phone Location Data

Xiping Yang [1], Zhixiang Fang [1,2,*], Yang Xu [3,*], Shih-Lung Shaw [1,2,4], Zhiyuan Zhao [1], Ling Yin [5], Tao Zhang [6] and Yunong Lin [7]

[1]   State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, 129 Luoyu Road, Wuhan 430079, China; 0yangxiping0@163.com (X.Y.); sshaw@utk.edu (S.-L.S.); zhaozhiyuan@whu.edu.cn (Z.Z.)
[2]   Collaborative Innovation Center of Geospatial Technology, 129 Luoyu Road, Wuhan 430079, China
[3]   Senseable City Laboratory, SMART Centre, 1 Create Way, Singapore 138602, Singapore
[4]   Department of Geography, University of Tennessee, Knoxville, TN 37996, USA
[5]   Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, 1068 Xueyuan Road, Shenzhen 518005, China; yinling@siat.ac.cn
[6]   Business Support Center, Hubei Mobile, 2 Jinyinhu Road, Wuhan 430040, China; zhangtao2@hb.chinamobile.com
[7]   School of Mathematical Sciences, Peking University, 5 Yiheyuan Road Haidian District, Beijing 100871, China; 1100011751@pku.edu.cn
*   Correspondence: zxfang@whu.edu.cn (Z.F.); yangxu@smart.mit.edu (Y.X.); Tel.: +86-276-877-9889 (Z.F.)

Abstract: Investigating human mobility patterns can help researchers and agencies understand the driving forces of human movement, with potential benefits for urban planning and traffic management. Recent advances in location-aware technologies have provided many new data sources (e.g., mobile phone and social media data) for studying human space-time behavioral regularity. Although existing studies have utilized these new datasets to characterize human mobility patterns from various aspects, such as predicting human mobility and monitoring urban dynamics, few studies have focused on human convergence and divergence patterns within a city. This study aims to explore human spatial convergence and divergence and their evolutions over time using large-scale mobile phone location data. Using a dataset from Shenzhen, China, we developed a method to identify spatiotemporal patterns of human convergence and divergence. Eight distinct patterns were extracted, and the spatial distributions of these patterns are discussed in the context of urban functional regions. Thus, this study investigates urban human convergence and divergence patterns and their relationships with the urban functional environment, which is helpful for urban policy development, urban planning and traffic management.

1. Introduction

Cities comprise flows of information, goods and people. Among these urban flows, human movements are critical components that drive the pulses of cities. Examining people flows and their spatiotemporal dynamics has always been an important task for a wide range of disciplines, e.g., GIScience, transportation, epidemiology, etc. Traditionally, our ability to capture timely and spatially-detailed human mobility data has been constrained by available resources and data collection techniques [1–3]. However, recent advances in location-aware technologies have produced new data sources,

e.g., mobile phones, smart cards and social media that detail the movements of people in their daily lives. Consequently, studies have addressed various research challenges related to urban vitality [3–5], mobility prediction [6,7] and transportation modelling [8,9]. These studies have enhanced our understanding of human mobility patterns in urban contexts. In this study, we attempt to improve the research in this field and focus on analyzing spatiotemporal patterns of human convergence and divergence in cities.

Convergence to a location suggests that the number of people flowing to a location is larger than the number of outgoing people. Conversely, divergence from a location suggests that the number of people leaving the location is larger than the number of incoming people. An understanding of how people flows converge and diverge in space and time in cities, as well as their relationships with urban land use can provide insight regarding urban dynamics and potentially benefit urban planning and public transportation management in cities. Therefore, the main research questions of this study are as follows:

1.　What spatiotemporal patterns of human convergence and divergence exist in the daily urban context?
2.　What types of urban land use are generally associated with these patterns?

To address these two questions, this study uses a large-scale mobile phone dataset collected in Shenzhen, China, on a weekday to investigate spatiotemporal patterns of human convergence and divergence. Unlike call detail records (CDRs) that only capture individual footprints during actual communication [10,11], the mobile phone dataset used in this study tracks individuals regularly over time (approximately once every hour) at the cell phone tower level, which enables us to investigate human convergence and divergence patterns with relatively fine and regular spatiotemporal resolution. These identified patterns reflect the essential characteristics of human travel patterns at different locations within the city and have implications for transportation planning, emergency response and epidemic control.

2. Literature Review

The development of information and communication technologies has profound implications for human sociology and physical mobility and makes it possible to collect large sets of georeferenced data from location-based devices, such as mobile phones, which creates new opportunities for understanding human mobility patterns and their relationship with urban functional environments [12–14].

Human mobility is closely related to urban transport and planning and is an important research topic in urban studies. For example, an individual's home and workplace can be identified from mobile phone data, and origin-destination flow matrices can be constructed to investigate commuting patterns [15–17]. Real-time traffic speeds and travel times can be measured using a cellular phone-based system [18]. In addition, real-time urban dynamics can be captured using mobile phone data to monitor human spatiotemporal distributions and provide insight into the real-time intensity of human activities in different urban areas [4,19–21]. Human mobility hotspots and dense areas can be detected by analyzing the trajectories and densities of cell phone users in urban environments [22–25].

Guo et al. [26] extracted pick-up and drop-off details from taxi trajectory data and proposed a hierarchical clustering method to map human flows with similar origins and destinations. Human mobility source-sink areas can also be identified based on temporal variations in pick-up and drop-off locations [27]. Mobility networks can also be created from human movements, reflecting the spatial interactions of different urban areas and communities, or areas with close connections can be detected and used to evaluate and optimize urban planning [28,29].

There is a strong relationship between human mobility and the functional environment [27,30]. The spatial distribution of different urban functional regions (e.g., residential, industrial or commercial) determines human activity locations, such as living, working, shopping and leisure. The spatial

separation of these functional regions and the demands of human activities lead to human flows in urban space. Functional differences associated with different types of land use appear as different human mobility patterns. Thus, land use information can be used to estimate travel demands in different urban areas (i.e., a land use-transport interaction model) [31]. The temporal population variation reflects the underlying function of the location. Thus, some studies have built temporal feature vectors for human activities at the grid cell level using human sensing data and machine learning methods to classify those vectors and infer urban land use information [32–34]. The classification accuracy decreases as the heterogeneity of land use increases, but additional information (e.g., spatial interaction patterns and points of interest) can be incorporated to identify different functional regions and improve the accuracy [35,36].

These studies demonstrate the powerful potential of emerging big data in research regarding human mobility patterns and the relationships between human mobility patterns and the urban functional environment. This study adds to this knowledge base by investigating the spatiotemporal patterns of human convergence and divergence in a city environment.

3. Study Area and Dataset

The study area for this research is Shenzhen, which is located in southern China. Shenzhen has experienced rapid development associated with reform policies over the past 30 years, and the area has attracted a large number of immigrant workers seeking job opportunities. The total area of Shenzhen is approximately 1996 square kilometers, and the population is more than 15 million, reflecting the highest population density among Chinese cities [37].

The mobile phone location dataset used in this study was collected by a mobile phone company that includes approximately 60% of the entire mobile phone market in Shenzhen. It covers 16 million mobile phone users over a single workday and records the cell phone tower locations each cell phone connects to approximatively every hour. Thus, each cell phone has 24 records each day containing a user ID, recording time and longitude and latitude of the cell phone tower. The user ID was encrypted for privacy protection before the dataset was released for research purposes. Table 1 shows an example of an individual user's mobile phone records for a day. In total, 5940 cell phone towers (CPTs) with unique Tower ID numbers were extracted from the dataset. Figure 1 shows the spatial kernel density of the cell phone towers.
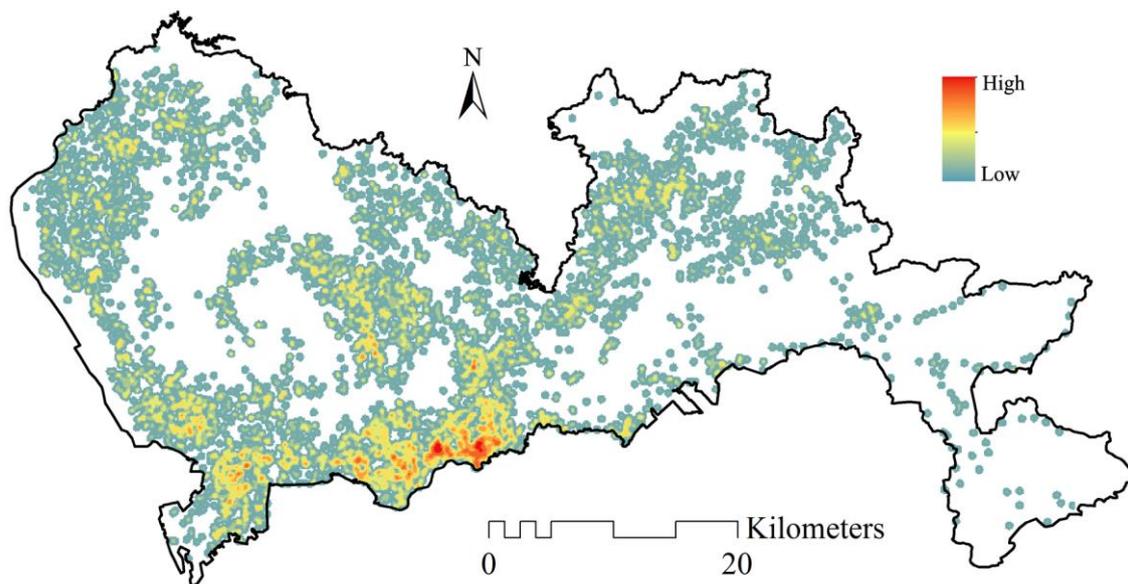

Figure 1. Spatial kernel density of the cell phone towers (CPTs).

Table 1. Example of an individual's cell phone records during a day.

User ID	Record Time	Longitude	Latitude
8d5b2b5******	00:25:36	113.***	22.***
8d5b2b5******	01:26:40	113.***	22.***
8d5b2b5******	02:20:53	113.***	22.***
8d5b2b5******	. . .	. . .	. . .
8d5b2b5******	23:33:50	113.***	22.***

The sign *** ignores the minutes of a Longitude or a Latitude and the sign ****** ignores last six numbers of a User ID due to privacy protection.

The other dataset used in this study comprised urban functional region data, which was generated from the comprehensive plan of Shenzhen city (2010–2020) [38]. This dataset includes ten functional region types: administrative (government agencies), commercial, industrial, residential, education, transport, tourism (scenic places and parks), sports, water and other (including agricultural, shrubs, bare land, etc.). Figure 2 shows the spatial distribution of urban functional regions.
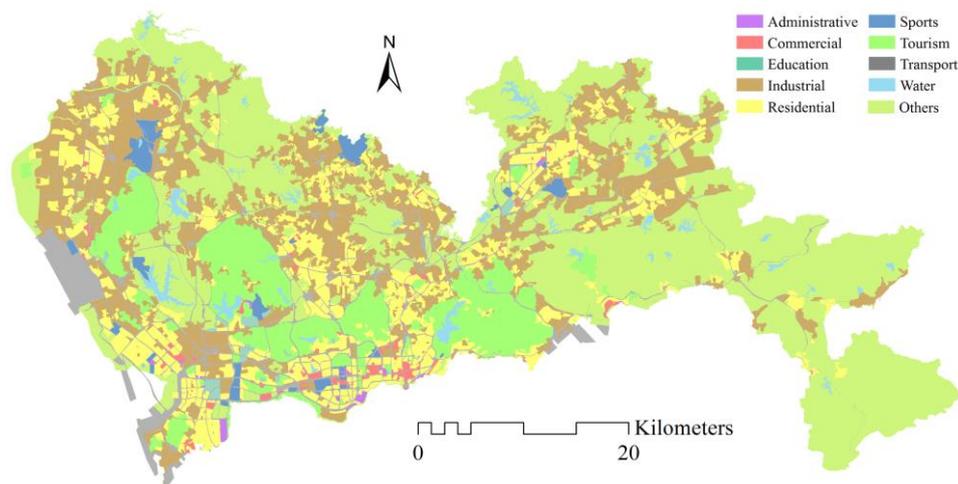
Figure 2. Spatial distribution of urban functional regions.

4. Methodology

The method used to identify the spatiotemporal patterns of human convergence and divergence included three main steps. First, we extracted the net flow from human space-time trajectories in each time slot to indicate human convergence and divergence. Then, we classified the netflow into ten classes according to quantile rules and categorized each grid cell to represent the human convergence and divergence intensity. Finally, a time series matrix was constructed based on the netflow classes, and the grid cells were grouped into clusters according to their temporal patterns.

4.1. Extracting Indicators of Human Convergence and Divergence

Using a concept of time geography [39], we constructed the space-time trajectory of each cell phone by connecting location records in chronological order. As shown in Figure 3, the cell phone trajectory can be represented as follows:

$$Tr = [p_1(x_1, y_1, t_1, Id_1), \cdots, p_i(x_i, y_i, t_i, Id_i), \cdots, p_n(x_n, y_n, t_n, Id_n)] \tag{1}$$

where x_i, y_i and Id_i represent the longitude, latitude and TowerID of record point p_i, respectively, and t_i represents the time when the point update occurred. For adjacent space-time points with different

record locations, we can extract a movement from cell phone tower Id_i to Id_{i+1} over time period t_i to t_{i+1}.

$$[p_i(x_i, y_i, t_i, Id_i), p_{i+1}(x_{i+1}, y_{i+1}, t_{i+1}, Id_{i+1})], \ Id_i \neq Id_{i+1} \tag{2}$$
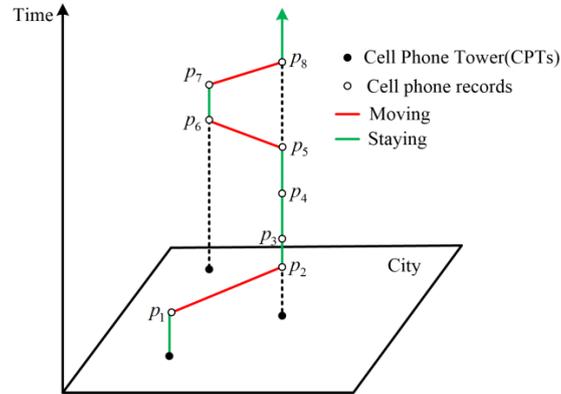
Figure 3. Space-time trajectory of an individual cell phone record.

Table 1 shows that the time window of the location records was updated approximately every hour, e.g., the first point was recorded between 00:00 and 01:00 and the second between 01:00 and 02:00. A movement can be extracted between 00:00 and 02:00, and the time window from 00:00–02:00 is considered time slot T_1. Thus, we can extract one movement for every two adjacent hours, and the day can be divided into 23 time slots, with T_j denoting the time window $(j-1)$:00–$(j+1)$:00.

One issue is that there may be signal switches between CPTs, which may be incorrectly interpreted as movements, particularly in areas with high tower densities [40,41]. We adopted Thiessen polygons to represent the service area of a cell phone tower in the early stage of this study. We found that some cell phone towers are located very close to each other. Overall, 396 cell towers are very close to nearby towers, and the distance between towers can be less than 10 m. For example, two cell towers may be located in the same high-rise building. These close cell phone towers can cause frequent signal jumps between the towers. We chose to use regular grid cells to aggregate very close cell phone towers, thereby reducing the influence of signal switches. We divided the city using different grid sizes from 100 m × 100 m–2 km × 2 km with an increment of 100 m and found that the 500 m × 500 m grid cells of cell phone towers accounted for 90.2% of the major human activity areas, which was much larger than the percentage in grid cells less than 500 m × 500 m. In addition, we found that the movements within grid cells increase linearly, and movements between grid cells decrease linearly with grid size. The 500 m × 500 m grid cells ignored approximately 16% of movements. Although 600 m × 600 m grid cells cover 98% of major human activity areas, they ignored approximately 20% of movements. Therefore, we chose grid cells of 500 m × 500 m as the analysis unit. The resolution provided a relatively fine scale for studying human mobility. Grid cells not containing a CPT were excluded because human movements could not be calculated between grid cells without cell phone towers. In total, 2801 grid cells were used as basic analysis units, and each was tagged with a unique Grid ID.

We filtered movements between CPTs to generate movements between grid cells by ignoring movements for which the origin and destination CPTs were in the same grid cell. Thus, we extracted a grid cell-based flow matrix (p, q, f_{pq}, T_j), where p and q are the origin and destination Grid IDs, respectively, f_{pq} represents the number of people moving from p to q, and T_j represents the time slot. For each grid cell p, the inflow and outflow during a time slot are computed as follows.

$$inflow_p = \sum_q f_{qp}, \ outflow_p = \sum_q f_{pq} \tag{3}$$

Additionally, the netflow of the grid cell is computed as follows.

$$netflow_p = inflow_p - outflow_p \qquad (4)$$

Netflow was used as an indicator of human convergence and divergence in a grid cell during time slot T_j. Compared to the call activity of CDRs, which reflects activity intensity, netflow reflects the difference in inflow and outflow, which indicates the change in the number of people in a cell during a time slot [42]. A positive netflow indicates that the number of people in the grid cell increased during the time slot, i.e., convergence, and a negative netflow indicates a decreasing number of people, i.e., divergence.

4.2. Classification of Human Convergence and Divergence Using Quantile Rules

This study examined human convergence and divergence, and their varying intensities over a day. We aggregated netflow values from all time slots and then grouped them into different classes, where $n_{i,j}$ represents the netflow of grid cell i during time slot T_j. The netflow set $N = \{n_{i,j}\}$ of the whole study region included 2801×23 values, with the distribution shown in Figure 4a. Most netflow values (95.4%) were between -1000 and 1000, which indicates that few locations have extremely large netflows. Additionally, the city can be considered a relatively homogeneous system.
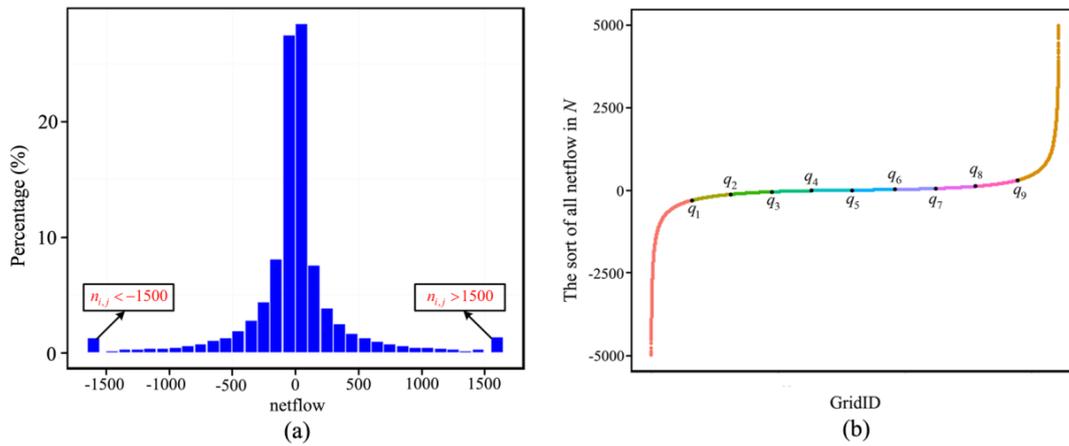

Figure 4. (**a**) Distribution of set N (bin width = 100); (**b**) sorting and break points of set N.

Netflow was then sorted in ascending order and grouped into ten classes by quantiles, producing the quantile vector $Q = [q_1, q_2, \ldots, q_9]$, where $q_1, q_2, \ldots, q_9$ represent the netflow values of nine break points in quantiles 10%, 20%, ..., and 90%, respectively (Figure 4b). In this paper, we generated the quantile vector of break points $Q = [-317, -128, -53, -18, -1, 15, 51, 122, 314]$. We use Q to classify each $n_{i,j}$ of N into different groups and assign it a level label to represent the intensity of convergence or divergence as shown in Table 2. The greater the strength of convergence or dispersion is, the larger the absolute level value is assigned. In Classes 5 and 6, convergence and divergence are relatively small, and we consider both at the same level of 0. After classification, we generate the corresponding set $L = \{l_{i,j}\}$, which indicates the intensity of human mobility of grid cell i in time slot T_j.

Table 2. Classification and labeling rules for $n_{i,j}$, where $q_1, q_2, \ldots, q_9$ represent the netflow values of nine break points in quantiles 10%, 20%, ..., and 90%, respectively.

Class	Classification	Level (*l*)	Status	Class	Classification	Level (*l*)	Status
1	$n_{i,j} < q_1$	-4	Divergence	6	$q_5 \leq n_{i,j} < q_6$	0	No
2	$q_1 \leq n_{i,j} < q_2$	-3	Divergence	7	$q_6 \leq n_{i,j} < q_7$	1	Convergence
3	$q_2 \leq n_{i,j} < q_3$	-2	Divergence	8	$q_7 \leq n_{i,j} < q_8$	2	Convergence
4	$q_3 \leq n_{i,j} < q_4$	-1	Divergence	9	$q_8 \leq n_{i,j} < q_9$	3	Convergence
5	$q_4 \leq n_{i,j} < q_5$	0	No	10	$n_{i,j} \geq q_9$	4	Convergence

4.3. Cluster Analysis of the Temporal Patterns of Human Convergence and Divergence

We transformed L into a time series matrix, V, to extract the spatiotemporal patterns of human convergence and divergence:

$$V = \begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_m \end{bmatrix} = [L_1, L_2 \cdots, L_{23}] = \left\{ \begin{array}{c} l_{1,1}, l_{1,2}, \cdots, l_{1,23} \\ l_{2,1}, l_{2,2}, \cdots, l_{2,23} \\ \cdots \\ l_{m,1}, l_{m,2}, \cdots, l_{m,23} \end{array} \right\} \tag{5}$$

where V_i represents the i-th row of the matrix, which indicates the variation in grid cell i over the day. There are 2801 rows in the matrix. L_j represents the j-th column of the matrix, which indicates the level in 2801 grid cells at time slot T_j, so there are 23 time slots. Table 3 provides examples of the matrix. The temporal characteristics of V incorporate the human mobility spatiotemporal dynamics of different areas of the city. For example, residential and commercial regions or workplaces located downtown or on the outskirts of the city may have different temporal patterns.

$$S = \sqrt{\sum_{t=1}^{23} (V_i^t - V_j^t)^2} \tag{6}$$

In the cluster analysis, our main goal is to extract these grid cells with similar levels of variation in human mobility, so we focus on clustering the rows in the matrix. As shown in Equation (6), the similarity between any two rows is calculated based on the Euclidean distance. An X-means clustering algorithm was adopted to cluster the time series matrix according to temporal characteristics. This algorithm is an improved method based on k-means and can automatically determine the number of clusters using Bayesian information criteria to overcome the drawbacks of k-means in choosing the number of clusters. It also accelerates the computation by using a *kd*-tree method to address the massive number of records [43]. Additionally, it is an unsupervised clustering method that is suitable for multidimensional variable datasets. The well-known data mining tool WEKA was employed to execute the X-means algorithm [44]. Based on the algorithm, eight clusters were extracted from V using X-means clustering, and they were denoted as $C1, C2, \ldots, C8$. A cluster analysis identified grid cells with similar human convergence and divergence variation patterns, and we discuss the characteristics of each cluster in Section 5.2.

Table 3. Examples of the matrix.

GridID	1	2	3	4	5	6	7	… …	17	18	19	20	21	22	23
211	2	1	1	0	1	2	4	… …	−3	−4	−2	−3	−1	−1	1
1056	−1	0	0	0	1	−3	−3	… …	2	3	3	2	1	1	1
⋮								⋮							
2135	1	1	0	0	1	2	2	… …	−2	−3	2	−3	−2	−1	1

5. Results and Discussion

5.1. Convergence and Divergence in each Time Slot

Figure 5 shows human convergence and divergence for selected time slots. Areas where people converged and diverged in different time slots are clearly distinguishable. Changes in human mobility intensity can also be observed. The level of most grid cells is close to zero at midnight (T_3), aside from a few areas in the urban centers. As dawn arrives, human mobility increases due to the morning commuting peak (T_8) and then declines as people start their work (T_{10}). The mobility intensity in some locations increases at noon (T_{12}) due to activities related to lunch, especially in the northern regions of the city. Then, it decreases again during the afternoon work hour (T_{15}) to a level below

that of the morning work hour (T_{10}). The evening commute (T_{18}) displays an opposite trend as T_8, with most grid cells exhibiting a high convergence during T_8 as people flow into locations that exhibit divergence at T_8, and this state can last until the evening hour (T_{21}). These patterns represent a typical urban workday dynamic that is related to human activity patterns, and it demonstrates the potential of mobile phone data for studying human mobility. These data can be used to understand aggregate mobility patterns on more detailed spatial and temporal scales.
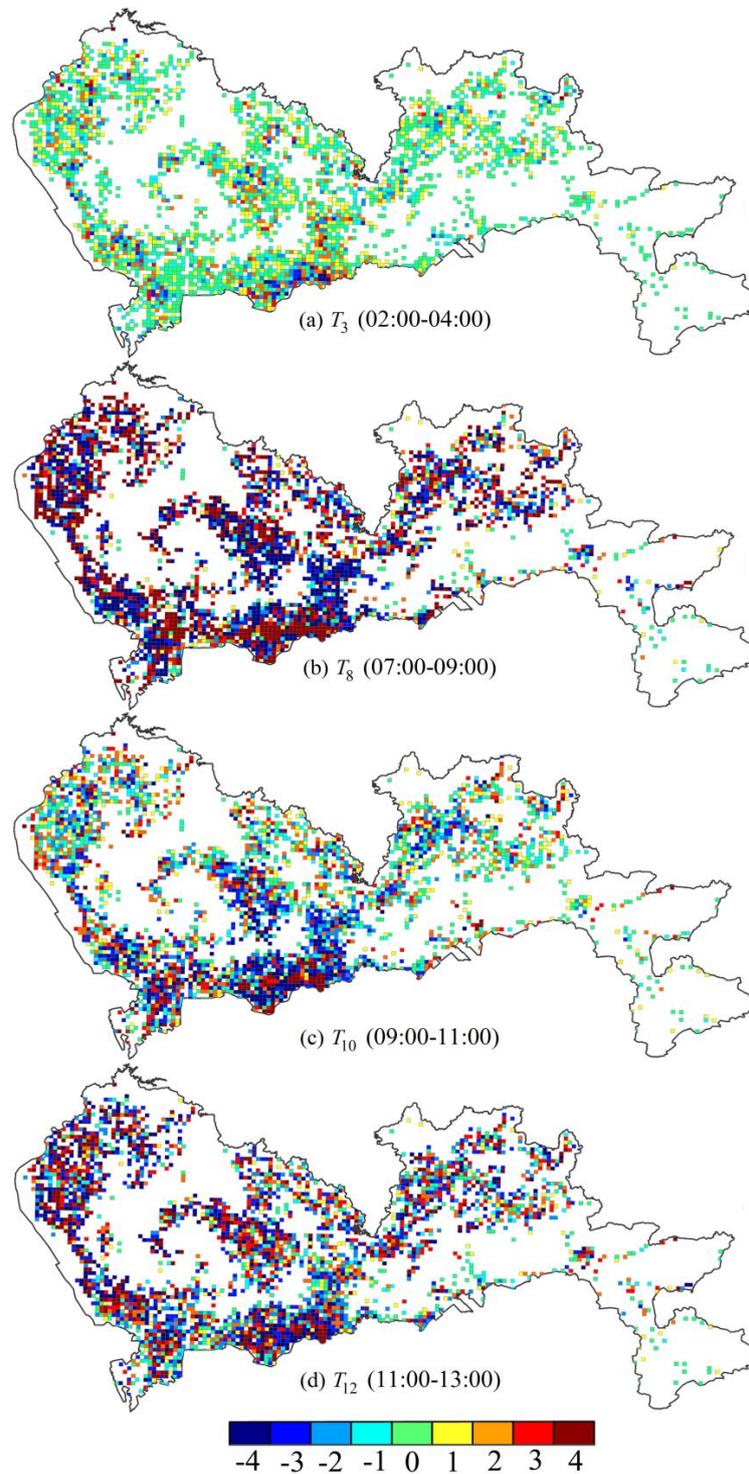
(a) T_3 (02:00–04:00)

(b) T_8 (07:00–09:00)

(c) T_{10} (09:00–11:00)

(d) T_{12} (11:00–13:00)

-4 -3 -2 -1 0 1 2 3 4
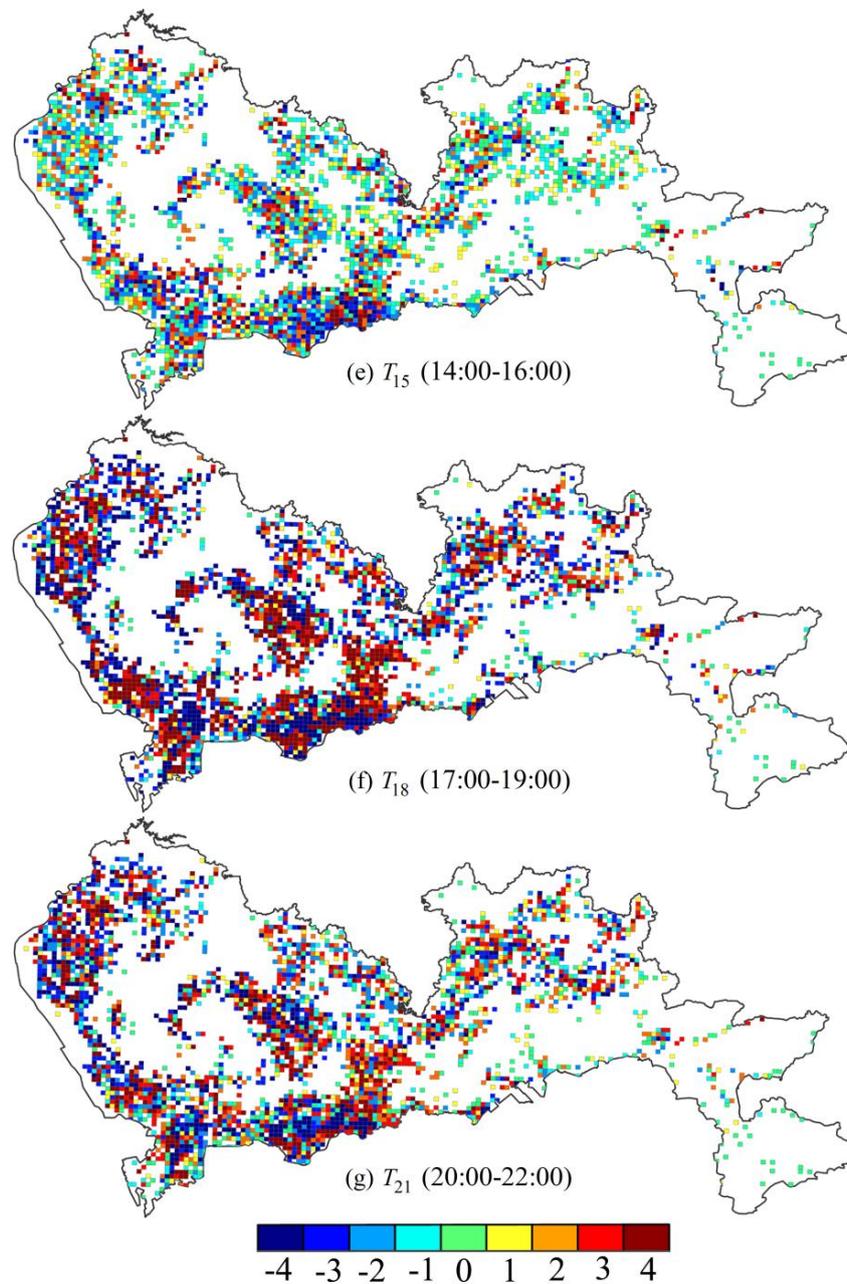
Figure 5. *Cont.*

Figure 5. Human convergence and dispersion in selected time slots. (**a**) Spatial distribution of human convergence and divergence during time slot T_3; (**b**) Spatial distribution of human convergence and divergence during time slot T_8; (**c**) Spatial distribution of human convergence and divergence during time slot T_{10}; (**d**) Spatial distribution of human convergence and divergence during time slot T_{12}; (**e**) Spatial distribution of human convergence and divergence during time slot T_{15}; (**f**) Spatial distribution of human convergence and divergence during time slot T_{18}; (**g**) Spatial distribution of human convergence and divergence during time slot T_{21}.

5.2. Temporal Patterns of Human Convergence and Divergence

Figure 6 illustrates the temporal patterns of the average values of each cluster. Distinct temporal characteristics can be observed between the clusters.

Grid cells in *C1* illustrate the high intensity of human convergence during most time slots, while *C8* cells display divergence during most of the day, except during the morning commute (T_6–T_8) when

the cells display high-intensity convergence. Grid cells in *C2* show convergence from T_6–T_{18} followed by high-intensity divergence from T_{19} until midnight (T_{23}).

C3 and *C4* have similar mobility patterns, with divergence mainly occurring from T_6–T_{10} and convergence after T_{17}. The major difference between these clusters is that the mobility intensity in *C4* is significantly higher than that in *C3*. *C3* also exhibits a clear convergence-divergence pattern from T_{11}–T_{14}.

Cluster *C5* shows a distinct convergence pattern during the morning and evening commutes, which last approximately two time slots, and divergence in the remaining time slots of the day.

C7 shows an opposite human mobility pattern to that of *C3*, with convergence mainly occurring from T_7–T_9 and divergence after T_{17}.

Compared to other clusters, there is no apparent temporal pattern in the grid cells of *C6*, and the mobility intensity is generally low.

The spatial distributions and mobility intensities of these human convergence and divergence patterns are associated with the spatial distribution of different land use types (e.g., residential, industrial, commercial, etc.) and the socioeconomic features of the geographical contexts [4,45,46].
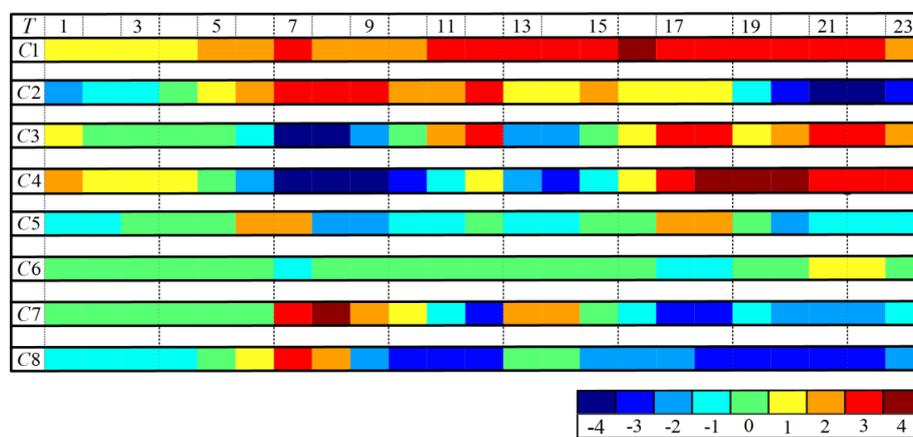

Figure 6. Clustering patterns of human convergence and divergence.

5.3. Spatial Distribution of Derived Clusters

We further analyzed the spatial distribution of the identified clusters by combining functional regions to gain better understanding of human convergence and divergence in the urban context. To simplify the maps, hollow cells were used to represent grid cells. In addition, we calculated the average percentages of different land uses in each cluster. We first calculated the proportion of each land use in each grid cell. Then, for grid cells belonging to a certain cluster, we calculated the average proportion of each land use. Table 4 lists the average percentages of the different land use types in each cluster.

Table 4. The distribution of land use in each cluster. Com, commercial land; Ind, industrial land; Res, residential land; Tra, transport land; Adm, administrative land; Edu, education land; Tou, tourism land; Spo, sport land; Wat, water land; Oth, other land (%).

Clusters	Com	Ind	Res	Tra	Adm	Edu	Tou	Spo	Wat	Oth
C1	0.3	31.3	30.3	15.2	0.2	0.3	8.0	1.3	0.5	12.6
C2	11.6	36.3	29.4	12.5	1.1	0.7	3.3	2.0	0.3	2.8
C3	0.6	32.0	50.4	6.6	0.0	0.7	3.0	0.4	0.0	6.3
C4	1.6	12.5	67.6	8.8	0.1	0.2	6.4	0.2	0.1	2.5
C5	3.4	31.1	40.1	12.5	0.3	1.0	4.4	0.8	0.1	6.3
C6	1.7	28.8	27.9	8.5	0.4	1.5	9.5	2.7	0.4	18.6
C7	2.4	58.4	11.8	9.8	0.6	1.7	3.5	1.6	0.0	10.2
C8	1.7	41.7	16.5	18.4	1.7	1.1	7.6	1.8	0.1	9.4

Figure 7 shows the spatial distribution of *C*1 and *C*8. It is counterintuitive that some areas continue to converge (*C*1) or diverge (*C*8) during most time slots (Figure 6). Most grid cells in these clusters are along the main roads of Shenzhen, and the average percentage of transportation land use in each grid cell in the two clusters is 15.2% and 18.4%, which are higher than the values in other clusters (Table 4). *C*1 cells tend to be on the boundary between industrial and residential regions, with industrial and residential land use accounting for 31.3% and 30.3%, respectively, of all land use in the cells (Table 4). *C*8 cells are mainly distributed along roads in industrial and downtown regions, and industrial and residential land use accounts for 41.7% and 16.5%, respectively, of land use in the cells. Thus, a large number of people flow into these regions during the morning commute (T_7 and T_8). The regions include some important intra-urban traffic junctions, as well as several inter-urban transportation hubs connected to nearby cities, e.g., several high-speed intersections, two railway stations and Futian Port (which connects to Hong Kong). Therefore, it is likely that the human mobility patterns in *C*1 and *C*8 are related to urban transportation. A possible explanation for the continuous convergence and divergence is that our dataset does not include interactions with nearby cities and neglects outflow from the city and inflow from other cities through these grid cells; thus, there is continuous positive or negative netflow during the day. This indicates that these areas may be main hubs that are closely connected to regions outside the city. This observation provides a reference for urban planners to locate and optimize urban bus public transit, so that people can be easily transferred from these places. Therefore, it is likely that *C*1 and *C*8 are often located along main urban roads.
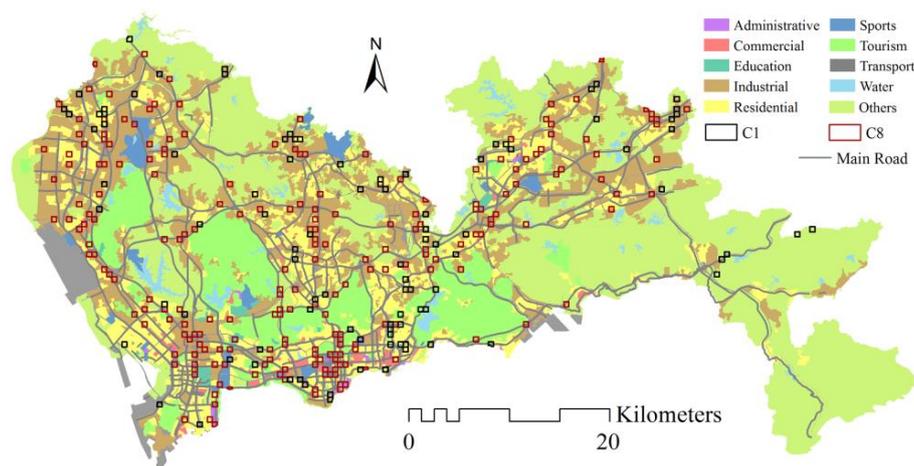

Figure 7. Spatial distributions of identified functional clusters *C*1 and *C*8.

Figure 8 shows the spatial distributions of grid cells in clusters *C*2 and *C*5. *C*2 grid cells are located in main commercial and industrial regions in the city, i.e., concentrated job locations that attract many people during the morning commute. The average commercial land use in this cluster is 11.6%, which is the maximum among all clusters (Table 4). The commercial regions also include many shopping malls, restaurants, financial institutions and recreational venues (bars, karaoke, entertainment, etc.). Therefore, these locations also attract numerous people for shopping, meals, entertainment and other activities during the daytime, with high-intensity divergence after T_{19}. Grid cells in *C*5 are mainly located near small business districts and workplaces inside residential regions, and the commercial, industrial and residential land uses are 3.4%, 31.1% and 40.1% in this cluster, respectively (Table 4). Land use in residential regions is mixed and includes shopping malls, restaurants and recreational venues. Therefore, human mobility in these locations does not exhibit a consistent pattern, and the human mobility intensity is low. For example, these locations attract people for work during morning times, while people living in residential regions diverge to workplaces simultaneously. Thus, convergence and divergence both occur during the morning commute time (T_6–T_9). The convergence and divergence pattern in *C*2 is likely to occur in main urban

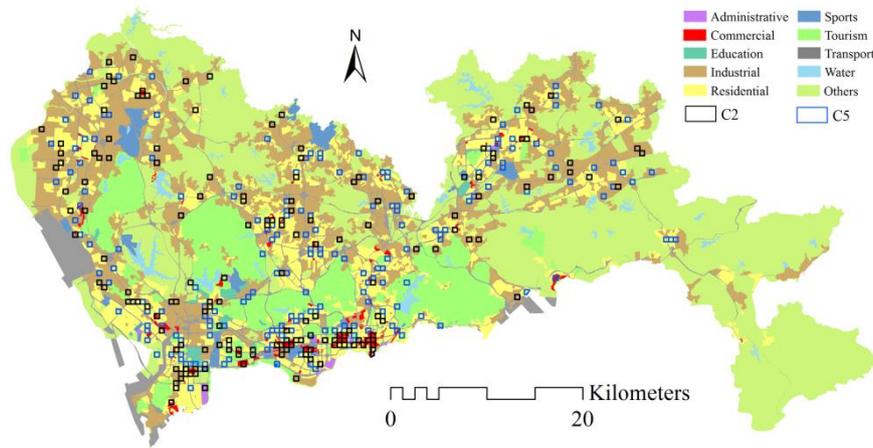commercial regions, whereas it tends to occur near business districts and workplaces within residential regions in *C*5.

Figure 8. Spatial distributions of identified functional clusters *C*2 and *C*5.

Figure 9 shows the spatial distributions of clusters *C*3 and *C*4. Grid cells in both clusters are mainly located in urban residential regions. The cells in *C*3 are mainly located in the northern part of the city, while the cells in *C*4 are located in the southern part of the city. As shown in Table 4, residential land is dominant in *C*3 and *C*4, accounting for 50.4% and 67.6% of land use in the clusters, respectively. As discussed in Section 5.2, there are also some human mobility differences between the clusters. For example, divergence lasts longer in *C*4 than in *C*3 during the morning (Figure 6). The cluster differences may be caused by differences between economic development and human mobility space in the northern and southern parts of the region. The southern region is the core of the urban business district in Shenzhen, and the economy in the southern region is more developed than that of the northern region. The southern population density is also higher than that in the northern region. The more developed economy and high population density may be the underlying reasons for the cluster pattern differences. However, many immigrant workers live in the northern part of Shenzhen, and they tend to live near their workplaces to save commuting time [47]. This short commute distance also makes it convenient for them to return home at noon for lunch or to take short breaks for activities, which may also contribute to the convergence-divergence pattern differences between T_{11} and T_{14} (Figure 6). Thus, the cells in *C*3 and *C*4 are likely located in urban residential regions, with *C*3 mainly located in the northern part of the city and *C*4 generally located in the southern part.
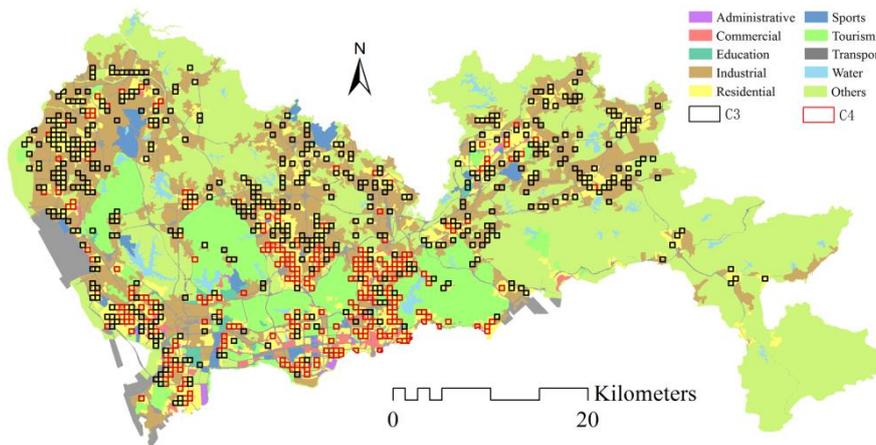

Figure 9. Spatial distributions of identified functional clusters *C*3 and *C*4.

Figure 10 shows the spatial distribution of *C*7. The grid cells in this cluster are mainly scattered across urban industrial regions. As shown in Table 4, the percentage of industrial land in this cluster is 58.4%, which is the dominant land use; thus, a large number of people converge in these areas to engage in work during the morning commute and then diverge from these areas to return home or travel to other locations when they finish their daily work. Thus, the human convergence and divergence pattern in *C*7 contrasts that in *C*3, although human mobility in both clusters show typical daily travel patterns related to work. Therefore, the human mobility pattern in *C*7 is likely associated with urban industrial regions.
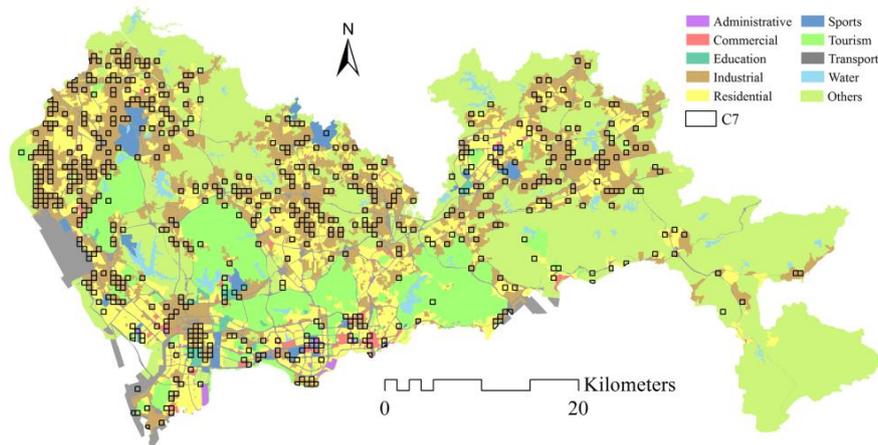

Figure 10. Spatial distribution of identified functional cluster *C*7.

Based on the spatial distribution, grid cells in *C*6 are not confined to a specific functional area, but scattered across different regions of Shenzhen (Figure 11), including urban administrative, education, sports and tourism regions. People have the freedom to choose the timing at which they arrive and leave these regions; thus, no consistent temporal patterns are formed in the regions. We can see that the difference between residential land (27.9%) and industrial land (28.8%) is small (Table 4). Many grid cells in this cluster are also located on the border of residential and industrial regions, so it is possible that a mixture of patterns occurs in these grid cells, e.g., during the morning commute, a grid cell containing industrial and residential land use would attract people to work, but people living in the grid cell may leave for work, resulting in an overall low netflow intensity. Some grid cells are also located in suburban areas with very low population densities, which may be another reason for the low intensity of human mobility.
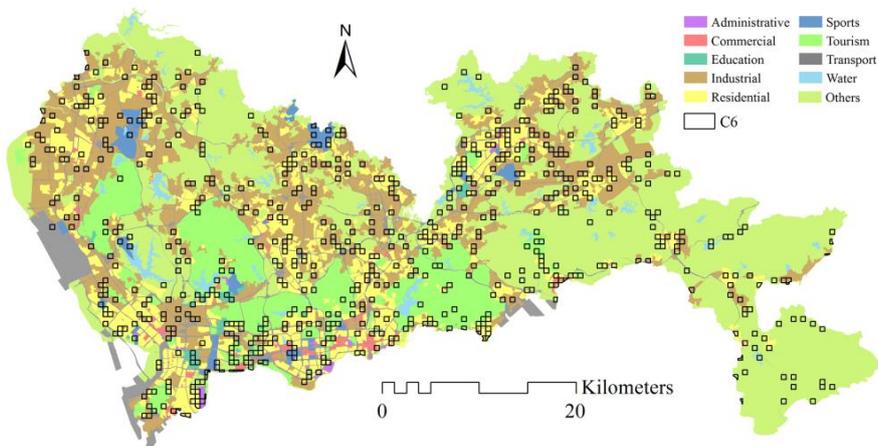

Figure 11. Spatial distribution of identified functional cluster *C*6.

The clusters identified in this study provide insight into the human dynamics at different locations in the city and potential land use characteristics associated with these different human mobility patterns. For example, *C*1 and *C*8 are likely located along main urban roads, whereas *C*2 tends to be located in urban commercial regions. In residential-dominant regions, a geographical difference in human mobility can be identified between the northern and the southern parts of Shenzhen. Although the study area and dataset are different, our findings are similar to those of a study that explored the interdependence between land use and traffic patterns using GPS-enabled taxi data in Shanghai [27]. In addition, these human mobility patterns are closely related to socioeconomic development and human activity areas [47]. These findings provide preliminary knowledge about human convergence and divergence patterns in urban areas based on different land use information.

This knowledge can help urban planners and policy makers to improve the efficiency of urban operations. Additionally, it can be used as input in Markov or training models to predict real-time urban traffic flows [31,48,49]. For example, when a new residential area is planned, human mobility patterns can be predicted based on its economic characteristics, thereby providing initial knowledge regarding the temporal travel demands of local residents. In addition, the findings can be used as a reference to estimate human convergence and divergence patterns using urban land use data in other cities without human tracking data. Conversely, urban land use information can be inferred based on these human mobility patterns [32,33]. In addition, based on the temporal convergence and divergence patterns of human mobility in different urban regions, managers can optimize urban public bicycle dock locations or real-time bicycle schedules in convergent and divergent areas to maintain a balance between supply and demand [50]. Similarly, taxi companies can allocate taxis in locations with high human convergence and divergence activities at specific times of a day [51]. Therefore, these findings can be used to improve urban public transport efficiency, which helps promote intelligent urban mobility [52,53].

6. Conclusions

The emergence of new location-aware data sources (e.g., mobile phone data) has provided opportunities and challenges associated with understanding human activities in the urban context (e.g., real-time monitoring of urban dynamics, human mobility patterns, etc.). This article explores the spatiotemporal patterns of human convergence and divergence using a big mobile phone location dataset from Shenzhen, China. From the location sequences of individual cell phone trajectories, we derived two measures (inflow and outflow) at the grid cell level (500 m $\times$ 500 m) to represent the numbers of incoming and outgoing trips at different locations in the city at different times of the day. Using the difference between inflow and outflow, we generated a time series for each grid cell, which reflects the direction and intensity of people flows and describes the temporal patterns of human convergence and divergence. Then, a clustering algorithm was employed to categorize distinct human convergence and divergence types within the city. We then investigated the spatial distributions of grid cells in different categories and examined how the identified patterns were associated with particular urban functional region types. This yielded additional insight into the relationships between people flows and the functional environment.

Eight distinct spatiotemporal clusters were identified, and the spatial distributions of these patterns were discussed based on the urban functional areas. Grid cells in clusters *C*1 and *C*8 were likely located along main urban roads in transportation-dominant regions (e.g., intra- and inter-urban traffic hubs); *C*2 and *C*5 were generally located in commercial-dominant urban regions; *C*3 and *C*4 were mainly located in residential-dominant regions; *C*7 was typically located in industrial-dominant regions; and *C*6 was scattered in different functional regions throughout the city. There was also a geographical (north–south) difference in human convergence and divergence in urban residential regions, and this difference mimicked the pattern of urban socioeconomic development. Distinct human convergent and divergent activities occurred at noon in northern residential and industrial regions, which may be due to low human mobility in those areas. These findings enhance our knowledge of human

mobility in different urban functional regions and provide a reference for policy makers to improve policy effectiveness.

There are some limitations of this study. First, one main limitation of this work is the potential impact of MAUP (modifiable area unit problem). Signal switches are a source of inherent bias in mobile phone data, and they may affect studies of human mobility patterns. The sample interval of the mobile phone data used in this study is approximately one hour, so we cannot accurately identify signal switches between cell phone towers. Most current studies employed Voronoi tessellations to represent the service areas of cell phone towers. However, there are many extremely close cell phone towers (separated by less than 10 m) in the study area (e.g., there are several cell phone towers in one office building in the urban center), so Voronoi tessellation does not prevent signal switching between these close cell phone towers. This study adopted 500 m × 500 m grid cells to divide the city and aggregate close cell phone towers to reduce the influence of signal switches between these cell phone towers. However, it is difficult to address the problem completely because the exact service area of a cell phone tower is uncertain. In addition, we excluded grid cells that did not contain cell phone towers because it is not feasible to calculate human movements between grid cells without cell phone towers. This may exclude some human activity areas. Although these movements were ignored, the analysis results provide useful information for understanding aggregate human mobility patterns in an urban functional context. Future studies can further analyze spatial interpolation differences between Voronoi tessellations and grid cells. Another limitation is that the dataset only covers one workday; thus, we were unable to investigate differences in weekly and seasonal patterns of human mobility. This study proposes a method for extracting daily spatiotemporal patterns of human convergence and divergence. The proposed method can be employed to extract human mobility patterns from long-term data, which is helpful for comparing human mobility on different days.

In future research, we will employ the identified patterns to optimize urban transportation and planning. For example, the urban public transport system could be optimized (i.e., the locations of bus stops or timetables of bus lines) based on the identified human mobility patterns. We will also further examine the relationship between human flow matrices and land use to provide better understanding of spatial interactions among different land use types. We believe that these analyses will deepen our knowledge of human activities in the urban context and provide many benefits to the development of urban systems.

Author Contributions: This research was mainly formulated and designed by Zhixiang Fang, Shih-Lung Shaw, Xiping Yang and Yang Xu. Ling Yin provided the dataset. Xiping Yang and Zhiyuan Zhao performed the experiments. Xiping Yang and Yang Xu wrote the manuscript. Tao Zhang and Yunong Lin reviewed the manuscript and provided comments.

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

References

1. Kwan, M.P. GIS methods in time-geographic research: Geocomputation and geovisualization of human activity patterns. *Geogr. Ann. Ser. B Hum. Geogr.* **2004**, *86*, 267–280. [CrossRef]
2. Chen, J.; Shaw, S.L.; Yu, H.; Lu, F.; Chai, Y.; Jia, Q. Exploratory data analysis of activity diary data: A space-time GIS approach. *J. Trans. Geogr.* **2011**, *19*, 394–404. [CrossRef]
3. Yuan, Y.; Raubal, M. Extracting dynamic urban mobility patterns from mobile phone data. In *Geographic Information Science*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 354–367.
4. Ratti, C.; Frenchman, D.; Pulselli, R.M.; Williams, S. Mobile landscapes: Using location data from cell phones for urban analysis. *Environ. Plan. B Plan. Des.* **2006**, *33*, 727–748. [CrossRef]

5. Silva, T.H.; Vaz de Melo, P.O.; Almeida, J.M.; Salles, J.; Loureiro, A.A. A comparison of foursquare and instagram to the study of city dynamics and urban social behavior. In Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing, Beijing, China, 12–16 August 2013.

6. Gonzalez, M.C.; Hidalgo, C.A.; Barabasi, A.L. Understanding individual human mobility patterns. *Nature* **2008**, *453*, 779–782. [CrossRef] [PubMed]

7. Song, C.; Qu, Z.; Blumm, N.; Barabási, A.L. Limits of predictability in human mobility. *Science* **2010**, *327*, 1018–1021. [CrossRef] [PubMed]

8. Caceres, N.; Wideberg, J.; Benitez, F. Deriving origin destination data from a mobile phone network. *Intell. Transp. Syst.* **2007**, *1*, 15–26. [CrossRef]

9. Pelletier, M.P.; Trépanier, M.; Morency, C. Smart card data use in public transit: A literature review. *Transp. Res. C Emerg. Technol.* **2011**, *19*, 557–568. [CrossRef]

10. Ranjan, G.; Zang, H.; Zhang, Z.L.; Bolot, J. Are call detail records biased for sampling human mobility? *ACM SIGMOBILE Mob. Comput. Commun. Rev.* **2012**, *16*, 33–44. [CrossRef]

11. Zhao, Z.; Shaw, S.L.; Xu, Y.; Lu, F.; Chen, J.; Yin, L. Understanding the bias of call detail records in human mobility research. *Int. J. Geogr. Inf. Sci.* **2016**, *30*, 1–25. [CrossRef]

12. Kwan, M.P. Analysis of human spatial behavior in a GIS environment: Recent developments and future prospects. *J. Geogr. Syst.* **2000**, *2*, 85–90. [CrossRef]

13. Ahas, R.; Aasa, A.; Silm, S.; Aunap, R.; Kalle, H.; Mark, Ü. Mobile positioning in space-time behaviour studies: Social positioning method experiments in Estonia. *Cartogr. Geogr. Inf. Sci.* **2007**, *34*, 259–273. [CrossRef]

14. Lu, Y.; Liu, Y. Pervasive location acquisition technologies: Opportunities and challenges for geospatial studies. *Comput. Environ. Urban Syst.* **2012**, *36*, 105–108. [CrossRef]

15. Isaacman, S.; Becker, R.; Cáceres, R.; Kobourov, S.; Martonosi, M.; Rowland, J.; Varshavsky, A. Identifying important places in people's lives from cellular network data. In *Pervasive Computing*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 133–151.

16. Calabrese, F.; Di Lorenzo, G.; Liu, L.; Ratti, C. Estimating origin-destination flows using mobile phone location data. *IEEE Pervasive Comput.* **2011**, *10*, 36–44. [CrossRef]

17. Exploring Universal Patterns in Human Home-Work Commuting from Mobile Phone Data. Available online: http://dx.doi.org/10.1371/journal.pone.0096180 (accessed on 12 September 2016).

18. Bar, G.H. Evaluation of a cellular phone-based system for measurements of traffic speeds and travel times: A case study from Israel. *Transp. Res. C Emerg. Technol.* **2007**, *15*, 380–391.

19. Calabrese, F.; Colonna, M.; Lovisolo, P.; Parata, D.; Ratti, C. Real-time urban monitoring using cell phones: A case study in Rome. *IEEE Transp Intell. Transp. Syst.* **2011**, *12*, 141–151. [CrossRef]

20. Alhasoun, F.; Almaatouq, A.; Greco, K.; Campari, R.; Alfaris, A.; Ratti, C. The city browser: Utilizing massive call data to infer city mobility dynamics. In Proceedings of the 3rd International Workshop on Urban Computing, New York, NY, USA, 24 August 2014.

21. Trasarti, R.; Olteanu-Raimond, A.M.; Nanni, M.; Couronné, T.; Furletti, B.; Giannotti, F.; Smoreda, Z.; Ziemlicki, C. Discovering urban and country dynamics from mobile phone data with spatial correlation patterns. *Telecommun. Policy* **2015**, *39*, 347–362. [CrossRef]

22. Rubio, A.; Sanchez, A.; Frias-Martinez, E. Adaptive non-parametric identification of dense areas using cell phone records for urban analysis. *Eng. Appl. Artif. Intell.* **2013**, *26*, 551–563. [CrossRef]

23. Hoteit, S.; Secci, S.; Sobolevsky, S.; Ratti, C.; Pujolle, G. Estimating human trajectories and hotspots through mobile phone data. *Comput. Netw.* **2014**, *64*, 296–307. [CrossRef]

24. Sagl, G.; Delmelle, E.; Delmelle, E. Mapping collective human activity in an urban environment based on mobile phone data. *Cartogr. Geogr. Inf. Sci.* **2014**, *41*, 272–285. [CrossRef]

25. Schlaich, J.; Otterstätter, T.; Friedrich, M. Generating trajectories from mobile phone data. In Proceedings of the 89th Annual Meeting Compendium of Papers, Transportation Research Board of the National Academies, Washington, DC, USA, 10–14 January 2010.

26. Zhu, X.; Guo, D. Mapping large spatial flow data with hierarchical clustering. *Transp. GIS* **2014**, *18*, 421–435. [CrossRef]

27. 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]

28. Ratti, C.; Sobolevsky, S.; Calabrese, F.; Andris, C.; Reades, J.; Martino, M.; Claxton, R.; Strogatz, S.H. Redrawing the map of great britain from a network of human interactions. *PLoS ONE* **2010**, *5*, e14248. [CrossRef] [PubMed]

29. Gao, S.; Liu, Y.; Wang, Y.; Ma, X. Discovering spatial interaction communities from mobile phone data. *Transp. GIS* **2013**, *17*, 463–481. [CrossRef]

30. Webster, F.V.; Bly, P.H.; Paulley, N.J. *Urban land Use and Transport Interaction: Policies and Models*; Gower Publishing: Brookfield, VT, USA, 1988.

31. Wegener, M. Land-use transport interaction models. In *Handbook of Regional Science*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 741–758.

32. Toole, J.L.; Ulm, M.; González, M.C.; Bauer, D. Inferring land use from mobile phone activity. In Proceedings of the ACM SIGKDD International Workshop on Urban Computing, Beijing, China, 12–16 August 2012.

33. Pei, T.; Sobolevsky, S.; Ratti, C.; Shaw, S.L.; Li, T.; 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]

34. Zhi, Y.; Li, H.; Wang, D.; Deng, M.; Wang, S.; Gao, J.; Duan, Z.; Liu, Y. Latent spatio-temporal activity structures: A new approach to inferring intra-urban functional regions via social media check-in data. *Geo-Spat. Inf. Sci.* **2016**, *19*, 94–105. [CrossRef]

35. 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.

36. 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]

37. Shenzhen Statistical Yearbook 2012. Available online: http://www.sztj.gov.cn/nj2012/indexeh.htm (accessed on 12 September 2016).

38. The Comprehensive Plan of Shenzhen City (2010–2020). Available online: http://www.szpl.gov.cn/xxgk/csgh/csztgh/201009/t20100929_60694.htm (accessed on 12 September 2016).

39. Hägerstraand, T. What about people in regional science? *Pap. Reg. Sci.* **1970**, *24*, 7–24. [CrossRef]

40. Iovan, C.; Olteanu-Raimond, A.M.; Couronné, T.; Smoreda, Z. Moving and calling: Mobile phone data quality measurements and spatiotemporal uncertainty in human mobility studies. In *Geographic Information Science at the Heart of Europe*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 247–265.

41. Vajakas, T.; Vajakas, J.; Lillemets, R. Trajectory reconstruction from mobile positioning data using cell-to-cell travel time information. *Int. J. Geogr. Inf. Sci.* **2015**, *29*, 1941–1954. [CrossRef]

42. Kang, C.; Liu, Y.; Ma, X.; Wu, L. Towards estimating urban population distributions from mobile call data. *J. Urban Technol.* **2012**, *19*, 3–21. [CrossRef]

43. Pelleg, D.; Moore, A.W. X-means: Extending k-means with efficient estimation of the number of clusters. In Proceedings of the Seventeenth International Conference on Machine Learnings, Stanford, CA, USA, 19 June–2 July 2000.

44. Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; Witten, I.H. The weka data mining software: An update. *ACM SIGKDD Explor. Newsl.* **2009**, *11*, 10–18. [CrossRef]

45. Gao, S. Spatio-temporal analytics for exploring human mobility patterns and urban dynamics in the mobile age. *Spat. Cogn. Comput.* **2015**, *15*, 86–114. [CrossRef]

46. Sagl, G.; Loidl, M.; Beinat, E. A visual analytics approach for extracting spatio-temporal urban mobility information from mobile network traffic. *ISPRS Int. J. Geo-Inf.* **2012**, *1*, 256–271. [CrossRef]

47. Xu, Y.; Shaw, S.-L.; Zhao, Z.; Yin, L.; Fang, Z.; Li, Q. Understanding aggregate human mobility patterns using passive mobile phone location data: A home-based approach. *Transportation* **2015**, *42*, 625–646. [CrossRef]

48. Lu, B.; Huang, M. Traffic flow prediction based on wavelet analysis, genetic algorithm and artificial neural network. In Proceedings of the International Conference on Information Engineering and Computer Science, Wuhan, China, 25–26 December 2010.

49. Necula, E. Dynamic traffic flow prediction based on GPS data. In Proceedings of the IEEE International Conference on TOOLS with Artificial Intelligence, Limassol, Cyprus, 10–12 November 2014.

50. Xu, Y.; Shaw, S.-L.; Fang, Z.; Yin, L. Estimating potential demand of bicycle trips from mobile phone data—An anchor-point based approach. *ISPRS Int. J. Geo-Inf.* **2016**, *5*, 131. [CrossRef]

51. Demissie, M.G.; Phithakkitnukoon, S.; Sukhvibul, T.; Antunes, F. Inferring passenger travel demand to improve urban mobility in developing countries using cell phone data: A case study of Senegal. *IEEE Transp. Intell. Trans. Syst.* **2016**, 17, 2466–2478. [CrossRef]

52. Garau, C.; Masala, F.; Pinna, F. *Benchmarking Smart Urban Mobility: A Study on Italian Cities*; Springer: Berlin/Heidelberg, Germany, 2015.

53. Garau, C.; Masala, F.; Pinna, F. Cagliari and smart urban mobility: Analysis and comparison. *Cities* **2016**, *56*, 35–46. [CrossRef]