Taylor & Francis
Taylor & Francis Group

International Journal of Geographical Information Science

ISSN: 1365-8816 (Print) 1362-3087 (Online) Journal homepage: http://www.tandfonline.com/loi/tgis20

Spatiotemporal model for assessing the stability of urban human convergence and divergence patterns

Zhixiang Fang, Xiping Yang, Yang Xu, Shih-Lung Shaw & Ling Yin

To cite this article: Zhixiang Fang, Xiping Yang, Yang Xu, Shih-Lung Shaw & Ling Yin (2017): Spatiotemporal model for assessing the stability of urban human convergence and divergence patterns, International Journal of Geographical Information Science, DOI: 10.1080/13658816.2017.1346256

To link to this article: http://dx.doi.org/10.1080/13658816.2017.1346256

Published online: 02 Jul 2017.

INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE, 2017
https://doi.org/10.1080/13658816.2017.1346256

ARTICLE

Spatiotemporal model for assessing the stability of urban human convergence and divergence patterns

Zhixiang Fang[a,b], Xiping Yang[c], Yang Xu[d], Shih-Lung Shaw[e,b] and Ling Yin[f]

[a]State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, Hubei, PR China; [b]Collaborative Innovation Center of Geospatial Technology, Wuhan, PR China; [c]School of Geography and Tourism, Shaanxi Normal University, Xi'an, PR China; [d]Senseable City Laboratory, SMART Centre, Singapore, Singapore; [e]Department of Geography, University of Tennessee, Knoxville, TN, USA; [f]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, PR China

ABSTRACT

Understanding the stability of urban flows is critical for urban transportation, urban planning and public health. However, few studies have measured the stability of aggregate human convergence or divergence patterns. We propose a spatiotemporal model for assessing the stability of human convergence and divergence patterns. A mobile phone location data set obtained from Shenzhen, China, was used to assess the stability of daily human convergence and divergence patterns at three different spatial scales, i.e. points (cell phone towers), lines (bus lines) and areas (traffic analysis zones [TAZs]). Our analysis results demonstrated that the proposed model can identify points and bus lines with time-dependent variations in stability, which is useful for delineating TAZs for transportation planning, or adjusting bus timetables and routes to meet the needs of bus riders. Comparisons of the results obtained from the proposed model and the widely used entropy measure indicated that the proposed model is suitable for assessing the differences in stability for various types of spatial analysis units, e.g. cell phone towers. Therefore, the proposed model is a useful alternative approach of measuring spatiotemporal stability of aggregate human convergence and divergence patterns, which can be derived from the space–time trajectories of moving objects.

ARTICLE HISTORY
Received 1 December 2016
Accepted 18 June 2017

KEYWORDS
Convergence; divergence; human mobility; stability; mobile phone data

1. Introduction

Understanding human mobility patterns can provide insights into the driving forces that affect human movements as well as the interactions between humans and the functional configuration of urban environment, thereby providing benefits in various fields such as urban transportation (Wang *et al*. 2012), urban planning (Ratti *et al*. 2006), urban policy development and public health (Mao *et al*. 2016). Previous studies have mainly analyzed human mobility patterns to understand these driving forces and interactions from either the individual or aggregate perspective. Many studies have analyzed the spatiotemporal mobility behavior of individuals (Ratti *et al*. 2006, Yuan and Raubal 2012, Liu *et al*. 2015) and their activities (e.g. work, shopping or entertainment)

CONTACT Zhixiang Fang zxfang@whu.edu.cn; Xiping Yang 0yangxiping0@163.com

(Phithakkitnukoon *et al*. 2010) in order to predict reproducible patterns (Gonzalez *et al*. 2008, Simini *et al*. 2012), including residential mobility and daily-to-weekly mobility (Sevtsuk and Ratti 2010). Aggregate human mobility patterns have been investigated using various approaches, such as identifying 'sink' and 'source' areas (Liu *et al*. 2012) or hot spot areas (Hoteit *et al*. 2014, Hu *et al*. 2014, Lichman and Smyth 2014, Scholz and Lu 2014), investigating human convergence and divergence patterns (Yang *et al*. 2016), and identifying activity groups (Shen and Cheng 2016) or mobility clusters (Hammond and Thompson 2002). Studies of individual and aggregate human mobility can elucidate the nature of mobility at various spatial and temporal scales, but the stability of human mobility has not received sufficient attention in previous investigations.

Human convergence and divergence patterns are important and common manifestations of aggregate human mobility patterns in urban space, and their stability is critical for understanding the forces that drive human movement as well as the efficiency of the interactions between humans and the functional configuration in order to facilitate urban policy development. Several studies have focused on the stability of individuals in cities (Hanson 2005) and the temporal stability (Hammond and Thompson 2002) of walking trips (Mitra *et al*. 2010), as well as the space–time structure of human mobility (Sun *et al*. 2011). However, it is still necessary to measure the stability of aggregate human mobility (i.e. convergence or divergence). In this study, we propose an approach for measuring the stability of aggregate human convergence and divergence to address this deficiency.

The research question addressed in this study is how to model the stability of human convergence and divergence using mobile phone big data. The appearance of convergence or divergence in an urban space unit comprises a sequence, which is not immutable and changes over time. The stability of this sequence may indicate changes in the travel demand for a place. Quantitative analyses based on the stability of this sequence can provide insights into the fluctuations in travel demand in an urban space, thereby yielding valuable information for traffic managers and urban planners when designing policies, such as planning bus lines and operating times. Several approaches have been used to investigate human mobility, such as questionnaires, travel dairies and geo-referenced human movement data sets obtained from different types of location-aware devices (e.g. mobile phones, smart cards and GPS-enabled taxis) (Lu and Liu 2012, Shaw *et al*. 2016). However, it is difficult to use questionnaires or travel diaries to investigate human convergence and divergence because they contain coarse space–time information and their sample sizes are limited (Ratti *et al*. 2006, Yuan and Raubal 2012). By contrast, geo-referenced human movement big data provide new opportunities for understanding human convergence and divergence patterns on unprecedented spatial and temporal scales (Xu *et al*. 2015). Thus, we used mobile phone location data to investigate the stability of human convergence and divergence patterns because of the advantages of these data for capturing human mobility in the urban environment.

Our study makes the following three main contributions. First, we propose a model for characterizing the stability of human divergence or divergence processes and sequences. Second, we demonstrate the feasibility of our model for assessing the stability of human convergence and divergence in three types of space units, i.e. locations, traffic analysis zones (TAZs) and bus transit lines, thereby providing a reference for optimizing and rescheduling bus line timetables and routes in order to improve the efficiency of public transport systems. Third, we compare the different capabilities of our model and the widely used entropy method for measuring the stability of human convergence and divergence.

The remainder of this paper is organized as follows. Section 2 reviews previous studies of aggregate human mobility based on big data and stability measurements. In Section 3, we propose an assessment model for characterizing the stability of human convergence or divergence processes and sequences. In Section 4, we present the study data set and analyses of the results obtained for three types of space units. The different capabilities of our model and the entropy method are compared and analyzed in Section 5. We give our conclusions in Section 6.

2. Related work

In this section, we review related research into aggregate human mobility using big data and stability measurements.

Three groups of approaches have been used to investigate aggregate human mobility. The first group aims to detect human mobility or activity hotspots to determine the mobility demand or activities in different places. Thus, Hoteit *et al*. (2014) detected human mobility hotspots based on human trajectories; Scholz and Lu (2014) defined a six-stage life cycle to model the dynamics of activity hot spots; and Hu *et al*. (2014) and Lichman and Smyth (2014) used kernel density estimation to generate a human mobility smooth surface based on the point locations obtained from moving objects. Liu *et al*. (2012) borrowed the 'source area and sink area' concept from ecology to characterize daily travel patterns, where they classified study areas into six traffic source-sink areas based on the temporal vectors of taxi pick-ups and drop-offs. Zhu and Guo (2014) proposed a hierarchical clustering method for mapping large spatial flow data (taxi trips), which can generalize origin–destination flows to represent the main human inflow and outflow areas.

The second group of approaches aims to identify concentrated human activity areas such as human activity groups, regions of interest, semantically similar areas and spatially interacting communities. For example, Schneider *et al*. (2013) found 17 unique networks of mobility motifs that captured up to 90% of the population based on surveys and mobile data sets for different countries. Pappalardo *et al*. (2015) identified the existence of two distinct classes of individuals, i.e. returners and explorers, and showed that their mobility patterns and social interactions were correlated. Hu *et al*. (2015) extracted urban areas of interest using geotagged photos. Shen and Cheng (2016) modeled activities as visits to spatiotemporal regions of interest and proposed a framework for identifying activity groups based on space–time profiles. Steiger *et al*. (2016) proposed a geographic, hierarchical self-organizing map to explore spatiotemporal and semantic clusters in Twitter data. In addition, Sobolevsky *et al*. (2013) delineated geographical regions using networks of human interactions, while Gao *et al*. (2013) discovered spatially interacting communities based on mobile phone data. All of these methods are helpful for understanding active areas of interest for urban planning.

The third group of approaches aims to investigate the spatiotemporal dynamics or rhythms of human dynamics. For example, some studies have utilized the call volume (Erlang value) or number of calls as indicators to monitor the spatiotemporal dynamics of cities in real time (Liu *et al*. 2009, Calabrese *et al*. 2011), as well as for estimating the population distribution (Kang *et al*. 2012, Deville *et al*. 2014) and identifying the rhythms and variations in human dynamics for different cities or times (hourly, daily, weekly and monthly) (Sevtsuk and Ratti 2010, Ahas *et al*. 2015) using long-term data. In addition, visual analytics approaches have been proposed to analyze the spatiotemporal dynamics of human mobility (Sagl *et al*. 2012, Gao 2015). Recently, Yang *et al*. (2016) investigated human convergence and divergence patterns using mobile phone data. These three groups of approaches are very useful for understanding aggregate human mobility, but they are not capable of measuring the stability of aggregate human mobility (i.e. convergence or divergence).

In terms of stability measurement, the Shorrocks trace index is used to summarize mobility across cluster classes (Shorrocks 1978) and time (Hammond and Thompson 2002). Mitra *et al*. (2010) used this index to summarize the stability of the spatial patterns during school walking trips over time. Previously, it was stated that: 'The Shorrocks index, however, does not yield complete information about mobility within the distribution' and 'does not differentiate between large and small movements within the distribution' (Hammond and Thompson 2002, pp. 378). Entropy (Clausius 1850) was first introduced to measure reversible changes in thermal energy, which was extended to quantify the spatiotemporal stability of a link by Zayani *et al*. (2012). A high entropy value indicates a high level of disorder, whereas a low value indicates greater organization. However, entropy measurements only provide an overview of order quantification. In addition, the temporal stability of the structure of human mobility in urban space (Sun *et al*. 2011) has been investigated using principal component analysis. Very few studies have focused on the stability of human convergence and divergence. Therefore, for the first time, we propose an assessment model for measuring the stability of human convergence and divergence using mobile phone data.

3. Proposed assessment model

In this section, we introduce three definitions of aggregate human mobility: human convergence process (HCP), human divergence process (HDP) and daily human convergence and divergence sequence (HCDS). Next, we explain the model for assessing the stability of each human divergence or divergence process. Finally, we describe the overall assessment model for daily HCDSs based on the stability of divergence and divergence processes.

3.1. *Definitions of human convergence and divergence*

First, we introduce the concepts and definitions of human convergence and divergence. In the case of mobile phone data, all the locations of humans are recorded by cell phone towers. The outgoing flow (*outflow*) of a cell phone tower is defined as the cumulative movement number departing from the cell phone tower, whereas the incoming flow (*inflow*) of this tower is defined as the cumulative movements arriving at the cell phone tower. Figure 1 illustrates the *outflow* and *inflow* for cell phone tower A during time slot T_i. The *netflow* of the cell phone tower during this time slot T_i is defined as:

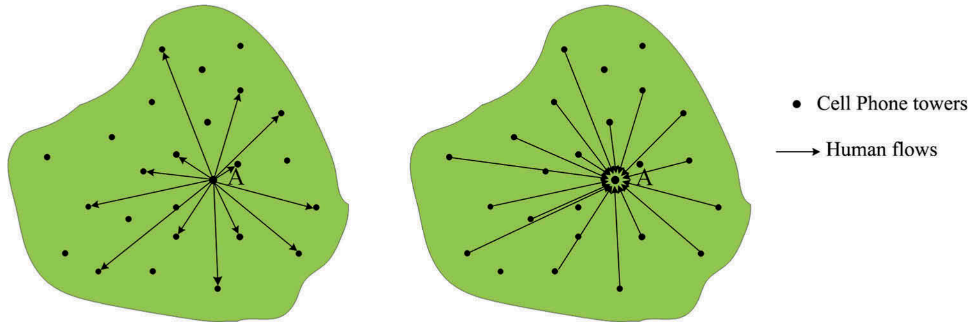

Figure 1. Outflow and inflow of cell phone tower A during time slot T_i.

$$netflow_{T_i} = inflow_{T_i} - outflow_{T_i}. \tag{1}$$

A positive *netflow* indicates that the number of people served by the cell phone tower is increasing during the time slot, where we treat this status as convergence. By contrast, a negative *netflow* indicates that the number of people served by the cell phone tower is decreasing during the time slot, where we treat this status as divergence. The *cumulative netflow* of the cell phone tower in time slot T_j can be calculated as:

$$N_j = \sum_{j=1}^{j} netflow_{T_i}, \tag{2}$$

where N_j represents the *cumulative netflow* value for the cell phone tower in the time period from T_1 to T_j. Figure 2 shows the variation in the *cumulative netflow* for a cell phone tower over time slots during a day. It is obvious that this variation presents the human dynamics of the cell phone tower. For each tower, there is a time series of *cumulative netflow* variation (Figure 2), which represents its human dynamics. Based on

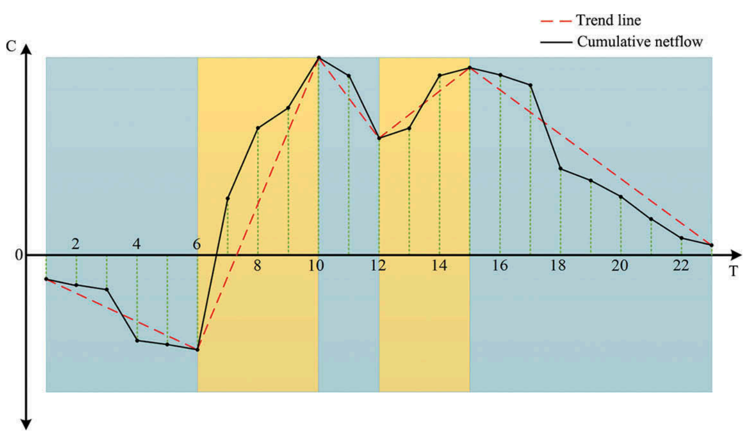

Figure 2. Variations in the *cumulative netflow* for the cell phone tower over time. The dashed red line indicates the trend in human convergence or divergence, and the solid black line indicates the variation in the *cumulative netflow*.

this representation, we introduce several definitions of human convergence and divergence, as follows.

Definition 1 Time-dependent *cumulative netflow* sequence: Let $N = \langle N_1, N_2, \cdots N_j, \cdots N_T \rangle$ be the time sequence of a *cumulative netflow* belonging to a cell phone tower, where N_j represents the *cumulative netflow* within time unit T_j. A subsequence N' includes part of N, which is represented as $N' = \langle N_i, N_{i+1}, N_{i+2}, \cdots, N_j \rangle, 1 \leq i \leq j \leq T,$ where T represents the total time slots in a day.

Definition 2 HCP: If a subsequence of N, i.e. $N' = \langle N_i, N_{i+1}, N_{i+2}, \cdots, N_j \rangle$, meets the following conditions in Equation (3), which means that the *cumulative netflow* is increasing monotonically from T_i to T_j, then it can be defined as a HCP. There are two exceptional cases: case 1 is $i = 1$, which can ignore the first condition in Equation (3); and case 2 is $j = T$, where T is the maximum time slot in the assessment, which can ignore the third condition in Equation (3). For example, the subsequences between time slots T_6 and T_{10}, and T_{12} and T_{15} in Figure 2 are HCPs.

$$\begin{cases} N_i \leq N_{i-1} \\ N_i \leq N_{i+1} \leq N_{i+2} \leq \cdots \leq N_j \\ N_j \geq N_{j+1} \end{cases}. \tag{3}$$

Definition 3 HDP: If a subsequence of N, i.e. $N' = \langle N_i, N_{i+1}, N_{i+2}, \cdots, N_j \rangle$, meets the following conditions in Equation (4), which means that the *cumulative netflow* is decreasing monotonically from T_i to T_j, then the subsequence of N can be defined as a HDP, except for the two cases related to HCP in the definition above. For example, the subsequences between time slots T_1 and T_6, T_{10} and T_{12}, and T_{15} and T_{23} in Figure 2 are HDPs.

$$\begin{cases} N_i \geq N_{i-1} \\ N_i \geq N_{i+1} \geq N_{i+2} \geq \cdots \geq N_j \\ N_j \leq N_{j+1} \end{cases}. \tag{4}$$

Definition 4 HCDS: The variation in the *cumulative netflow* of a phone tower can be defined as a sequence of alternations between HCP and HDP. The HCDS is used to represent the human dynamics of this cell phone tower. For example, as shown in Figure 2, the changes in the *cumulative netflow* with yellow shading indicate the HCP, whereas those with blue shading indicate the HDP. The HCDS of this cell phone tower can be modeled as: HDP → HCP → HDP → HCP → HDP. We propose an assessment model for quantifying the stability of HCP, HDP and HCDS.

3.2. *Stability of HCP or HDP*

Before defining the stability of human convergence and divergence, we first consider their characteristic variations. Figure 3 shows four examples of HCPs. T_s and T_e represent the start and end time slot of each convergence process, and N_s and N_e represent the corresponding *cumulative netflow* at start time slot T_s and end time slot T_e, respectively. We assume that these processes have the same N_s and we consider some different situations for this HCP, as

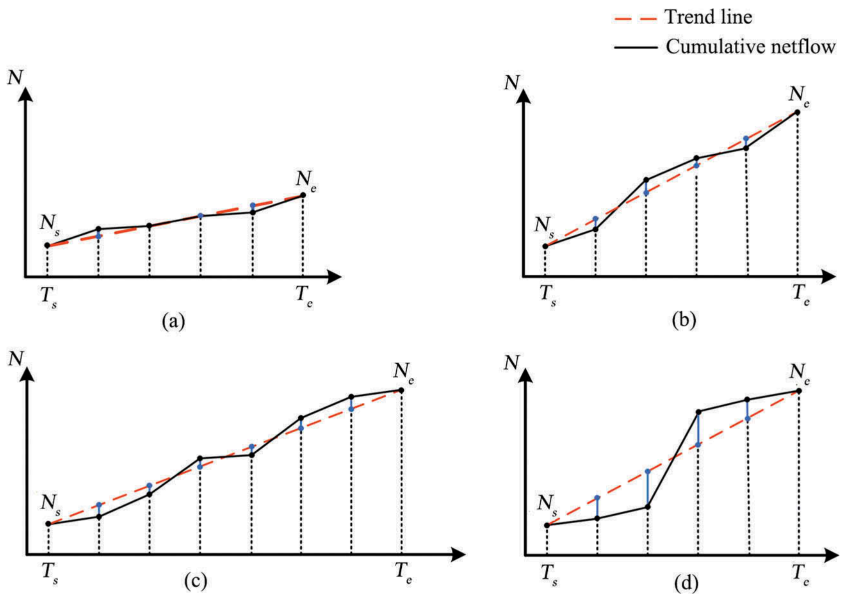

Figure 3. Examples of human convergence processes (HCPs). The red dashed line indicates the trend of human convergence and the black solid line indicates the variation in the *cumulative netflow*.

follows. (1) All four processes have the same value of N_s but the duration of the convergence process in Figure 3(c) is longer than that of the other three processes. (2) The *cumulative netflow* N_e at the end time slot is the same in Figure 3(b)–(d), but in Figure 3(a), N_e is smaller than that in the other three processes. (3) Both the duration and *cumulative netflow* are the same in Figure 3(b,d), but the fluctuation of the HCP in Figure 3(b) is weaker than that in Figure 3(d). Therefore, these specific characteristics should be considered when quantifying the stability of the human dynamics of a place.

Due to the differences discussed above, we measure the stability by combining three factors, i.e. the duration (T_e–T_s), magnitude (N_e –N_s) and fluctuation (which indicates the extent of variation during a human convergence or divergence process). If two HCPs have the same magnitude ($N_e - N_s$), the HCP with a longer duration is considered to be more stable than the other HCP with a shorter time duration. If two HCPs have the same duration, the HCP with the smaller magnitude is considered to be more stable than the HCP with a larger magnitude. If two HCPs have the same duration and magnitude, the HCP with a smaller fluctuation is considered to be more stable than the other HCP with a greater fluctuation. Thus, the duration, magnitude and fluctuation are used to quantify differences when assessing the stability of HCPs.

The trend in the variation of the *cumulative netflow* ($N_e - N_s$) in a HCP is represented as a red dashed trend line in Figure 3, which indicates the intensity of human gathering during the time period between T_e *and* T_s. The slope of the trend line can be used to quantify the intensity of the HCP:

$$v = \frac{N_e - N_s}{T_e - T_s}, \tag{5}$$

$$f = e^{-k \cdot |v|}, \tag{6}$$

where v is the slope of the trend line, k is a scale parameter and the exponential function is employed to normalize v, and thus the range of f is between zero and one. When f is near to 1, the HCP is constant during the time period, so the convergence process is very stable. When f is near to 0, the HCP undergoes extreme growth during the time period, so the HCP is highly unstable. In this manner, f can measure the stability of the HCP by integrating variations in the *cumulative netflow* and the duration.

To quantify the differences in the fluctuations of the HCPs in Figure 3(b,d), which have the same N_s, N_e and time period, we calculate the values (plotted as blue nodes on the trend line) and summarize the difference between the trend values and real *cumulative netflow* values to denote the scope of the change in HCP. The trend line is formulated as follows:

$$y = \frac{N_e - N_s}{T_e - T_s} \cdot (x - T_s) + N_s, \tag{7}$$

Based on Equation (7), we calculate a trend value for each time slot T_i. Therefore, each time slot has a different value (Equation 8) between the real *cumulative netflow* value N_i and the calculated trend value Y_i. The standard deviation of these different values in a HCP is employed to quantify the fluctuation in the convergence process. If the HCP has smaller standard deviations (Equation 9) for these differences, then the fluctuation process is more stable. Similarly, an exponential function is employed to normalize the standard deviation using Equation (10):

$$\Delta \bar{N}_i = N_i - y_i, \tag{8}$$

$$\sigma_{\Delta N} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\Delta N_i - \Delta \bar{N})^2}, \tag{9}$$

$$g = e^{-\frac{\sigma_{\Delta N}}{\Delta T}}, \Delta T = T_e - T_s + 1, \tag{10}$$

where $\Delta \bar{N}$ represents the average value of ΔN_i and ΔT represents the duration of the convergence process. g ranges between zero and one, and the fluctuation in the HCP process is more stable when the value of g is larger.

After quantifying the stability of the trend line f and the fluctuation g for the HCP, we use Equation (11) to define the overall stability level of the HCP in a mutually independent manner:

$$s = f \cdot g = e^{-k \cdot |v|} \cdot e^{-\frac{\sigma_{\Delta N}}{\Delta T}}, \tag{11}$$

where s represents the stability of the HCP process, which ranges between zero and one, and the HCP process is the more stable when the value of s is closer to one.

The stability of a HDP can be calculated in a similar manner to a HCP. The only difference in the process used for calculating a HCP and HDP is the slope of the trend lines. The slope of the trend line is positive in a HCP, whereas the slope is negative in a HDP. However, this is not important for assessing stability in the proposed model because we use the absolute value of the slope of the trend line.

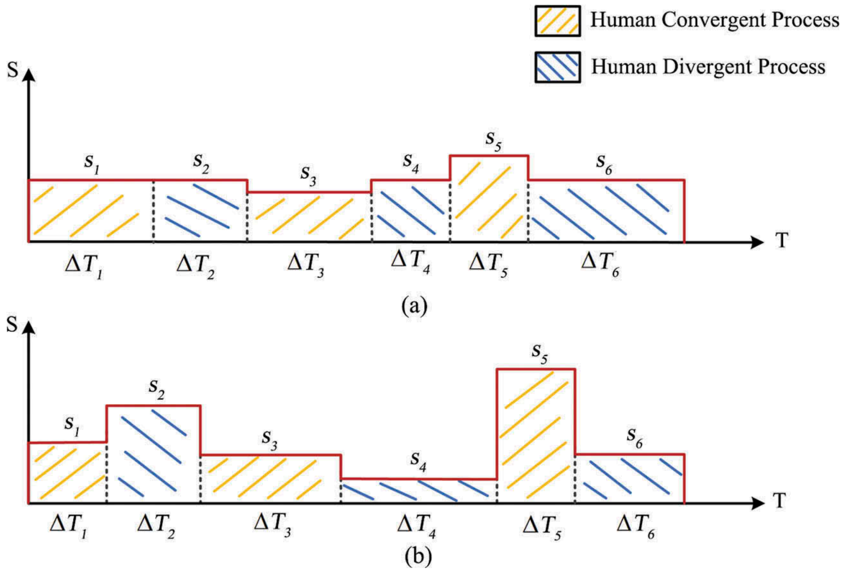

Figure 4. Examples of human convergence and divergence sequences. Area filled with blue lines is human divergent process, and area with orange lines is human convergent process. The red line is the boundary of convergence and divergence process.

3.3. *Stability of daily convergence and divergence sequences*

Based on the definitions of HCP and HDP, we propose a model for assessing the stability of the daily HCDS for each cell phone tower. Figure 4 shows two examples of daily HCDSs. In the daily HCDS, the HCP is the rectangle filled with yellow lines and the HDP is that filled with blue lines. The stability of each HCP or HDP can be calculated using Equation (11), where s denotes the stability. Therefore, s_i and ΔT_i in Figure 4 represent the stability value and duration time for the corresponding HCP or HDP, respectively. A HCP or HDP is more stable if the stability value (s_i) is larger or the duration (ΔT_i) is longer, or both. Therefore, we use the area covered by these rectangles for HCP or HDP to indicate the relative changes in stability (A) for each HCDS, which is formulated as:

$$A = \sum_{i=1}^{m} A_i = \sum_{i=1}^{m} s_i \cdot \Delta T_i, \tag{12}$$

$$\sum_{i=1}^{m} \Delta T_i = T, 0 \leq A \leq T, \tag{13}$$

where m represents the total number of all processes, T is the total number of time slots and s_i ranges between zero and one, so the range of A is between zero and T. In addition to the summed area of these processes, the fluctuation between HCP and HDP can also indicate the variation in the stability of the HCDS. For example, the two sequences in Figure 4 have the same area but the fluctuation in Figure 4(b) is more obvious than that in Figure 4(a), which indicates that the HCDS in Figure 4(b) is less stable than that in Figure 4(a). In order to distinguish this difference, we employ the length of the upper boundary (red line in Figure 4) to represent the fluctuation. A HCDS with a longer upper boundary has a greater fluctuation between a neighboring HCP and HDP; therefore, this HCDS is more unstable. The length can be calculated as follows:

$$P = s_1 + \sum_{i=2}^{m} |s_i - s_{i-1}| + s_m + \sum_{i=1}^{m} \Delta T_i, P \geq T. \tag{14}$$

In order to combine the effects of the summed area and the length of the upper boundary on the stability of the HCDS, we assess the stability of the HCDS by using the ratio between them, as follows:

$$Q = \frac{A}{P} = \frac{\sum_{i=1}^{m} s_i \cdot \Delta T_i}{s_1 + \sum_{i=2}^{m} |s_i - s_{i-1}| + s_m + \sum_{i=1}^{m} \Delta T_i}, 0 \leq Q \leq 1, \tag{15}$$

where A is less than or equal to P, so the stability Q lies between zero and one. Equation (15) shows that if a HCDS has a larger stability value and a shorter upper boundary, then it will be more stable.

4. Experiment and results

4.1. *Data set and preprocessing*

The mobile phone location data set used in this study was collected by a major telecom operator in Shenzhen, which is a city with the highest population density and the fourth highest economic output in China. A recent census showed that Shenzhen City covers a total area of approximately 2000 km^2, and it has a population of more than 15 million residents (Shenzhen Statistical YearBook 2012).

The mobile phone data set used in this study was acquired through a research collaboration. The data set comprised one workday of records for approximately 16 million cell phones. In contrast to call detail records that passively capture individual footprints only during actual communication such as phone calls and text messages (Yin *et al*. 2015, Zhao *et al*. 2016), the data set used in this study represents individual locations (at the cell phone tower level) approximately once every hour. Each record comprised an anonymous user ID, recording time, and the longitude and latitude of the cell phone tower used. Table 1 shows an example of a cell phone user's records from one day. The time of each record was at a fine scale of seconds. In total, 5940 cell phone towers were extracted from the data set and each tower was labeled with a unique Tower-ID number (see Figure 5 in

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

User ID	Record time	Time window	Longitude	Latitude
bfa8m7******	00:25:36	00:00–01:00	113.***	22.***
bfa8m7******	01:26:40	01:00–02:00	113.***	22.***
bfa8m7******	02:20:53	02:00–03:00	113.***	22.***
⋮	⋮	⋮	⋮	⋮
bfa8m7******	23:33:50	23:00–24:00	113.***	22.***

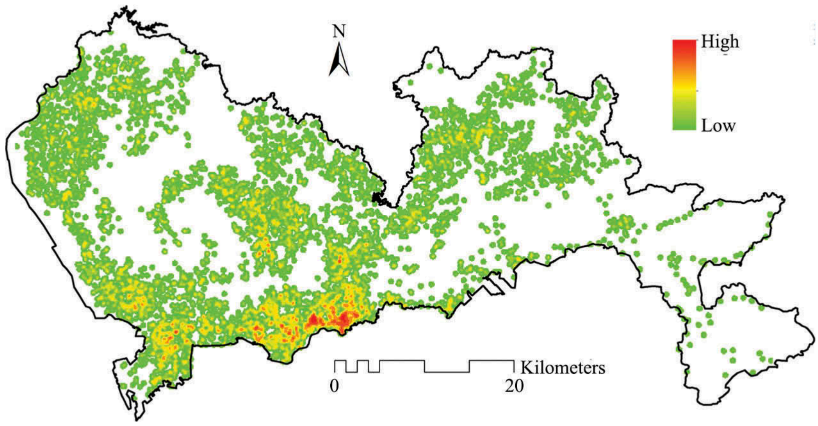

Figure 5. Spatial kernel density of the distribution of cell phone towers.

which, for privacy protection, individual records have been smoothed using a spatial kernel density surface generated based on the locations of the cell phone towers).

For each cell phone, the space–time trajectory can be constructed as follows:

$$Tr = [p_1(x_1, y_1, t_1, Id_1), p_2(x_2, y_2, t_2, Id_2), \cdots, p_n(x_n, y_n, t_n, Id_n)], \tag{16}$$

where x_i, y_i and Id_i are the longitude, latitude and Tower-ID of recording point p_i, and t_i represents the time when the location of the cell phone was updated. Each record was allocated to a time window of approximately 1 hour (Table 1). For every pair of adjacent time windows, if the locations of the two recording points were not identical ($x_i \neq x_{i+1}$ or $y_i \neq y_{i+1}$), then this indicated that a movement occurred between the two time windows. We recorded the original and terminal Tower-ID of the movement and identified the two adjacent time windows as a time slot. For example, p_1 was recorded by cell phone tower Id_i during time window 00:00–01:00 and p_2 was recorded by cell phone tower Id_j during the next time window 01:00–02:00. If $Id_i \neq Id_j$, then a movement could be extracted from Id_i to Id_j, and the two adjacent time windows 00:00–02:00 could be denoted as the time slot T_1 . In this manner, we extracted all the movements for time slots T_1 (00:00–02:00), T_2 (01:00–03:00), T_3 (02:00–04:00), …, T_{23} (22:00–24:00). Thus, there were 23 time slots in one day. For each time slot, we used all the extracted movements to create *inflow*, *outflow* and *netflow* data for the cell phone tower at each time slot, which we used to measure the stability of human convergence and divergence.

4.2. *Stability of HCP or HDP*

We extracted 22,234 HCPs and 21,132 HDPs from the data set and calculated the stability of these HCPs and HDPs for each cell phone tower according to Equation (11). The selection of the scale parameter k in Equation (6) depends on the *netflow* values. In this data set, more than 95% of the *netflow* values varied between −1000 and 1000. If $k = 1$, the value of f is extremely small, so we used $k = 0.01$ to ensure that $k^*|v|$ was between −10 and 10 because the value of f could be normalized in a range of [0, 1]. Figure 6 shows the distribution of the stability for the HCPs and HDPs. More than 42% of the HCPs and 37.6%

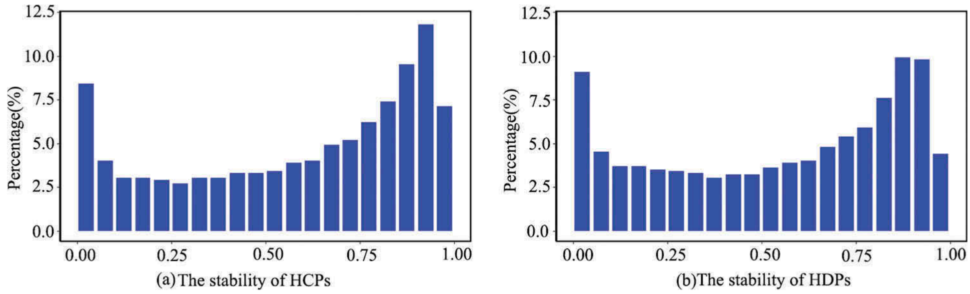

Figure 6. Distribution of stability for human convergence processes (HCPs) and human divergence processes (HDPs) (bin width = 0.05).

of the HDPs were greater than 0.75 for these cell phone towers. In order to determine the similarity between the percentages in Figure 6(a,b), we used Pearson's correlation coefficient R (Pearson 1920), which is defined as:

$$R = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot \sum_{i=1}^{n}(y_i - \bar{y})^2}}, \tag{17}$$

where x_i and y_i are the different stability percentages, and $\bar{x}$ and $\bar{y}$ are the average percentages. The R value calculated between the percentages in Figure 6(a,b) is 0.944, and thus the percentage distribution for HCP and HDP was relatively close. This indicates the symmetry of urban human flow, i.e. the inflow for a cell phone tower always comes from the outflow of another cell phone tower. Hence, the results obtained in this study are consistent with expectations.

To investigate the differences in stability for the HCPs and HDPs in different time periods during the day, we selected three time periods: morning (06:00–12:00), afternoon (12:00–17:00) and evening (17:00–23:00). Figure 7 shows the stability distributions for the HCPs and HDPs in the morning, afternoon and evening, where their Pearson's R values are 0.989, 0.620 and 0.925, respectively, which means that the HCPs and HDPs had similar percentage variation patterns, especially in the morning and evening.

Some specific stability distribution patterns are shown in Figure 7. For example, the percentages of both unstable HCPs and HDPs in the morning were the highest among the three periods, followed by the evening. The stability percentage distribution was more homogeneous in the afternoon than the other two time periods. The relatively high HCP and HDP percentages in the morning and evening represent the high intensity of human mobility associated with the commuting period, which was more unstable than that in the afternoon.

4.3. *Assessing the HCDS stability for people covered by TAZs*

Based on the assessments of HCPs and HDPs for cell phone towers, we employed the proposed approach to assess the HCDS stability for people covered by urban TAZs. TAZs

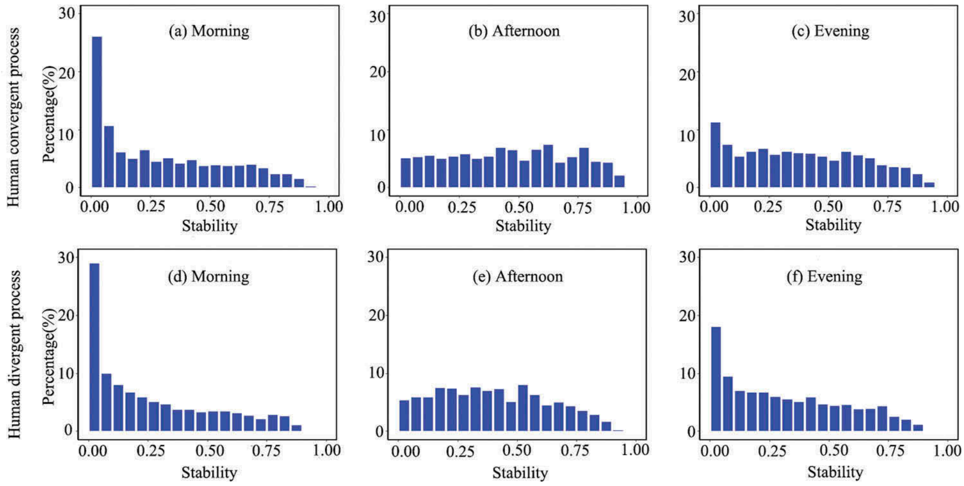

Figure 7. Distributions of stability for human convergence processes (HCPs) and human divergence processes (HDPs) in the morning, afternoon and evening periods (bin width = 0.05).

are the primary spatial units used in analyses, where trips begin and end in transportation planning, and they play important roles in travel demand forecasting models (e.g. trip generation and trip distribution) (You *et al*. 1998, Dong *et al*. 2015). Assessing the stability of TAZs is helpful for understanding the aggregate human mobility in these zones, and it can support the division of TAZs during transportation planning.

To assess the stability of HCDS, we aggregated the *inflow*, *outflow* and *netflow* derived from the space–time trajectories of mobile phones and calculated the stability in each TAZ using the proposed method. Figure 8 shows the statistical distribution of stability for individual TAZs, which demonstrates that over 85% of the TAZs had stability values smaller than 0.5. Figure 9 shows the spatial distribution of stability, where there are several obvious patterns. First, most of the TAZs with the highest stability were located in the eastern part of the city, as shown by the red color in Figure 9. These areas included green industrial zones and nature preservation zones. Second, the HCDS was more unstable in the southern part of the city because this is the economic center of the city, which includes most of the companies, residential areas, shopping malls,

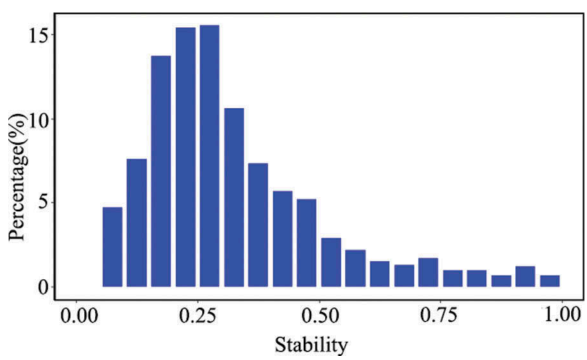

Figure 8. Statistical distribution of stability for traffic analysis zones.

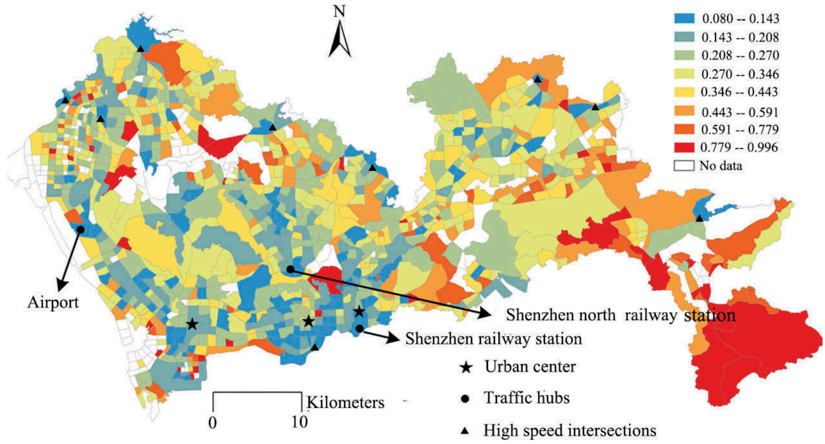

Figure 9. Spatial distribution of stability for traffic analysis zones.

restaurants, financial institutions and recreational venues (bars, cinemas, etc.). These locations usually attract numerous people to work, shop, eat, visit entertainment centers and engage in other activities during the daytime. In addition, the HCDS was very unstable at three major traffic hubs (airport, Shenzhen railway station and Shenzhen north railway station) and some highway intersections. These locations are very important for connecting with areas outside the city, which suggests that the human convergence and divergence at these locations changed greatly during the daytime. According to the spatial patterns shown in Figure 9, the stability value at individual cell phone towers can be used to support the subdivision of TAZs. Previously, Dong *et al.* (2015) extracted four characteristics of human mobility (i.e. real-time user volume, inflow, outflow and incremental flow) from mobile phone data to obtain TAZ subdivisions. Based on these data, transportation planners can introduce separate policies for stable and unstable TAZs in order to make transportation systems more efficient.

4.4. *Assessing the HCDS stability for people covered by a bus transit system*

We also used the proposed assessment model to evaluate the HCDS stability of people covered by a bus transit system, which could help to understand the potential demand for a bus transit system and facilitate policy decisions, such as adjusting the bus timetable or routes according to the number of people served.

We used mobile phone data to assess the stability of the people who were most likely served by each bus stop. According to the design criteria for public transport stops (National Standards of China, GB50220-95, 1995) and the distance people normally walk to access a bus stop (Delbosc and Currie 2011), we selected 500 m as the area covered by a bus stop. Due to the limitation of the mobile phone data set, it was impossible to know the exact bus stops to which people walked and the people served by each bus stop. Therefore, for each bus stop, we could only use the cell phone towers located within 500 m to assess the stability of the bus stop. It was possible that some towers were covered by two or more bus stops, which could lead to some estimation errors, but the results of this analysis are still useful for understanding human convergence and divergence in a bus transit system at the aggregate level.

We used a distance of 500 m to search for mobile phone towers near a bus stop (Figure 10), and we assumed that the people near each tower could be covered by the bus stop. The *netflow* of each bus stop was calculated by aggregating the *inflow, outflow* and *netflow* derived from the space–time trajectories of the mobile phones that entered or left the area within 500 m of the bus stop. We then calculated the stability of each bus stop in the same manner as the cell phone towers.

Figure 11 shows the stability of all the bus stops during one day. To understand the HCDS stability, we grouped the stability values of all the bus stops into five levels with an interval of 0.2. In Figure 11, groups A and E represent the lowest and highest levels of stability, respectively. Most of the bus stops with the lowest stability were located in large industrial zones (i.e. a, b, c and f), high-tech parks (i.e. d and g) and urban central business districts (i.e. h and j).

A bus line comprised bus line segments according to the representation of a bus line system in the GIS for transportation. The HCDS stability of a bus line segment between two bus stops was set as the average HCDS stability value for these bus stops. Based on the assigned HCDS stability values, we also grouped the stability values into five levels in the same manner as the bus stops. Figure 12 illustrates the HCDS stability results obtained for all the bus lines, which shows that 62.4% of the bus line segments had stability level C, and thus most of the bus line segments had a relatively stable human convergence and divergence status. Very few bus line segments had the highest level of

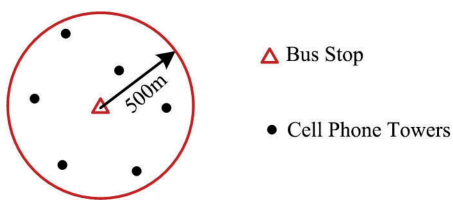

Figure 10. Area covered by bus stops.

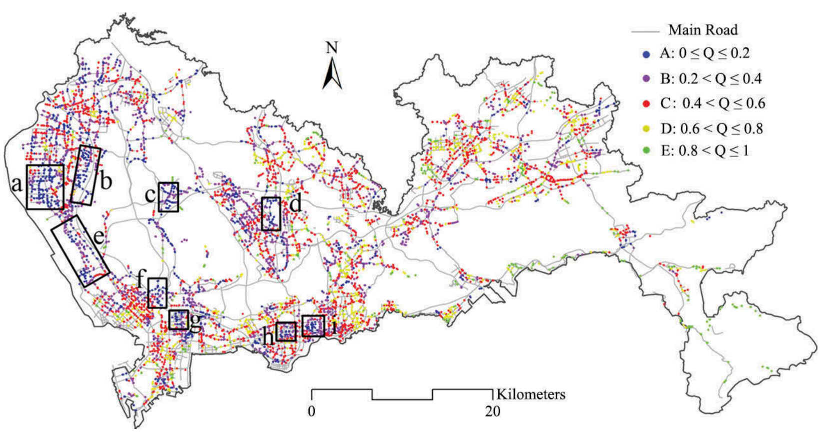

Figure 11. Human convergence and divergence sequence (HCDS) stability for urban bus stops.

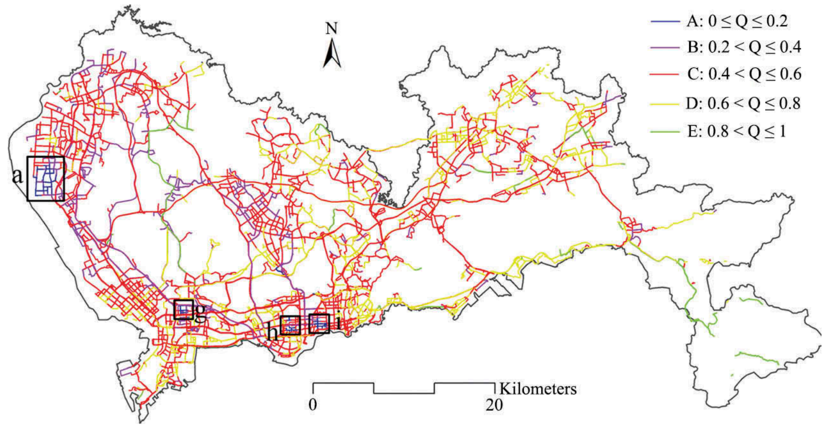

Figure 12. Human convergence and divergence sequence (HCDS) stability for bus line segments.

stability, and they were located in the relatively rural part of this study area (i.e. the green lines in Figure 12). The highly unstable areas of the bus line segments were located in the center of the city (i.e. g, h and i) and close to Shenzhen airport (i.e. a).

According to the rhythm of daily life for urban citizens, we divided the whole day into seven time periods (see Table 2), which included the main meaningful activity periods within a day. These seven time periods were used to examine the detailed HCDS stability for bus lines, which could help bus companies to understand the adjustments needed to their routine timetable and routes in order to meet the requirements of bus users.

Tables 3 and 4 show the percentages of bus stops and the service distances of the bus line segments in five levels for the seven time periods, which demonstrate that 75.2% of the bus stops and 84.4% of the service distances of the bus line segments were located at level E in T_1 because most people were sleeping at home during this time period. In addition, 19.3% and 45.5% of bus stops were at levels A and B, respectively, during the morning commuting period T_2, which were higher than those in the evening commuting period T_6 (5.8% at level A and 26.8% at level B). The HCDS stability of the bus line segments also exhibited similar patterns

Table 2. The seven time periods used in this study.

	Time period	Main activity		Time period	Main activity
T_1	00:00–06:00	Sleep time	T_5	14:00–17:00	Afternoon work time
T_2	06:00–09:00	Morning commute time	T_6	17:00–20:00	Evening commute time
T_3	09:00–12:00	Morning work time	T_7	20:00–24:00	Free time
T_4	12:00–14:00	Lunch time			

Table 3. Percentage of bus stops at the five levels for the seven time periods.

	A (%)	B (%)	C (%)	D (%)	E (%)
T_1	0.1	0.3	2.0	22.4	75.2
T_2	19.3	45.5	28.6	5.2	1.4
T_3	5.1	23.1	41.4	26.4	4.0
T_4	4.6	16.3	39.2	32.7	7.2
T_5	3.6	19.3	47.1	25.8	4.2
T_6	5.8	26.8	41.7	22.1	3.6
T_7	3.1	17.3	39.2	32.4	8.0

Table 4. Percentage of service distances of the bus line segments at the five levels for the seven time periods.

	A (%)	B (%)	C (%)	D (%)	E (%)
T_1	0.4	0.7	1.7	12.8	84.4
T_2	5.0	42.9	44.3	6.9	0.9
T_3	0.7	9.9	40.3	44.9	4.2
T_4	0.8	5.2	35.6	49.5	8.9
T_5	0.7	6.6	48.8	40.7	3.2
T_6	1.6	19.5	48.1	28.4	2.4
T_7	0.7	8.8	37.3	45.0	8.2

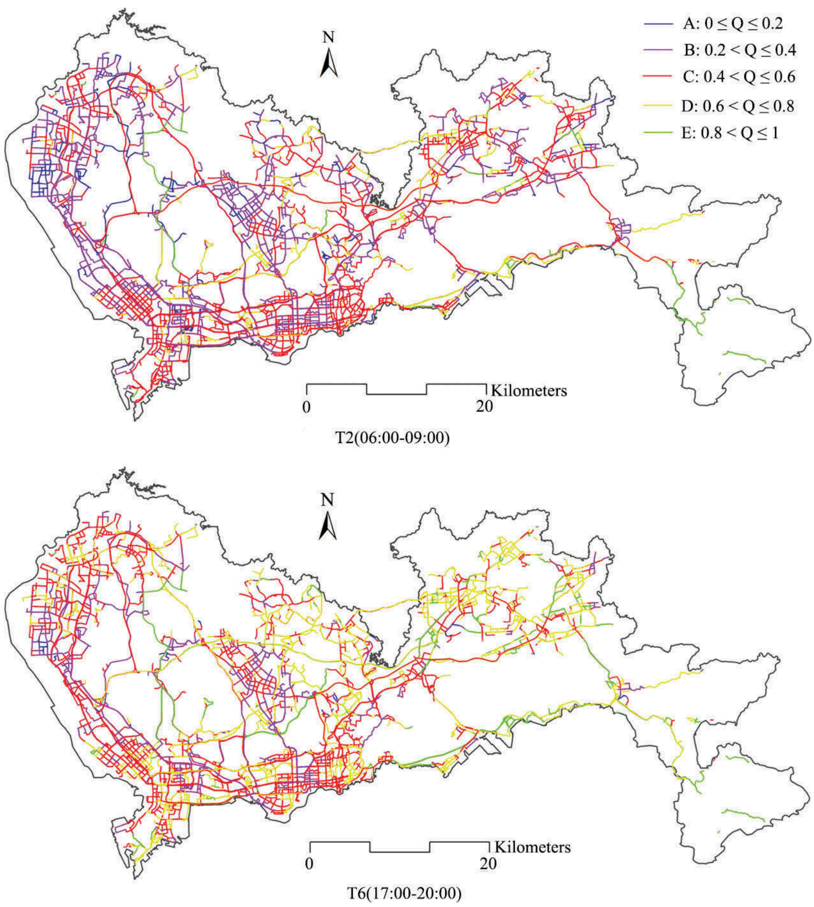

Figure 13. Human convergence and divergence sequence (HCDS) stability of bus line segments in T_2 and T_6.

(see Figure 13). These results indicate that human mobility is more unstable in the morning commuting period than the evening commuting period, which suggests that bus companies need to pay more attention to the travel demands in these unstable areas (e.g. the lines at level B) by monitoring flows and responding to unstable flows with different strategies in terms of the bus line departure interval. Thus, they can improve the performance of bus line services by saving energy and making full use of labor resources. By contrast, in time periods T_3, T_4, T_5 and T_7, most of the bus stops and bus line segments were at levels C and D, which indicates that human convergence and divergence were relatively stable during these time periods. These patterns suggest that the bus line departure intervals in these time periods do not need to be adjusted dynamically, but instead the departure intervals should be decided by the actual travel flows in these areas.

5. Discussion

In order to assess the suitability of the proposed model in identifying spatiotemporal stability of aggregate human convergence and divergence patterns, we compare the results of the proposed model with the entropy measure. Entropy is a widely used index for examining the stability of a system, and it has been employed recently to reflect the heterogeneity of human movements (Song *et al*. 2010). In this study, we compared the difference in performance between entropy and the proposed assessment model for determining the stability of daily human dynamics in a place.

Based on the defined time series of the *cumulative netflow* for a cell phone tower (see Figure 2), the entropy of human dynamics for a cell phone tower can be calculated as follows:

$$sum\_N = \sum_{i=1}^{23} |N_i|, \tag{18}$$

$$p_i = \frac{|N_i|}{sum\_N}, \tag{19}$$

$$E = -\sum_{i=1}^{23} p_i \cdot \log_2 p_i, \tag{20}$$

where E represents the entropy index. A larger entropy value for a cell phone tower indicates that each *netflow* N_i is close to the average value of sum_N, which indicates that the time series of the *cumulative netflow* for a tower is more stable.

Figure 14 illustrates the distribution of the entropy value and the proposed stability model, where this figure shows that they do not have a linear relationship, which indicates that the entropy and proposed stability model differ in their ability to reflect the HCDS stability.

In order to compare the differences between entropy and the proposed stability model, we selected cell phone towers with comparable entropy values and stability values, and then visualized the time series for the *cumulative netflow* to check the human dynamics around these cell phone towers. It should be noted that no specific rule was employed for selecting the entropy or stability values. Figure 15(a) shows the time-dependent *cumulative netflow* for the entropy values at 3.9, 4.0 and 4.1. Figure 15 (b) shows the time-dependent *cumulative netflow* for the stability values at 0.5, 0.6 and 0.7. An obvious pattern is visible in Figure 15(a) where the *cumulative netflow* varies significantly whereas the entropy value remains the same. However, the proposed model can reflect the stability of the *cumulative netflow* better than the entropy value because the time series of the *cumulative netflow* with a high stability value derived from

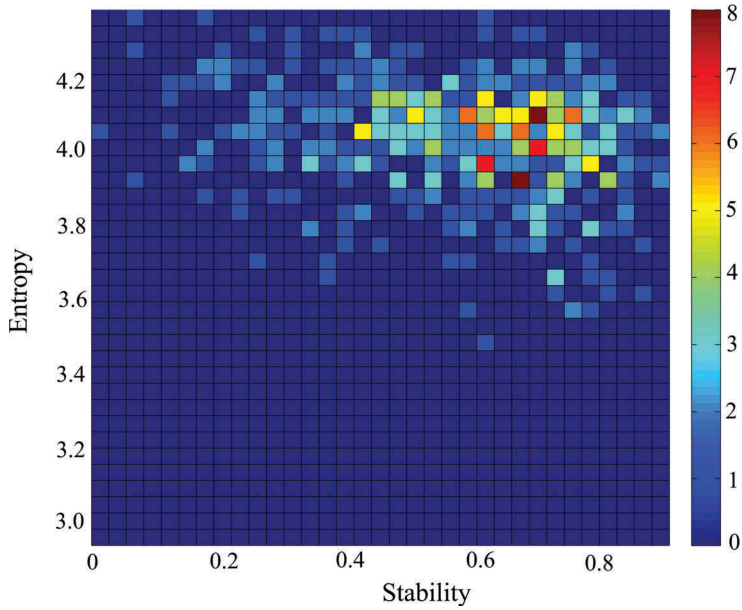

Figure 14. Stability and entropy of cell phone towers.

the proposed model does not exhibit large variations (e.g. levels D and E in Figure 15(a)), whereas those with a low value have larger variations (e.g. level A in Figure 15(a)). This result indicates that entropy performed poorly at reflecting the variations in the time series of the cumulative *netflow* between different towers. Figure 15(b) shows the capacity of the proposed model for measuring variations in the time series of the *cumulative netflow*, but these lines also exhibit some variations with different entropy values. Therefore, Figure 15 demonstrates that entropy is a better option for comparing variations in the aggregate human mobility for an individual cell phone tower during different time periods, whereas the proposed stability model is better for comparing the differences in aggregate human mobility between different cell phone towers. To some extent, these two indicators are complementary to each other and they can be used together to assess the stability of aggregate human mobility.

6. Conclusion

In this study, we proposed a model for assessing the stability of the daily HCDS, which we used to assess the HCDS stability for a cell phone tower, TAZs and bus lines. The experimental results obtained in this study demonstrate the capacity of the proposed model for finding potentially unstable areas of human mobility. These areas can then be used to facilitate decision-making, such as time-varying bandwidth assignment for a mobile communication system based on the HCDS stability of cell phone tower, combining or separating TAZs according to the HCDS stability of TAZs, responding to unstable flows with different strategies for bus line departure intervals and helping to find a suitable departure interval according to the HCDS stability of bus stops and bus lines. In addition, we compared the capacity of entropy and the proposed model for representing HCDS, and

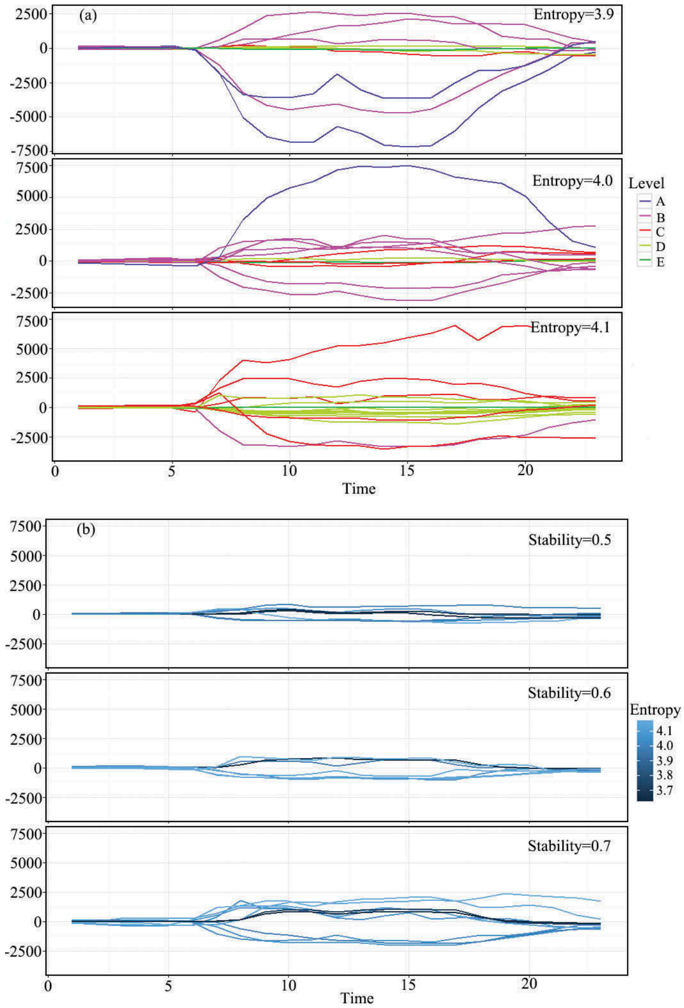

Figure 15. Variations in the *cumulative netflow* for cell phone towers with the same: (a) entropy = 3.9, 4.0 and 4.1; or (b) stability = 0.5, 0.6 and 0.7.

we concluded that entropy can compare the variations for a tower in different time period series, whereas the proposed stability model can compare the variations for different towers. Both measures can be used in a complementary manner.

The proposed model is capable of assessing the stability of human convergence and divergence with different spatial units (e.g. cell phone towers, bus lines and TAZs). Thus, the model can be applied to other cities or data sets to inform decision-making during transport planning. The results of this study are based on mobile phone data that included most of the population of a city. Therefore, our results reflect certain aspects of the aggregate characteristics of human convergence and divergence. In the future, it would be useful to apply the same model to other types of data sets, such as transit smart card data and floating car data, in order to capture a more comprehensive overview of the human convergence and divergence patterns in a city. In addition, there is a common problem when the signal switches between adjacent cell phone towers. We focused on aggregate population flows rather than individual mobility patterns, but we consider that our results are not highly sensitive to this issue because signal switches tend to balance out with each other at an aggregate level. However, this is a data issue that researchers should consider when conducting mobility research using mobile phone data.

In future research, we may need to integrate mobile phone location data with other data sources, such as CCTV video monitoring data, smart transit card data from a city or social media data to obtain more accurate flow assessments. In addition, our future research may focus on investigating the relationships between the HCDS stability and the spatial and social characteristics of urban systems, such as land-use functions and points of interest.

Acknowledgements

The authors also would like to thank the anonymous reviewers and Shawn Laffanfor for their valuable comments and suggestions.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The research was supported in part by the National Natural Science Foundation of China (Grants #41231171, #41371420, #40971233), the innovative research funding of Wuhan University (2042015KF0167).

References

Ahas, R., *et al*., 2015. Everyday space–time geographies: using mobile phone-based sensor data to monitor urban activity in Harbin, Paris, and Tallinn. *International Journal of Geographical Information Science*, 29 (11), 2017–2039. doi:10.1080/13658816.2015.1063151

Calabrese, F., *et al*., 2011. Real-time urban monitoring using cell phones: a case study in Rome. *Intelligent Transportation Systems, IEEE Transactions On*, 12 (1), 141–151. doi:10.1109/TITS.2010.2074196

Clausius, R., 1850. Über die bewegende Kraft der Wärme und die Gesetze welche sich daraus für die Wärmelehre selbst ableiten lassen. *Annalen der Physik*, 79 (368–397), 500–524. doi:10.1002/andp.18501550306

Delbosc, A. and Currie, G., 2011. Using Lorenz curves to assess public transport equity. *Journal of Transport Geography*, 19 (6), 1252–1259. doi:10.1016/j.jtrangeo.2011.02.008

Deville, P., *et al*., 2014. Dynamic population mapping using mobile phone data. *Proceedings of the National Academy of Sciences*, 111 (45), 15888–15893. doi:10.1073/pnas.1408439111

Dong, H., *et al*., 2015. Traffic zone division based on big data from mobile phone base stations. *Transportation Research Part C: Emerging Technologies*, 58, 278–291. doi:10.1016/j.trc.2015.06.007

Gao, S., *et al*., 2013. Discovering spatial interaction communities from mobile phone data. *Transactions in GIS*, 17 (3), 463–481. doi:10.1111/tgis.12042

Gao, S., 2015. Spatio-temporal analytics for exploring human mobility patterns and urban dynamics in the mobile age. *Spatial Cognition & Computation*, 15 (2), 86–114. doi:10.1080/13875868.2014.984300

Gonzalez, M.C., Hidalgo, C.A., and Barabasi, A.-L., 2008. Understanding individual human mobility patterns. *Nature*, 453 (7196), 779–782. doi:10.1038/nature06958

Hammond, G.W. and Thompson, E., 2002. Mobility and modality trends in US state personal income. *Regional Studies*, 36, 375–387. doi:10.1080/00343400220131142

Hanson, S., 2005. Perspectives on the geographic stability and mobility of people in cities. *Proceedings of the National Academy of Sciences*, 102 (43), 15301–15306. doi:10.1073/pnas.0507309102

Hoteit, S., *et al.*, 2014. Estimating human trajectories and hotspots through mobile phone data. *Computer Networks*, 64, 296–307. doi:10.1016/j.comnet.2014.02.011

Hu, Y., *et al.*, 2015. Extracting and understanding urban areas of interest using geotagged photos. *Computers, Environment and Urban Systems*, 54, 240–254. doi:10.1016/j.compenvurbsys.2015.09.001

Hu, Y., Miller, H.J., and Li, X., 2014. Detecting and analyzing mobility hotspots using surface networks. *Transactions in GIS*, 18 (6), 911–935. doi:10.1111/tgis.12076

Kang, C., *et al.*, 2012. Towards estimating urban population distributions from mobile call data. *Journal of Urban Technology*, 19 (4), 3–21. doi:10.1080/10630732.2012.715479

Lichman, M. and Smyth, P., Modeling human location data with mixtures of kernel densities. ed. *Proceedings of the 20th ACM SIGKDD international conference on knowledge discovery and data mining*, 2014, 35–44. doi:10.1145/2623330.2623681

Liu, L., Biderman, A., and Ratti, C., Urban mobility landscape: real time monitoring of urban mobility patterns. In: *Proceedings of the 11th international conference on computers in urban planning and urban management*, 2009, 1–16. http://senseable.mit.edu/papers/pdf/20090618_Liu_etal_UrbanMobility_CUPUM.pdf [Accessed May 2017].

Liu, Y., *et al.*, 2012. Urban land uses and traffic 'source-sink areas': evidence from GPS-enabled taxi data in Shanghai. *Landscape and Urban Planning*, 106 (1), 73–87. doi:10.1016/j.landurbplan.2012.02.012

Liu, Y., *et al.*, 2015. Social sensing: a new approach to understanding our socioeconomic environments. *Annals of the Association of American Geographers*, 105 (3), 512–530. doi:10.1080/00045608.2015.1018773

Lu, Y. and Liu, Y., 2012. Pervasive location acquisition technologies: opportunities and challenges for geospatial studies. *Computers, Environment and Urban Systems*, 36 (2), 105–108. doi:10.1016/j.compenvurbsys.2012.02.002

Mao, L., *et al.*, 2016. Mapping intra-urban transmission risk of dengue fever with big hourly cellphone data. *Acta Tropica*, 162, 188–195. doi:10.1016/j.actatropica.2016.06.029

Mitra, R., Buliung, R.N., and Faulkner, G.E.J., 2010. Spatial clustering and the temporal mobility of walking school trips in the Greater Toronto Area, Canada. *Health & Place*, 16, 646–655. doi:10.1016/j.healthplace.2010.01.009

National Standards of China, GB50220-95, 1995. Code for transport planning on urban road. Available from: http://www.doc88.com/p-7824234712083.html

Pappalardo, L., *et al.*, 2015. Returners and explorers dichotomy in human mobility. *Nature Communications*, 6, 8166. doi:10.1038/ncomms9166

Pearson, K., 1920. Notes on the history of correlation. *Biometrika*, 13 (1), 25–45. doi:10.2307/2331722

Phithakkitnukoon, S., *et al.*, 2010. Activity-aware map: identifying human daily activity pattern using mobile phone data. *International Workshop on Human Behavior Understanding*, 14–25. doi:10.1007/978-3-642-14715-9_3

Ratti, C., *et al.*, 2006. Mobile landscapes: using location data from cell phones for urban analysis. *Environment and Planning B: Planning and Design*, 33 (5), 727–748. doi:10.1068/b32047

Sagl, G., Loidl, M., and Beinat, E., 2012. A visual analytics approach for extracting spatio-temporal urban mobility information from mobile network traffic. *ISPRS International Journal of Geo-Information*, 1 (3), 256–271. doi:10.3390/ijgi1030256

Schneider, C.M., *et al.*, 2013. Unravelling daily human mobility motifs. *Journal of the Royal Society Inerface*, 10, 20130246. doi:10.1098/rsif.2013.0246

Scholz, R.W. and Lu, Y., 2014. Detection of dynamic activity patterns at a collective level from large-volume trajectory data. *International Journal of Geographical Information Science*, 28 (5), 946–963. doi:10.1080/13658816.2013.869819

Sevtsuk, A. and Ratti, C., 2010. Does urban mobility have a daily routine? Learning from the aggregate data of mobile networks. *Journal of Urban Technology*, 17 (1), 41–60. doi:10.1080/10630731003597322

Shaw, S.-L., Tsou, M.-H., and Ye, X., 2016. Editorial: human dynamics in the mobile and big data era. *International Journal of Geographical Information Science*, 30 (9), 1687–1693. doi:10.1080/13658816.2016.1164317

Shen, J. and Cheng, T., 2016. A framework for identifying activity groups from individual space time profiles. *International Journal of Geographical Information Science*, 30 (9), 1785–1905. doi:10.1080/13658816.2016.1139119

Shenzhen Statistical YearBook, 2012. *Shenzhen Statistical YearBook*. Available from: http://www.sztj.gov.cn/nj2012/indexeh.htm.

Shorrocks, A.F., 1978. The measurement of mobility. *Econometrica: Journal of the Econometric Society*, 46(5), 1013–1024. http://www.jstor.org/stable/1911433 [Accessed May 2017]. doi:10.2307/1911433

Simini, F., *et al*., 2012. A universal model for mobility and migration patterns. *Nature*, 484 (7392), 96–100. doi:10.1038/nature10856

Sobolevsky, S., *et al*., 2013. Delineating geographical regions with networks of human interactions in an extensive set of countries. *PLoS One*, 8 (12), e81707. doi:10.1371/journal.pone.0081707

Song, C., *et al*., 2010. Limits of predictability in human mobility. *Science*, 327 (5968), 1018–1021. doi:10.1126/science.1177170

Steiger, E., Resch, B., and Zipf, A., 2016. Exploration of spatiotemporal and semantic clusters of Twitter data using unsupervised neural networks. *International Journal of Geographical Information Science*, 30 (9), 1694–1716. doi:10.1080/13658816.2015.1099658

Sun, J.B., *et al*., 2011. Exploring space-time structure of human mobility. *Physica A*, 390, 929–942. doi:10.1016/j.physa.2010.10.033

Wang, P., *et al*., 2012. Understanding road usage patterns in urban areas. *Scientific Reports*, 2. doi:10.1038/srep01001

Xu, Y., *et al*., 2015. Understanding aggregate human mobility patterns using passive mobile phone location data: a home-based approach. *Transportation*, 42 (4), 625–646. doi:10.1007/s11116-015-9597-y

Yang, X., *et al*., 2016. Understanding spatiotemporal patterns of human convergence and divergence using mobile phone location data. *ISPRS International Journal of Geo-Information*, 5, 177. doi:10.3390/ijgi5100177

Yin, L., *et al*., 2015. Re-identification risk versus data utility for aggregated mobility research using mobile phone location data. *PLoS ONE*, 10 (10), e0140589. doi:10.1371/journal.pone.0140589

You, J., Nedović-Budić, Z., and Kim, T.J., 1998. A GIS-based traffic analysis zone design: technique. *Transportation Planning and Technology*, 21 (1–2), 45–68. doi:10.1080/03081069708717601

Yuan, Y. and Raubal, M., 2012. Extracting dynamic urban mobility patterns from mobile phone data. *Geographic Information Science*. Springer, 354–367. doi:10.1007/978-3-642-33024-7_26

Zayani, M.-H., Gauthier, V., and Zeghlache, D., 2012. Quantifying spatiotemporal stability by means of entropy: approach and motivations. Telecom Sud Paris Research Report #RS2M10001. https://arxiv.org/ftp/arxiv/papers/1207/1207.2564.pdf [Accessed May 2017].

Zhao, Z., *et al*., 2016. Understanding the bias of call detail records in human mobility research. *International Journal of Geographical Information Science*, 30 (9), 1738–1762. doi:10.1080/13658816.2015.1137298

Zhu, X. and Guo, D., 2014. Mapping large spatial flow data with hierarchical clustering. *Transactions in GIS*, 18 (3), 421–435. doi:10.1111/tgis.12100