Landscape and Urban Planning 106 (2012) 73–87

Contents lists available at SciVerse ScienceDirect

Landscape and Urban Planning

journal homepage: www.elsevier.com/locate/landurbplan

Urban land uses and traffic 'source-sink areas': Evidence from GPS-enabled taxi data in Shanghai

Yu Liu[a,b,*], Fahui Wang[b], Yu Xiao[a], Song Gao[a]

[a] Institute of Remote Sensing and Geographical Information Systems, Peking University, Beijing 100871, China
[b] Department of Geography and Anthropology, Louisiana State University, Baton Rouge, LA 70803, USA

ARTICLE INFO

Article history:
Received 29 June 2011
Received in revised form 24 January 2012
Accepted 14 February 2012
Available online 14 March 2012

Keywords:
GPS-enabled taxi data
Urban land use
Traffic 'source-sink area'
Urban form
Shanghai

ABSTRACT

Most of the existing literature focuses on estimating traffic or explaining trip lengths from land use. This research attempts to reveal intraurban land use variations from traffic patterns. Using a seven-day taxi trajectory data set collected in Shanghai, we investigate the temporal variations of both pick-ups and drop-offs, and their association with different land use features. Based on the balance between the numbers of drop-offs and pick-ups and its distinctive temporal patterns, the study area is classified into six traffic 'source-sink' areas. These areas are closely associated with various land use types (commercial, industrial, residential, institutional and recreational) as well as land use intensity. The study shows that human mobility data from location aware devices provide us an opportunity to derive urban land use information in a timely fashion, and help urban planners and policy makers in mitigating traffic, planning for public services and resources, and other purposes.

1. Introduction

Examining the interdependence between land use and travel behavior (especially diurnal movement) is a long tradition in urban studies (e.g., Chen, Chen, & Barry, 2009; Goodchild & Janelle, 1984; Goodchild, Klinkenberg, & Janelle, 1993; Khisty & Lall, 2003; Maat, van Wee, & Stead, 2005; Zandviliet & Dijst, 2006). Most literature focuses on how to estimate traffic or explain trip lengths from urban land use. For instance, much of the traffic demand forecasting relies on modeling origin-destination (O-D) traffic volumes from land use data (Black, 2003). There is also a large body of literature on explaining commuting patterns by urban land use structure. The wasteful commuting proposition (Hamilton, 1982; Small & Song, 1992; White, 1988) questions the assumption in the traditional urban economic model (Mills, 1972; Muth, 1969) that people choose their residential location as a tradeoff between housing space and commuting time, and stimulates much work on linking travel to urban form (Gordon, Kumar, & Richardson, 1989). The jobs-housing balance approach emphasizes that long commutes tend to be associated with spatial separation and severe imbalance between jobs and residences (Cervero, 1989; Wang, 2000), but others challenge the validity of the approach in predicting commute lengths (Giuliano & Small, 1993). More recent work incorporates both spatial structure (i.e., land use) and socioeconomic characteristics in explaining commuting (Antipova, Wang, & Wilmot, 2011; Shen, 2000; Wang, 2001). This school of work is motivated by possible planning and policy remedy in attempt to alter urban land use for journey-to-work reduction, although the results are far short of the expectation (Weber & Sultana, 2007).

The above brief review provides a glimpse of the rich literature on examining the influence of urban land use on travel. In contrast, very little work has been reported on the reversed linkage in the interdependence between land use and traffic, i.e., how to extract the information on land use from traffic patterns. There is just as much value for doing so. First and foremost, land use data are expensive and time consuming to compile, and timely updated land use data needed by urban planners and researchers are scarce. Many are left with no choice but to estimate land use from remote sensing images (e.g., Jensen, 1983; Lu & Weng, 2005; Xiao et al., 2006), but "effective real-world operational examples", especially for urban areas, are a rarity (Rogan & Chen, 2004) due to various challenges. For cities in a developing country such as China, land uses and spatial structure are ever changing in a fast pace. Some areas may experience no transformation in land use type but significant change in land use intensity with important implication in traffic patterns, service planning and others. It is imperative for planners and policy makers to monitor citywide land use changes and detect areas of rapid growth or decline in a timely fashion. This is a task that can be accomplished neither by the costly traditional

* Corresponding author at: Institute of Remote Sensing and Geographical Information Systems, Peking University, Beijing 100871, China. Tel.: +86 10 62751182; fax: +86 10 62751187.
E-mail addresses: liuyu@urban.pku.edu.cn (Y. Liu), fwang@lsu.edu (F. Wang), xiao_yu@pku.edu.cn (Y. Xiao), songgaogeo@gmail.com (S. Gao).

0169-2046/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.landurbplan.2012.02.012

survey method nor remote sensing that cannot detect detailed land use patterns. Secondly, documenting changes of land use at a finer time interval over a period of time can also help researchers identify major phases of urban structure changes (e.g., from a monocentric to a polycentric metropolis, from a bedroom satellite city to a suburban job center, from the downtown primacy to suburb dominance in office markets). Finally, similar to the debate on the effect of land use on travel patterns, our research on this topic will provide supporting evidence on the interdependence between them, but from an inverted angle.

The lack of research along this line of work may be attributable to paucity of detailed intra-urban origin-destination traffic data until recently a large volume of human motion data have become available. With the rapid development of information and communication technologies (ICT), data from location aware devices (LAD), including mobile phones and Global Positioning System (GPS) receivers, have been collected and made available. Such data have been widely applied in human mobility pattern research (González, Hidalgo, & Barabási, 2008; Jiang, Yin, & Zhao, 2009; Kang, Ma, Tong, & Liu, 2012; Rhee, Shin, Hong, Lee, Chong, 2008; Song, Koren, Wang, & Barabási, 2010; Song, Qu, Blumm, & Barabási, 2010). Identified patterns from LAD-generated data provide an important approach to understanding urban dynamics (Chowell et al., 2003; Phithakkitnukoon, Horanont, Lorenzo, Shibasaki, & Ratti, 2010; Ratti, Pulselli, Williams, & Frenchman, 2006). Much research has been conducted to investigate the spatio-temporal patterns of urban scale human motion. Related applications include studies in cities such as Tallinn of Estonia (Ahas, Aasa, Silm, & Tiru, 2010), Milan (Ratti et al., 2006) and Rome of Italy (Sevtsuk & Ratti, 2010), and Hong Kong (Jiang & Liu, 2009), Wuhan (Li, Zhang, Wang, & Zeng, 2011) and Shenzhen of China (Sun, Yuan, Wang, Si, & Shan, 2011).

Some research has been conducted on collecting GPS data to investigate travel behavior and the underlying geographic environment (Stopher, Fitzgerald, Zhang, 2008; Wolf, Oliveira, & Thompson, 2003). Recently, GPS-enabled vehicles have been widely adopted to collect real time traffic data (Dai, Ferman, & Roesser, 2003; Kühne et al., 2003; Lü, Zhu, Wu, Dai, & Huang, 2008; Tong, Coifman, Merr, 2009). In many applications, such data are often obtained from GPS-enabled taxis instead of private automobiles due to privacy concerns. In addition to monitoring real-time traffic situations, large volume of GPS-enabled taxi trajectory data have been used for travel time estimation, dynamic accessibility evaluation, and travel behavior analysis (Berkow, Monsere, Koonce, Bertini, & Wolfe, 2009; Bricka & Bhat, 2006; Liu, Andris, & Ratti, 2010). A study by Qi et al. (2011) uses the dynamic taxi data to reflect coarse urban land uses, but does not systematically examine the relationship between the diurnal motion patterns and the underlying urban structure.

This research uses a massive data set of over 6600 taxis for seven days in Shanghai, China, to reveal the association between traffic patterns and urban land uses. On any workday, one generally leaves home to work in the morning and goes back in the afternoon or evening. Commutes exhibit a strong daily rhythm for a city. We borrow a concept, '*source-sink*,' developed by Pulliam (1988) in ecological studies to characterize the daily travel patterns. In ecology, a source patch has higher birth rates than death rates and thus a growing population, and a sink patch has higher death rates than birth rates and thus a declining population. Here, a traffic source area has more pick-ups than drop-offs from the taxi data and thus a net traffic outflow; and a traffic sink area otherwise. Therefore, a residential area and workplace can be viewed as a source and sink in the morning, respectively; and their roles switch in the afternoon. By classifying each area in terms of its temporal balance of generation or attraction of traffic, we can examine the spatial patterns of these classifications and their association with different types of land uses. Urban development is an ongoing process, and changes are rapid particularly in cities in a fast-growing economy such as China. Transportation planning and implementing traffic mitigation measures demand timely data beyond the traditional data sources such as census or land use survey. Large volume of real-time data collected by GPS-enabled vehicles fulfills this demand. This research demonstrates the potential of using such data to monitor residents' travel patterns and reveal the dynamics of urban land use changes.

Urban land use types include commercial, industrial, residential, transportation and others, and can be grouped into two broad categories (employment and population) in terms of its implication on traffic generation and transportation planning. The "population (evening-time) density pattern reflects the variation of residential land use", and "employment (daytime) density captures business-related land uses including industrial, commercial and others" (Wang, Antipova, & Porta, 2011). The identification of traffic source-sink areas and related traffic intensities by GPS-enabled taxi data is an important step toward timely monitoring land use changes and planning effective transportation mitigation responses.

The remainder of this paper is organized as follows. Section 2 describes the data source and processing. Section 3 discusses the methods for analyzing temporal patterns of trips and classifies the study area into various source-sink areas. Section 4 examines the association of the classifications with various land uses. The paper is concluded with a brief summary and some discussion in Section 5.

2. Data preparation and study area

Shanghai is the most populous city in China. Its socio-spatial structure has gone significant changes in the post-reform era (Li & Wu, 2006; Wu & Li, 2005). Some research has examined the impact of urban form on residents travel behavior (Chai, Weng, & Shen, 2008; Pan, Shen, & Zhang, 2009; Song & Ikeda, 2005). But those studies used survey data with small samples, and self-reported travel data may not be representative of general population and often lack accuracy in reported variables of a trip. This research uses a large volume of trip data collected from location-aware devices to examine the spatial structure of Shanghai.

Like most cities in China, taxis play an important role in intra-urban transportation in Shanghai. According to the Shanghai Municipal Transport and Port Authority (http://www.jt.sh.cn/), taxi-based trips account for about 20% of daily trips in 2009. Many taxi companies have installed GPS receivers in their fleets to monitor the real-time movement of each taxi. This research uses a data set of more than 6600 taxis for seven consecutive days (from June 1st to June 7th in 2009) from an anonymous taxi company. The data recorded each taxi's location, velocity, and status (vacancy or occupancy of passengers) in about every 10 s with the position accuracy of about ±10 m, which is acceptable in investigating intra-urban travel patterns. All trajectories were cleaned by removing invalid points caused by data recording or transfer errors. Fig. 1 demonstrates a one-day trajectory of a taxi, where red lines denote the paths with passengers in the taxi, and blue lines indicate the taxi unoccupied. Based on the data, we can identify the locations where passengers were picked up and dropped off, and thus the origin and destination of a completed trip. Each trip can be simplified to be a vector from (x_o,y_o,t_o) to (x_d,y_d,t_d), where (x,y) denotes the location and t time of a pick-up (with subscript "o") event and a drop-off (with subscript "d") event, respectively. There were a total of 1,552,635 trips extracted from the data. These trips form a representative sample of intra-urban movement.

Naturally the trips were more concentrated in central urban areas than suburban or urban fringes. This can be demonstrated

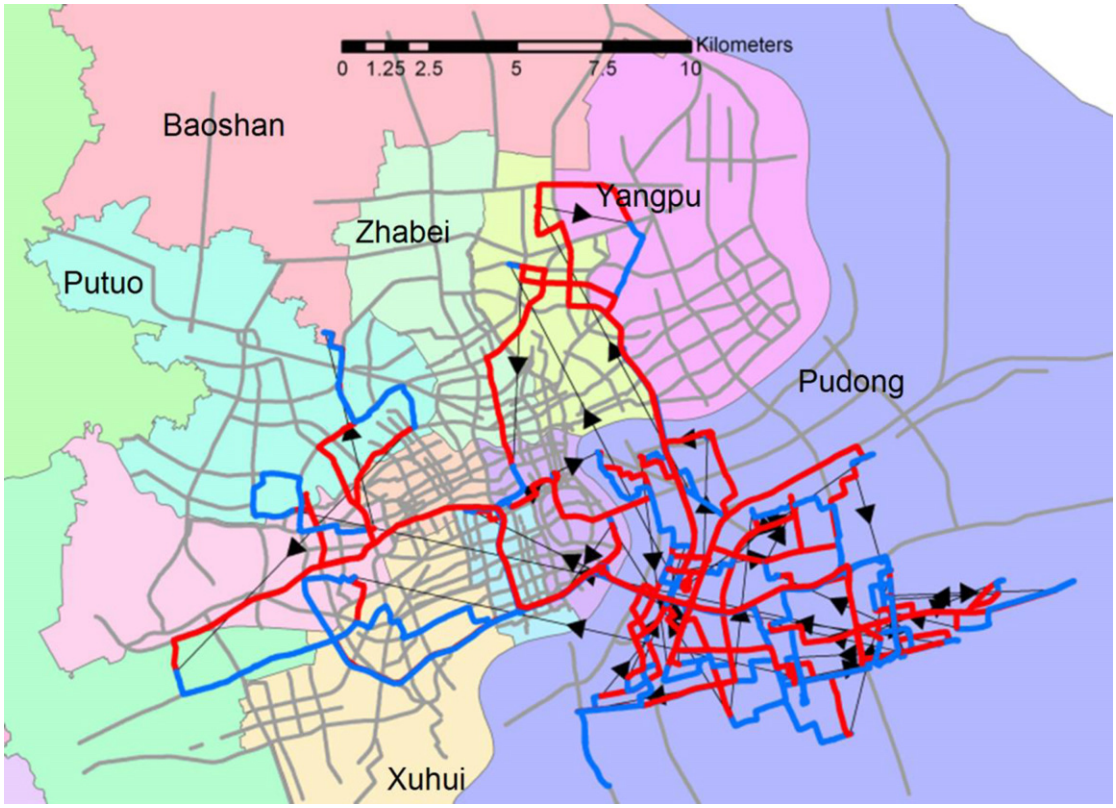

Fig. 1. A taxi's one-day trajectory in Shanghai on June 1, 2009 [a red line along the streets denotes the status with passengers, and a straight line with an arrow denotes an O-D pair of a trip]. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

by density analyses on pick-up points (PUPs) and drop-off points (DOPs). Fig. 2a and b depicts the kernel density estimations of all PUPs and DOPs in the seven days. The spatial distributions of both PUPs and DOPs are positively correlated with the population distribution (Fig. 2c). Fig. 2c shows estimated population density based on the LandScan™ 2008 High Resolution Global Population Data Set (http://www.ornl.gov/sci/landscan/) because of lack of more recent census data in the study area. The Pearson correlation coefficients for PUPs vs. population and DOPs vs. population are 0.785 and 0.783, respectively.

Based on the distributions of PUPs and DOPs as well as population, we select a 35 km × 50 km rectangle region that covered the major urban areas of Shanghai as the study area for this research (see the box in Fig. 2c). The rectangle area was discretized into 1750 1 km × 1 km cells. The right-top corner of the rectangle contains 313 no-data cells that were excluded from the analysis. In order to investigate the temporal characteristics of trips, we also discretize the seven days into 168 1-h intervals. PUPs and DOPs represent origins and destinations of various trips, and are critical for us to understand traffic patterns and related land use in the city. For each cell, the numbers of PUPs and DOPs in each hour was computed. In summary, two three-dimensional matrices, $\boldsymbol{P}[i, j,t]$ and $\boldsymbol{D}[i,j,t]$, were obtained for pick-ups and drop-offs, respectively. Note that $i = 1,\ldots, 35$; $j = 1,\ldots,50$, and $t = 1,\ldots,168$. In this research, we use R (=35 rows), C (=50 columns), and T (=168 h) to represent the spatio-temporal extent of the data set.

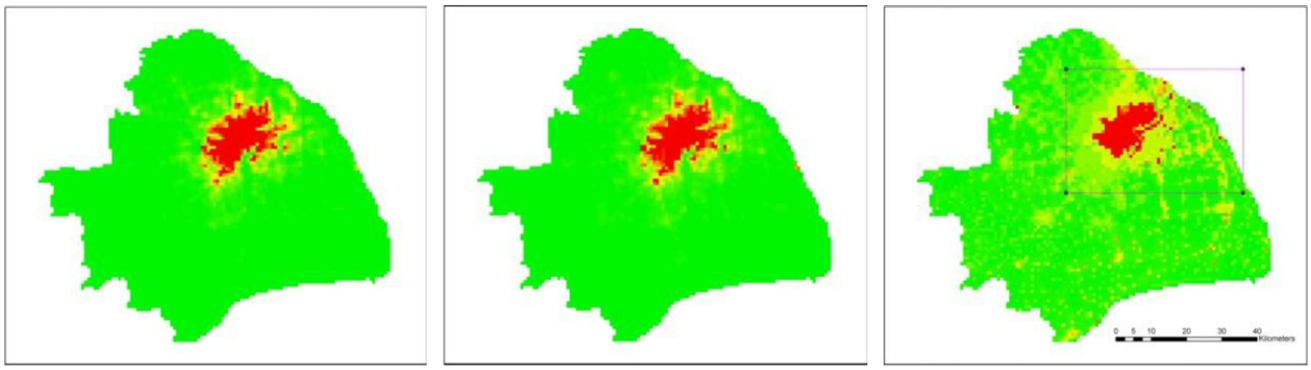

Fig. 2. (a) Spatial distribution of pick-up points, (b) spatial distribution of drop-off points, (c) population distribution (c is based on the LandScan 2008™ High Resolution Global Population Data Set, copyrighted by UT-Battelle, LLC, operator of Oak Ridge National Laboratory under Contract No. DE-AC05-00OR22725 with the United States Department of Energy).

3. Spatio-temporal patterns of taxi pick-ups and drop-offs

3.1. Temporal pattern of pick-ups and drop-offs

For each hour in the seven days, we compute the total numbers of pick-ups and drop-offs in the study area such as

$$S_p[t] = \sum_{i=1,j=1}^{R,C} \boldsymbol{P}[i,j,t],\ t = 1,\ldots,T \quad (1)$$

$$S_d[t] = \sum_{i=1,j=1}^{R,C} \boldsymbol{D}[i,j,t],\ t = 1,\ldots,T \quad (2)$$

where i and j specify the coordinates of a pixel. The results are plotted in Fig. 3a. The temporal sequences represented by curves of S_p and S_d are very similar. Seven 24-h cycles can be clearly identified. That is to say, the temporal distributions of both pick-ups and drop-offs roughly repeat in the seven days. The Fourier transform is a mathematical operation to find the underlying frequencies of a time-series data set, and the 24-h period length can be further confirmed by the discrete Fourier transform (DFT) analysis on S_p and S_d. Since S_p and S_d are similar, we simply perform the DFT on $S_p + S_d$. The power of DFT is plotted in Fig. 3b, where the peak corresponds to 24, indicating that the dominant period length is 24 h. Additionally, we can find three peaks in the daytime on weekdays. They correspond with three temporal intervals, i.e., morning, noon, and early evening, with large numbers of trips.

The temporal distributions of pick-ups and drop-offs can be further examined in detail at the cell (pixel) level. For each 1 km × 1 km cell in the study area, two 168-dimensional vectors, denoted by V_p and V_d, can be constructed to represent the temporal variations of trips in the area. Five sample points are selected from the study area to represent various locations (land uses) (Fig. 4), and their corresponding V_p and V_d are depicted in Fig. 5. Table 1 summarizes the statistics of V_p and V_d associated with the five sample locations. Their temporal patterns differed significantly. For example, the average numbers of pick-ups and drop-offs were roughly equal for cells A and B. In either cell C or D, however, the average number of pick-ups was much fewer than the average number of drop-offs. Cell E had far lower numbers of pick-ups and drop-offs than the other four locations.

To investigate the temporal characteristics of the five points, we obtain the peak point of each location on every day over the seven days. A peak point is the time corresponding to the maximum number of pick-ups (or drop-offs) in 24 h. Fig. 6 plots the distribution of the peak points, some of which occurred regularly at the same hour of a day. The temporal patterns of cells A-D from Fig. 5a–d, where there are large numbers of pick-ups and drop-offs, exhibit a 24-h cycle but of different patterns, suggesting that the areas play different roles in trip generation and attraction. In the downtown area (cell A, Fig. 5a), there is only one peak of drop-offs in the morning as commuters reported to work, and there are two peaks of pick-ups (one in the morning when people left home to work and another in late afternoon till early evening as people returned home). The downtown area in a Chinese city including Shanghai usually has a significant number of residents. The trend is reversed in the central city residential subdivision (cell B, Fig. 5b) with a morning peak of pick-ups and an early evening peak of drop-offs. The two major airports of Shanghai, Hongqiao Airport and Pudong Airport (cells C and D shown in Fig. 5c and d), exhibit a similar pattern with a morning peak of drop-offs and steady numbers of pick-ups throughout a day into the evening time. The suburban area (cell E, Fig. 5e) has small numbers of both pick-ups and drop-offs, and its temporal pattern is less regular (Fig. 6). In summary, different locations and land uses have distinctive traffic patterns revealed in taxi pick-ups and drop-offs. This will be explored in depth in the next sub-section.

3.2. Spatio-temporal patterns of differences between pick-ups and drop-offs

From the previous discussion, the temporal patterns of taxi pick-ups (V_p) and drop-offs (V_d) varied a great deal from place to place. For example, comparing Fig. 5a with b, we can find that during 7:00–8:00 am of each day ($t = 8, 32, \ldots, 152$), V_p and V_d were roughly the same for point A, but $V_p > V_d$ for point B. On the contrary, during 7:00–8:00 pm ($t = 18, 44, \ldots, 164$), $V_p > V_d$ for point A and $V_p < V_d$ for point B. This observation enlightens us that the difference between the numbers of pick-ups and drop-offs may reveal the distinctive functions of different areas.

Due to the periodic properties of $\boldsymbol{P}$ and $\boldsymbol{D}$, in the following research, we aggregate $\boldsymbol{P}$ and $\boldsymbol{D}$ into two 24-h matrices using

$$\boldsymbol{P}'[i,j,t] = \sum_{k=0}^{6} \boldsymbol{P}[i,j,k\times 24+t],\ i = 1,\ldots,\mathrm{R};\ j = 1,\ldots,C,\quad t = 1,\ldots,24 \quad (3)$$

$$\boldsymbol{D}'[i,j,t] = \sum_{k=0}^{6} \boldsymbol{D}[i,j,k\times 24+t],\ i = 1,\ldots,R;\ j = 1,\ldots,C,\quad t = 1,\ldots,24 \quad (4)$$

where i and j denote a pixel, t and k stand for a 1-h period and one day, respectively. $\boldsymbol{P}'$ and $\boldsymbol{D}'$ represent the diurnal variations of pick-ups and drop-offs. For each hour t, we compute the difference between them, $\boldsymbol{D}'[i,j,t]$-$\boldsymbol{P}'[i,j,t]$ ($i = 1,\ldots,R$ and $j = 1,\ldots,C$), a two-dimensional matrix that can be visualized as an image. It is termed drop-offs pick-ups balance matrix (DPBM). Meanwhile, for each pixel $[i,j]$, we use drop-offs pick-ups balance vector (DPBV) for $\boldsymbol{D}'[i,j,t]$-$\boldsymbol{P}'[i,j,t]$ ($t = 1,\ldots,24$). Obviously, DPBM represents the spatial variation of trips for a given time, while DPBV indicates the temporal pattern associated with a position.

Fig. 7a–c depicts the DPBMs during 7:00–8:00 am (morning time), 13:00–14:00 pm (noon time), and 19:00–20:00 pm (evening time), respectively. Note that the images were smoothed using the resample function with cubic interpolation for better visualization effects. The contour lines in each image denote zero values, which indicate the numbers of pick-ups and drop-offs were equal.

Fig. 7a–c shows interesting spatial patterns. In morning time, the number of drop-offs is greater than that of pick-ups in the downtown area (shown in yellow color in Fig. 7a), as the downtown area is a major destination of many trips. Meanwhile, the number of trips from this area is relatively small. Borrowing the terms from ecological studies as described earlier, the downtown can be viewed as a major *sink area* of trips with a large volume of inflow traffic ($\boldsymbol{D}' > \boldsymbol{P}'$) in the morning. Similarly, the two airports and two railway stations are also major sink areas in the morning. On the other hand, the areas in blue in Fig. 7a are *source areas* with a high volume of outflow traffic ($\boldsymbol{D}' < \boldsymbol{P}'$). These areas are generally the residential areas of Shanghai. In evening time, the pattern is reversed (Fig. 7c). Most sink areas in morning time become source areas in evening time, and source areas in morning time are transformed to sink areas in evening time. Fig. 7b shows the pattern of $\boldsymbol{D}' - \boldsymbol{P}'$ at noon, where the sink areas are more scattered and the numbers of both pick-ups and drop-offs are also smaller than Fig. 6a. Note that the four transportation facilities are still sinks in noontime but sources in evening time.

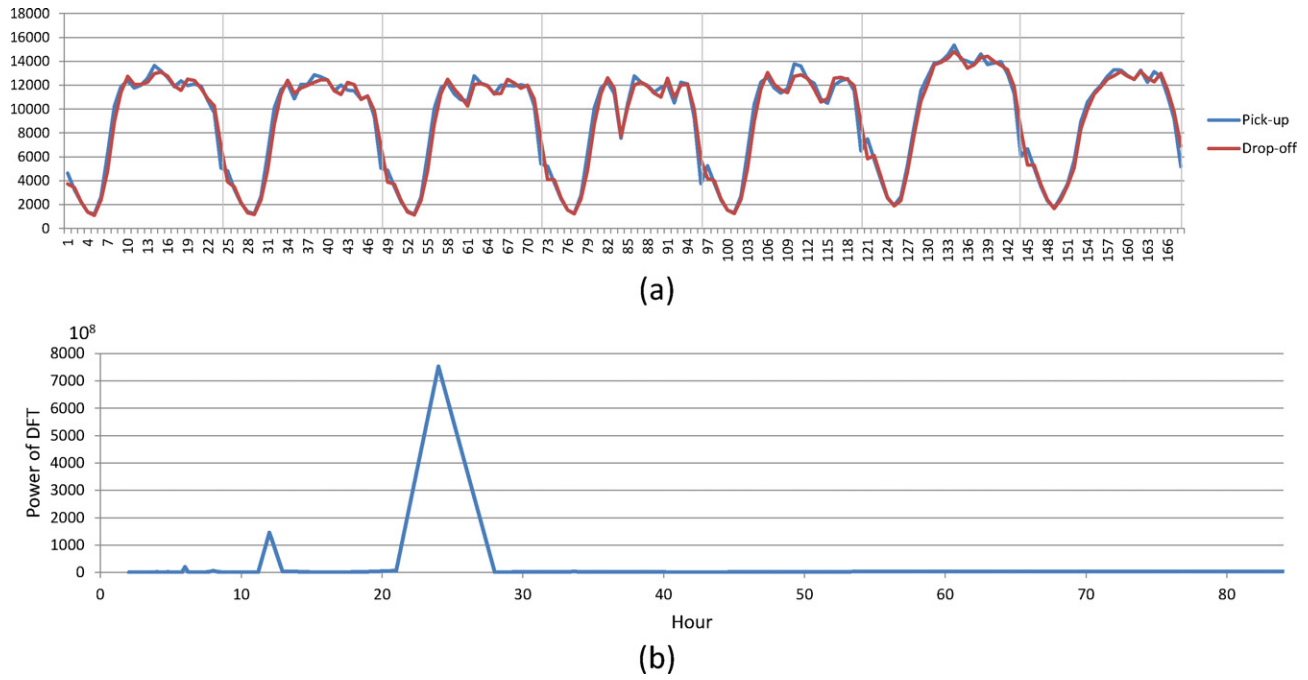

Fig. 3. (a) Curves of S_d and S_p representing temporal variations of pick-ups and drop-offs of the entire study area, (b) power of discrete Fourier transform (DFT) on $S_p + S_d$.

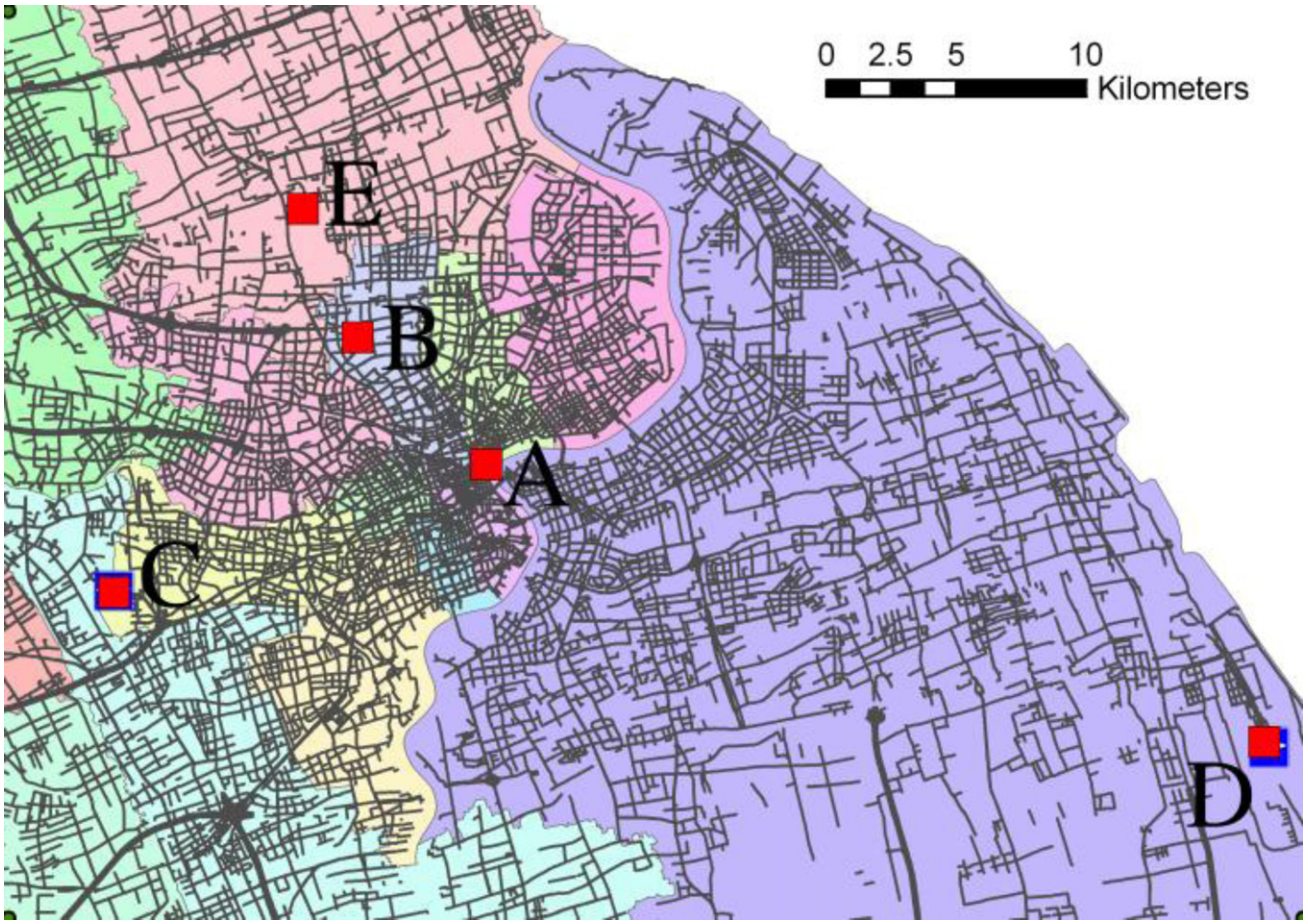

Fig. 4. Five sample points in the study area (A: downtown; B: residential; C: Hongqiao Airport; D: Pudong Airport; E: suburban).

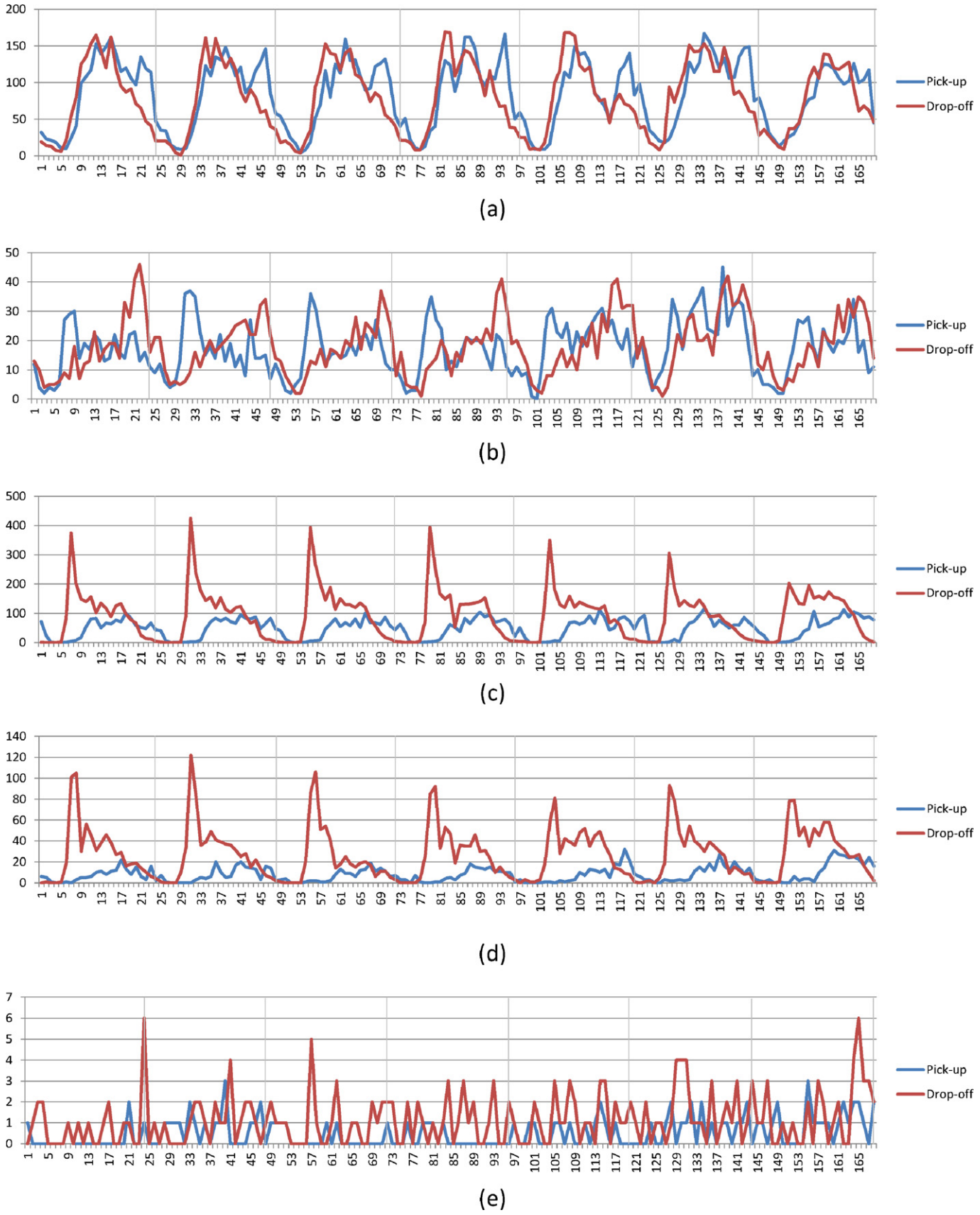

Fig. 5. Curves of V_p and V_d representing temporal variations of pick-ups and drop-offs of 5 sample points A–E.

Table 1
Statistics of diurnal pick-up points (PUP) and drop-off points (DOP) at five sample locations.

	A		B		C		D		E	
	PUP	DOP	PUP	DOP	PUP	DOP	PUP	DOP	PUP	DOP
Average	84.31	78.35	17.28	18.02	48.55	92.01	7.97	26.92	0.51	1.13
Std. dev.	46.82	48.83	9.24	10.17	34.42	85.64	7.45	25.45	0.69	1.24

3.3. Classifications of source-sink areas by temporal balances of taxi pick-ups and drop-offs

As shown in Fig. 7, the spatio-temporal pattern of balance between drop-offs and pick-ups ($\boldsymbol{D}' - \boldsymbol{P}'$) reflects the variation of land uses across the study area. In other words, the temporal variation of $\boldsymbol{D}' - \boldsymbol{P}'$ for each pixel reveals its role in trip generation. Based on the DPBV values such as $(\boldsymbol{D}' - \boldsymbol{P}')[i,j,t]$ for pixel $[i,j]$ at time t (=1,. . .,24), we classify all pixels in the study area into different classes of traffic source-sink areas using the k-means clustering method. In other words, a pixel is grouped into a class (cluster) so that the within-cluster total variances of DPBV is minimized (MacQueen, 1967).

The center vectors of five classes are plotted in Fig. 8a. The five classes of areas exhibit different temporal patterns, and are named as "*strong source-sink*", "*weak source-sink*", "equilibrium", "*weak source-sink*", and "*moderate source-sink*". A sink-source area is a sink area in morning time and a source area in evening time. In contrast, a source-sink area is a source area in morning time and a sink area in evening time. Terms "strong", "moderate" and "weak" reflect the high, medium and low values of the trips, respectively. An equilibrium area has a relatively flat curve, indicating that the numbers of pick-ups and drop-offs are roughly balanced in each hour. Equilibrium areas can be further divided into two classes: *high equilibrium* and *low equilibrium*. High-equilibrium areas have high numbers of pick-ups and drop-offs, and low-equilibrium areas have low numbers of pick-ups and drop-offs. Both equilibrium areas have similar values of DPBV, $(\boldsymbol{D}' - \boldsymbol{P}')$, near 0. In order to differentiate them, Fig. 8b is based on the total counts of drop-offs and pick-ups ($\boldsymbol{D}' + \boldsymbol{P}'$). The spatial distribution of the resulting six classes is shown in Fig. 8c.

The DPBV is parallel to the concept of "jobs to resident workers ratio" in the jobs-housing balance approach discussed in Section 1. In an area where jobs outnumber resident workers, it is more likely a source-sink area with a negative DPBV in the morning and a positive DPBV in the evening. The opposite can be said in an area where there are more resident workers than jobs.

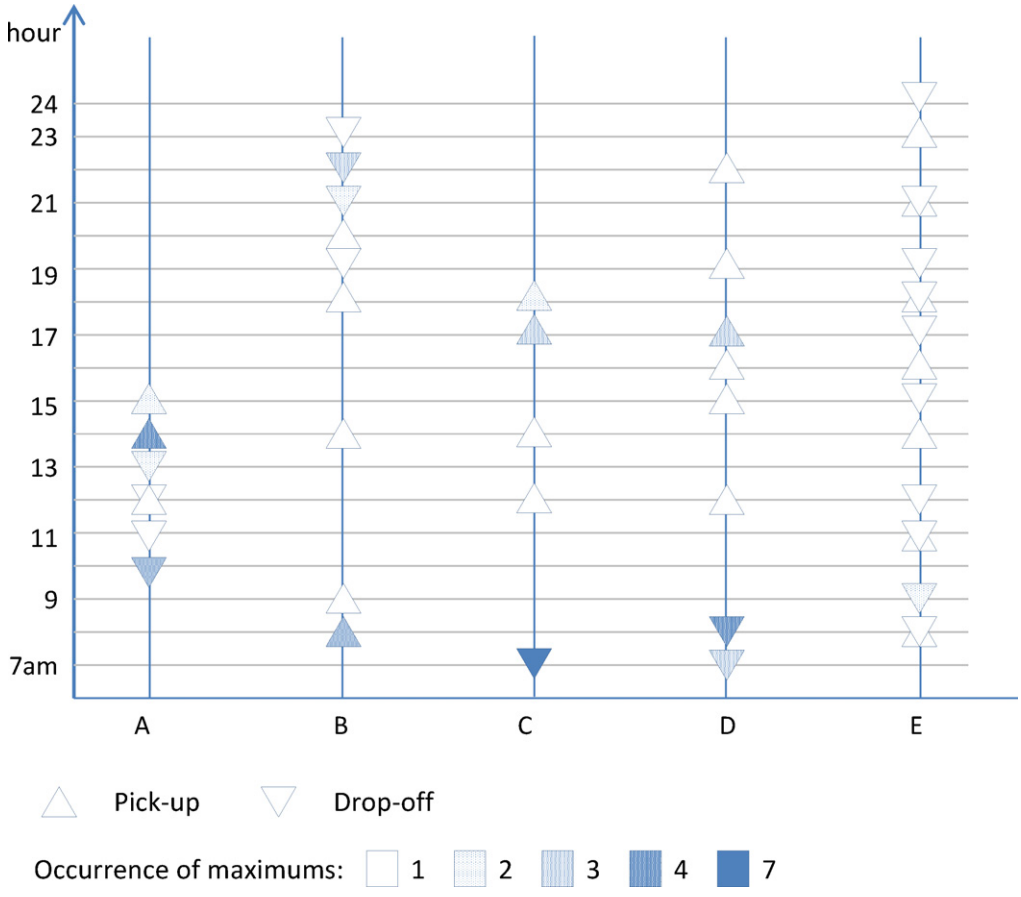

Fig. 6. Repetitive occurrences of maximum pick-ups and drop-offs for 5 sample points in 7 days.

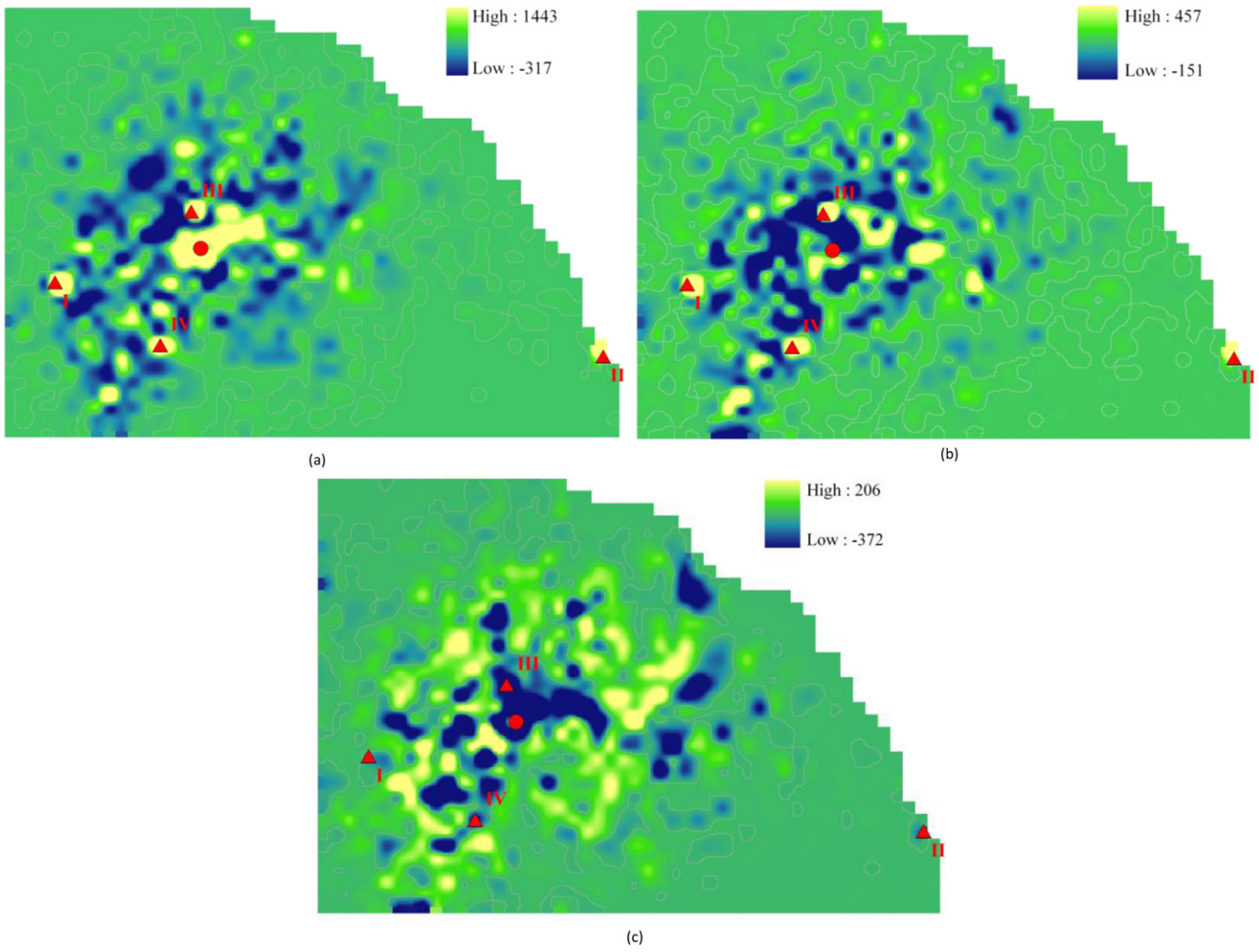

Fig. 7. Images of drop-offs pick-ups balance vector (DPBV), ($\boldsymbol{D}^t - \boldsymbol{P}^t$), in different times: (a) 7:00–8:00 am, (b) 13:00–14:00 pm, (c) 19:00–20:00 pm. The circle denotes downtown, and 4 triangles mark I: Hongqiao Airport, II: Pudong Airport, III: Shanghai Railway Station, IV: Shanghai South Railway Station. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

4. Association of source-sink area classifications with land uses

After a close examination of Fig. 8a, the following conjectures are proposed on possible land use implications of various traffic source-sink areas. A typical commuting trip begins at home and stops at workplace in morning time, and returns from workplace to home again in evening time. Therefore, a weak or moderate source-sink area indicates a residential dominant area. On the other side, a weak or strong source-sink area reflects an employment concentration area such as commercial, industrial, or institutional land uses. The class "high equilibrium" corresponds to mixed land uses of high-density development areas, and the class "low equilibrium" is associated with remote suburbs or even rural areas in the outskirts of Shanghai, where the trip number is low. The presence of strong source-sink areas and absence of source-sink areas indicate that the high-density residential areas do not generate as much traffic as high-density employment area.

4.1. Urban land use features and their densities

Two data issues warrant some discussion prior to our analysis. First, urban trips include home-workplace commutes, and other trips of other purposes such as shopping, recreation, education and social activities. Some trips by taxi are for non-commuting purposes (Li, Yuan, Xie, Cao, & Wu, 2007). For an individual trip maker, it may not end at his/her workplace, but the place being visited for other purposes may well be the workplace of others. At an aggregate level for all trips in the study area, data on taxi trips are valuable in revealing the urban structure and land use patterns. One may argue that taxi-based trips might not reflect a complete picture of urban travels that are made of multiple modes. For our purpose of revealing land use patterns, taxi trips carry significant signals. Chinese cities (particularly large cities) have more concentrated activities and higher-density settlements than most western cities, and thus taxi generally plays a more prominent role in urban transportation in both mode share and spatial range. Taxis tend to be confined in down area, airports and major transit terminals in a western city, but are usually available citywide in China. The taxi ridership also represents a wide spectrum of demographics and income groups in China.

Another issue is the ambiguous roles of major transportation facilities in generating taxi trips in Shanghai. Presence of a bus or subway station may reduce the number of taxi trips in an area or its nearby area because of convenient access to public transportation. On the other hand, a public transit station may

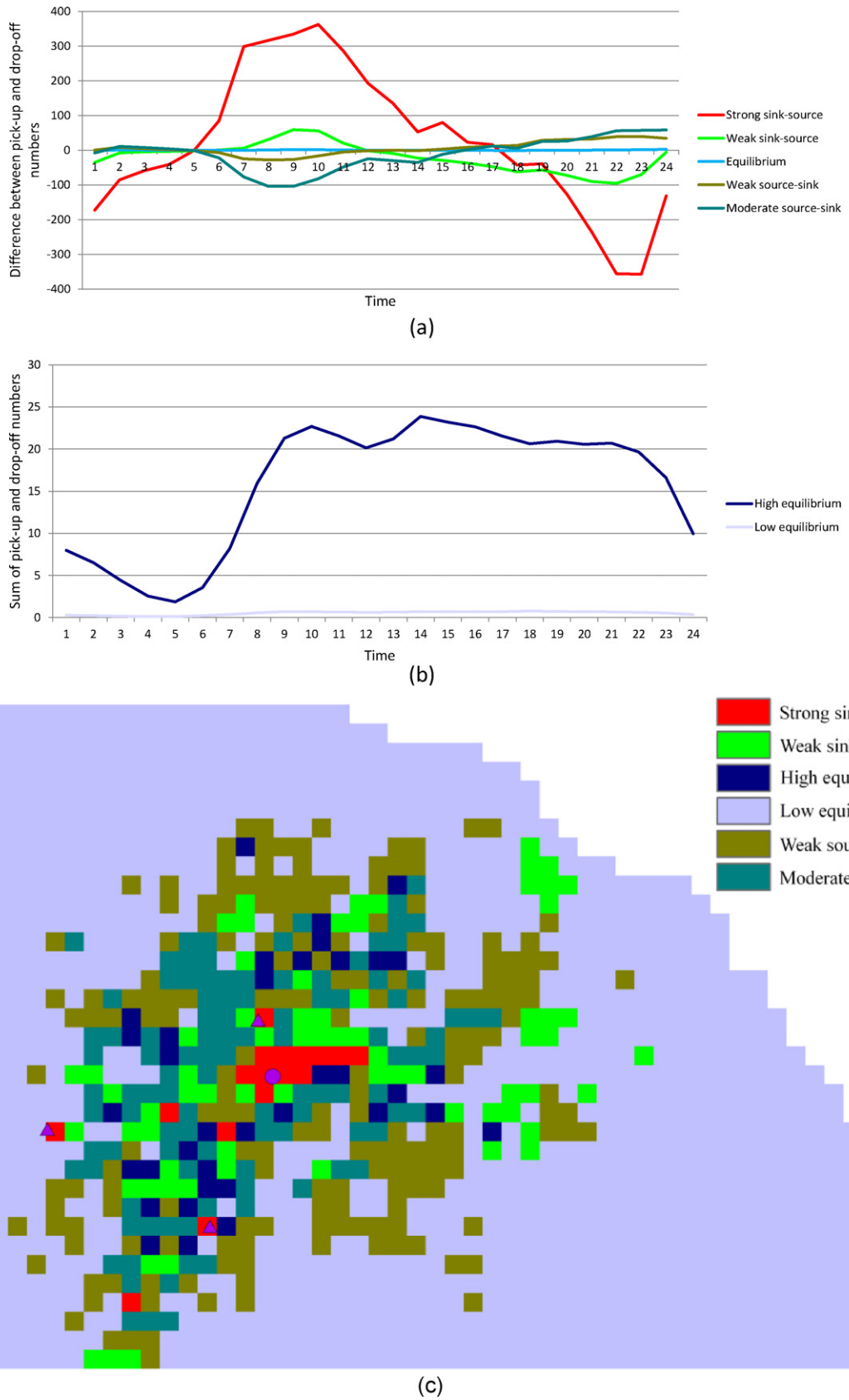

Fig. 8. (a) Center vectors of source-sink areas based on ($\boldsymbol{D}' - \boldsymbol{P}'$), (b) center vectors of two types of equilibrium areas based on ($\boldsymbol{D}' + \boldsymbol{P}'$), (c) spatial distribution of six types of source-sink areas.

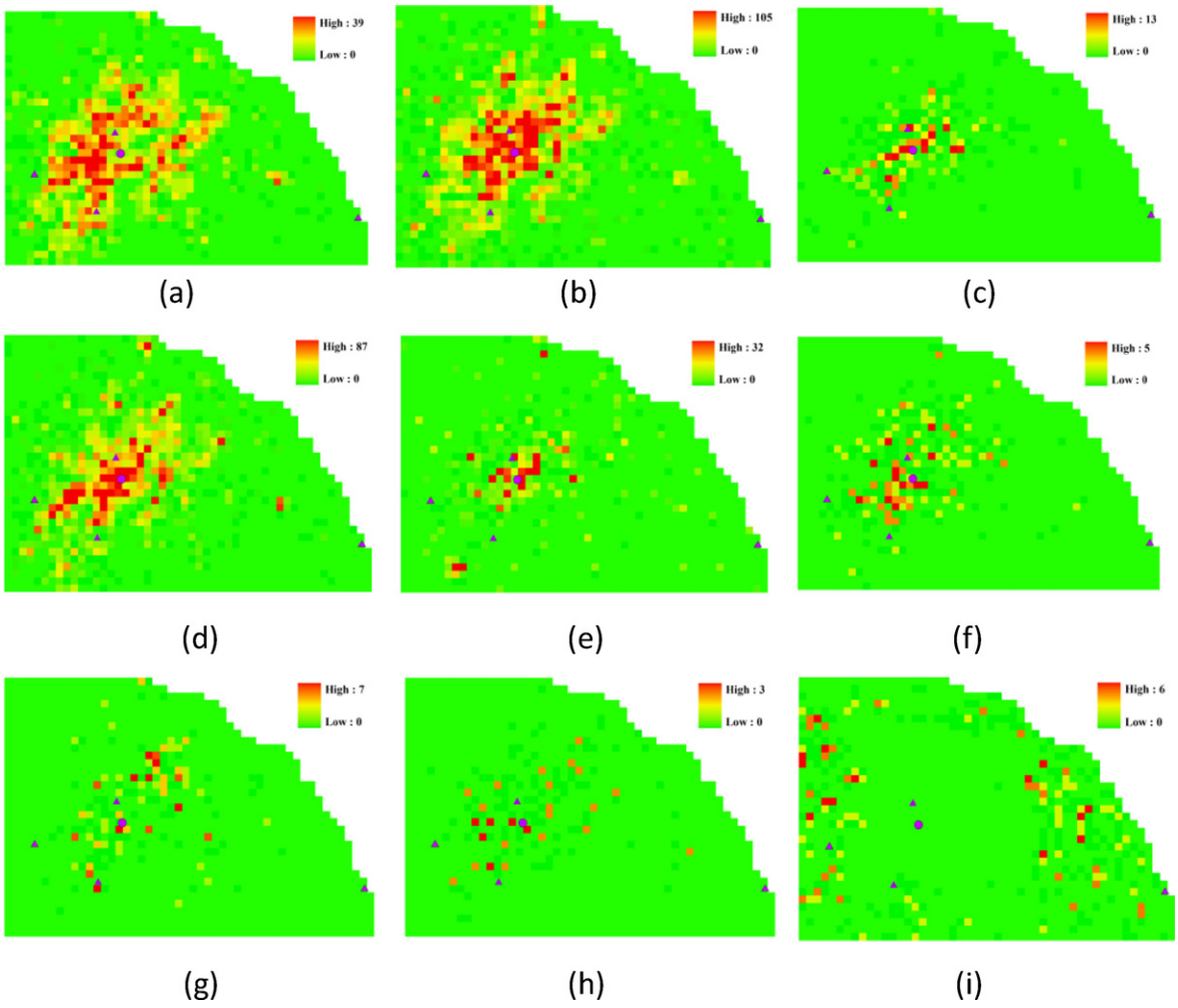

Fig. 9. Spatial distributions of nine types of point geographic features: (a) residential subdivisions, (b) shopping places, (c) hotels, (d) restaurants, (e) government offices, (f) hospitals, (g) schools, (h) parks, and (i) factories.

also serve as a transfer point linking short taxi trips to public transportation, and thus also lead to more taxi trips. For this reason, transportation facilities such as transit stations are usually not the real origins or destinations but transfer points of trips, and they are not included in the subsequent analysis of association between land use features and source-sink area classifications.

Based on the data collected by a Web map provider in 2007, this research has extracted nine types of geographical features to represent urban land uses. The nine types are: residential subdivision (*ju-zhu xiao-qu*), government office, hospital, school, park, shopping place (including shopping mall, super market, etc.), hotel, restaurant, and factory. These point features are potentially origins and destinations of taxi trips. Each feature class was input into a geographical information system (GIS) as a point feature, and the system computed the number of each point features inside every 1 km × 1 km cell to match the framework of source-sink areas. Fig. 9 shows the resulting density maps that depict the spatial distributions of various land uses in the study area.

In order to examine the spatial variation of a geographic feature's density and compare across different features, we use the *normalized density* $\boldsymbol{L}_k[i, j]$ to capture the relative intensity of kth ($k = 1,\ldots,9$) feature for pixel $[i, j]$ such as,

$$L_k[i,j] = \frac{N_k[i,j]}{\sum N_k[l,m]/M}$$

where the numerator $N_k[i, j]$ denotes the number of points inside cell $[i, j]$, the dominator is the average number of a

Table 2
Primary rules for classifying the six traffic source-sink areas based on the distributions of four classes of point features.

	Residential	Commercial	Institutional and recreational	Industrial
Strong sink-source	4.5–9.5	>27.5	>12.5	N/A
Weak sink-source	<8.5	>27.5	<12.5	N/A
High Equilibrium	8.5–9.5	>27.5	<12.5	N/A
Low Equilibrium	<1.5	N/A	N/A	N/A
Weak source-sink	6.5–9.5	>2.5	<4.5	N/A
Moderate source-sink	>9.5	<12.5	<4.5	N/A

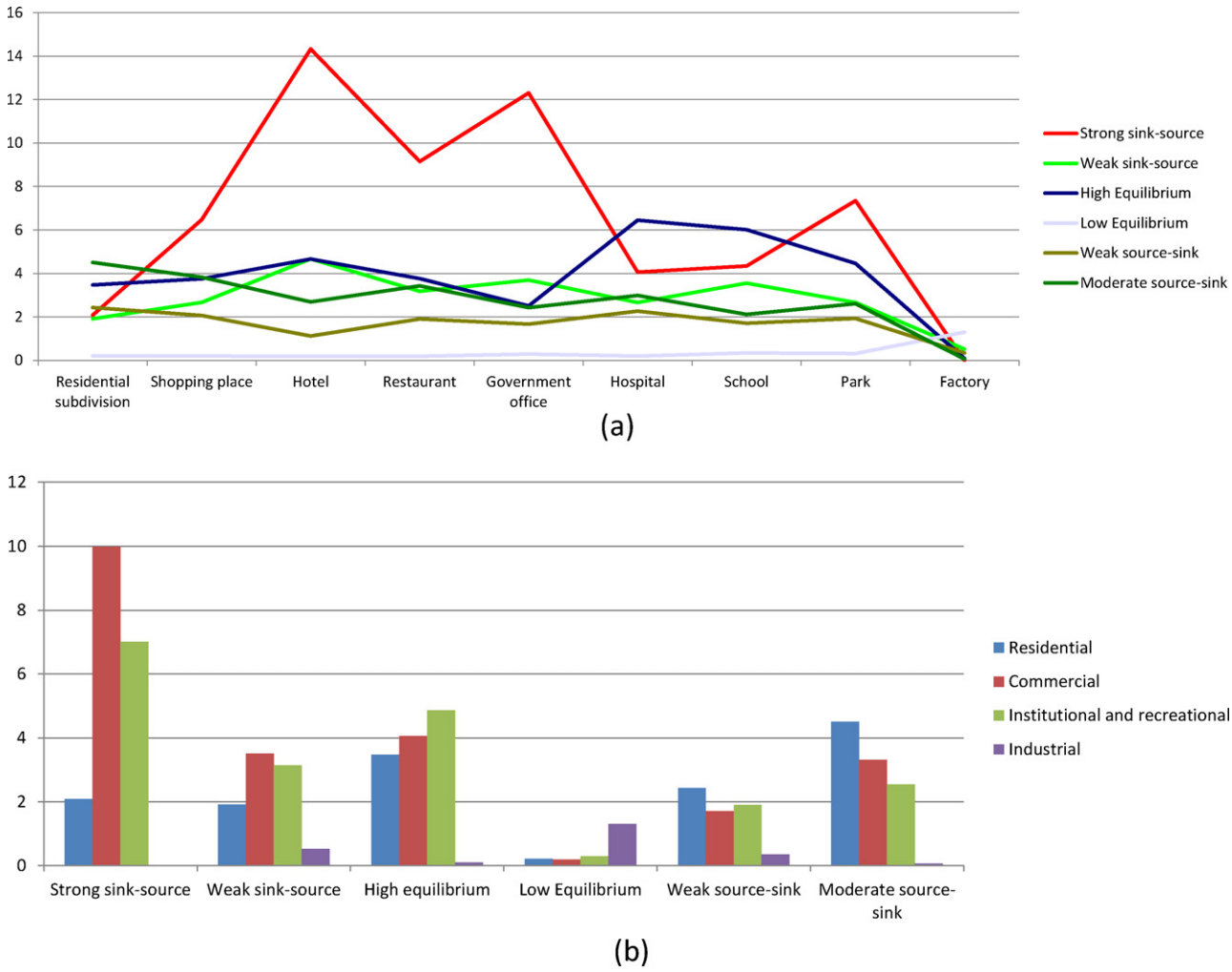

Fig. 10. Relative densities of land use features for various trip patterns: (a) 9 feature classes and (b) 4 aggregated classes.

geographic feature in a cell (i.e., its total number divided by the number of pixels in the study area M (=1437)).

4.2. Characterizing traffic source-sink areas by land use features

For each type of the six traffic source-sink areas, the average of normalized density L_k[i, j] for the kth feature (across all cells that are classified to this particular type of traffic area) is computed and plotted in Fig. 10a. With regard to the pattern of strong source-sink areas, the normalized densities of shopping places, hotels, restaurants, government offices, and parks are greater than those of other features such as residential subdivisions. In other words, these geographic features tend to be sinks in morning time but sources in afternoon/evening time. For moderate source-sink areas, the highest density is in that of residential subdivisions, indicating that indeed moderate source-sink areas tend to be associated with residential-dominated areas. The pattern of weak source-sink areas is consistent with (though of lower densities) that of strong source-sink areas; and the pattern of weak source-sink is also similar to but with lower densities than that of moderate source-sink areas. Except for factories, the normalized densities of all feature classes are higher in high-equilibrium areas (mixed land uses) than in low-equilibrium areas. Most factories are located in the suburban areas in Shanghai, and their normalized densities are higher in low-equilibrium areas (with mixed agriculture-industrial land uses) than those in high-equilibrium areas.

In order to further clarify the association of traffic source-sink areas with land uses, we consolidate the nine feature classes into four land use types: residential, commercial (including shopping place, hotel, and restaurant), institutional and recreational (including government office, hospital, school, and park), and industrial. As shown in Fig. 10b, the patterns discussed above become clear: (1) the strong source-sink areas are associated with commercial and institutional and recreational land uses, so are the weak source-sink areas, (2) both moderate and weak source-sink areas are foremost associated with residential areas, (3) high-equilibrium areas are most likely mixed residential-commercial-institutional-recreational land uses, and (4) low-equilibrium areas are often linked to industrial more than any other land uses.

Two approaches are used here to quantitatively examine the relationship between the six types of traffic source-sink areas and land uses. First, the *classification tree* method helps us to explain responses organized in a tree structure on a categorical dependent variable (Breiman, Friedman, Olshen, & Stone, 1984). In this research, the four predictor-variables are the actual normalized densities of the four consolidated feature classes (land uses) in all pixels, and the predicted outcomes are the six traffic source-sink areas. The method is used to map the ranges of four predictor variables to the six traffic areas. After training a classification tree, a

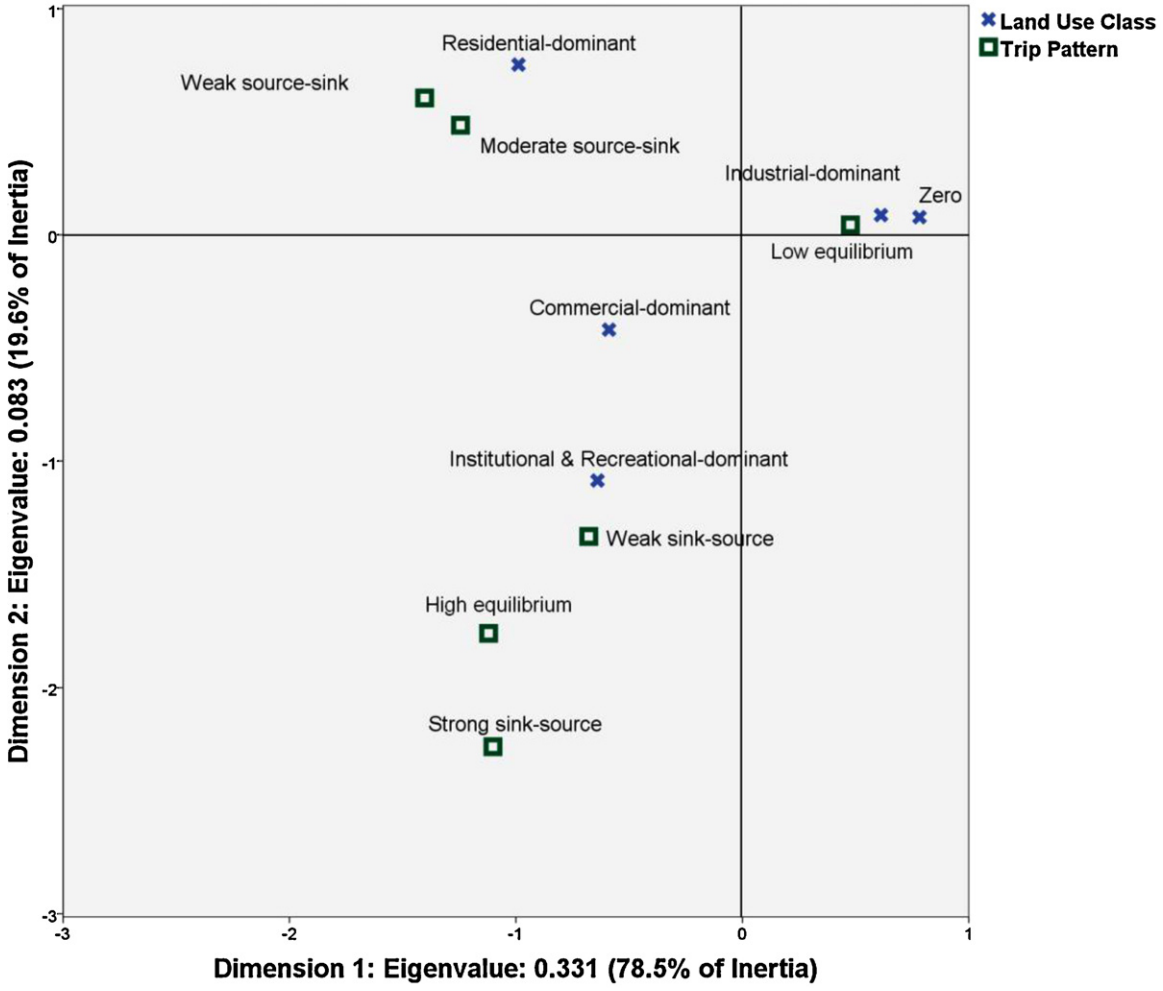

Fig. 11. Correspondence analysis plot of trip patterns and land use classes.

total accuracy of 83.5% (Kappa coefficient = 0.623) is achieved with the primary rules listed in Table 2. It is clear that the rules are consistent with the dependence relations between trips and land uses. Note that the industrial land use is not involved in any rules, since the spatial distributions of factories and the other classes of points, e.g. shopping places, are complementary and dependent among them (cf. Fig. 9).

Another approach, correspondence analysis, is used to examine the linkage between source-sink areas and land use features. *Correspondence analysis* represents associations in a table of frequencies or counts to a two-dimension graph (Greenacre, 1983). We categorize all pixels into five types: residential-dominant (RD), commercial-dominant (CD), institutional & recreational-dominant (IRD), industrial- dominant (ID), and zero (Z), according to the maximum normalized densities of the four aggregated land use classes. If no point is inside a pixel, then the pixel is denoted as zero. Hence, two classification schemata are available. They are based on temporal trip patterns and point distributions. With these two classification schemata, the graphical output (Fig. 11) of a correspondence analysis shows that weak source-sink and moderate source-sink areas are quite close to the class of RD, indicating that these two types of traffic areas tend to be located at residential-dominant areas. Similarly, low-equilibrium area is close to classes Z and ID, which are in the suburbia. Last, strong and weak source-sink areas as well as high-equilibrium area have closer associations to CD and IRD than to other land use classes. The correspondence analysis confirms the findings obtained from the classification tree method.

4.3. Urban structure revealed by traffic source-sink areas

Our final analysis revisits the issue of urban form, and examines how the distribution of traffic source-sink areas corresponds to their location in terms of distance from the city center. This can be confirmed by the spatial distribution of the six trip-based land use classes. As the urban development of Shanghai is confined by the coastal line as well as the Yangtze River, the shape of the city resembles more to an eclipse than a perfect circle (shown in Fig. 2c). Fig. 12a shows six ellipses with the same center and eccentricity as the minor semi-axes increases from 1 km to 13 km by an increment of 2 km. In each of the six zones, the proportions of the traffic source-sink areas are computed. Fig. 12b shows the distribution of various source-sink areas within each zone. From inner to outer zones, the composition of traffic areas changes from predominantly strong source-sink, to weak source-sink & moderate source-sink, to moderate source-sink, to weak source-sink, and mostly low-equilibrium areas. It indicates declining land use intensity gradually toward outer zones, and corresponding change in land use type from mostly commercial and institutional toward more residential areas. This pattern is consistent with the findings

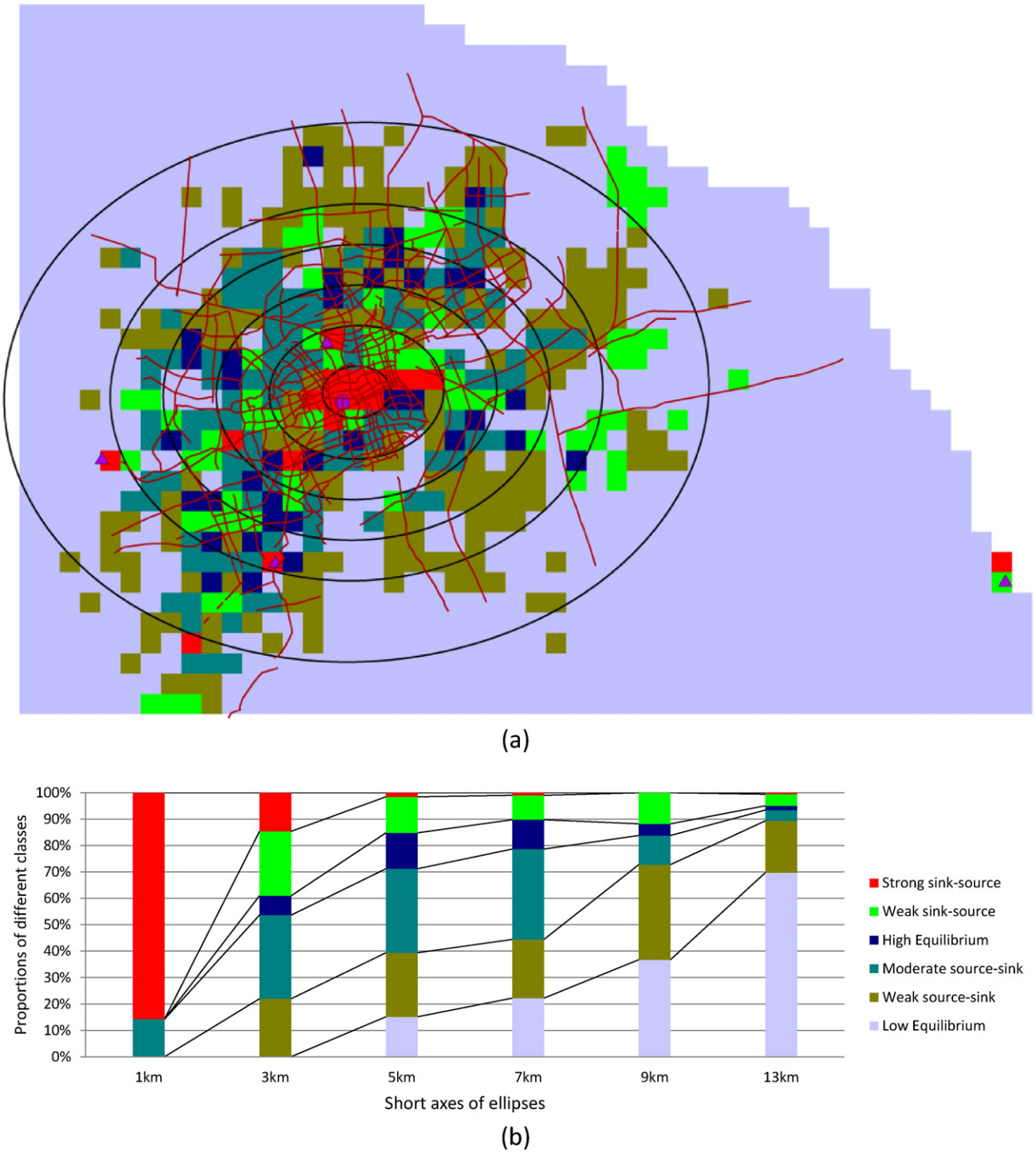

Fig. 12. "Concentric" urban structure revealed by trip patterns: (a) six ellipses with semi minor-axes of 1 km, 3 km, 5 km, 7 km, 9 km, and 13 km, (b) proportions of various source-sink areas in zones.

by Li, Wu, and Gao (2007), which revealed the concentric pattern in Shanghai using census and survey data.

5. Conclusions

Much of the existing literature on the interdependence between urban land use and travel patterns focuses on estimating traffic or explaining trip lengths from land use. This research attempts to examine the reversed linkage by revealing intra-urban land use variations from O-D traffic patterns. Using a seven-day taxi trajectory data set in Shanghai, we investigate the temporal variations of both pick-ups and drop-offs, and identify how different types of land use features play different roles at different times of a day in trip generation.

Overall, the distributions of taxi pick-ups and drop-offs well reflect the urban activity space, and their temporal variations exhibit a strong daily routine with three peaks in the morning, noon and early evening. The temporal patterns of pick-ups and

drops-offs vary a great deal from place to place, and are dependent on the function of the place. The balance between them, namely drop-offs pick-ups balance vector (DPBV), and its variation over a 24-h time span signal the function at a particular location. Borrowing the terms from ecological studies, a sink area of trips has more inflow traffic than outflow and thus a positive DPBV, and a source area the opposite. A typical residential area is a source area in morning time but a sink area in evening time, and a non-residential area (e.g., commercial, industrial, institutional-recreational) is the opposite. By examining the temporal patterns of DPBV across the study area of 1437 cells of $1\,km \times 1\,km$, we have identified six types of traffic areas: strong source-sink, weak source-sink, low equilibrium, high equilibrium, weak source-sink, and moderate source-sink. A source-sink area is a sink area in morning time and a source area in evening time, a source-sink area is reversed, and an equilibrium area has the numbers of pick-ups and drop-offs roughly balanced in each hour. Terms "strong", "moderate" and "weak" refer to traffic volume.

Based on the data from a local map supplier, four land use types are identified: residential, commercial (including shopping place, hotel, and restaurant), institutional and recreational (including government office, hospital, school, and park), and industrial. A close examination of the normalized densities of the four land use features in each source-sink area reveals their association patterns. Both strong and weak source-sink areas are primarily commercial and institutional and recreational land uses. Both moderate and weak source-sink areas tend to be associated with residential areas. High-equilibrium areas are most likely mixed land uses of high intensity, and low-equilibrium areas are often linked to industrial more than any other land uses. The classification tree and correspondence analysis methods further confirm the relationships. By dividing the study area to several "concentric" eclipses, we are able to show that land use type changes from mostly commercial and institutional toward more residential areas, and that land use intensity declines toward outer zones. Both are reflected in the varying DPBV and composition of various traffic source-sink areas. We believe that increasing availability of human mobility data from location aware devices will enable us to derive urban land use information and detect changes in a timely fashion, and help us identify and understand the process of urban structure change. The information will also be valuable for urban planners and policy makers in mitigating traffic, planning for public services and resources, and other purposes.

Acknowledgements

This research is supported by NSFC (Grant Nos. 40928001 and 41171296) and the National High Technology Development 863 Program of China (Grant No. 2011AA120301). The authors would also like to thank the three reviewers for their constructive comments.

References

Ahas, R., Aasa, A., Silm, S., & Tiru, M. (2010). Daily rhythms of suburban commuter's movements in the Tallinn metropolitan area: Case study with mobile positioning data. *Transportation Research Part C*, *18*, 45–54.

Antipova, A., Wang, F., & Wilmot, C. (2011). Urban land uses, socio-demographic attributes and commuting: A multilevel modeling approach. *Applied Geography*, *31*, 1010–1018.

Berkow, M., Monsere, C., Koonce, P., Bertini, R., & Wolfe, M. (2009). Prototype for data fusion using stationary and mobile data: Sources for improved arterial performance measurement. *Transportation Research Record*, *2099*, 102–112.

Black, W. R. (2003). *Transportation: A Geographical Analysis*. New York: Guilford.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and Regression Trees*. Monterey, CA: Wadsworth & Brooks/Cole Advanced Books & Software.

Bricka, S., & Bhat, C. R. (2006). A comparative analysis of GPS-based and travel survey-based data. *Transportation Research Record*, *1972*, 9–20.

Cervero, R. (1989). Jobs-housing balance and regional mobility. *Journal of the American Planning Association*, *55*, 136–150.

Chai, Y., Weng, G., & Shen, J. (2008). A study on commercial structure of Shanghai based on residents' shopping behavior. *Geographical Research*, *27*(4), 897–906 [in Chinese]

Chen, C., Chen, J., & Barry, J. (2009). Diurnal patterns of transit ridership: A case study of the New York City subway system. *Journal of Transport Geography*, *17*, 176–186.

Chowell, G., Hyman, J. M., Eubank, S., & Castillo-Chavez, C. (2003). Scaling laws for the movement of people between locations in a large city. *Physical Review E*, *68*(1-7).

Dai, X., Ferman, M. A., & Roesser, R. P. (2003). A simulation evaluation of a real-time traffic information system using probe vehicles. In *Proceedings of IEEE Intelligent Transportation Systems* 1, (pp. 475–480).

González, M. C., Hidalgo, C. A., & Barabási, A.-L. (2008). Understanding individual human mobility patterns. *Nature*, *453*, 779–782.

Goodchild, M. F., & Janelle, D. G. (1984). The city around the clock: Space-time patterns of urban ecological structure. *Environment and Planning A*, *16*, 807–820.

Goodchild, M. F., Klinkenberg, B., & Janelle, D. G. (1993). A factorial model of aggregate spatio-temporal behavior: Application to the diurnal cycle. *Geographical Analysis*, *5*, 277–294.

Gordon, P., Kumar, A., & Richardson, H. W. (1989). The influence of metropolitan spatial structure on commuting time. *Journal of Urban Economics*, *26*(2), 138–151.

Greenacre, M. (1983). *Theory and applications of correspondence analysis*. London: Academic Press.

Giuliano, G., & Small, K. A. (1993). Is the journey to work explained by urban structure. *Urban Studies*, *30*, 1485–1500.

Hamilton, B. W. (1982). Wasteful Commuting. *The Journal of Political Economy*, *90*(5), 1035–1053.

Jensen, J. R. (1983). Biophysical remote sensing. *Annals of the Association of American Geographers*, *73*(1), 111–132.

Jiang, B., & Liu, C. (2009). Street-based topological representations and analyses for predicting traffic flow in GIS. *International Journal of Geographical Information Science*, *23*(9), 1119–1137.

Jiang, B., Yin, J., & Zhao, S. (2009). Characterizing the human mobility pattern in a large street network. *Physical Review E*, *021136*(1–11), 80.

Kang, C., Ma, X., Tong, D., & Liu, Y. (2012). Intra-urban human mobility patterns: An urban morphology perspective. *Physica A*, *391*(4), 1702–1717.

Khisty, C. J., & Lall, B. K. (2003). *Transportation engineering: An introduction* (3rd ed.). Upper Saddle River, NJ: Pearson.

Kühne, R., Schäfer, R.-P., Mikat, J., Thiessenhusen, K.-U., Böttger, U., & Lorkowski, S. (2003). New approaches for traffic management in metropolitan areas. In *Proceedings, 10th IFAC Symposium* Tokyo, Japan, August, (pp. 4–6).

Li, Q., Zhang, T., Wang, H., & Zeng, Z. (2011). Dynamic accessibility mapping using floating car data: A network-constrained density estimation approach. *Journal of Transport Geography*, *19*(3), 379–393.

Li, Y., Yuan, Z., Xie, H., Cao, S., & Wu, X. (2007). Analysis on trips characteristics of taxi in Suzhou based on OD data. *Journal of Transportation Systems Engineering and Information Technology*, *7*(5), 85–89 [in Chinese].

Li, Z., Wu, F., & Gao, X. (2007). Polarization of the global city and sociospatial differentiation in Shanghai. *Scientia Geographica Sinica*, *27*(3), 304–311 [in Chinese].

Li, Z., & Wu, F. (2006). Residential disparity in urban China: A case study of three neighborhoods in Shanghai. *Housing Studies*, *21*(5), 695–717.

Liu, L., Andris, C., & Ratti, C. (2010). Uncovering cabdrivers' behaviour patterns from their digital traces. *Computers, Environment and Urban Systems*, *34*(6), 541–548.

Lu, D., & Weng, Q. (2005). Urban land-use and land-cover mapping using the full spectral information of Landsat ETM+ data in Indianapolis, Indiana. *Photogrammetric Engineering & Remote Sensing*, *71*(11), 1275–1284.

Lü, W., Zhu, T., Wu, D., Dai, H., & Huang, J. (2008). A heuristic path-estimating algorithm for large-scale real-time traffic information calculating. *Science in China, Series E: Technological Sciences*, *51*, 165–174.

Maat, K., van Wee, B., & Stead, D. (2005). Land use and travel behaviour: Expected effects from the perspective of utility theory and activity-based theories. *Environment and Planning B*, *32*, 33–46.

MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability* (pp. 281–297). University of California Press.

Mills, E. S. (1972). *Studies in the structure of the urban economy*. Baltimore: Johns Hopkins University.

Muth, R. (1969). *Cities and housing*. Chicago: University of Chicago.

Pan, H., Shen, Q., & Zhang, M. (2009). Impacts of urban forms on travel behavior: Case studies in Shanghai. *Urban Transport of China*, *7*(6), 28–32 [in Chinese].

Phithakkitnukoon, S., Horanont, T., Lorenzo, G. D., Shibasaki, R., & Ratti, C. (2010). Activity-aware map: Identifying human daily activity pattern using mobile phone data. Human Behavior Understanding. *Lecture Notes in Computer Science, vol. 6219*. Berlin: Springer, pp. 14–25.

Pulliam, H. R. (1988). Sources, sinks, and population regulation. *American Naturalist*, *132*, 652–661.

Qi, G., Li, X., Li, S., Pan, G., Wang, Z., & Zhang, D. (2011). Measuring social functions of city regions from large-scale taxi behaviors. *Proceedings of IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)*, 384–388.

Ratti, C., Pulselli, R. M., Williams, S., & Frenchman, D. (2006). Mobile landscapes: Using location data from cell phones for urban analysis. *Environment and Planning B*, *33*, 727–748.

Rhee, I., Shin, M., Hong, S., Lee, K., & Chong, S. (2008). On the Levy-walk nature of human mobility. *Proceedings of IEEE INFOCOM, 2008*, 924–932.
Rogan, J., & Chen, D. (2004). Remote sensing technology for mapping and monitoring land-cover and land-use change. *Progress in Planning, 61*, 301–325.
Sevtsuk, A., & Ratti, C. (2010). Does urban mobility have a daily routine? Learning from the aggregate data of mobile networks. *Journal of Urban Technology*, *17*, 41–60.
Shen, Q. (2000). Spatial and social dimensions of commuting. *Journal of the American Planning Association*, *66*, 68–82.
Small, K. A., & Song, S. (1992). Wasteful commuting: A resolution. *Journal of Political Economy*, *100*, 888–898.
Song, C., Qu, Z., Blumm, N., & Barabási, A.-L. (2010). Limits of predictability in human mobility. *Science, 327*, 1018–1021.
Song, C., Koren, T., Wang, P., & Barabási, A.-L. (2010). Modelling the scaling properties of human mobility. *Nature Physics*, *6*, 818–823.
Song, Y., & Ikeda, T. (2005). A study on present Shanghai urban inhabitants' leisure activities and sites. *Journal of Asian Architecture and Building Engineering*, *4*(2), 301–306.
Stopher, P. R., Fitzgerald, C., & Zhang, J. (2008). Search for a global positioning system device to measure person travel. *Transportation Research C*, *16*(3), 350–369.
Sun, J., Yuan, J., Wang, Y., Si, H., & Shan, X. (2011). Exploring space-time structure of human mobility in urban space. *Physica A*, *390*(5), 929–942.
Tong, D., Coifman, B., & Merr, C. J. (2009). New perspectives on the use of GPS and GIS to support a highway performance study. *Transactions in GIS*, *13*, 69–85.
Wang, F. (2000). Modeling commuting patterns in Chicago in a GIS environment: A job accessibility perspective. *Professional Geographer*, *52*, 120–133.
Wang, F. (2001). Explaining intraurban variations of commuting by job accessibility and workers' characteristics. *Environment and Planning B*, *28*, 169–182.
Wang, F., Antipova, A., & Porta, S. (2011). Street centrality and land use intensity in Baton Rouge, Louisiana. *Journal of Transport Geography*, *19*, 285–293.
Weber, J., & Sultana, S. (2007). Journey-to-work patterns in the Age of Sprawl: Evidence from two Midsize Southern Metropolitan areas. *The Professional Geographer*, *59*, 193–208.
White, M. J. (1988). Urban commuting journeys are not wasteful. *Journal of Political Economy*, *96*, 1097–1110.
Wolf, J., Oliveira, M., & Thompson, M. (2003). The impact of trip underreporting on VMT and travel time estimates: Preliminary findings from the California statewide household travel survey GPS study. *Transportation Research Record*, *1854*, 189–198.
Wu, F., & Li, Z. (2005). Sociospatial differentiation in subdistricts of Shanghai. *Urban Geography*, *26*(2), 137–166.
Xiao, J., Shen, Y., Ge, J., Tateishia, R., Tanga, C., Liang, Y., et al. (2006). Evaluating urban expansion and land use change in Shijiazhuang, China, by using GIS and remote sensing. *Landscape and Urban Planning*, *75*(1-2), 69–80.
Zandviliet, R., & Dijst, M. (2006). Short-term dynamics in the use of places: A space-time typology of visitor populations in the Netherlands. *Urban Studies*, *43*(7), 1159–1176.