Transportation
https://doi.org/10.1007/s11116-020-10094-z

Understanding the modifiable areal unit problem and identifying appropriate spatial unit in jobs–housing balance and employment self-containment using big data

Xingang Zhou[1,2] · Anthony G. O. Yeh[3]

Abstract

Jobs–housing balance (JHB) and employment self-containment (ESC) have been used to examine the jobs–housing relationship. However, the effect of the modifiable areal unit problem (MAUP) on ESC and JHB has received little attention. This study aims to examine the effect of the MAUP on the spatial variation of ESC and JHB by utilizing mobile positioning data from Shenzhen, China. Journey-to-work trips are examined at the individual level and then aggregated into different spatial areal units. It is found that the average ESC increases with the increase in spatial areal units and that the relationship between JHB and ESC is amplified when the spatial areal unit increases with spatial aggregation. A 2 km grid is found to be an ideal spatial unit for the analysis of ESC in Shenzhen because it is the turning point in which the increase in ESC started to slow down, and the decrease in the coefficient of variation began to diminish. In addition, workers were more likely to commute by non-motorized transport modes when their jobs were within 2 km. This study helps elucidate the effect of the MAUP on ESC and JHB as well as determine the appropriate grid size for analysis. This study further suggests that the ideal spatial unit for the analysis of ESC and JHB may be related to the transport mode of the city under study.

Keywords MAUP · Mobile positioning data · Employment self-containment · Jobs–housing balance · Work trips

✉ Anthony G. O. Yeh
hdxugoy@hkucc.hku.hk

Xingang Zhou
zxg@tongji.edu.cn

[1] College of Architecture and Urban Planning, Tongji University, Shanghai 200092, China

[2] Key Laboratory of Ecology and Energy-Saving Study of Dense Habitat, Ministry of Education, Tongji University, Shanghai 200092, China

[3] Department of Urban Planning and Design, The University of Hong Kong, Pokfulam Road, Hong Kong, Hong Kong SAR, China

Published online: 09 March 2020

Introduction

Employment self-containment (ESC) is the percentage of resident workers who live and work locally over the number of resident workers living in an area (Cervero 1996). High ESC helps decrease automobile dependence and is more environmentally friendly (Curtis and Olaru 2010). Jobs–housing balance (JHB) and ESC indicators are generally used to measure the jobs–housing relationship and to understand the relationship between land use and commuting. ESC reflects actual jobs–housing balance that reflects whether workers choose local jobs. Policymakers have used JHB policies to increase ESC. Most previous studies on ESC and JHB utilized travel survey or census data and were based on traffic analysis zones (TAZ) or census tracts (Cervero 1996; Martinus and Biermann 2017; Niedzielski et al. 2013). The spatial areal units in these studies were of different sizes, and comparative analyses of ESC and JHB in different areal units were affected by the modifiable areal unit problem (MAUP). A spatial areal unit is "modifiable," or may be defined in different ways, and this condition affects analysis results (Fotheringham and Wong 1991). The MAUP is a typical problem in transportation research that uses spatially aggregated data (Hong et al. 2014; Zhang and Kukadia 2005). The proportions of intrazonal and interzonal trips vary with different geographical zones (Ding 1998). Moreover, the sizes of spatial areal units exert a positive effect on the percentage of intrazonal trips (Viegas et al. 2009).

Whether JHB effectively reduces urban commuting has been debated in the literature (Murphy and Killen 2011; Horner 2007; Levinson 1998). Journey-to-work data for Los Angeles indicate that JHB exerts a minor effect on commuting distances (Giuliano and Small 1993). 6-Year commuting patterns of Kaiser Permanente employees reveal that JHB does not significantly affect commuting distances (Wachs et al. 1993). By contrast, Horner (2002) showed that commuting distance is correlated with internal JHB at the metropolitan level. Yan et al. (2019) correlated jobs–residence balance and commuting time–cost with multiple scales and confirmed a positive impact of JHB. Comparing the results of different empirical studies is difficult and may be meaningless (Handy 1996; Yang et al. 2019).

Although the MAUP is a well-known problem in the geography literature (Openshaw 1984), it has got little attention in transportation studies. Few studies show that the MAUP affects jobs–housing relationship indicators (Horner and Murray 2002; Niedzielski et al. 2013; Yang et al. 2019). Yang et al. (2019) examined the MAUP when studying the relationship between the built environment and travel behavior in Beijing, China, and found the ideal unit with which the built environment was most associated with the specific trip. Understanding the effect of the MAUP in jobs–housing relationship helps to find the ideal spatial analysis areal unit.

It is necessary to analyze the relationships between JHB and ESC with different spatial areal units and test what the ideal and meaningful one. Spatially detailed commuting trips with a large sample size could be inferred from mobile positioning data, which can be further aggregated into any grid size for analysis and can be used to analyze ESC with different spatial areal units. The current study used mobile positioning data in Shenzhen, a mega city in China, to analyze spatial variations in ESC at multiple spatial scales and to understand how the MAUP affects ESC and JHB. Uniform grids were utilized to examine the spatial variation in ESC and the relationship between ESC and JHB. Different uniform grids were selected as spatial areal units to maintain uniformity and examine the spatial variation in ESC and JHB, as well as the relationship between these two variables.

Literature review

Employment self-containment (ESC)

Self-containment is an important sustainable strategy in urban planning, and urban planners aim to achieve high ESC and encourage nonmotorized transport modes (Martinus and Biermann 2017). However, only a few studies in the literature have examined journey-to-work trips from the perspective of ESC (Cervero 1996; Yigitcanlar et al. 2007).

Jobs–housing balance (JHB)

JHB represents jobs–housing relationship in an area (Cervero 1996), and several jobs–housing indicators have been developed in the literature (Horner and Mefford 2007). Different spatial areal units, such as census tracts, TAZ, or large spatial units (e.g., buffers with radius of 5–7 miles), have been used in previous studies (Peng 1997). Few studies have used big data to quantify JHB (Zhang et al 2017; Zhou et al 2018b; Long and Thill 2013).

The modifiable areal unit problem

The MAUP originated from the fact that spatial analysis units may be defined in different ways (Openshaw 1984). The scale problem is that analysis results vary when data used for the study are aggregated due to the uncertainty in the number of analysis units. The zoning problem is the variation in results of analysis when different analysis units are utilized with a constant number of units because of the uncertainty in the way that the analysis zone is subdivided into a number of analysis units (Wong 2009).

The MAUP is not only an active research area in spatial analysis but also a typical problem in transportation research. Zhang and Kukadia (2005) diagnosed the MAUP in the effect of urban form on travel modes and discovered that regression coefficients decrease as the spatial scale increases. Mitra and Buliung (2012) reported that zoning affects the relationship between built environment and active school transportation. Several studies have also examined the effect of the MAUP on the relationship between built environment and travel behavior (Clark and Scott 2014; Forsyth et al. 2007; Mitra and Buliung 2012). However, few studies have analyzed the effect of the MAUP on ESC and JHB (Horner and Murray 2002; Niedzielski et al. 2013; Yang et al. 2019). Understanding the effect of the MAUP in jobs–housing relationship helps minimize its effects on the spatial analysis results.

Examining work trips by using mobile positioning big data

Previous transport studies that used journey-to-work trips were usually based on household or roadside travel survey data, which are collected once every several years with low update frequencies (Iqbal et al. 2014). Traditional travel survey data, as a type of small data, are limited by sample sizes. Massive spatiotemporal big data, such as mobile positioning data (Song et al. 2010), taxi GPS trajectory data (Liu et al. 2012), and smart card data (Huang et al. 2018; Li et al. 2020a, b), have been used to track human mobility. However, travel information obtained from smart card and taxi GPS trajectory data are limited by the travel modes. With the increase in the functions of mobile phone, mobile phones have a higher penetration rate and a higher percentage of residents have mobile phones than before. In Shenzhen, the ownership of mobile phone is around 90%. As a result of the high ownership rate, work trips with a large sampling size of the workers could be inferred and tracked from mobile phones (Kung et al 2014). The sampling size of mobile positioning data is larger than that of household travel diary survey data using questionnaires and can be used to study human mobility with larger samples (Xu et al. 2015). Kung et al. (2014) inferred residences and work locations of mobile phone users according to their most visited locations during sleeping and working periods.

The individual work trips derived from mobile positioning data can be aggregated into different spatial analysis units to examine the effect of the MAUP on ESC and JHB, as well as the relationship between the two variables.

Research methodology

Identifying work trips from mobile positioning data

Identification of home and work locations

To examine ESC and JHB, the home and work locations of workers were identified from the mobile positioning data. After that, journey-to-work trips were generated. A three-dimensional coordinate system was utilized to help determine the activity patterns of residents in space–time at the individual level. A mobile phone user's mobility pattern could be demonstrated as spatiotemporal paths in three dimensions where the X–Y axes reflected the geographic movements and the Z axis was the temporal dimension in different time of the day, to help understand the activity pattern (Shaw and Yu 2009). The movement was represented by a spatiotemporal sequence of triples (x_k, y_k, t_k) that record the geographic coordinate of the resident at time t. The data model "activity place–activity starting time–activity duration" (Long and Thill 2013) was used to represent the movement of residents in the city, and the analytical framework in Zhou et al. (2018a) and Kung et al (2014) was used to identify mobile phone users' home and work locations. Activity place refers to the cellphone tower service area. The average cellphone tower service area is about 0.28 km^2. Activity starting time refers to the approximate starting time when the cellphone user arrived at that place for the activity. Activity duration is the approximate duration of the mobile phone user's stay in the place of the activity. During sleeping time, if a resident stayed in one activity place for a duration that is longer than the threshold (4 h), then this activity place might be inferred as the home location of that resident. The threshold was determined by a sensitivity analysis so that the inferred residents were reliable samples. The other constraint was that a proportion of the buildings in the activity place were for residential use instead of pure employment use according to the building information data to avoid the bias of shift workers.

During working time, if a person stayed at an activity place that was not his or her residence longer than the threshold (5 h), then this place would be inferred as the workplace of that person. The other constraint to consider this place as the workspace was that a proportion of the buildings were for employment use instead of pure residential use according to the building information data. The estimation of sleeping time, from 0 a.m. to 7 a.m. and working time, from 8 a.m. to 12 a.m. and from 2 p.m. to 6 p.m., was based on the commuting habits of workers derived from the travel survey data (Zhou et al. 2018a). The thresholds were determined by a sensitivity analysis so that the inferred workers were reliable samples.

For mobile phone users with few records, their records could not reflect their journey-to-work trips. Hence, in the process of inferring the residences of mobile phone users, users with records below the threshold (4 h) during sleeping time were filtered out. This filtering process ensured an appropriate degree of certainty that the inferred residences were the residential places of the cellphone users. Similarly, in the process of inferring the workplaces of mobile phone users, users with records below the threshold (5 h) during working time were filtered out.

Commuting inferred from the mobile positioning data

With the identified home and work locations of workers, the journey-to-work patterns were explored. For each mobile phone user, the commuting distance between residential tower cell and workplace tower cell was calculated using Dijkstra's algorithm based on the road network data to represent the spatial proximity between residence and workplace.

Conversion of tower-to-tower OD matrix into zone-to-zone OD matrix

The tower-to-tower OD matrix was generated after home and work locations were identified from the mobile positioning data. Then, the tower-to-tower OD matrixes were aggregated into the zone-to-zone OD matrix to examine work trips and the spatial variation in JHB and ESC. The towers are assigned to different zones based on locations because if the towers are within the same zone, then the tower coverage areas are likely to be within the same zone as well. Mobile positioning data provide individual-level work trips; hence, aggregating them into different spatial areal units is possible. The following section presents the methods to examine the spatial variation in JHB and ESC based on the zone-to-zone OD matrix.

Calculating ESC and JHB

This study examines the relationship between ESC and JHB. ESC is calculated according to its definition as follows:

$$\mathbf{ESC}_i = \frac{\mathbf{WORK}_{ii}}{\sum_{j=1}^{n} \mathbf{WORK}_{ij}} \tag{1}$$

where zone $\boldsymbol{i}$ is the work trip origin, zone $\boldsymbol{j}$ is the work trip destination, $\mathbf{WORK}_{ii}$ represents the number of resident workers working in zone $\boldsymbol{i}$, $\mathbf{WORK}_{ij}$ represents the number of commuters from zone $\boldsymbol{i}$ to zone $\boldsymbol{j}$, and $\sum_{j=1}^{n} \mathbf{WORK}_{ij}$ is calculated as the number of employees who live in zone $\boldsymbol{i}$.

JHB is measured with the jobs/workers ratio (JWR) using mobile positioning data in this study.

$$\boldsymbol{JWR}_i = \frac{\boldsymbol{JOB}_i}{\mathbf{WORK}_i} \tag{2}$$

where $\boldsymbol{JOB}_i$ represents available jobs of zone $\boldsymbol{i}$, and $\mathbf{WORK}_i$ represents how many workers residing in zone $\boldsymbol{i}$.

Figure 1 illustrates the scale issue of the MAUP with ESC and JHB. The work trips in a city are shown in panels (a, b) at fine and coarse scales with two zoning systems. The total numbers of workers are same as the total number of jobs. ESC in the zones is shown as the diagonal elements of the tables in panels (c, d). JWR in the zones is shown as the diagonal elements of the tables in panels (e, f). Comparison of ESC at fine (c) and coarse (d) scales shows that the average ESC increases or remains the same as the spatial analysis unit expands. Comparison of JWR at fine (e) and coarse (f) scales shows that the average JWR might increase, decrease or remain the same as the spatial areal unit increases.

In this study, ESC and JHB were analyzed with different spatial areal units through the use of mobile positioning data to understand the effect of the MAUP on ESC and JHB.

Study area and data

Study area

Shenzhen in the southern part of China is selected as the study area because it is similar with other Chinese cities in spatial structure, i.e. with core city and mixed industrial and housing suburbs. Urban expansion has led to long-distance commuting and increasing

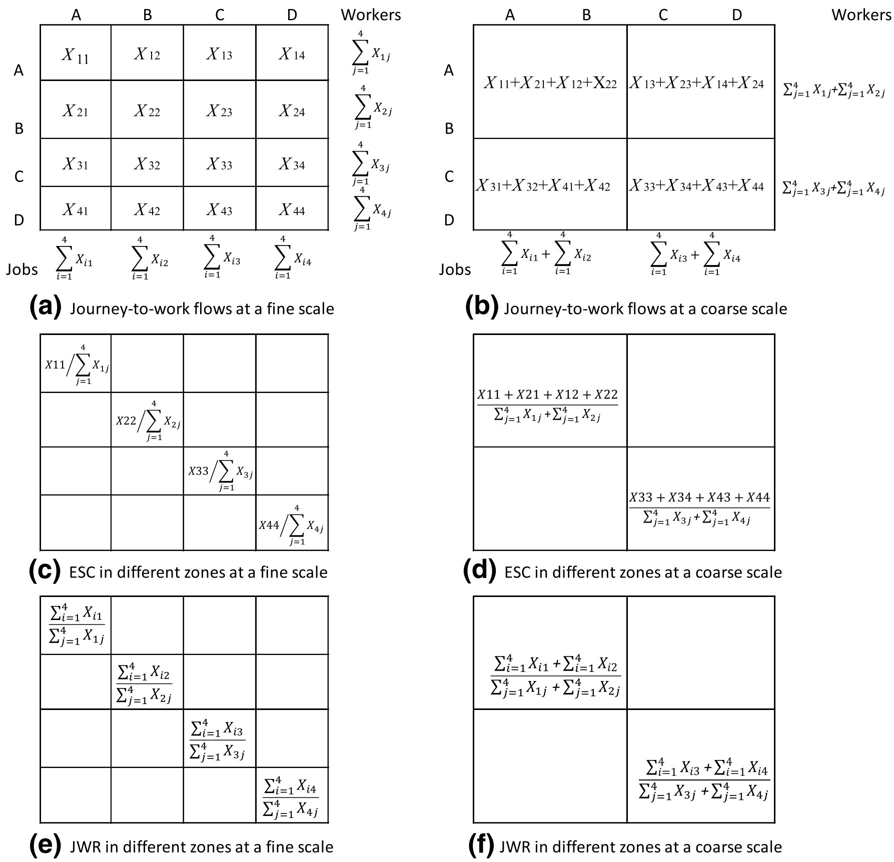

Fig. 1 Illustration of the scale problem of the MAUP with ESC and JHB

dependence on automobiles. The southeastern parts of the city are excluded because these areas are mainly used for recreation or conservation. The main built-up areas of Shenzhen are selected as the study area (Fig. 2). The population density is high, with more than 15 Mio. residents in an area of approximately 2000 km^2.

Data

Mobile positioning data

Approximately 0.3 billion mobile positioning records from 5908 mobile phone towers have been collected on a working day in March, 2012. When the phone was active, the ID of each mobile phone user and the locations of the base transceiver station were recorded with around 0.5 to 1-h interval. The records of more than 12.4 Mio. cellphone users were provided by a leading cellphone operator with over 76.8% of the total mobile customers in Shenzhen. The residences of around 7.2 Mio. cellphone users were inferred from the mobile positioning data. Considering that Shenzhen has around 15 Mio. residents, the sampling size of about 48% in this study was significantly higher than that for travel survey data of about 2%. The population density of Shenzhen was high, and towers were near one another in the study area. Cellphone users' locations were approximated by Voronoi polygons through base transceiver stations.

Household travel survey data

The household travel survey data included detailed travel information, such as trip starting time, arrival time, and trip purposes of 480,692 trips made by 192,015 individuals, obtained with random sampling in October 2010. The data were provided by Shenzhen City Planning and Land Resources Bureau. The travel survey data were based on household questionnaire surveys, which were time-consuming and

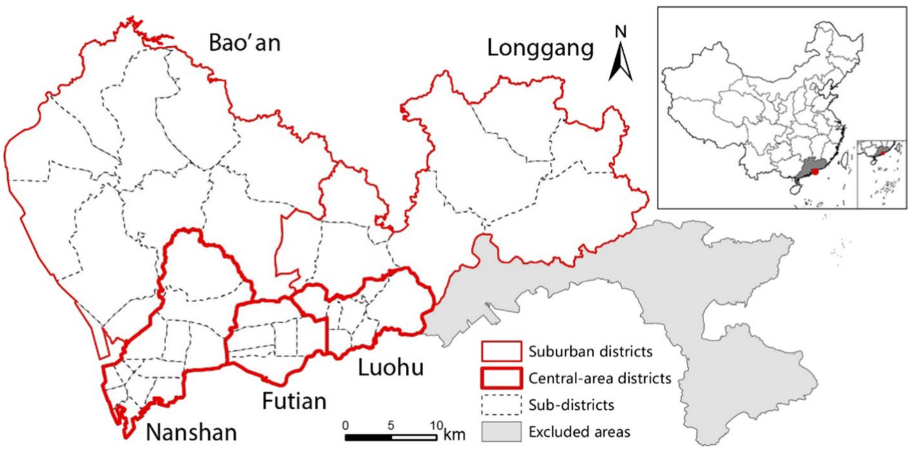

Fig. 2 Study area

labor-intensive. As a result, the sampling size for the travel survey data was relatively low, at approximately 2%, and lower than that of the mobile positioning data which can be 48% in this study.

Spatial variation in ESC with modifiable areal units

After the home and work location of the workers were inferred from the mobile positioning data with the help of the travel survey data, the work trips were generated. Figure 3 shows the distribution of the workers' commuting distance in whole Shenzhen. The average commuting distance was shorter in the suburbs (5.34 km) than that in the central city (6.12 km).

Table 1 demonstrates the commuting patterns and the ESC derived from mobile positioning data at the district level in whole Shenzhen.

ESC among the five districts in Shenzhen was compared at a macro level (Fig. 4A). The ESC in the suburbs including Bao'an and Longgang was higher than that in the central–city areas including Nanshan, Futian, and Luohu. It was not feasible to conclude that workers residing in the suburbs were more likely to work locally than those residing in the central–city areas because the sizes of the districts in the suburbs were larger than those in the central–city areas. Figure 4B shows ESC at the sub-district level in the five districts in Shenzhen. In existing literature, the sub-district is usually employed to measure JHB in Chinese cities (Zhao et al. 2011). The sub-districts in a city are usually of different sizes and shapes, and the average sub-district size in Shenzhen is approximately 35.0 km^2. Large

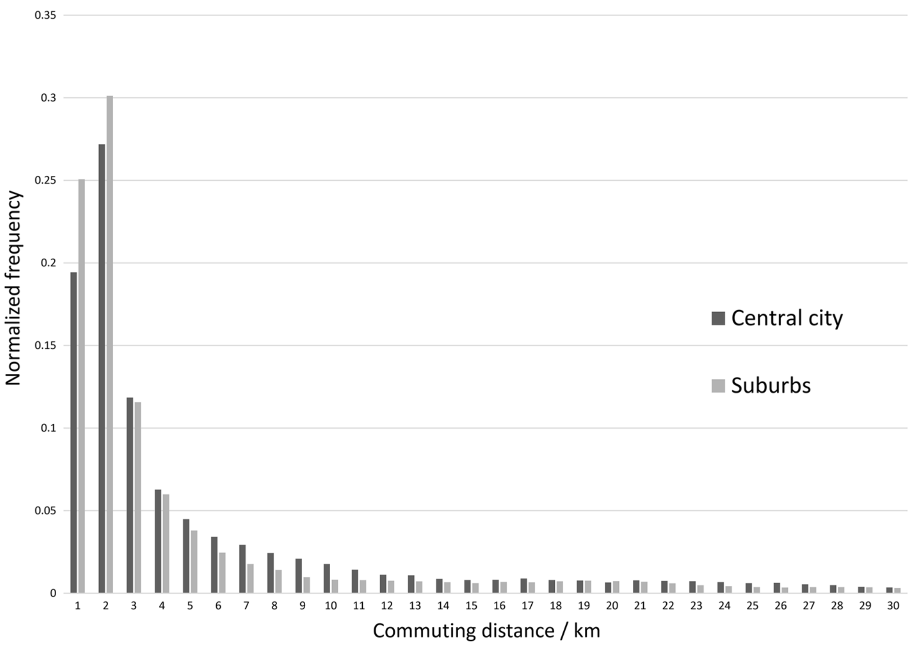

Fig. 3 Distribution of commuting distances in Shenzhen

Table 1 Commuting patterns at the district level in whole Shenzhen

	Bao'an (%)	Longgang (%)	Luohu (%)	Futian (%)	Nanshan (%)	Total percentage of outgoing work trips (%)
Bao'an	**90.1**	0.9	0.7	3.8	4.5	100.0
Longgang	2.3	**86.9**	4.1	5.6	1.1	100.0
Luohu	1.2	4.5	**71.7**	19.9	2.6	100.0
Futian	1.7	0.9	6.8	**84.5**	6.1	100.0
Nanshan	4.1	0.4	1.8	10.4	**83.3**	100.0

Bold values indicate ESC

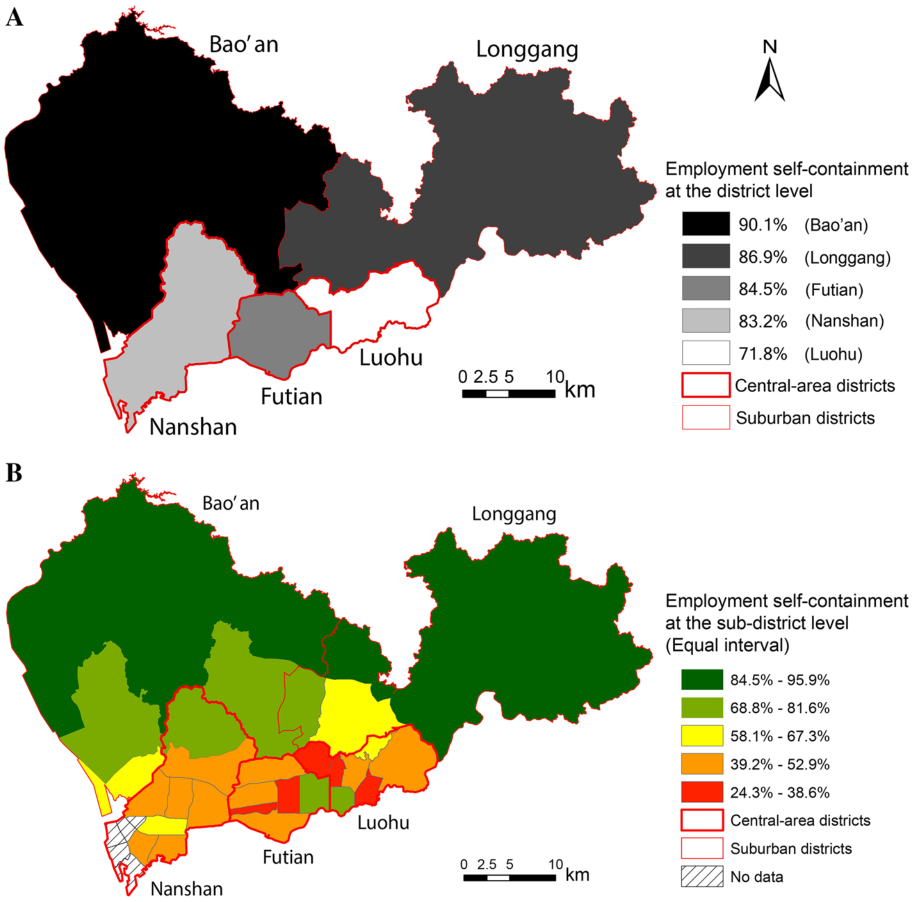

Fig. 4 Employment self-containment at the district and sub-district level in Shenzhen

sub-districts tended to be more self-contained than small ones. Average ESC at the sub-district level was lower than that at the district level.

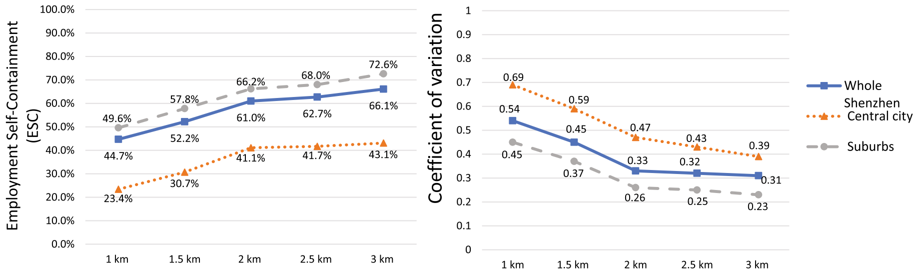

Fig. 5 Average employment self-containment increased, and the coefficient of variation decreased with the increase in the spatial areal unit

Table 2 Correlation between JWR and ESC with different spatial areal units

Spatial areal units	1 km grid	1.5 km grid	2 km grid	2.5 km grid	3 km grid
Correlation coefficient	No significant effect	No significant effect	0.21*	0.36**	0.41***
R-squared			0.08	0.13	0.22

*Significant at the 0.05 level

**Significant at the 0.01 level

***Significant at the 0.001 level

Jobs–housing balance and employment self-containment with different spatial areal units

To perform a comparative analysis and maintain uniformity, uniform grids were selected as the spatial analysis unit to analyze the spatial variation in ESC. Figure 5 shows the spatial variation in ESC and the coefficient of variation with different spatial areal units. The smallest areal unit was set as a 1 km grid to reduce possible errors caused by the variation in mobile phone coverage areas.

Statistical analyses showed that the average ESC in the central areas was lower than that in the suburbs in Shenzhen. Suburban residents mostly work near home while central city people commute longer distances, which is different from the commuting pattern in most Western cities. The spatial variation in ESC between the suburbs and central–city areas was hypothesized to be partly because the jobs–workers ratio (JWR) had a stronger correlation with ESC in the suburbs than that in the central areas. JWR was used to explain the spatial variation in ESC. High JWR provided the potential to encourage workers to work locally, and the actual JHB was measured by ESC (Zhou et al. 2018b). The logarithmic value of JWR was the independent variable because the relationship between JWR and ESC was nonlinear. The correlation between JWR and ESC was examined with different spatial areal units (Table 2).

JWR did not exert a significant effect on ESC at the 1 and 1.5 km grid levels. However, it had a mild effect on ESC at the 2 km grid level. The regression coefficient and R-squared value increased as the spatial areal unit increased to 2.5 and 3 km grids. The regression coefficient increased as the spatial areal unit increased and became increasingly aggregated because extreme values, such as minimum or maximum, became less varied, with small spatial areal units aggregating to form large analysis units. Nonetheless, the regression and correlation coefficients might not represent the “real” effect of JWR on ESC with different spatial areal units. The correlation between JWR and ESC increased as the number of aggregated spatial areal units increased. According to the MAUP analysis, an ideal spatial areal unit must be selected for the analysis of ESC and JHB to become meaningful.

As shown in Fig. 5, the 2 km grid was the turning point at which the increase in ESC started to decelerate, and the decrease in the coefficient of variation began to diminish. The 2 km grid in this study might be the right geographic unit for examining the correlation between JWR and ESC because workers might commute by non-motorized modes if their work trips were shorter than 2 km. The average nonmotorized travel time was 17 min, and approximately 44% of workers travelled by walking and 10.2% by biking according to Shenzhen household travel survey data. The biking speed was 15 km/h and walking speed was 5 km/h approximately. Therefore, in this study, we use 2 km as the average work trip distance of the nonmotorized transport modes. Policies that increasing ESC at the 2 km grid level might induce workers to choose nonmotorized transport modes. As a result, 2 km grid is used as the ideal areal unit for this study in Shenzhen.

The average interzonal commuting distance in the grids was used to represent the distance the workers had to commute to their workplaces if they did not work locally. Figure 6 demonstrates the scatterplot of ESC and average inter-zonal commuting distance in the central–city districts (Futian, Nanshan and Luohu) and in the suburban districts (Bao’an and Longgang). The average ESC was 44.2% (vertical line) and average inter-zonal commuting distance was 3.59 km (horizontal line).

The ESC and the average inter-zonal commuting distance of the 2 km grids in Shenzhen are mapped in Fig. 7. In Fig. 7A, the blue grids represent relatively highly self-contained grids with ESC higher than the average of 44.2% and an average inter-zonal commuting distance of 3.02 km; these cells were mostly in the suburbs. Many workers residing in these grids might be manufacturing workers who resided in factory dormitories and had short commuting distances. In Fig. 7B, the workers residing in the dark blue grids, compared with those in the light blue grids, had to commute longer to other grids if they did not work locally. These grids might be in the suburbs adjacent to central areas, and workers commuting far away to central areas might be tertiary-sector workers. The choice to live in

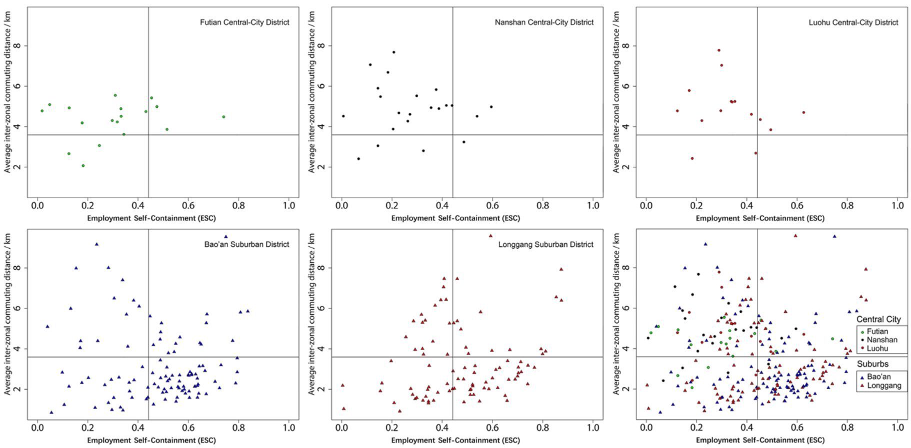

Fig. 6 The scatterplot of ESC and average inter-zonal commuting distance

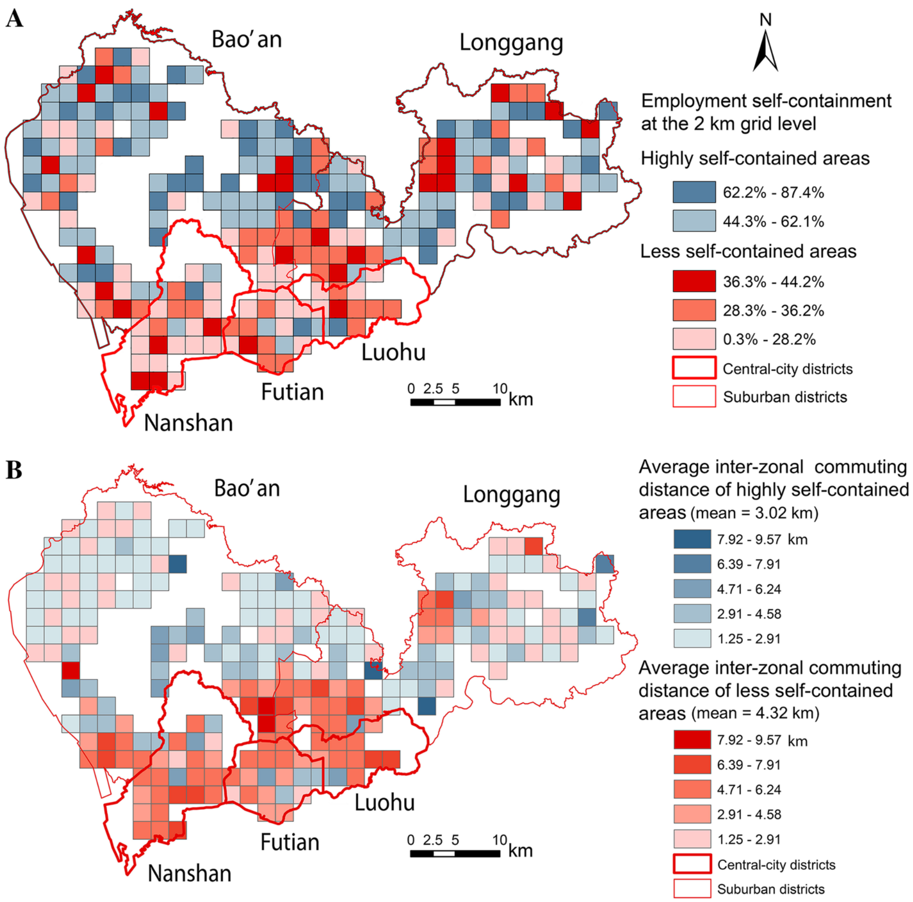

Fig. 7 Employment self-containment and average inter-zonal commuting distance

the suburbs far away from their workplaces may be because the housing prices in their residence areas were lower than those near their workplaces. The workers residing in the light blue grids only need to commute short distances to other grids if they did not work locally. In Fig. 7A, the red grids represent the less self-contained grids with ESC lower than the average of 44.2% and an average inter-zonal commuting distance of 4.32 km. Many workers residing in the red grids in the central areas might be service and office workers who had long commuting distances. One reason was that they had higher transit accessibility, and tended to reside far away from their workplaces where housing would be cheaper than those near their workplaces. In Fig. 7B, the workers residing in the dark red grids had to commute longer to other grids to work compared with those in the light red grids. In the suburbs, workers residing in the light red grids only needed to commute short distances to the grids nearby to work, while those residing in the dark red grids near the central–city areas had to commute longer distances likely toward the central–city areas.

Discussion

The analyses showed that ESC and JHB exhibited significant spatial variation with different spatial areal units. However, applying the analysis results from one spatial scale to another might result in biased findings because a spatial analysis unit may vary from a small area to the entire city. An ideal spatial analysis unit must be selected because an incorrect spatial analysis unit may lead to biased analysis results due to the MAUP. The MAUP occurs because the spatial analysis unit used is arbitrary and unable to represent the underlying process in the data. Understanding the biases caused by the MAUP can help make the analysis results reliable.

Although the MAUP has been well examined in the literature (Openshaw 1984; Fotheringham and Wong 1991), this study helps to elucidate the effect of the MAUP in transport research. First, mobile positioning data were used as a type of big data to aggregate journey-to-work trips into different spatial areal units to examine the spatial variation in ESC and JHB. Second, analyses at multiple spatial scales helped diagnose how the relationships between ESC and JHB are affected by the MAUP. Given that the spatial analysis units used in previous studies are inconsistent, this study provides a systematic analysis of how changes in spatial analysis units affect the analysis results.

As it is found that the MAUP will affect the analysis results, there is a need to find an appropriate spatial areal unit for the study of JHB and ESC. Travel behavior consideration and justification for a specific areal unit are important when selecting an ideal spatial analysis areal unit. ESC is proposed to encourage workers to work locally and select non-motorized transport modes besides JHB. Our analyses with different spatial areal units using mobile positioning data in Shenzhen shows that 2 km grid scale is the turning point where the increase in ESC started to slow down, and the decrease in the coefficient of variation began to diminish. Furthermore, the 2 km grid is the reasonable commuting distance for walking or biking in Shenzhen that workers might choose non-motorized transport modes. Therefore, around 2 km grid might be an ideal spatial areal unit for the analysis of ESC and JHB in Shenzhen.

Conclusion

Mobile positioning data could be utilized to track spatially detailed commuting trips with a large sample size, and these trips can be aggregated into different spatial analysis units to examine employment self-containment (ESC) and jobs–housing balance (JHB) at different spatial scales. The study shows how JHB and ESC vary with the increase in the size of the spatial areal unit. Large spatial areal units tend to have high ESC, whereas small spatial areal units are likely to have low ESC. The use of an inappropriate spatial areal unit may lead to inappropriate policies and decisions. Correlation analyses with an irregular spatial areal unit should consider the MAUP. Urban planning policies aiming to achieve high ESC should pay attention to the MAUP and select an appropriate areal unit for spatial analysis and formulation of urban policies. Achieving high ESC with a large spatial areal unit does not guarantee the same result as that obtained with a smaller spatial areal unit. Grids with uniform sizes and shapes should be used to explore spatial variations in ESC.

The study demonstrates that there is a turning point of the increase in spatial analysis unit at which the ESC started to slow down and the coefficient of variation began to diminish. In the case of Shenzhen, the turning point is found to be approximately 2 km, and it is related to nonmotorized modes of transport (primarily, walking or biking). The turning point is different in U.S. cities. For example, Fig. 8 shows that in Seattle, which is a high-density city in the U.S., approximately 45% of the workers drove alone to work every day (Seattle Department of Transportation 2010). A larger share of workers (44% and 10.2%) in Shenzhen commuted by walking and biking, respectively, than that in Seattle (8% and 3%).

This turning point may vary according to different transport modes that may vary from city to city. The hypothetical spatial analysis unit of the turning point of ESC in U.S. cities may be larger than that of Chinese cities (Fig. 9). This hypothesis derived from the present study could be further verified by carrying out similar studies of the changes in ESC with different spatial analysis units using mobile positioning data in U.S. cities.

The present study shows that the MAUP affects both JHB and ESC and their relationship. The key to selecting the most ideal spatial analysis unit is to understand how the spatial analysis unit affects the analysis results. This study contributes to the existing literature of MAUP in jobs–housing relationship to show that travel behavior consideration and justification for a specific areal unit are important in selecting an ideal spatial analysis areal unit. An appropriate and meaningful spatial areal unit for analyzing ESC and JHB in Shenzhen might be around 2 km grid and it would vary in different cities. Travel behavior consideration and justification need to be tested in other cities in China or other countries to find the appropriate grid size for analyzing JHB and ESC.

This study has several limitations. Firstly, the use of mobile positioning data might lead to error in the inference of the home and work locations of workers. For example, mobile phone location records during sleeping time were used to infer the residences of workers, which may result in errors if a mobile phone user stayed at a friend's home. Moreover, a mobile phone user might have stayed in a hospital as a patient during working hours and

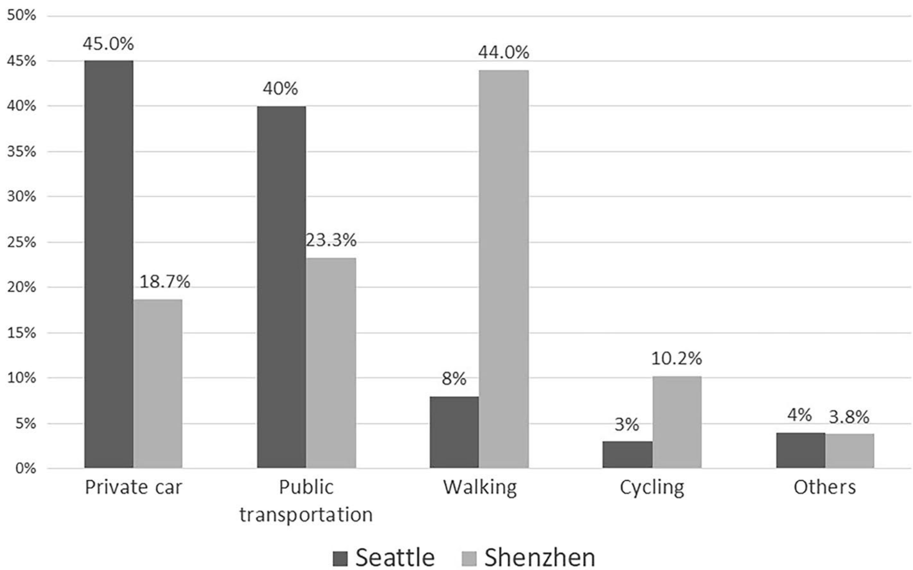

Fig. 8 Share of all commuters by mode in Seattle and Shenzhen in 2010 (*Source*: Seattle household travel survey and Shenzhen travel survey, respectively)

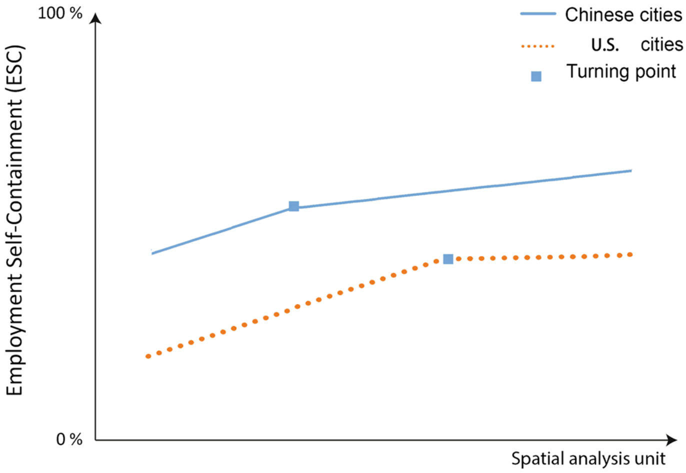

Fig. 9 Hypothetical turning point of the spatial analysis unit of employment self-containment in Chinese and U.S. cities because of the differences in transport modes

thus might have been identified as a worker in the hospital. However, obtaining individual work trips on that day for validation is difficult. One-week mobile positioning data, if available, could better identify the residences and workplaces of workers. If a mobile phone user stays at one location during sleeping time most of the days of the week, then that location is likely to be the residence. Likewise, if a mobile phone user stays at one location during working hours most of the days from Monday to Friday, then that location is likely to be the workplace. Second, short work trips may not have been recorded when the residence and workplace of a mobile phone user were within the same mobile phone tower service area. Work trips were recorded only when the current mobile phone tower servicing a mobile phone user shifted to another. Workers with residences and workplaces within the same mobile phone coverage areas were excluded to minimize possible errors in the identification of workers. However, the above possible errors would not significantly affect the calculation of ESC and JHB if the spatial unit of analysis is 2 km as identified from the study.

Acknowledgements This research was jointly supported by National Natural Science Foundation of China (41471378, 41801147), Chan To-Haan Endowed Professorship Fund of the University of Hong Kong. Dr. Xingang Zhou would like to thank the support from Shanghai Pujiang Talent Program (18PJC114).

References

Cervero, R.: Jobs–housing balance revisited: trends and impacts in the San Francisco bay area. J. Am. Plan. Assoc. **62**(4), 492–511 (1996)

Clark, A., Scott, D.: Understanding the impact of the modifiable areal unit problem on the relationship between active travel and the built environment. Urban Stud. **51**(2), 284–299 (2014)

Curtis, C., Olaru, D.: The relevance of traditional town planning concepts for travel minimization. Plan. Pract. Res. **25**(1), 49–75 (2010)

Ding, C.: The GIS-based human-interactive TAZ design algorithm: examining the impacts of data aggregation on transportation-planning analysis. Environ. Plan. B Plan. Des. **25**(4), 601–616 (1998)

Forsyth, A., Oakes, J.M., Schmitz, K.H., Hearst, M.: Does residential density increase walking and other physical activity? Urban Stud. **44**(4), 679–697 (2007)

Fotheringham, A.S., Wong, D.W.: The modifiable areal unit problem in multivariate statistical analysis. Environ. Plan. A **23**(7), 1025–1044 (1991)

Giuliano, G., Small, K.A.: Is the journey to work explained by urban structure? Urban Stud. **30**(9), 1485–1500 (1993)

Handy, S.: Methodologies for exploring the link between urban form and travel behavior. Transp. Res. D Transp. Environ. **1**(2), 151–165 (1996)

Hong, J., Shen, Q., Zhang, L.: How do built-environment factors affect travel behavior? A spatial analysis at different geographic scales. Transportation **41**(3), 419–440 (2014)

Horner, M.W.: Extensions to the concept of excess commuting. Environ. Plan. A **34**(3), 543–566 (2002)

Horner, M.W.: A multi-scale analysis of urban form and commuting change in a small metropolitan area (1990–2000). Ann. Reg. Sci. **41**(2), 315–332 (2007)

Horner, M.W., Mefford, J.N.: Investigating urban spatial mismatch using job–housing indicators to model home–work separation. Environ. Plan. A Econ. Space **39**(6), 1420–1440 (2007)

Horner, M.W., Murray, A.T.: Excess commuting and the modifiable areal unit problem. Urban Stud. **39**(1), 131–139 (2002)

Huang, J., Levinson, D., Wang, J., Zhou, J., Wang, Z.-J.: Tracking job and housing dynamics with smartcard data. Proc. Natl. Acad. Sci. **115**(50), 12710 (2018)

Kung, K.S., Greco, K., Sobolevsky, S., Ratti, C.: Exploring universal patterns in human home–work commuting from mobile phone data. PLoS ONE **9**(6), e96180 (2014)

Iqbal, M.S., Choudhury, C.F., Wang, P., González, M.C.: Development of origin–destination matrices using mobile phone call data. Transp. Res. Part C **40**, 63–74 (2014)

Levinson, D.M.: Accessibility and the journey to work. J. Transp. Geogr. **6**(1), 11–21 (1998)

Li, S.Y., Lv, D.J., Liu, X.P., Tan, Z.Z., Gao, F., Huang, G.P., Wu, Z.F.: The varying patterns of rail transit ridership and their relationships with fine-scale built environment factors: big data analytics from Guangzhou. Cities **99**, 102580 (2020a)

Li, S.Y., Lv, D.J., Huang, G.P., Zhang, X.H., Gao, F., Yu, Y.T., Liu, X.P.: Spatially varying impacts of built environment factors on rail transit ridership at station level: a case study in Guangzhou, China. J. Transp. Geogr. **82**, 102631 (2020b)

Liu, Y., Kang, C., Gao, S., Xiao, Y., Tian, Y.: Understanding intra-urban trip patterns from taxi trajectory data. J. Geogr. Syst. **14**(4), 463–483 (2012)

Long, Y., Thill, J.-C.: Combining smart card data and household travel survey to analyze jobs–housing relationships in Beijing. Comput. Environ. Urban Syst. **53**(1), 19–35 (2013)

Martinus, K., Biermann, S.: Strategic planning for employment self-containment in metropolitan subregions. Urban Policy Res. **36**, 1–12 (2017)

Mitra, R., Buliung, R.N.: Built environment correlates of active school transportation: neighborhood and the modifiable areal unit problem. J. Transp. Geogr. **20**(1), 51–61 (2012)

Murphy, E., Killen, J.E.: Commuting economy: an alternative approach for assessing regional commuting efficiency. Urban Stud. **48**(6), 1255–1272 (2011)

Niedzielski, M.A., Horner, M.W., Xiao, N.: Analyzing scale independence in jobs–housing and commute efficiency metrics. Transp. Res. Part A Policy Pract. **58**, 129–143 (2013)

Openshaw, S.: The Modifiable Areal Unit Problem. Geo Abstracts University of East Anglia, Norwick (1984)

Seattle Department of Transportation: Only 35% of downtown commuters drive alone. https://sdotblog.seattle.gov/2011/07/20/only-35-of-downtown-commuters-drive-alone/ (2010). Accessed 20 Dec 2019

Song, C., Qu, Z., Blumm, N., Barabási, A.-L.: Limits of predictability in human mobility. Science **327**(5968), 1018–1021 (2010)

Peng, Z.R.: The jobs–housing balance and urban commuting. Urban Stud. **34**(8), 1215–1235 (1997)

Shaw, S.-L., Yu, H.: A GIS-based time-geographic approach of studying individual activities and interactions in a hybrid physical–virtual space. J. Transp. Geogr. **17**(2), 141–149 (2009)

Viegas, J.M., Martinez, L.M., Silva, E.A.: Effects of the modifiable areal unit problem on the delineation of traffic analysis zones. Environ. Plan. B **36**(4), 625–643 (2009)

Wachs, M., Taylor, B.D., Levine, N., Ong, P.: The changing commute: a case-study of the jobs–housing relationship over time. Urban Stud. **30**(10), 1711–1729 (1993)

Wong, D.: The modifiable areal unit problem (MAUP). In: Fotheringham, A.S., Rogerson, P.A. (eds.) The SAGE Handbook of Spatial Analysis, pp. 105–123. SAGE Publications, Thousand Oak (2009)

Xu, Y., Shaw, S., Zhao, Z., et al.: Understanding aggregate human mobility patterns using passive mobile phone location data: a home-based approach. Transportation **42**, 625–646 (2015)

Yan, L.X., Wang, D., Zhang, S.W., Xie, D.C.: Evaluating the multi-scale patterns of jobs–residence balance and commuting time-cost using cellular signaling data: a case study in Shanghai. Transportation **46**(3), 777–792 (2019)

Yang, L., Hu, L., Wang, Z.: The built environment and trip chaining behaviour revisited: the joint effects of the modifiable areal unit problem and tour purpose. Urban Stud. **56**(4), 795–817 (2019)

Yigitcanlar, T., Dodson, J., Gleeson, B., Sipe, N.: Travel self-containment in master planned estates: analysis of recent Australian trends. Urban Policy Res. **25**(1), 129–149 (2007)

Zhang, M., Kukadia, N.: Metrics of urban form and the modifiable areal unit problem. J. Transp. Res. Rec. **1902**, 71–79 (2005)

Zhao, P., Lü, B., Roo, G.D.: Impact of the jobs–housing balance on urban commuting in beijing in the transformation era. J. Transp. Geogr. **19**(1), 59–69 (2011)

Zhou, X., Yeh, A.G.O., Li, W., Yue, Y.: A commuting spectrum analysis of the jobs–housing balance and self-containment of employment with mobile phone location big data. Environ. Plan. B Urban Anal. City Sci. **45**(3), 434–451 (2018a)

Zhou, X., Yeh, A.G.O., Yue, Y.: Spatial variation of self-containment and jobs–housing balance in Shenzhen using cellphone big data. J. Transp. Geogr. **68**, 102–108 (2018b)

Zhang, P., Zhou, J., Zhang, T.: Quantifying and visualizing jobs–housing balance with big data: a case study of Shanghai. Cities **66**, 10–22 (2017)