

# Evaluating the multi-scale patterns of jobs-residence balance and commuting time–cost using cellular signaling data: a case study in Shanghai

Longxu Yan<sup>1</sup> · De Wang<sup>1</sup> · Shangwu Zhang<sup>1</sup> · Dongcan Xie<sup>1</sup>

Published online: 1 June 2018

© Springer Science+Business Media, LLC, part of Springer Nature 2018

**Abstract** Assessing jobs-housing balance (JHB) and commuting efficiency is crucial to urban and transport planning. However, the scale dependency problem, meaning that metrics may be biased as they rely on an arbitrary search radius or pre-defined jurisdictional division, remains unresolved. This paper proposes to apply clustering method to multi-scale indicators for evaluating aggregate patterns of jobs-housing relationship and commuting time–cost. We based our analysis on individual commuters extracted from cell phone data, a reliable substitute for travel surveys. After the commuter origin-destinations are carefully inferred, we draw two types of curves to represent the multi-scale characteristics of jobs-residence relationship and commuting efficiency. One is the multi-scale index curve (MSIC) of jobs-residence ratio that measured within multiple scales. Another is the probability density curve (PDC) of commuting time–cost. Then we applied the affinity propagation clustering method to the curves and detected six patterns of MSIC and ten of PDC. By comparing our result with the conventional methods we showed that the latter's classification accuracy never exceeds 60% when an arbitrary radius or metric is used. Moreover, we correlated the jobs-residence ratio with the average commuting time–cost and confirmed a positive effect of JHB, which, however, also varies dramatically with different search radius. Additionally, we conducted an overlay analysis of the MSIC and PDC patterns and obtained eight joint-patterns as well as spatial divisions which contributes effectively to our understanding of the commuting issues in Shanghai.

**Keywords** Jobs-residence balance · Commuting time–cost · Aggregate patterns · Multi-scale analysis · Affinity propagation clustering · Shanghai

---

✉ De Wang  
dewang@tongji.edu.cn

<sup>1</sup> College of Architecture and Urban Planning, Tongji University, 1239 Siping Road, Shanghai, China

## Introduction

The concept of jobs-housing balance (JHB) describes the relationship between the configuration of workers and jobs within a given geographical area. An area is considered balanced when the resident workers can obtain a job within a reasonable travel distance and when the available housing types can complement a variety of employees' housing demands (Cervero 1991; Giuliano 1992). Planners, academics, and policy-makers generally consider JHB as positive because reducing the spatial separation between homes and jobs increases the potential for sustainability outcomes such as shorter commutes and reduced use of motorized modes (Cervero 2002; Cervero and Duncan 2006; Guo and Chen 2007), whereas an imbalance in jobs and housing creates longer commute times, more single driver commutes, loss of job opportunities for workers without vehicles, traffic congestion, and poor air quality (Zhou et al. 2012a).

However, the complicated causal effect between JHB and commuting efficiency should be acknowledged. Researchers show that the positive effect of JHB could be statistically significant but not very large (Giuliano and Small 1993; Levine 1998; Wachs and Kumagai 1973), or only existing under certain circumstances such as extreme jobs-housing ratio values (Peng 1997) and private car ownership (Blumenberg and Ong 2001; Zhou et al. 2012b), or could be not significant at all (Wachs et al. 1993). Even the standard of JHB is not consistently defined across studies. Consider the most frequently used metric, i.e., the jobs-housing ratio (Cervero 1988; Cervero 1989) or jobs-workers ratio (Horner 2007; Merlin 2014; Zhou et al. 2013), which have been applied at regional or intra-metropolitan level in many studies. To Cervero (1989), the ratio's upper ceiling should be 1.5, and if it gets bigger, there can be an undersupply of houses to meet the needs of the local workforce (Cervero 1989). Frank (1996) defined the balanced jobs-housing relationship within census tracts as a ratio between 0.8 and 1.2 (Frank 1996). And Peng (1997) found that only when the jobs-housing ratio is less than 1.2 or larger than 2.8 do VMT (vehicle miles traveled) vary noticeably as the jobs-housing ratio changes. In fact, it is inappropriate and even misleading to consider jobs and housing balanced only when the residents live and work in the same census tract or neighborhood.

These contradictions reflect one of the most limitations of the JHB measurement, the scale dependence problem, meaning that the metrics vary with analytical scale, such as an arbitrary search radius or pre-defined jurisdictional division. Researchers may argue that a proper search radius is 9.7–12.9 (Livingston 1989) or 4.8–16.1 km (Deakin 1989). Some researchers have noticed the limitation and tried to provide alternative metrics. Peng (1997) proposed a better choice which use several buffers as floating catchment areas to calculate the JHB ratio. But he did not provide analytical methods to handle the multi-scale indicators that resulted from the floating radius. In fact, the effect of scale and spatial unit definition on excess commuting (EC) values was identified at the inception of the commuting efficiency framework by Small and Song (1992) and was later systematically studied by Horner and Murray (2002). In the latter study, they showed that EC varies by 21.86 percentage points between the most (286 zones) and the least disaggregated (25 zones) scales and by 10–25.53 percentage points for different partitioning schemes. Niedzielski et al. (2013) tested some more related metrics and found that the newer ones are scale independent because their values are not dependent on the areal units. However, despite the widespread use of jobs-housing ratio, the side effect of scale dependence problem is not investigated yet. More importantly, new methods are

required to help the indicators like jobs-housing ratio maintain their comprehensibility as well as avoid the scale dependency limitation.

This paper proposes to handle the scale dependence problem in a straight-forward way, which is to measure the metrics at every scale (search radius) and use a distribution-based clustering method to identify aggregate patterns. Although the importance of other metrics of commuting efficiency, such as excess commuting metrics (Murphy and Killen 2011; Niedzielski 2006), is acknowledged, to focus on the above problem, here the commuting efficiency is simply defined as the realistic transportation cost of commuting trips. We applied our method on two indicators, i.e., the jobs-residence ratio and commuting time–cost, to provide a better understanding of the intervening relationship between them and to show how our method can be generalized to other indicators that limited by the scales. Besides, since the official travel survey data in China suffers from update rate and tight regulation, in this paper we based our analyses on the potential commuters extracted from cellular signaling data (CSD), which captures the location of a user once his cell phone communicates actively or passively with the cellular network via a tower. And through this study we also showed the potential of such a dataset for individual-level analysis of commuting issue.

Therefore, this paper consists of three sections: (1) extract home and job location of potential commuters from the CSD; (2) generate the jobs-residence ratio multi-scale index curve (MSIC) and commuting time–cost probability density curve (PDC) of each unit, and use affinity propagation clustering method to detect patterns and evaluate the bias of conventional method/metrics; (3) test the correlation between jobs-residence ratio and average commuting time–cost and evaluate the side effect of scales on it; and (4) conclude joint patterns with overlay analysis and discuss the key commuting issues revealed in Shanghai.

## Data preliminaries

### Data overview

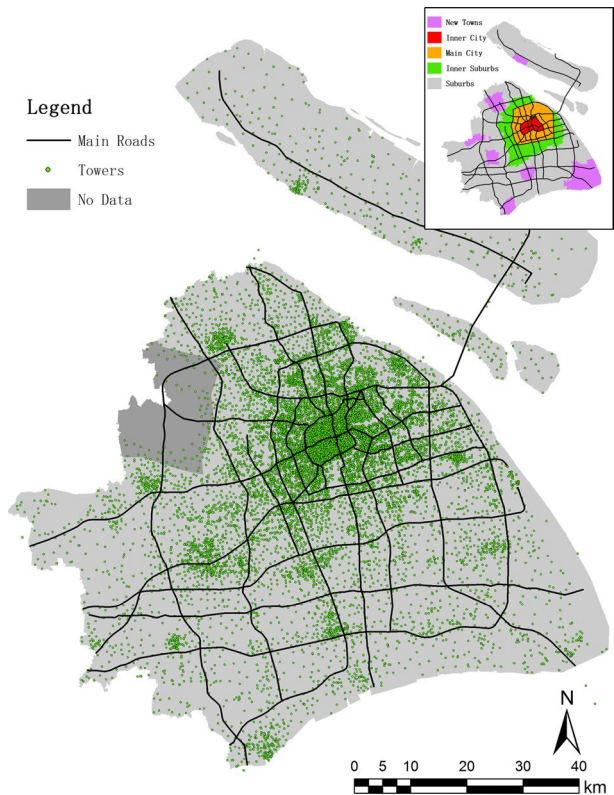
The usability of CSD comes from studies at both individual level and aggregate level. researchers have explored ways to extract activity sightings from these datasets (Bayir et al. 2010; Calabrese et al. 2013; Gonzalez 2013). Endorsement from validations has made these datasets reliable substitutions of traditional surveys (Jiang et al. 2013; Widhalm et al. 2015). At aggregate level, similar research subjects mainly focus on identifying and assessing origin–destination matrices (Duan et al. 2011; Iqbal et al. 2014; Rokib et al. 2015; Sohn and Kim 2008) and traffic flows (Alexander et al. 2015; Becker et al. 2011; Çolak et al. 2016; Holleczek et al. 2014; Tettamanti and Varga 2014; Toole et al. 2015). This paper follows the feasible methods in these studies with adjustments based on validation from the census data.

Our dataset consists of anonymous signaling records collected from more than 15 million phones, from one of the largest mobile operator in Shanghai within two weeks in march 2014. Once a cell phone communicates with the cellular network via a tower, a record may be generated and formed by (1) an anonymous device ID; (2) a time stamp; (3) an ID of the connected tower, allowing us to reconstruct the user's trajectory; and (4) the communication event type, as the examples shown in Table 1. Compared with census data in Shanghai, which only have about 5 thousand census tracts, the CSD provides traces of urban activities in a better resolution with 37 thousand towers (Fig. 1).

**Table 1** An example of cellular signaling record

(1) MSID	(2) Time stamp	(3) Tower ID	(4) Event ID
00000077D60EEDE43166F9C9CBF321C4	20140316131406 <sup>a</sup>	632258594	110
00000077D60EEDE43166F9C9CBF321C4	20140316131418	632258289	110
—	—	—	—
000000E974A986C39B1C18A63709EB48	20140316175854	617265265	61

<sup>a</sup>The format of the time stamp is “YYYYMMDDHHMMSS”

**Fig. 1** The 37 thousand towers involved in the dataset

### Infer potential commuters

Similar to many studies (González et al. 2008; Jiang et al. 2017; Widhalm et al. 2015), we inferred the home and job locations based on the accumulated time a user spends on each tower. The most frequently visited towers during the day (10:00–16:00) and night (20:00–06:00) are assigned to the job and home locations. The time interval is slightly different from existing papers considering the traffic peaks hours in Shanghai are 06:00–10:00 and 16:00–20:00 respectively. Then we calculated the proportion of the accumulated time spent at the home and job location tower as an index of the identification confidence. In formula,

$$IC\_home_n = \frac{dura_{ni}}{\sum dura_{ni}} \text{ in } 20 : 00 - 06 : 00$$

$$IC\_job_n = \frac{dura_{ni}}{\sum dura_{ni}} \text{ in } 10 : 00 - 16 : 00$$

where  $IC\_home_n$  and  $IC\_job_n$  are the Identification-Confidence (IC) of the home and job location of user  $n$ ;  $dura_{ni}$  is the accumulated time that user  $n$  spent at tower  $i$  during the days or nights; and  $\sum dura_{ni}$  is the total time that user  $n$  appears in the records during the day or night. Then we sampled the potential commuters based on the IC index and the appearing days as criteria, whose threshold values are chosen by a series of correlation tests with the aggregated census data. The final criteria are, (1)  $IC\_home \geq 0.5$  and  $0.15 \leq IC\_job \leq 0.75$ ; (2)  $appearing\ days \geq 2$ ; and (3) home tower should be different from job tower. The final sample consists of 4.34 million commuters, whose home distribution has a correlation of 0.943 with the 6th national population census data and job distribution has a correlation of 0.790 with the 2nd national economic census data. The latter is acceptable considering the census year is 2008, given the absence of other available validation sets.

## Method

### MSIC of the jobs-residence ratio

We use a multi-scale index curve (MSIC) to represent the jobs-residence relationship in each unit. The MSIC is defined as an array which synthesizes the values of the adjusted jobs-residence ratios measured at every search radius of the unit. To avoid the problems caused by the varying shapes and sizes of jurisdictional division, we use 1 km grids as analytic units. The adjusted jobs-residence ratio is calculated by,

$$R_{ir} = \frac{J_{ir} - W_{ir}}{J_{ir} + W_{ir}}$$

where  $R_{ir}$  is the adjusted jobs-housing ratio within the distance  $r$  of unit  $i$ ,  $J_{ir}$  is the total jobs within the distance  $r$  of unit  $i$ , and  $W_{ir}$  is the total workers who live within the search radius  $r$  of unit  $i$ . According to the formula, an area is considered employment-oriented when  $R_{ir}$  is approaching 1, and residence-oriented while  $R_{ir}$  gets closer to  $-1$ . And we measured the ratios with a step of 1 km in radius  $r$ .

### PDC of the commuting time–cost

We use a commuting time–cost probability density curve (PDC) to represent the overall characteristics of commuting efficiency in each unit. Similar to existing study (Xu et al. 2017), we utilized the time–cost captured from map service website instead of that extracted from the CSD data, considering the sparsity of the records makes it inappropriate to use the time between the last record at home the first record at workplace. Specifically, we drew the PDCs by four steps,

(1) Using the same grid as analytic units, each telecom tower is assigned to the nearby units with weights calculated by a standard 2-dimensional gaussian function. This assignment method is better than spatial joining or voronoi diagram for the reason that a phone

is not always choosing the closest tower to connect, but the closer towers do have higher probabilities to be connected.

(2) Commuting time–costs are captured from one of the largest map-service website in China (<http://ditu.amap.com>) by public transportation navigations. We chose the public transportation means instead of private cars because it has the largest share (31.3%), almost 10% larger than private car, according to “Shanghai transportation annual report 2015”.

(3) Home-job telecom tower pairs and the corresponding time–cost are assigned to fish-net units by the weights calculated in the first step, which are integrated to estimate the PDC of commuting time–cost.

## Extract patterns with a clustering method

Here patterns are defined as clusters in which the curves have the most similarity to each other. Each cluster has a most suitable exemplary curve which represents the multi-scale characteristics of that cluster. We use the affinity propagation algorithm proposed by Frey and Dueck (2007) to detect the clusters/patterns and the representatives. Basically, the algorithm identifies exemplars among data points and forms clusters based on the input measures of the *Affinity* between pairs of data points. There are several different ways to calculate the *Affinity* between two curves, such as euclidean distance, manhattan distance, or correlation coefficient. To make the most of the curves values, we calculated Euclidean distance between MSICs and *Relative Entropy* between PDCs as the *Affinity* inputs. Specifically, the *Relative Entropy* ( $D$ ) of one discrete probability distribution ( $P$ ) to another ( $Q$ ) is calculated by,

$$D_{(P||Q)} = \sum P_i \ln(P_i/Q_i)$$

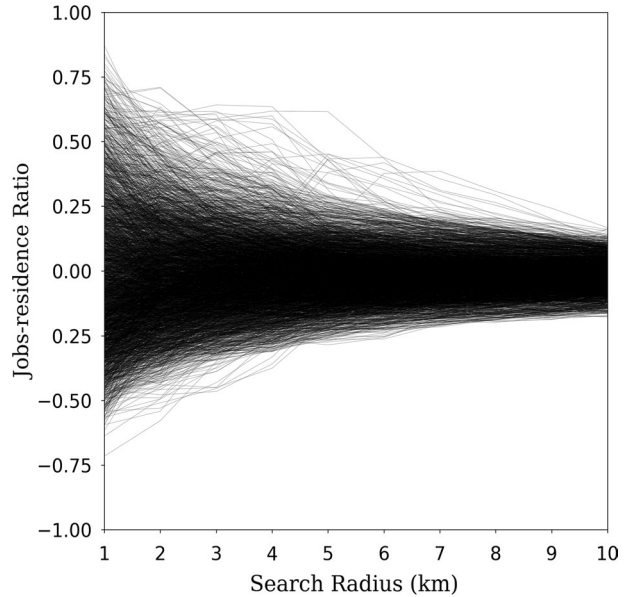
The *Affinity* values are compiled into the *Affinity matrix*, which is then put into affinity propagation algorithm to detect the clusters. We chose a small number of *preference*, which is another priori parameter representing the suitability of samples serve as an exemplar (Frey and Dueck 2007), to generate a small number of clusters/patterns.

## Result

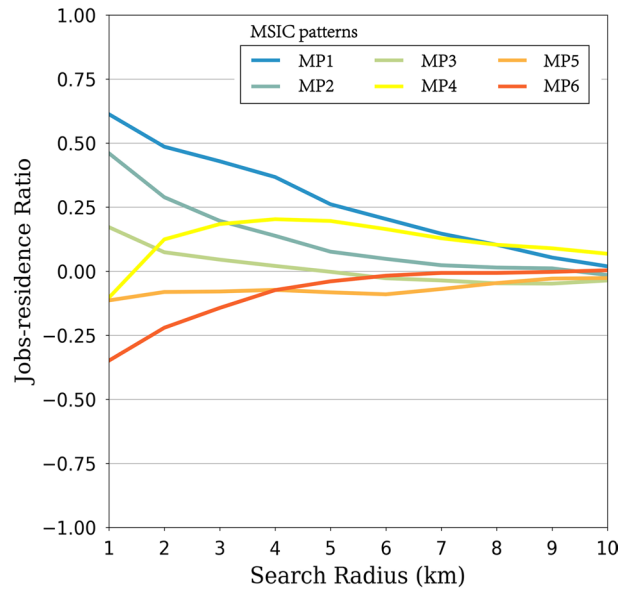
### MSIC patterns and the bias of conventional method

The MSICs reveal the scale-dependent nature of the jobs-residence ratio. As in Fig. 2, the jobs-residence ratios are generally imbalanced when measured within a small radius but are getting closer to 0 as the search radius increases, since every identified commuter will be connected to a job if the analytic radius is large enough. Moreover, different trends could be recognized from the revealed six patterns in Fig. 3, that most curves, such as those represented by MP1, MP2, MP3, and MP6, are simply approaching balance when search radius increases, while in some area, such as those represented by MP4, the jobs-residence relationship varies dramatically from residence-oriented to employment-oriented. It is also noticeable that all the representative curves have the steepest slope within a search radius of 2 km, indicating the metric within 2 km could be unreliable to provide standards for planning strategies.

**Fig. 2** The MSICs of all units. The radius of 1 km means to measure the ratio inside each grid, and 2–10 km represents the search radius from the units



**Fig. 3** The six MSIC patterns. The label from MP1 to MP6 and the color ramp from blue to red represent the transition from employment-oriented to residence-oriented. (Color figure online)



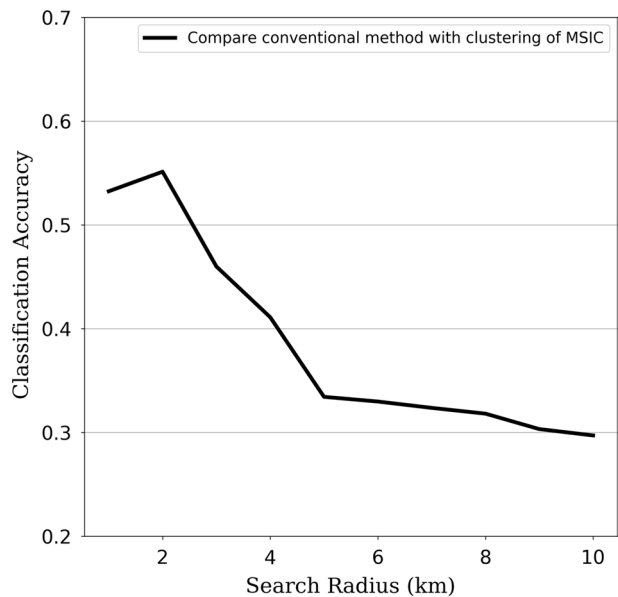
In addition, we found that the representative curves cross over each other frequently within radius from 3 to 10 km, which is commonly used by researchers (Deakin 1989). For example, the red MP6 curve is the most residence-oriented when the analytical radius is smaller than 4 km. However, it becomes the most balanced area when radius increase larger than 6 km. Obviously, an arbitrary radius may lead to entirely different result from another when analyzing jobs-residence relationship at intra-metropolitan level. Thus, our finding may shatter all the “perfect” search radius in existing studies.

To measure the bias of the conventional method, we compared our result with the K-means clustering based on the jobs-residence ratio with pre-defined search radius. We set K to the number of clusters in our multi-scale result, which is 6. Then, we considered the representative curves in our multi-scale result as markers, and if one unit is classified into the same cluster in the K-means result, we counted this unit as a right classified point for that search radius. Thus the percentage of right classified units is calculated as accuracy to show how robust such a pre-defined search radius is when compared to our multi-scale method. Not surprisingly, the best accuracy is only 55% (Fig. 4), indicating that a pre-defined radius can hardly bear comparison with the multi-scale method. Moreover, as the radius increases to 10 km, the accuracy declines all the way to 30%, which is reasonable because the unstable ratio at small radius contributes more to the euclid distance we use for affinity calculation. However, it still demonstrates how an arbitrary radius may lead to a different result from the multi-scale result.

### PDC patterns and the bias of conventional metrics

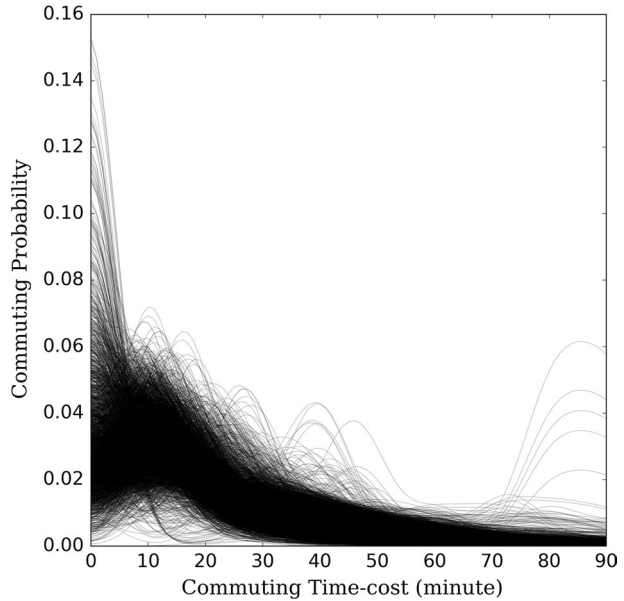
A straightforward demonstration of the city's commuting efficiency is provided in Fig. 5, where the curves generally appear in a bell-shape with one peak and a long tail. Distinct patterns could be inferred directly from the curves, for instance, based on the time–cost point the curves peak at. In fact, the affinity propagation algorithm archived this goal perfectly. In Fig. 5, together the ten exemplars illustrate a clear and complete picture of the commuting efficiency with distinct peaks moving away from one and another. The CP1 to CP4 which peak at time–cost smaller than 20 min could be recognized as “compact” patterns with excellent commuting efficiency, while some other curves which peak at 35 min (e.g., CP8) or have a fat tail when time–cost larger than 40 min (e.g., CP7 and CP9), are definitely representing bad commuting performance. The worst situation is CP10, which shows a completely reversed pattern, indicating most of the commuters live in those units spend more than one hour traveling to workplace every day.

**Fig. 4** The percentage of right classified units comparing conventional arbitrary-scale metric with our MSIC method



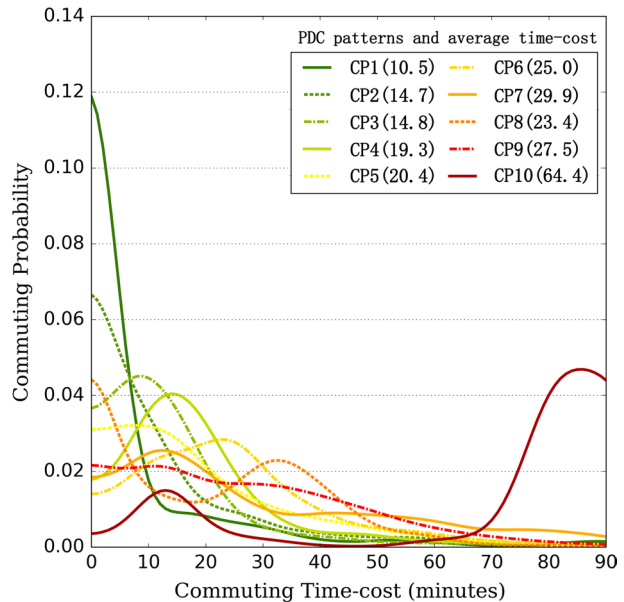


**Fig. 5** The PDCs of all units. To obtain smooth curves, we use a 1-dimensional Gaussian interpolation function with a standard deviation of 5 min



Moreover, Fig. 6 shows the advantage of our methods explicitly. On the one hand, the curves imply a lot more information than conventional metrics such as the mean time-cost. As shown in Fig. 6, despite the distinct distribution revealed by the curves, their average time-costs are much closer to each other. For example, CP8 is the third worst situation among the representative curves but its average time-cost is even better/smaller than CP6

**Fig. 6** The ten representative commuting PDCs. The label from CP1-10 and the color ramp from green to red represents distinct patterns from “compact” to “flat” distribution. In the parentheses we attached the average time-cost of each curve. (Color figure online)



and CP7. On the other hand, we also measured the gap between conventional metrics and our method by comparing the clustering results. Similarly, the representative PDCs are considered as markers, and if one unit belongs to the same cluster in the K-means result, then it is counted as a right point. The result shows an accuracy of 43% for the mean time–costs and 57% for median time–costs. Obviously, our method prevails over such metrics in assessing commuting efficiency.

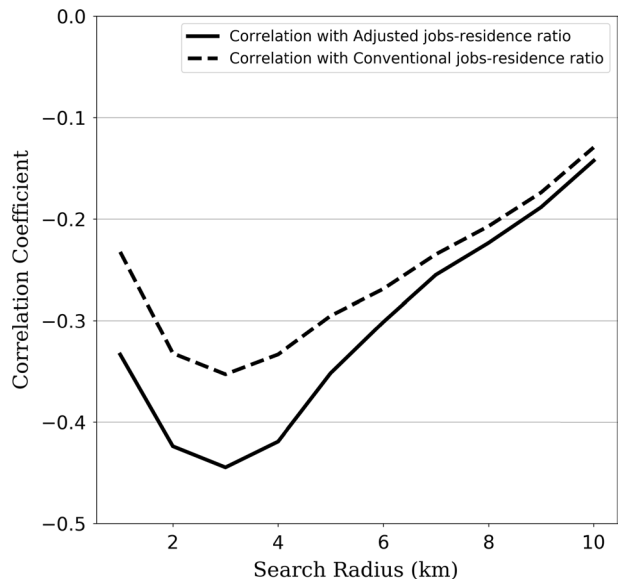
### Correlation between jobs-residence ratio and average commuting time–cost

We correlated the jobs-residence ratio with the average commuting time–cost and confirmed the potential positive effect of JHB. According to Cervero and Duncan (2006), every 10% increase in the number of jobs in the same occupational category within 4 miles of one’s residence is associated with a 3.29% decrease in daily VMT (vehicle miles traveled). We found similar result: all other things being equal, every 0.1 increase in jobs-residence ratio within 6.5 km (about 4 miles) is associated with 0.7 min decrease in mean time–cost, with a correlation coefficient of 0.25. However, scale dependence problem still exists. We plotted the correlation coefficients between the mean time–cost and the two types of jobs-residence ratio at every search radius in Fig. 7. The correlation coefficient changes from  $-0.45$  to  $-0.12$  across radiuses, and the coefficient of jobs-residence ratio varies even more dramatically from  $-0.15$  to  $-0.72$ . Again this raises doubt about all the measurements of the positive effect of jobs-residence ratio given a pre-defined radius.

### Joint-patterns and the commuting issue in Shanghai

In fact, a simple overlay analysis combining the two patterns together may provide a better understanding of the commuting issue in Shanghai than the statistics. In Fig. 8 it depicts the overlay matrix of MSICs and PDCs, where only some of the units, that with comparatively balanced MSIC (such as MP4 and MP5), are having excellent commuting

**Fig. 7** The correlation between the mean time–cost and the two types of jobs-residence ratio within every search radius. The adjusted ratio is the one we use in this paper. The conventional ratio refers to the number of jobs per resident workers



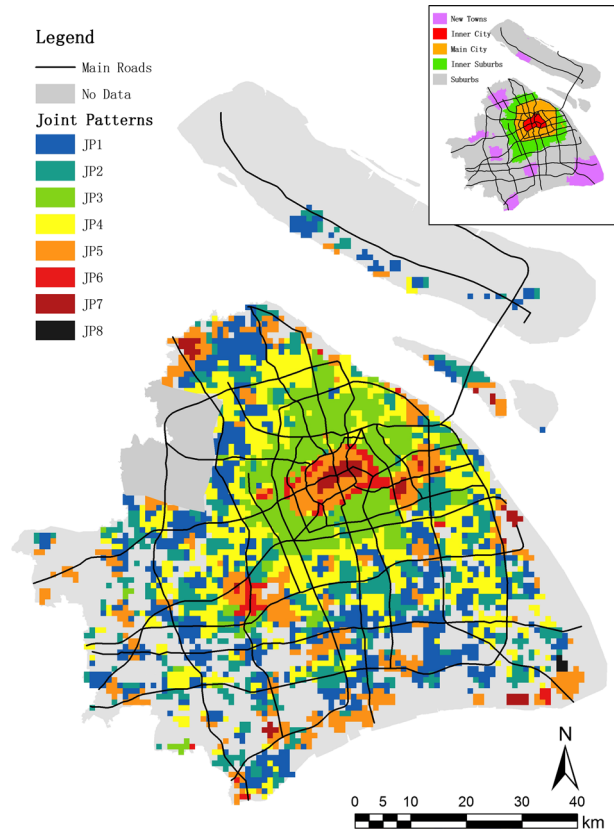
		MSIC patterns						
		Employment developed			Residence developed			
		MP1	MP2	MP3	MP4	MP5	MP6	
Commuting PDC patterns	Compact	CP1	1	19	17	6	6	1
		CP2	14	74	23	166	89	17
		CP3	12	93	79	27	128	7
		CP4	6	45	49	129	129	5
		CP5	35	2	2	258	449	19
		CP6	4	19	21	44	65	10
		CP7	4	5	0	22	79	8
		CP8	2	14	11	10		3
	Flat	CP9	0			48		279
		CP10	0	0	1	0	0	4

**Fig. 8** The overlay matrix of MSICs and PDCs. The grey figures indicate the number of units. All the units are divided into eight blocks, each of that refers to a joint pattern. Every joint pattern is filled with the same color as its spatial division in Fig. 9

performance. We provided the matrix without any further clustering process for planners for detecting important commuting problems in Shanghai. It turns out to be a very effective tool because of its ability to link the sub-regions with visualized curves representing the multi-scale characteristics of jobs-residence relationship as well as commuting efficiency. We concluded eight joint-patterns and mapped their spatial division in Fig. 9. The joint-patterns and the explanations are described as follows,

1. The suburbs are generally covered by JP1 and JP2. JP1 refers to the units which have an excellent jobs-residence relationship, such as clusters MP4 and MP5, and a relatively compact PDC. JP2 has the same PDC but a consistently residence-oriented MSIC. Both of them represent the self-organization nature of the suburban development. In addition, none of the new towns shows a different pattern, which indicates that the goal of developing new towns to help improve the commuting efficiency of the main city is not successful.
2. JP3 and JP5 constitute two rings area outside the main center with green and orange color in Fig. 9, which illustrate the worst jobs-residence relationship in Shanghai. JP3 is matched but not balanced while JP5 balanced but mismatched. The crucial point derives from the opposite jobs-residence configuration, the relatively long distance between the two sub-regions, and the vast population involved, all of which may lead to enormous traffic flows from the green ring to the orange ring to fill the vacant jobs. It could be recognized as a result of the rapid expansion of the housing development and the lack of consideration of employment development in the green area during the latest two decades.
3. JP4 has a relatively balanced MSIC but a flat PDC, which implies a severe mismatch issue. According to a survey in 2014 (Zhu 2014), most of these commuters are those work in the main city or new town but cannot afford the housing price there. However,

**Fig. 9** The spatial division of the eight joint patterns



their commuting efficiency are acceptable since the PDC still have a prominent peak at 30 min.

4. JP7 indicates a very strong main center of Shanghai, whose shape is more like a corridor than a node as depicted in Fig. 9, with extremely employment-oriented MSICs at every analytic radius. This employment center in Shanghai may have similar jobs-residence ratio as Beijing (Zheng et al. 2015), but a lot more imbalanced than other cities such as Xi'an (Zhou et al. 2014). On the positive side, most of these units have a relatively compact PDC, which means most of the resident workers there could find suitable jobs not far from homes.
5. Both JP6 and JP8 rarely appear but they imply severe commuting issues. JP6 has an employment-oriented MSIC and a flat PDC, which implies a severe mismatch issue at the edge of the main core. JP8 is a very special case which only exists in the service area of the Lingang harbor.

## Conclusions and discussion

This paper proposes to apply clustering method to multi-scale indicators for evaluating aggregate patterns of jobs-housing relationship and commuting time-cost. Taking Shanghai as a case study, the analysis in this paper uncovers new aspects of the scale

dependence problem surrounding the jobs-housing balance and commuting efficiency framework and shows the potential of our method to be generalized to other indicators. Meanwhile, we based our analysis on individual commuters extracted from cell phone data and showed the potential of such a dataset for commuting researches.

The results show that our multi-scale method prevails over conventional ones with pre-defined search radius or metrics in three ways. First, the multi-scaled curves of MSIC and PDC represent naturally more information than conventional metrics. On the one hand, we confirmed the limitation of search radius smaller than 2 km, noticing the unstable ratio is not qualified to define any standard of JHB. On the other hand, the crossover points of MSICs which occur frequently at radius from 3 to 10 km have shattered all the proposed “perfect” search radius in existing studies. Second, we compared our result with the conventional methods and showed the bias of the latter. When our multi-scale result is considered as markers, we showed that the best accuracy of the jobs-residence ratio with an arbitrary radius is 55%, and the accuracies of the mean and median time-costs are 43% and 57%, indicating that a pre-defined radius/metrics can hardly bear comparison with the multi-scale method. Third, we correlated the jobs-residence ratio with the average commuting time-cost and confirmed the potential positive effect of JHB. However, the coefficient of jobs-residence ratio varies dramatically from  $-0.15$  to  $-0.72$ , which raises doubt about the stability of all the measured positive effect of jobs-residence ratio given a pre-defined radius.

Instead of statistics, through the cooperation with local planners, our methodology is proved to be an effective tool because of its ability to link the commuting problems with visualized and comprehensive curves. The case study of Shanghai uncovers some underlying aspects of the commuting issue in big Chinese cities. In the existing accomplishments, the institutional factors, such as the “danwei” system (Wang and Chai 2009; Zhou et al. 2014) or the housing reform policies (Wang et al. 2011), and the land use configuration, such as those in industrial parks and development zones (Zhou et al. 2016a, b), are the most recognizable factors affecting the jobs-housing relationship and commuting efficiency in China. In this case, however, the problematic areas of JP3 and JP5, with opposite jobs-residence configurations, could hardly be adequately explained by factors as mentioned above. It seems that urban expansion and the lack of consideration of employment development at the edge of the main city are two critical factors. This is also different from the common situation of the US big cities in the post-suburban era, where they experienced not only population but also employment decentralization (Lucy et al. 1997). The situation of Shanghai is more like another strand, such as Hong Kong, that despite the population decentralization, the jobs-housing imbalance under urban sprawl is mostly caused by the remained concentration of jobs in the central districts (Hui and Lam 2005; Loo and Chow 2008). This may have led to unsustainable transport trends such as longer commuting distances and severe peak-hour traffic congestion in Shanghai. We suggest that all these possible causes and effects should be carefully examined before implementing any targeted policy.

#### **Compliance with ethical standards**

**Conflict of interest** On behalf of all authors, the corresponding author states that there is no conflict of interest.

## References

- Alexander, L., Jiang, S., Murga, M., González, M.C.: Origin–destination trips by purpose and time of day inferred from mobile phone data. *Transp. Res. Part C Emerg. Technol.* **58**, 240–250 (2015)
- Bayir, M.A., Demirbas, M., Eagle, N.: Mobility profiler: a framework for discovering mobility profiles of cell phone users. *Pervasive Mob. Comput.* **6**(4), 435–454 (2010)
- Becker, R.A., Caceres, R., Hanson, K., Ji, M.L., Urbanek, S., Varshavsky, A., Volinsky, C.: A tale of one city: using cellular network data for urban planning. *Pervasive Comput. IEEE* **10**(4), 18–26 (2011)
- Blumenberg, E.A., Ong, P.M.: Cars, buses, and jobs: welfare participants and employment access in Los Angeles (2001)
- Calabrese, F., Mi, D., Lorenzo, G.D., Ferreira, J., Ratti, C.: Understanding individual mobility patterns from urban sensing data: a mobile phone trace example. *Transp. Res. Part C Emerg. Technol.* **26**(1), 301–313 (2013)
- Cervero, R.: America's suburban centers: a study of the land-use-transportation link. TRID. No. UMTA-TX-06-0054-88-1 (1988)
- Cervero, R.: Jobs-housing balancing and regional mobility. *J. Am. Plann. Assoc.* **55**(2), 136–150 (1989)
- Cervero, R.: Jobs-housing balance as public policy. *Urban Land* **50**(10), 10–14 (1991)
- Cervero, R.: Built environments and mode choice: toward a normative framework. *Transp. Res. Part D Transp. Environ.* **7**(4), 265–284 (2002)
- Cervero, R., Duncan, M.: Which reduces vehicle travel more: jobs-housing balance or retail-housing mixing? *J. Am. Plann. Assoc.* **72**(4), 475–490 (2006)
- Çolak, S., Lima, A., González, M.C.: Understanding congested travel in urban areas. *Nat. Commun.* **7**, 10793 (2016)
- Deakin, E.: Land use and transportation planning in response to congestion problems: a review and critique. *Transp. Res. Rec.* **1237**, 77–86 (1989)
- Duan, Z., Liu, L., Wang, S.: MobilePulse: dynamic profiling of land use pattern and OD matrix estimation from 10 million individual cell phone records in Shanghai. In: 2011 19th International Conference on Geoinformatics, pp. 1–6 (2011)
- Frank, L.D.: An analysis of relationships between urban form (density, mix, and jobs: housing balance) and travel behavior (mode choice, trip generation, trip length, and travel time). *Transp. Res. Part A* **1**(30), 76–77 (1996)
- Frey, B.J., Dueck, D.: Clustering by passing messages between data points. *Science* **315**(5814), 6–972 (2007)
- Giuliano, G. (1992). Is Jobs-Housing Balance a Transportation Issue? *University of California Transportation Center Working Papers*, 1305(1305), 305–312
- Giuliano, G., Small, K.A.: Is the journey to work explained by urban structure? *Urban Stud.* **30**(9), 1485–1500 (1993)
- González, M.C., Hidalgo, C.A., Barabási, A.L.: Understanding individual human mobility patterns. *Nature* **453**(7196), 82–779 (2008)
- Gonzalez, M.: Unraveling daily human mobility motifs. In: *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, xxxviii–xxxviii (2013)
- Guo, J.Y., Chen, C.: The built environment and travel behavior: making the connection. *Transportation* **34**(5), 529–533 (2007)
- Holleczeck, T., Yu, L., Lee, J.K., Senn, O., Ratti, C., Jaillet, P.: Detecting weak public transport connections from cellphone and public transport data. In: *Proceedings of the International Conference on Big Data Science and Computing*, pp. 1–8. (2014)
- 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., Murray, A.T.: Excess commuting and the modifiable areal unit problem. *Urban Stud.* **39**(1), 131–139 (2002)
- Hui, E.C.M., Lam, M.C.M.: A study of commuting patterns of new town residents in Hong Kong. *Habitat Int.* **29**(3), 421–437 (2005)
- Iqbal, M.S., Choudhury, C.F., Gonza'Lez, M.C., Wang, P.: Development of origin-destination matrices using mobile phone call data: a simulation based approach. In: *Transportation Research Board 93rd Annual Meeting* (2014)
- Jiang, S., Fiore, G.A., Yang, Y., Ferreira, J., Frazzoli, E.: A review of urban computing for mobile phone traces: current methods, challenges and opportunities. In: *ACM SIGKDD International Workshop on Urban Computing*, pp. 1–9 (2013)
- Jiang, S., Ferreira, J., Gonzalez, M.C.: Activity-based human mobility patterns inferred from mobile phone data: a case study of Singapore. *IEEE Trans. Big Data* **3**(2), 208–219 (2017)

- Levine, J.: Rethinking accessibility and jobs-housing balance. *J. Am. Plann. Assoc.* **64**(2), 133–149 (1998)
- Livingston, B.L.: Using jobs/housing balance indicators for air-pollution controls. Master's thesis. United States: N. p. (1989)
- Loo, B.P.Y., Chow, A.S.Y.: Changing urban form in Hong Kong: what are the challenges on sustainable transportation? *Int. J. Sustain. Transp.* **2**(3), 177–193 (2008)
- Lucy, W.H., Phillips, D.L., Lucy, W.H., Phillips, D.L.: The post-suburban era comes to richmond: city decline, suburban transition, and exurban growth. *Landsc. Urban. Plan.* **36**(4), 259–275 (1997)
- Merlin, L.A.: Measuring community completeness: jobs-housing balance, accessibility, and convenient local access to nonwork destinations. *Environ. Plan.* **41**(4), 736–756 (2014)
- 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.: A spatially disaggregated approach to commuting efficiency. *Urban Stud.* **43**(13), 2485–2502 (2006)
- 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)
- Peng, Z.R.: The jobs-housing balance and urban commuting. *Urban Stud.* **34**(8), 1215–1235 (1997)
- Rokib, S.A., Karim, M.A., Qiu, T.Z., Kim, A.: Origin-destination trip estimation from anonymous cell phone and foursquare data. In: *Transportation Research Board Annual Meeting*, (2015)
- Small, K.A., Song, S.: Wasteful commuting: a resolution. *J. Polit. Econ.* **100**(4), 888–898 (1992)
- Sohn, K., Kim, D.: Dynamic origin-destination flow estimation using cellular communication system. *IEEE Trans. Veh. Technol.* **57**(5), 2703–2713 (2008)
- Tettamanti, T., Varga, I.: Mobile phone location area based traffic flow estimation in urban road traffic. pp. 1–15, (2014)
- Toole, J.L., Colak, S., Sturt, B., Alexander, L.P., Evsukoff, A., González, M.C.: The path most traveled: travel demand estimation using big data resources. *Transp. Res. Part C Emerg. Technol.* **58**, 162–177 (2015)
- 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)
- Wachs, M., Kumagai, T.G.: Physical accessibility as a social indicator. *Socio-econ. Plann. Sci.* **7**(5), 437–456 (1973)
- Wang, D., Chai, Y.: The jobs–housing relationship and commuting in Beijing, China: the legacy of Danwei. *J. Transp. Geogr.* **17**(1), 30–38 (2009)
- Wang, E., Song, J., Xu, T.: From “spatial bond” to “spatial mismatch”: an assessment of changing jobs–housing relationship in Beijing. *Habitat Int.* **35**(2), 398–409 (2011)
- Widhalm, P., Yang, Y., Ulm, M., Athavale, S., González, M.C.: Discovering urban activity patterns in cell phone data. *Transportation* **42**(4), 597–623 (2015)
- Xu, Y., Li, R., Jiang, S., Zhang, J., González, M. C.: Clearer skies in Beijing-revealing the impacts of traffic on the modeling of air quality. (2017)
- Zheng, S., Xu, Y., Zhang, X., Yu, D.: Jobs-housing balance index and its spatial variation: a case study in Beijing. *Qinghua Daxue Xuebao/J. Tsinghua Univ.* **55**(4), 475–483 (2015)
- Zhou, J., Chen, X., Wei, H., Peng, Y.U., Zhang, C.: Jobs-housing balance and commute efficiency in cities of central and western China: a case study of Xi'an. *Acta Geogr. Sin.* **68**(10), 1316–1330 (2013)
- Zhou, J., Wang, Y., Cao, G., Wang, S.: Jobs-housing balance and development zones in China: a case study of suzhou industry park. *Urban Geogr.* **38**, 1–18 (2016a)
- Zhou, J., Wang, Y., Schweitzer, L.: Jobs/housing balance and employer-based travel demand management program returns to scale: evidence from Los Angeles. *Transp. Policy* **20**, 22–35 (2012a)
- Zhou, J., Wang, Y., Schweitzer, L.: Jobs/housing balance and employer-based travel demand management program returns to scale: evidence from Los Angeles. *Transp. Policy* **20**(1), 22–35 (2012b)
- Zhou, J., Zhang, C., Chen, X., Wei, H., Peng, Y.: Has the legacy of danwei persisted in transformations? the jobs-housing balance and commuting efficiency in Xi'an. *J. Transp. Geogr.* **40**, 64–76 (2014)
- Zhou, S., Liu, Y., Kwan, M.: Spatial mismatch in post-reform urban China: a case study of a relocated state-owned enterprise in Guangzhou. *Habitat Int.* **58**, 1–11 (2016b)
- Zhu, J.: Dilemmas and counter-measures of peri urbanization phenomena in metropolitan suburbs: the case study of Shanghai. *Urban Plan. Forum* **6**, 13–21 (2014)

**Longxu Yan** is a PhD candidate at the College of Architecture and Urban Planning (CAUP), Tongji University, and a visiting PhD student at the Senseable City Lab, MIT. His research focus on implementing innovative data-driven approaches in urban studies to provide deep insights into urban spatial structure, built environment and human mobility, city performance and housing market development. As an urban planner

and data scientist by training, he has conducted spatial analyses and modeling based on various big/open dataset at Tongji and MIT Senseable City Lab.

**De Wang** is a professor at the CAUP, Tongji University, where he directs the Technical Research Center for Digital Planning, Tongji Institute of Urban Planning and Design. His research focus on behavior and policy, urban mobility, and big data analysis.

**Shangwu Zhang** is a professor at the CAUP, Tongji University. He is also the vice dean of the CAUP. His research focus on urban and regional development, spatial strategic planning, and rural planning.

**Dongcan Xie** is a master student at the CAUP, Tongji University. His research focus on big/open data mining and spatial modeling.