Computers, Environment and Urban Systems 96 (2022) 101851

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

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

Analyzing the multiscale patterns of jobs-housing balance and employment self-containment by different income groups using LEHD data: A case study in Cincinnati metropolitan area

Zhiyuan Yao[a], Changjoo Kim[b,*]

[a] Data Science Center, University of California, Los Angeles, United States of America
[b] Department of Geography & GIS, University of Cincinnati, United States of America

ARTICLE INFO

Keywords:
Commuting
Jobs-housing balance
Employment self-containment
K-medoids
LEHD

ABSTRACT

Achieving a balanced jobs-housing relationship has been treated as a solution to relieve traffic congestion and air pollution. Jobs-housing ratio and employment self-containment (ESC) are applied to quantify different levels of jobs-housing balance. Though appropriated methods have been analyzed to explore jobs-housing relationship at multiple fixed scales, scale dependency problem remains unsolved and resulted in discord about jobs-housing relationship. Furthermore, the differences of the jobs-housing balance and ESC by various income groups received little attention. This paper proposed a K-medoids clustering method to explore the aggregate multi-scale patterns of jobs-housing ratio and ESC by higher, medium, and lower-income groups using 2016 longitudinal employment household dynamics (LEHD) data. To alleviate the scale effect, individual workers and jobs were simulated by Monte Carlo approach according to LEHD data provided within each block. Afterwards, we analyzed the jobs-housing relationship within multi-scale units by different aggregated groups. We found jobs-housing balance had a positive effect on ESC, while the effect varied by different income groups. Jobs-housing balance affects ESC more for lower-income group when the searching radius is <4 km, while beyond 4 km, jobs-housing balance has more effect on ESC of higher-income group. ESC of medium-income group is least affected. Employing the K-medoids clustering method, we classified the four patterns of jobs-housing and the two patterns of ESC for each income group. We further identified jobs-housing matched and mismatched areas through eight joint patterns of jobs-housing balance and ESC. Our work contributes to understanding of jobs-housing balance and commuting by scale variations and income.

1. Introduction

The spatial relationships between the distribution of jobs and house units determine the demands of journey-to-work trips (Zhou, Yeh, Li, & Yue, 2018). Jobs-housing balance has been applied to investigate this spatial relationship and estimate the demand within a certain geographic area. A balanced area tends to have shorter and less journey-to-work trips, while to what extent the balance can be measured by employment self-containment (ESC). ESC is the percentage of resident workers who live and work locally over the number of resident workers living in an area (Cervero, 1996). It reflects realistic journey-to-work trips and further indicates if the residential workers tend to work locally (Zhou & Yeh, 2021). A high ESC implies less unnecessary trips, contributes to less automobile dependency and is more environmentally friendly (Curtis & Olaru, 2010; Kelobonye, Mao, Xia, Swapan, & McCarney, 2019). Jobs-housing balance policies have been adopted with the aim of increasing ESC and decreasing the commuting distance/time.

However, the complex relationship between the jobs-housing balance and commuting should be acknowledged. Though several studies found jobs-housing balance would reduce commuting time (Cervero, 1989; Levinson & Kumar, 1994; Sultana, 2002; Zhao, Lü, & De Roo, 2011), others reported minor or no association between jobs-housing balance and commuting patterns (Giuliano, 1991; Giuliano & Small, 1993; Wachs, Taylor, Levine, & Ong, 1993). One possible limitation contributed to the inconsistence is the scale dependency problem, meaning the metrics to measure jobs-housing balance vary with the analytical scales. At regional or metropolitan scale, there is roughly

---

* Corresponding author at: 400B Braunstein Hall, Cincinnati, OH 45221-0131, United States of America.
E-mail addresses: zyao@ucla.edu (Z. Yao), changjkm@ucmail.uc.edu (C. Kim).

https://doi.org/10.1016/j.compenvurbsys.2022.101851
Received 18 November 2021; Received in revised form 14 June 2022; Accepted 18 June 2022

balanced jobs and housing, while at district or neighborhood level, there are job-rich areas or worker-rich areas or vice versa. The jobs-housing balance also varies by arbitrary boundaries, such as census tracts. To avoid the arbitrariness, Peng (1997) pointed out the meso-level was the best scale for measuring jobs-housing balance and recommended using floating catchment areas centered at traffic analysis zone (TAZ). He pointed out that the range calculated from the average distance and the standard deviation was an appropriate radius range for the floating catchment area. This method was adapted by various research to analyze the jobs-housing balance and commuting (Wang, 2000; Antipova, Wang, & Wilmot, 2011; Zhao et al., 2011; Stoker & Ewing, 2014; Hu & Wang, 2016), but these studies did not provide the analytical methods to examine the multi-scale metrics resulted from the floating radius. Studies analyzed excess commuting (EC) frame identified the scale dependency problem (Small & Song, 1994) and explored the influence of the spatial unit on the EC values systematically. Horner and Murray (2002) explored the impact of the spatial scales on the excess commuting at 11 different aggregated spatial scales for Boise, ID and found that as the size of spatial unit increased, the excess commuting tended to decrease, implying a better jobs-housing balance and high commuting efficiency, indicating a larger area tends to smooth the variations. Niedzielski, Horner, and Xiao (2013) extended the analysis of the scale issues for a suit of excessing commuting metrics and found while excess commuting was scale dependent, the capacity used, and capacity efficiency were independent of the scale. To alleviate the spatial dependency, studies employed Monte Carlo simulation to generate the random residential and workplace locations (Hu & Wang, 2015, 2016; Hu, Wang, & Wilmot, 2017). Recently, with the availability of big data, such as mobile phone location data, and the smart-card data, researchers applied individual data to alleviate the scale dependence when examined the jobs-housing balance under EC framework (Huang, Hu, Wang, & Li, 2021; Ma, Liu, Wen, Wang, & Wu, 2017; Zhang, Zhou, & Zhang, 2017; Zheng, Liu, Guo, & Lei, 2019; Zhou, Yeh, Li, & Yue, 2018; Zhou, Yeh, & Yue, 2018).

EC reveals the overall jobs-housing balance for a study area but fails to unveil the local variations that exert influence on individual's commuting pattern. Revealing the local variations, jobs-housing ratio and ESC are indicators that can represent the jobs-housing balance at local areas under certain spatial scale units (Debenham, Stillwell, & Clarke, 2003; Martinus & Biermann, 2018a; Zhou, Yeh, & Yue, 2018; Kelobonye et al., 2019). Studies have been done to explore the impact of jobs-housing balance on ESC and commuting considering local variations. Zhou and Yeh (2021) explored the effect of multiple spatial units (1 km – 3 km grids) on the jobs-housing ratio and ESC by utilizing mobile positioning data. They found the 2 km grid was the best ideal spatial unit for the analysis of ESC. But Yan, Wang, Zhang, and Xie (2019) indicated that jobs-housing balance in one scale of spatial unit revealed limited information. Rather than finding the best spatial unit or searching radius, they proposed a clustering method that evaluated the aggregate patterns of jobs-housing relationship using a multi-scale indicator in a series of spatial scales. Their result concluded that the clustering method prevailed over the conventional ones with pre-defined search radius or grids. Though aggregate patterns reveal more information naturally (Yan et al., 2019), few literature examines those of ESC and jobs-housing balance systematically.

Besides the variability of jobs-housing balance across spatial scales, the variability of jobs-housing balance were also identified across different income groups (Horner & Schleith, 2012; Schleith & Horner, 2014; Suárez, Murata, & Delgado Campos, 2016) and the commuting distance/time (Hu et al., 2017a). The income played a determining role in residential location selection, which was driven by the trade-off among the commuting cost, the living environment and housing affordability. On the one hand, high-income workers are more likely to have jobs that required advanced skill, and thus may need to travel longer for high-paying jobs. On the other hand, high-paying workers have more options to choose where to reside. High-income people who work at an inner core of a city can choose a high-rent apartment to avoid traffic congestion and save commuting time. In comparison, some other high-income people may be longing for the amenity of suburban environment and choose to locate at suburban with high commuting cost but are compensated by spacious houses. Compared with high-income workers, low-income workers have relative less options to choose where to locate and, in some case, where to work. Low-income workers who find low-paying jobs in major urban centers may be priced out due to the high cost of either housing or renting, while those who have less skilled jobs out of urban centers may have trouble in commuting as the scarcity of public transit in suburban areas. Spatial mismatch of skilled jobs, affordable housing, and accessible public transit may contribute to the less balanced jobs-housing relationship for low-income groups. Metropolitan strategic plans have targeted the increasing employment in sub-regions to relieve traffic congestion and commuting issues that threaten urban sustainability (Martinus & Biermann, 2018). Due to the employment complexity, housing affordability, residential choices, and commuting cost, high-income workers and low-come workers respond differently to these plans. Studies have been done to identify the varying effects and proposed corresponding policy suggestions to achieve a better jobs-housing balance. Cervero (1996) examined the ratios of jobs to employed residents in 23 large San Francisco Bay Area cities during the 1980s and found imbalance declined in dormitory regions while became worse in job-surplus areas. He applied the number of vehicles as a proxy of income and investigated the impact of jobs-housing balance on independent index (Thomas, 1969) which measured the ratio of internal commutes to external commutes. The study found that cities with higher income were associated with more external commutes, which implied a longer commuting. He pointed out high housing prices compelled some workers to live in another city. Blumenberg and King (2021) re-re visited jobs-housing balance in California cities from 2002 to 2015. They found the California cities were becoming less self-contained and were consonant with Cervero (1996) that cities with low costing had higher levels of self-containment. They concluded that housing affordability crisis pushed workers and jobs farther apart and recommended policies to better match job skills and housing prices to characteristics of workers. Both Cervero (1996) and Blumenberg and King (2021) explored the impact of income on jobs-housing balance and self-containment, but did not further examined how differences were of jobs-housing balance by various income groups. Hu et al. (2017) analyzed the community variability by wages groups from 1990 to 2010 in Baton Rouge. He found affluent neighborhoods tended to commute more but highest-wage neighborhoods retreated for less commuting. The study highlighted that both low-wage workers and highest-wage workers tended to have shorter commuting time. But he pointed out that the economic commuting distance of low-wage workers might be largely associated with poor transport mobility while the highest-wage workers were groups who truly enjoyed the efficient commute time. Schleith, Widener, and Kim (2016) analyzed jobs-housing balance for 26 metropolitan regions by different income groups using LEHD data. They found higher-income group had long commuting and low jobs-housing balance, while the low-income group had shorter commutes and high jobs-housing balance.

Though attention was paid to analyzing the varying jobs-housing balance by income groups, few studies investigated that of ESC. Zhou, Yeh, and Yue (2018) examined spatial variation of self-containment and jobs-housing balance by job types and locations. They found secondary-sector workers tended to reside near their workplaces because of relatively balanced jobs and housing. Their result showed the correlation between jobs-housing balance and ESC was stronger for secondary-sector worker than that of tertiary-sector workers. Jobs-housing policies that aimed at high ESC should consider job types. Policies that were effective for secondary-sector worker may not be effective for tertiary-sector workers. Martinus and Biermann (2018b) investigated self-containment across a range of occupations and industry types. They found agriculture grouping was the most self-contained industry (ESC >

65%) with highest percentage of occupations, such as managers, professional, clerical, and sales living and working within same region. Mining industry had the lowest value of ESC (ESC $< 5\%$), reflecting a large proportion of commute in/out workers. The large disparity of ESC by different occupations and industries indicated that the jobs-housing policies and ESC target which focused on increasing local jobs in subregions needed to consider the skills needed and industry structure of the targeted regions. These studies explored ESC from perspectives of job and industry types. Yet limited effort has been put into exploring the differences in jobs-housing balance and ESC by income groups.

Literature shows gaps of scale dependency issue and the role of income groups in measuring jobs-housing relationship. Identifying these gaps, this study asks how scale affects the measurement of jobs-housing balance and ESC and then future asks how jobs-housing balance affects ESC by different income groups. To answer these questions, this article aimed to 1) examine the jobs-housing balance and ESC within multiscale units by different income groups and 2) explore the aggregate patterns of jobs-housing balance and ESC. For aim 1, we explored the jobs-housing balance, ESC, and their relationship within various sizes of floating catchment areas (1 km - 10 km radius) by different income groups using Monte Carlo simulated individual data derived from the 2016 Longitudinal Employment Household Dynamic (LEHD) data. For aim 2, we applied K-medoids clustering methods to investigate the aggregate patterns of jobs-housing balance and ESC respectively considering incremental spatial scales, and further identified the joint-patterns of jobs-housing balance and ESC.

2. Study area

Cincinnati metropolitan statistical area (MSA), the study area of this research, spans Indiana, Kentucky and Ohio (Fig. 1). As estimated by United States Census Bureau, Cincinnati MSA ranks 28th by population (2,193,676) among 384 MSAs in the nation in 2016, and commuters accounted for around 43.84% of the population. The MSA is not static entity, the boundary changes as the commuting flows between counties changes. The 2010 Standards for Delineating Metropolitan and Micropolitan Statistical Areas (US Office of Management and Budget, 2010) indicates the counties in MSAs need to be contiguous and share at least 25% of commuting flows. From 2003 to 2013, the MSA included 15 counties, with 3 countries from Indiana, 7 from Kentucky and 5 from Ohio (Fig. 1), while after 2013, Cincinnati MSA replaced Franklin County, Indiana with Union County, Indiana (Fig. 2). But Union County shared a small portion of boundary with Butler County. It is apparent that when commuters from Union County travel to work, they inevitably

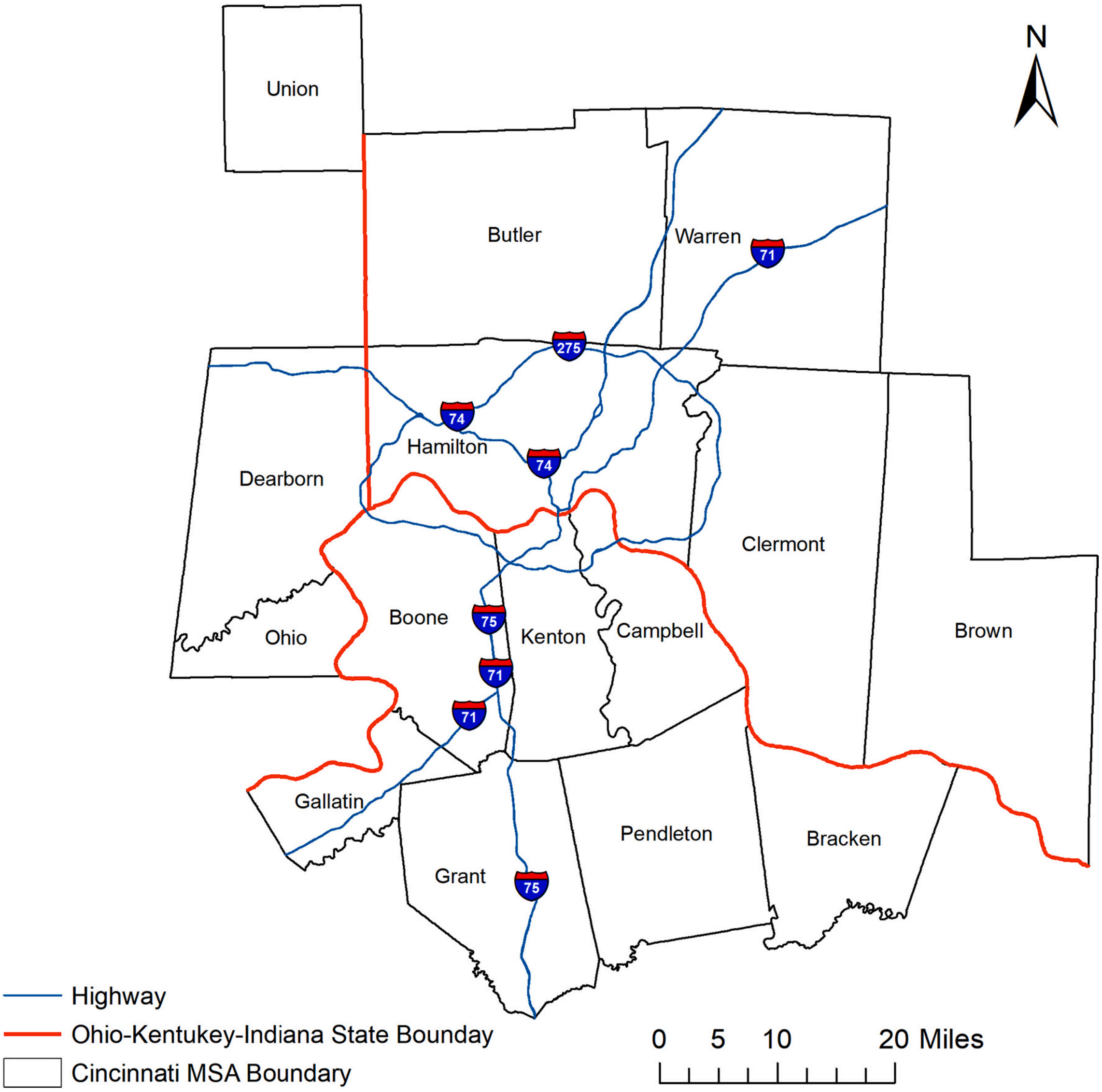

Fig. 1. Cincinnati boundary for 2013–2016.

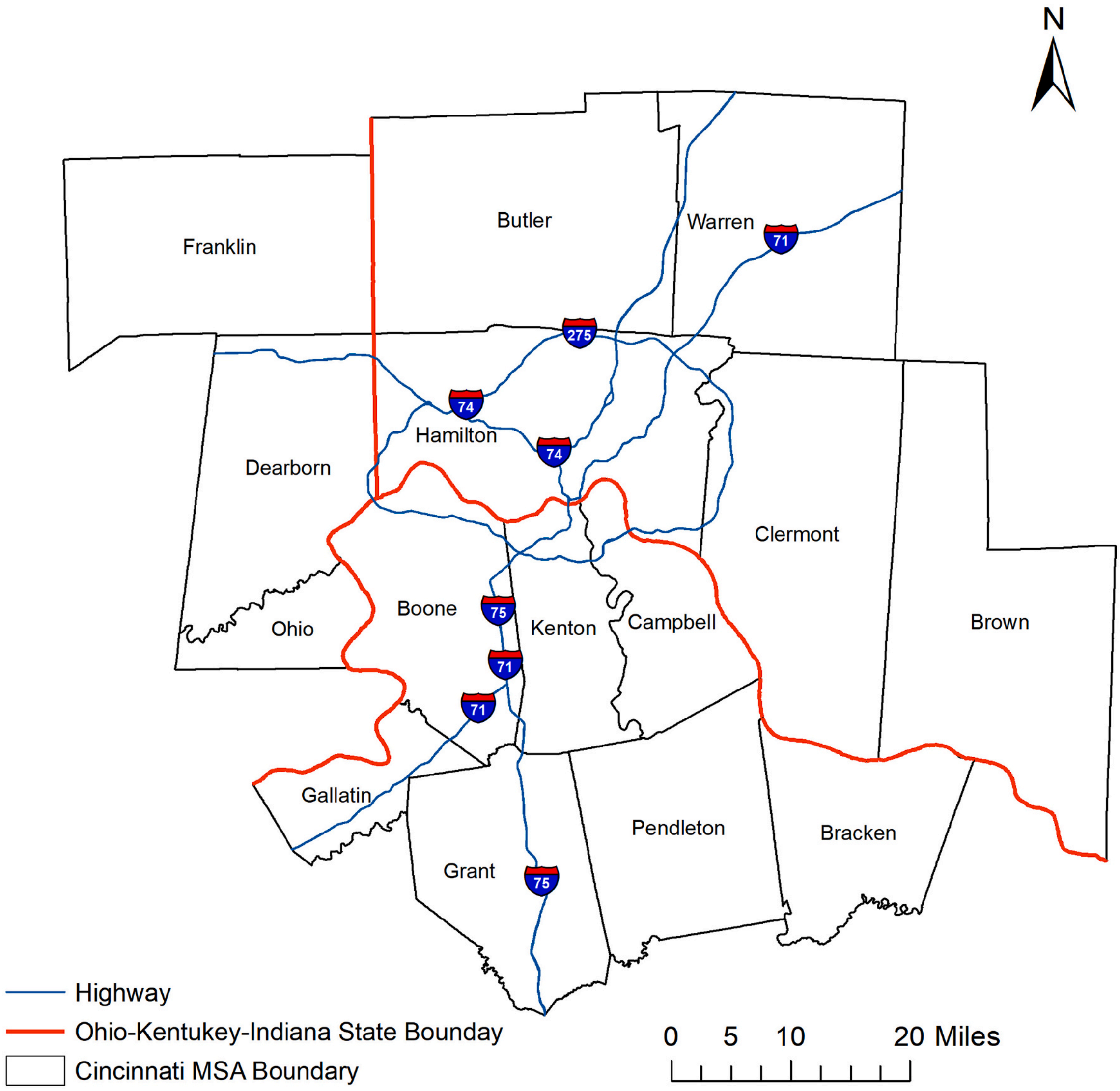

Fig. 2. Cincinnati boundary for this study.

travel across Franklin County. It is also the same for commuters who live in Butler Country but work in Union County. Though the flows between Franklin County and other counties in the metropolitan area are <25% of the commuting flows, Franklin County connects the Union County and the other counties geographically. The jobs of Franklin County may also affect the flows of commuters from Butler County or more counties. Considering the geographic location of Franklin County, we decided to choose the 2003–2013 boundary for Cincinnati Metropolitan Area, which includes Franklin County rather than Union County (Fig. 2).

3. Data

The data of jobs and residential workers came from Longitudinal Employment Household Dynamics (LEHD) program released by the U.S. Census, which provided datasets for studying local economy. LEHD Origin-Destination Employment Statistics (LODES) helps users to understand and analyze where people work and where workers live. LODES data are state-based and comprise three types: Origin-Destination (OD), Residence Area Characteristics (RAC), and Workplace Area Characteristics (WAC) at census block level. OD files provide the number of commuting flows from a home census block to a work census block. RAC data are job totaled based on home census blocks, while WAC data are job totaled by work census blocks. All tabulations for these three types are available by job dominance, i.e., primary and secondary. The number of jobs applied in this research only includes primary jobs, since we assume when measuring how far the job is from home, workers are more likely to consider their primary jobs (Schleith et al., 2016). OD, RAC, and WAC files include characteristics about average monthly earnings (Table 1). We applied lower-income, medium-income, and higher-income to differentiate the different income groups classified by LEHD. Lower-income group means the income in this group is the lowest in the three categories, and higher-income group indicates this group has the highest income among the three groups. We intended to use them to reveal relative relationships and not to represent

Table 1
LEHD data.

Data	Year	Source	Format	Income group	
LEHD	2016	Census Bureau	Excel	Lower (23.84%)	less than $1250/ month
				Medium (57.73%)	between $1251/month and $3333/month
				Higher (18.43%)	greater than $3333/ month
NLCD	2016	USGS	Raster	NA	

the low income (around $25,000 per year) or high income (around $100,000 per year) that are generally perceived. The lower-income workers and higher-income workers account for 23.84% and 18.43% respectively. Most of the workers belong to medium-income group (57.73%). We employed RAC, WAC, and OD data to explore the jobs and housing balance and calculate ESC.

Though the LEHD LODES data has been a great resource for transportation and regional planning, the limitations should be acknowledged. The main data source for LEHD data is state employment-insurance (UI) wage records. The UI records are reports filed by employers every quarter for each individual in covered employment through Quarterly Census of Employment and Wages (QCEW) Program. The data collection method indicates self-employed individuals and independent contractors are not covered (Abowd, Haltiwanger, & Lane, 2004; Stevens, 2002). The independent contractors include people working in 'gig economy' with temporary positions and short-term commitments, such as uber drivers, delivery drivers, labor jobs and etc. Another limitation is the data of the worksite locations for employer with multiple worksites. An employer may have multiple worksites within a state. Employers with multiple worksites include large corporations with several manufacturing plants (e.g., auto manufacturers), school districts with teachers and staff located in different schools, and retail establishments with multiple store locations. QCEW asks the multi-worksites employer to submit the locations of secondary worksites, but it is voluntary for some states, such as Ohio, Indiana, and Kentucky (Spear, 2011). This means though people work at secondary worksites, the work location in LEHD dataset may be the headquarter/primary worksite of the employer. This would lead to the wrong commuting flows derived from the LEHD data.

Another dataset we applied in this study was 2016 National Land Cover Database (NLCD), which illustrated 20 land cover types for the States, including low intensity, medium intensity, and high intensity developed area. The NLCD has a spatial resolution of 30 m * 30 m, and each pixel has only one type of land cover type. We extracted the low, medium, and high intensity developed areas as the residential areas, and medium and high intensity areas as job located areas in each census block. Table 1 summarized the data sources used in this research.

4. Multiscale jobs-housing ratio and employment self-containment by income groups

LEHD provides data of residents, jobs, and origin-to-destination flows at census block level. Alleviating the spatial dependency problem, we adopted the Monte Carlo method (Hu et al., 2017; Hu & Wang, 2016) to simulate individual locations, randomly distributing the residential workers of each block to the selected low, medium, and high intensity developed areas (Fig. 3) and allocating the jobs of each block to the selected medium, and high intensity developed areas for different income groups. Afterwards, we randomly paired the origin (workers) and destination (jobs) flows that followed a discrete frequency distribution and were consistent with the actual flows between origin and destination from LEHD. The simulated residential workers, jobs, and individual origin to destination flows were utilized to calculate the jobs-housing ratio and ESC for different income groups.

The jobs-housing balance and ESC at one scale of units reveal the jobs-housing relationship at a certain scale, while the multi-scaled analysis of aggregate patterns reveals more information of jobs-housing relationship across incremental scales jointly. Exploring the jobs-housing balance and ESC in multiscale units, we firstly utilized 500 m tessellation grids as analytical units; secondly, we created buffers centered at each grid with radius varying from 1 km to 10 km and finally calculated the jobs-housing balance and ESC within each size of buffers. We chose the 500 m * 500 m grid as the analytical unit as it was approximate to the average size of a block (520 m*520 m) in the study area. The jobs-housing index measuring the jobs-housing balance is calculated by jobs-workers ratio (JWR),

$$JWR_{ir} = \frac{J_{ir} - W_{ir}}{J_{ir} + W_{ir}} \tag{1}$$

$$r \in (1, 2\ldots 10), i \in (overall, low\ income, medium\ income, and\ high\ income)$$

where J_{ir} is the number of jobs within the radius r of i income group for each grid, W_{ir} is the number of residential workers within the radius r of i income group for each grid. JWR_{ir} ranges from −1 and 1. An area is considered as job rich area when the index is approaching to −1 and is considered as resident rich area when the index is approaching 1. In contrast, 0 represent the absolute balanced jobs and residential workers.

ESC measures the percentage of residential workers who work in the same area as they live. The higher ESC indicates an area provides

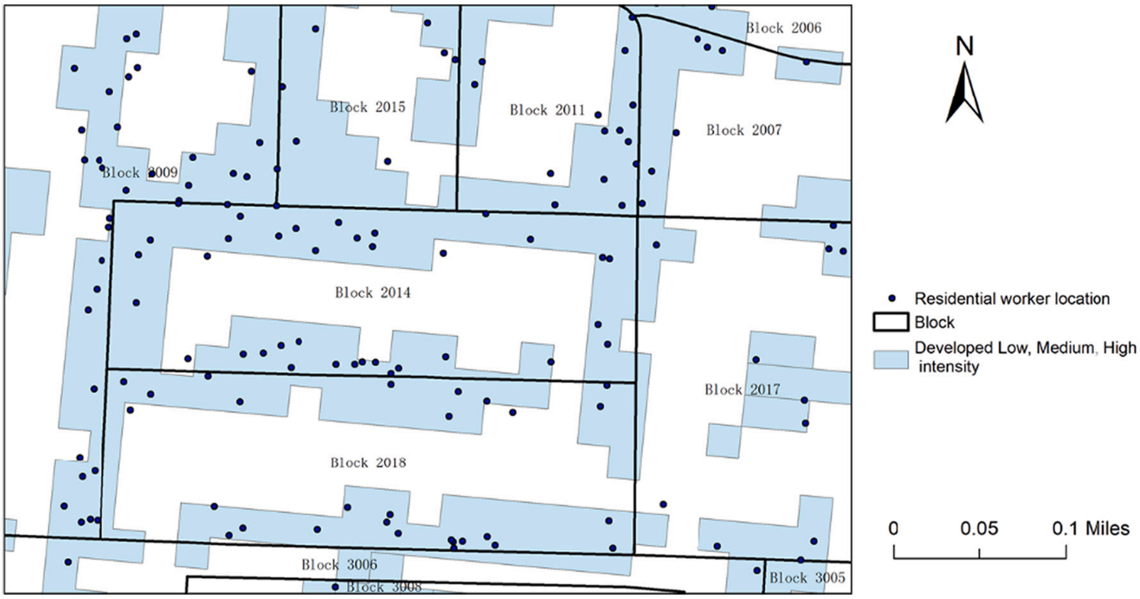

Fig. 3. Monte Carlo simulated residential locations in blocks.

enough jobs that residential workers can find jobs within where they live, while the smaller ESC implies the jobs in an area do not meet the residential workers' needs that they need to travel outside of the area to find jobs. The ESC is calculated for low, medium, and higher-income group as,

$$ESC_{ir} = \frac{\text{number of workers who work and live in the same area}}{\text{number of worker living in the area}} \tag{2}$$

$$r \in (1, 2\ldots10),\ i \in (overall, low\ income, medium\ income, and\ high\ income)$$

JWR and ESC were calculated within varying sizes of buffers centered at each grid for lower-, medium-, and higher-income groups.

We applied K-medoids clustering method to explore the aggregate patterns considering information with varying sizes (1 km to 10 km) of buffers. K-medoids (also called as Partitioning Around Medoids) algorithm was proposed by Kaufman and Rousseeuw (1990) which split the data set of n objects into K clusters with K as the prior. K-medoids clustering minimizes the dissimilarity between the points in this cluster and the medoid of the cluster. We used Euclidean distance in this research (Yan et al., 2019) to measure the dissimilarity between each pair of points. The medoids are selected randomly from the input data points at the beginning. The K-medoids clustering was realized by a R package: pam.

Since the number K is a prior, we needed to set it before running the clustering algorithm. We used Elbow method to identify the best number of K for the clustering of JWR and ESC at multiscale units. Elbow method first calculates the total within-cluster sum of square (WSS) and then plots the curve of WSS according to the number of clusters. The location of a bend (knee) in the plot is considered as the appropriate number K of clusters. Increasing the number of clusters will naturally decreases the WSS because the smaller the cluster is, the more similar the points in the cluster are. But the high number of K is more likely to lead to overfitting. The elbow (bend/knee) of the WSS plot identifies the point after which the WSS decreases slowly when K increases. We calculated the WSS when K ranged from 1 to 10 and plotted the curve of WSS for both JWR and ESC. Fig. 4 indicates that the appropriate number of K is 4 for overall JWR, and the appropriate number of K is 2 for the overall ESC (Fig. 5). We also adopted 4 clusters and 2 clusters for JWR and ESC when analyzing the aggregate patterns for different income groups (lower-, medium-, and higher-income). The WSS was calculated using the function *fviz-nvcluster* from the R package *clustersim*. The workflow of exploring the patterns of JWR and ESC is illustrated in the Fig. 6.

5. Result

5.1. Aggregated patterns of jobs-housing

The aggregate patterns of JWR are represented as identified clusters.

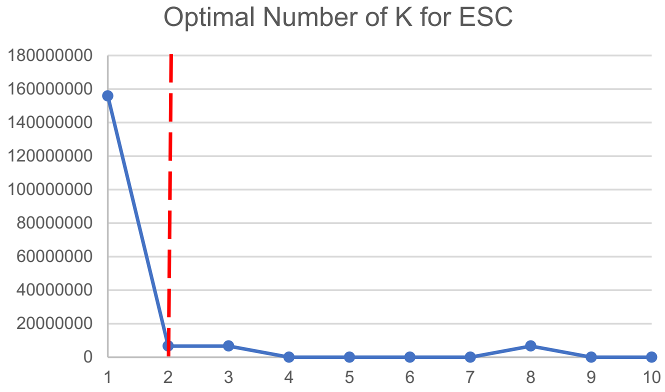

Fig. 4. Optimal number of K clusters of ESC.

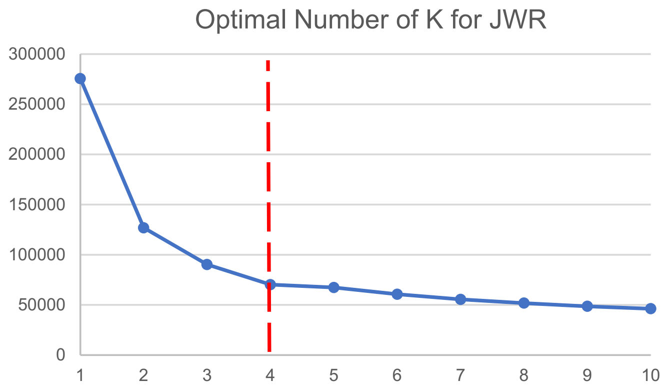

Fig. 5. Optimal number of K for clusters of JWR.

There are 4 clusters (aggregate patterns) identified for the overall jobs-housing balance, which are *Balanced, Job Rich, Resident Rich-Suburb*, and *Resident Rich-Rural.* The boxplots illustrate the distribution of JWR of all clusters, and the lines in the chart (Fig. 7) show the average JWR of each group at multiscale units. *Job Rich* cluster is job-oriented areas where there are more jobs than residential workers across spatial units. *Balanced* cluster is resident-oriented areas at small size of buffers but becomes balanced as the size of spatial units increases, while *Residential Rich-Suburb* and *Residential Rich-Rural* are resident-oriented areas across varying spatial units. The varying JWR at different spatial units reveals the spatial dependency in evaluating JWR. According to the boxplots, the median of JWR increases as the buffer radius increases, and the variation of JWR gradually decreases as the buffer size increases. Examining the lines in the chart, the jobs-housing is initially imbalance at the smaller size of units but approaches to balanced as the radius increases. *Job Rich* cluster has a relatively high JWR within small radius buffers, but gradually approaches to 0, indicating as the size of spatial unit increases, the jobs and housing tend to be balanced. The *Balanced* cluster is more resident-oriented at small radius of buffers but turns balanced for 5 km to 10 km radius of buffers, implying the numbers of jobs and workers are approximate the same after the buffer radius exceeds 5 km. The *Resident Rich-Suburb* and *Resident Rich-Rural* clusters have low JWR at small size of spatial units, but as the buffer radius increases, the JWR of the two groups increases but still fails to reach a balanced condition. The *Resident Rich-Suburb* cluster has a relative higher JWR than the *Resident Rich-Rural* cluster, implying suburb areas have more jobs than rural areas.

Fig. 8 illustrates spatial patterns of JWR clusters. The initial map has 42,601 grids. But we removed the grids that have no residential workers, leaving 23,467 grids in total. Out of all grids, *Job Rich* cluster accounts for 13.94%, *Balanced* cluster accounts for 15.65%, *Resident Rich* cluster accounts for 32.80% and *Resident Rich-Rural* accounts for 37.60% (Table 2). The *Job Rich* cluster grids are mainly located at the central cities of the metropolitan area, with some scattered in peripheral areas. From the southwest to northeast, the *Job Rich* cluster covers Florence, Burlington, Erlanger in Kentucky, Cincinnati, Springfield, Blue Ash, Montgomery, West Chester, Mason, and all the way north to Monroe and Middletown in Ohio. The scattered *Job Rich* grids are located at Oxford and Harrison in Indiana, Mt. Orab, Georgetown, and Huntington in OH, and Crittenden and Williamstown in Kentucky. The *Balanced* cluster encircles the *Job Rich* cluster. There are relative less jobs in this cluster than in *Job Rich* cluster, but the residential workers living in this area can find jobs within certain radii of buffer zones as these areas are adjacent to *Job Rich* cluster. *Resident Rich-Suburb* cluster is resident-oriented areas, encircling *Balanced* areas. More residential workers are in this cluster than jobs across all radii (1 km −10 km) of buffers centered at these grids. In the middle of the metropolitan area, the outer boundary of *Resident Rich-Suburb* cluster approximately corresponds with the boundary of urban areas. The rest of the grids constitutes the

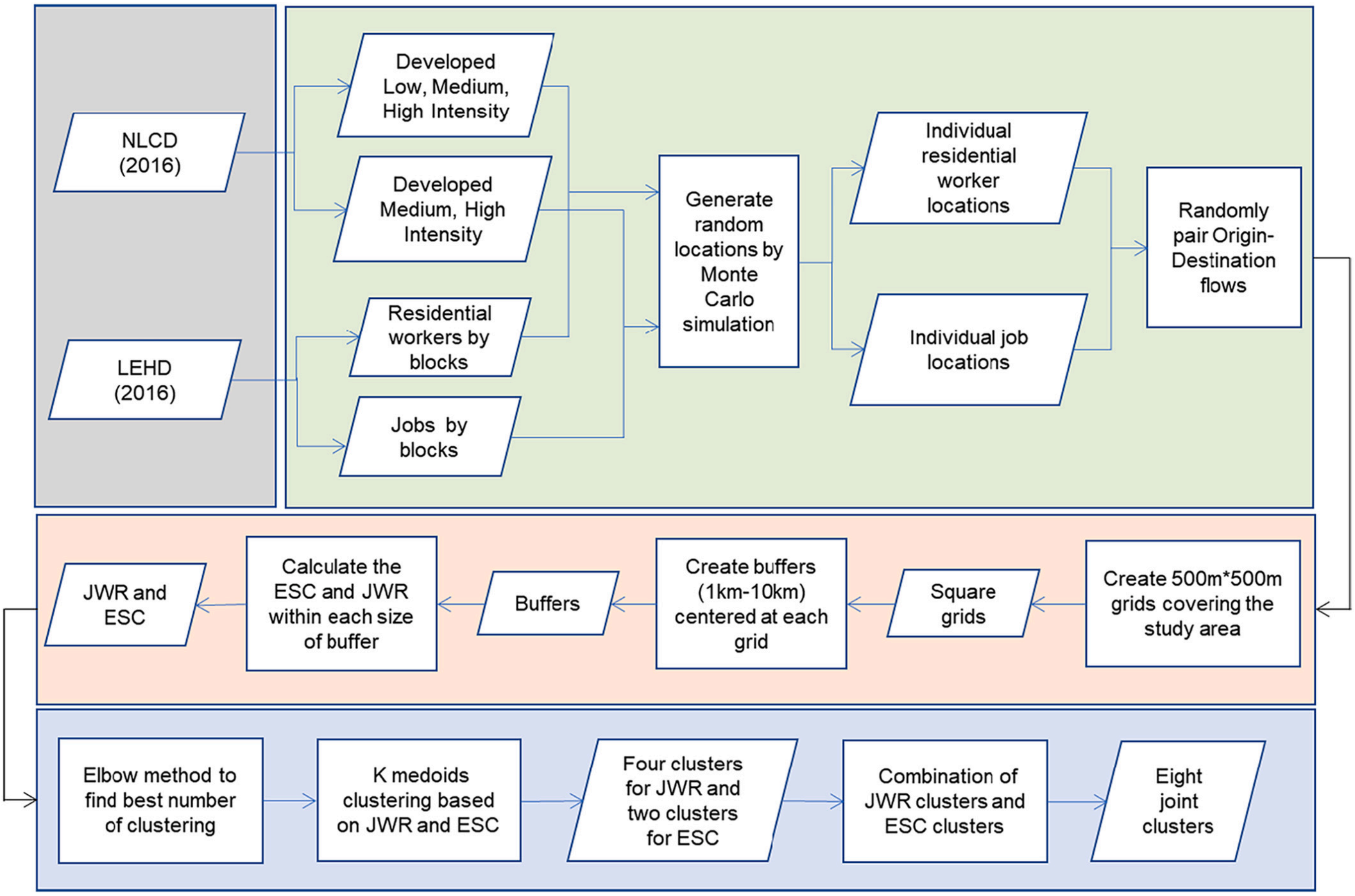

Fig. 6. The workflow of this study.

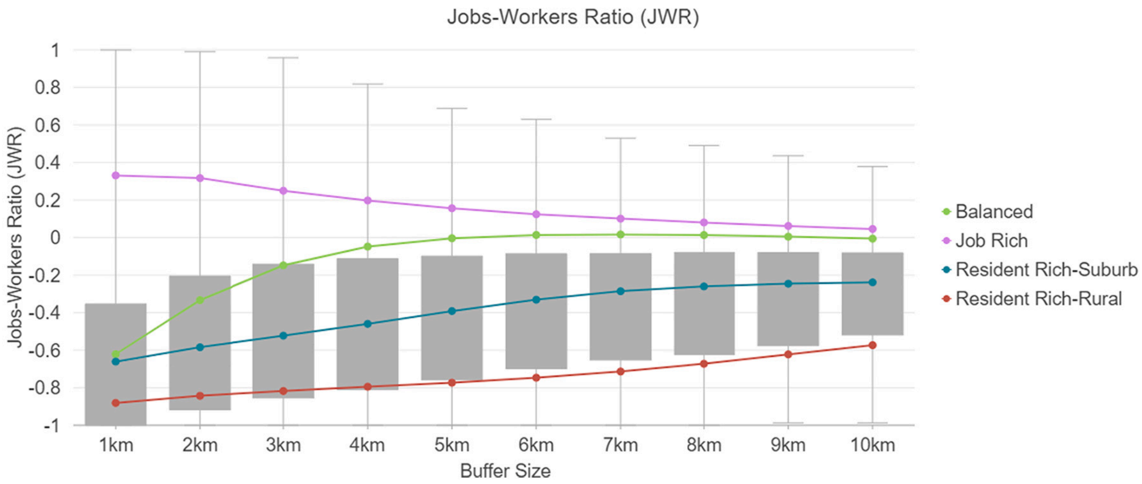

Fig. 7. The overall JWR of multiscale units.

Resident Rich-Suburb cluster.

We also identified the JWR patterns for lower-, medium-, and higher-income groups (Fig. 8) using K-medoids clustering method. Each income group was classified into 4 clusters (*Job Rich, Balanced, Resident Rich-Suburb, and Resident Rich-Rural*) according to the JWR calculated for each income group. The JWR patterns of different income groups vary in terms of the percentage of and spatial distribution of grids in different clusters (Table 2). For higher-income group, the *Job Rich* cluster (12.49%) grids are located at the central cities of metropolitan area and barely distribute in peripheral cities. The *Balanced* cluster grids account for 23.65%, encircling the *Job Rich* cluster. For medium-income group, the percentage of *Job Rich* cluster grids is 20.30%, covering both the central cities and peripheral cities. The *Balanced* cluster of medium-income group covers more areas and accounts for 28.80% of all grids. Observing the lower-income group, the *Job Rich* cluster grids are more scattered than those of higher-income and lower-income groups. The *Balanced* cluster grids make up 24.68% of all grids, filling up the gapped areas between the *Job Rich* cluster grids.

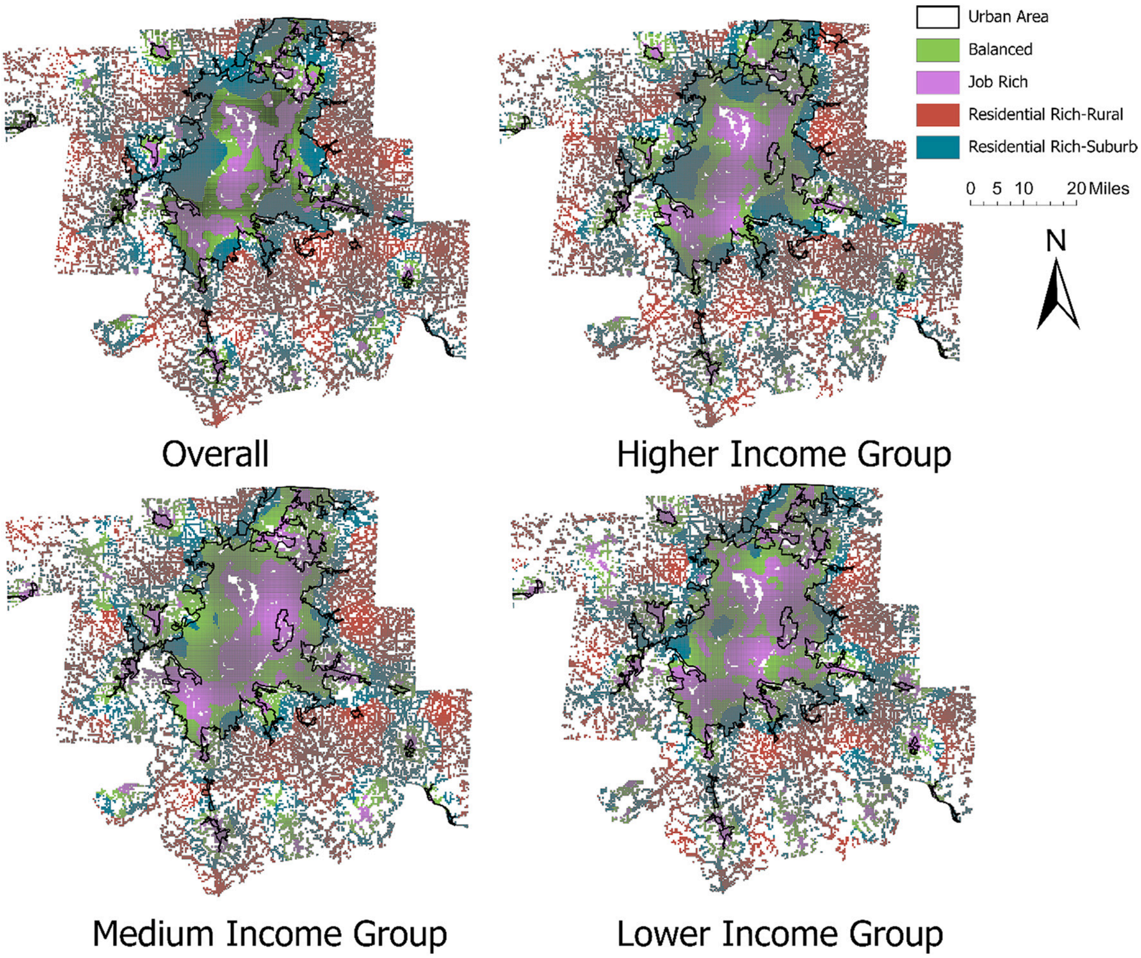

Fig. 8. The distribution of JWR clusters.

Table 2
The percentage of grid in different clusters by low, medium, and higher-income group.

JWR	Overall	Lower-income	Medium-income	High-income
Job Rich	13.94%	22.10%	20.30%	12.49%
Balanced	15.65%	24.68%	28.80%	23.65%
Resident Rich-Suburb	32.80%	28.43%	23.98%	30.59%
Resident Rich-Rural	37.60%	24.79%	26.92%	33.27%

Comparing 4 clusters of different income groups, the percentages of *Job Rich* cluster grids and *Balanced* cluster grids increase from higher-income group (36.14%) to lower-income group (49.10%). This indicates that lower-income group has relative more balanced areas compared with the higher-income group. Higher-income jobs are mainly located in the central cities and thus the *Job Rich* area of higher-income group are more concentrated in the central cities. Lower-income jobs are distributed more broadly and thus the *Job Rich* grids of lower-income group cover more area.

5.2. Aggregated patterns of employment self-containment

Two aggregate patterns/clusters of ESC were recognized: *High Matched* and *Low Matched* (Fig. 9). The boxplots in the chart illustrate the distribution of ESC in each size of buffers, and the lines represent the average of multi-scale ESC for the identified clusters (*High Matched and Low Matched*). Revealed by the boxplots, the median of ESC gradually increases from 0.0 to 0.18 as the size of buffer increases, indicating the spatial dependency issue in evaluating the ESC. When the size of measuring units is small, most of the workers commute out. As a result, the ESC is small. When the size of spatial units increases, more percentage of workers work within the measuring units, and thus ESC increases. The result echoes with previous studies (Kelobonye et al., 2019; Zhou & Yeh, 2021). Examining the lines in the chart, the ESCs of both clusters are roughly same in the small size of spatial units but show increasing disparity when the size of spatial units increases. The *High Matched* cluster gradually increases till 0.38 but the *Low Matched* cluster keeps a relatively low values for all sizes of spatial units. The increase of ESC implies more internal journey-to-work trips, but the share of the internal trips of *High Matched* area is way higher than that in *Low Matched* area. Comparing the ESC of *High Matched* and *Low Matched* cluster at varying sizes of spatial units, the result shows that residential workers living in the *High Matched* cluster have more job opportunities and tend to work in this residential area, indicating the potential for a less commuting distance (Biermann & Martinus, 2020). (See Fig. 10.)

The two patterns/clusters of ESC display distinct spatial distributions. The *High Matched* cluster grids are mainly located at the central cities of the metropolitan and some places in the peripheral urban areas, while the rest of the grids belongs to *Low Matched* cluster. The *High Matched* cluster has a relative high ESC, which means the residential workers living in this area have a relative better access to jobs and commute shorter to work. In contrast, most of residential workers living in the *Low Matched* areas need to travel outside of the residential area to work. The distributions of *High Matched* cluster and *Low Matched* cluster vary across different income groups. From higher-income group to the lower-income group, the percentage of *High Matched* cluster grids increases from 20.17% to 41.61% (Table 3), and the coverage of the grids

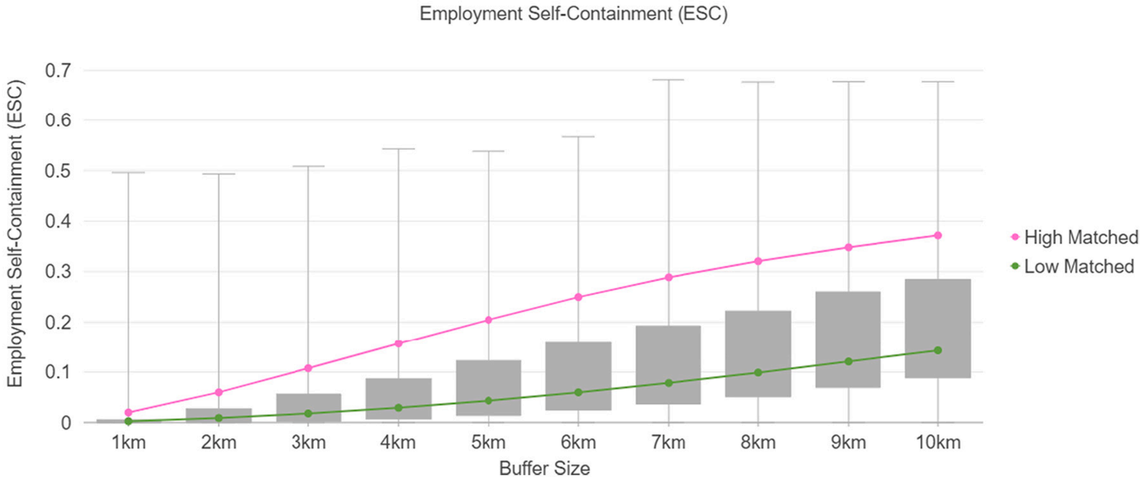

Fig. 9. The overall ESC of multiscale units.

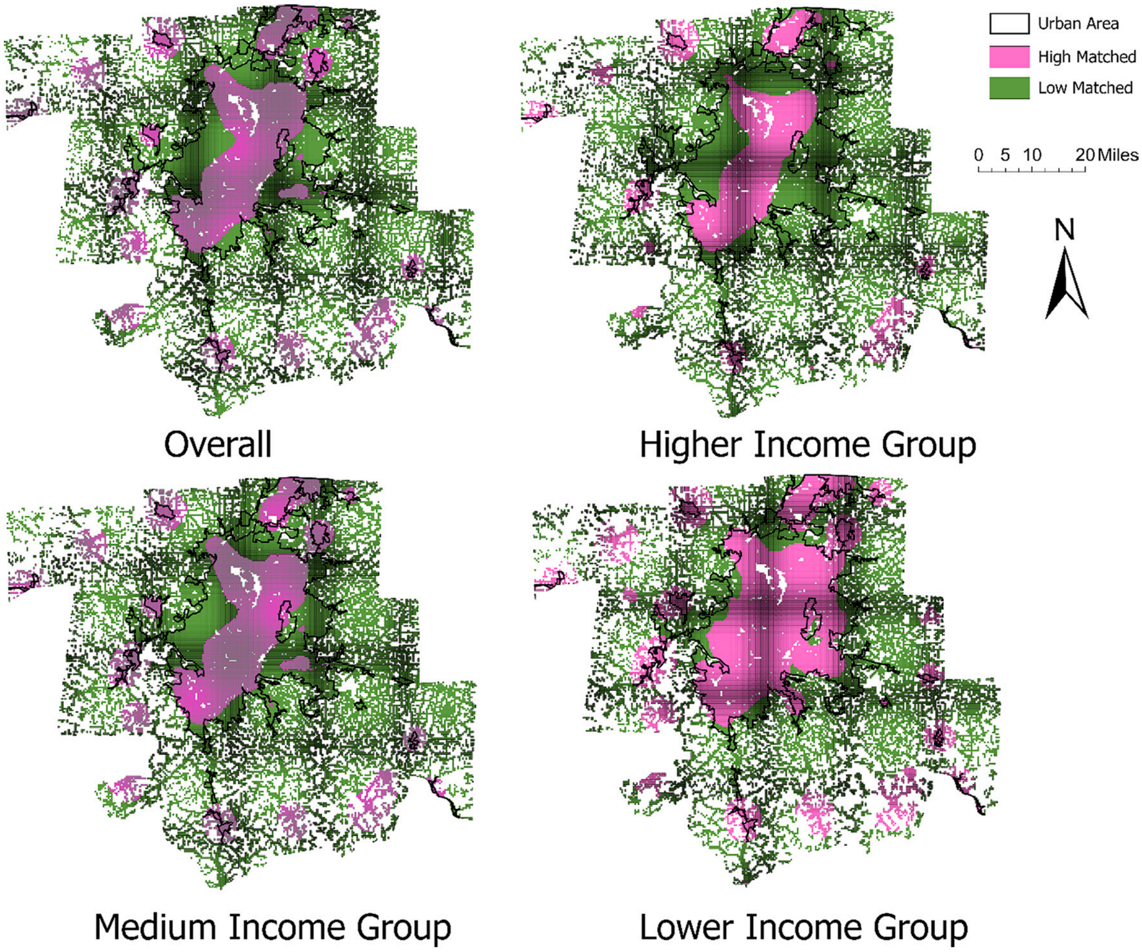

Fig. 10. The spatial distribution of ESC by different income groups.

Table 3
The percentage of different cluster by different income group for ESC.

ESC	Overall	Lower-income	Medium-income	Higher-income
High Matched	25.62%	41.61%	27.85%	20.17%
Low Matched	74.38%	58.39%	72.15%	79.83%

expands from the central cities to all urban areas (within the urban boundary). The lower-income group has more grids with higher ESCs, implying more percentage of workers of lower-income group commute a relative short distance.

5.3. Joint patterns of jobs-housing ratio and employment self-containment

We overlayed clusters of JWR and clusters of ESC and then identified

7 joint-patterns of JWR and ESC (Fig. 11). We did not choose to run a clustering process to identify the joint patterns because JWR and ESC are highly correlated (Figs. 12 and 13). It is recommended to conduct principal component analysis (PCA) before running the clustering (Vichi & Saporta, 2009). But it may lose information at each scale of units and have trouble explaining the patterns derived from the factors extracted from PCA. We chose to overlay the patterns of JWR and ESC to get joint patterns by different income groups.

Job Rich-High Matched cluster grids have more jobs than workers at varying scales, and most of the workers live and work within these grids. The more jobs mean more opportunities for residential workers, and it is reasonable these areas are classified as *High Matched* in terms of ESC. *Job Rich-High Matched* areas are mainly located at the central cities, where the residential have shorter commuting (Zhou, Yeh, Li, & Yue, 2018). *Job Rich-Low Matched* cluster grids are areas where there are more jobs than workers in all proposed scales. However, a substantial number of residential workers has trouble finding jobs and need to travel outside to work, implying a severe mismatch. The grids are mainly located at Indian Hills, Milford, and Monroe OH, where are residential locations for higher-income workers. Though there is a substantial number of jobs around them, but these jobs do not match these skills, and thus they need to travel longer to work. These areas are the worst mismatched in the metropolitan area, accounting for 3.39% of all grids.

Balanced-High Matched pattern is similar as the *Job Rich-High Matched* pattern. Though the areas do not have as many jobs as *Job Rich* area, workers living in the *Balanced* patterns can find jobs within a reasonable travel distance, and the ESC is relatively high. The *Balanced Rich-High Matched* area mainly encircles the *Job Rich-High Matched* area in central cities. *Balanced-Low Matched* pattern indicates though the jobs-housing is balanced, the actual jobs and workers are mismatched, meaning the residential workers need to travel outside of the units to work. Grids of this pattern are either adjacent to or encircle the *Job Rich-Low Matched* grids in suburb areas. The mismatch is mainly due to the higher-income workers cannot find matched jobs and instead need to travel long to work (Hu et al., 2017; Lin, Allan, & Cui, 2016; Stoker & Ewing, 2014). *Balanced-Low Matched* pattern is also severe mismatched area, making up 5.49% of all grids (Table 4).

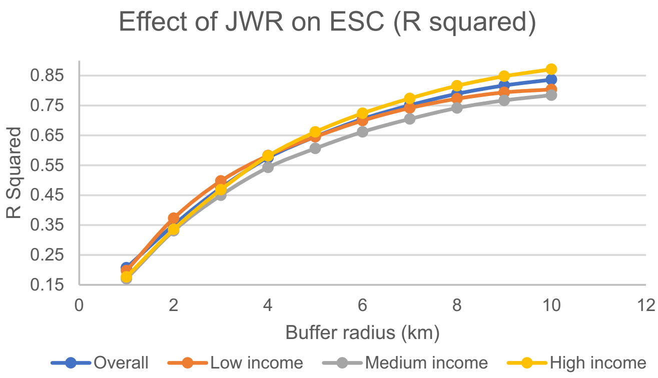

Fig. 12. The relationship between JWR and ESC.

Resident Rich-Suburb-High Matched pattern illustrates areas where there are more workers than jobs in the measuring spatial units (1 km −10 km buffers), and workers still find matched jobs within the area where they live. The grids are mainly located at Hamilton, and Middletown in Ohio and some small cities in the peripheral area of the metropolitan area. The grids in this pattern account for 4.86% of all grids (Table 4). *Resident Rich-Suburb-Low Matched* pattern describes the area where there are more workers than jobs, and a substantial number of workers travel long to work. These areas are mostly located at the suburb area. As the residential workers dominate these areas, it is

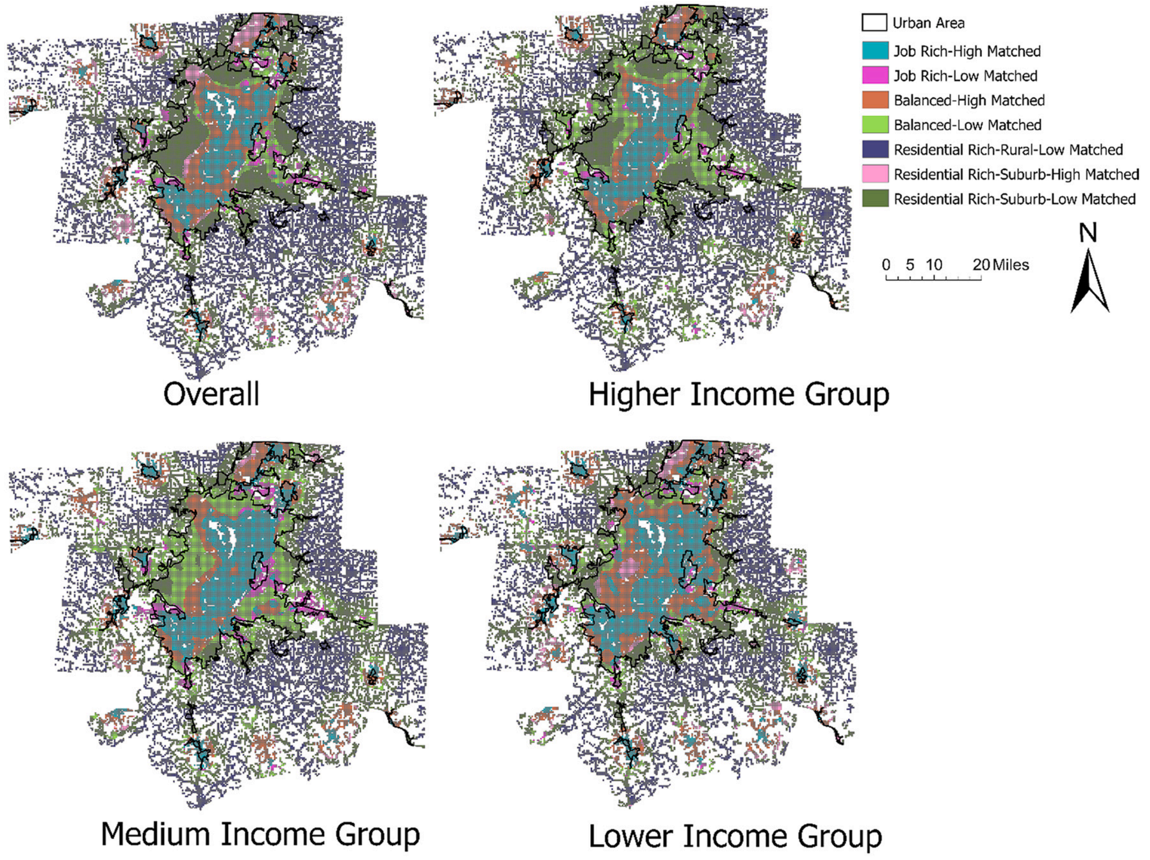

Fig. 11. The joint-patterns of JWR and ESC by different income groups.

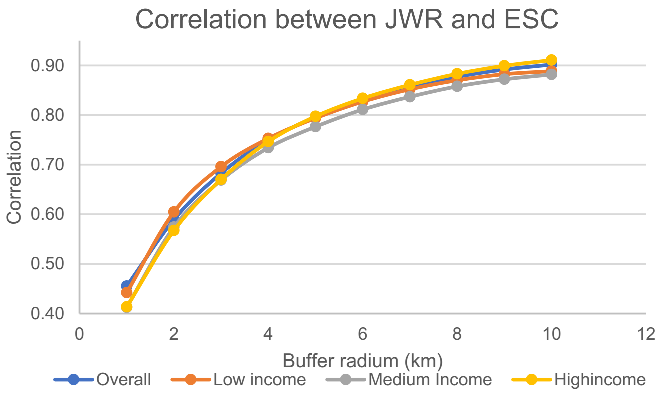

Fig. 13. The correlation between JWR and ESC.

Table 4
The percentage of different joint patterns by different income groups.

Joint Pattern	Overall	Lower-income	Medium-income	Higher-income
Job Rich-High Matched	10.54%	19.84%	15.66%	10.34%
Job Rich-Low Matched	3.39%	2.25%	4.64%	2.15%
Balanced-High Matched	10.18%	17.40%	11.07%	8.84%
Balanced-Low Matched	5.49%	7.28%	17.74%	14.80%
Residential Rich-Suburb-High Matched	4.86%	4.37%	1.12%	0.98%
Residential Rich-Suburb-Low Matched	27.93%	24.06%	22.86%	29.61%
Residential Rich-Rural-Low Matched	37.58%	24.79%	26.92%	33.27%

rational workers commute to job rich areas to work.

The last pattern is *Residential Rich-Rural Low Matched* pattern. Only a small number of jobs is distributed in the rural area compared with the number of residential workers. Workers need to travel out of the residential area to work, resulting in a relative low ESC. This pattern makes up most (37.58%) of the grids of the metropolitan area, covering almost all rural areas. *Resident Rich-Rural-High Matched* pattern is not identified in the joint patterns. It is understandable that in the rural area where the number of jobs is far less than the number of workers, workers need to travel out of the measured buffers.

The spatial distributions of the 7 patterns vary by different income groups, largely in patterns of *Job Rich-High Matched, Balanced-High* Matched, and *Balanced-Low* Matched. The percentage of *Job Rich-High Matched* grids increases from higher-income group (10.34%) to lower-income group (19.84%). Besides the increase in percentage, the distribution of *Job Rich-High Matched* grids changes from concentrated pattern in the central cities to a fragmented and leapfrog grid pattern. This is consistent with the JWR pattern of *Job Rich* cluster. The lower-income group has the most *Balanced-High Matched* grids among all three income groups, implying the lower-income workers have relative better matched jobs-housing and ESC. In contrast, the medium-income group has the most *Balanced-Low Matched* grids that distributes in suburb areas around central cities. Suburb areas are where most of residential workers reside. The medium-income workers living in suburb areas have a relative worse matched jobs-housing and ESC.

6. Discussion

Taking the Cincinnati Metropolitan area as a case study, this research aimed to examine the jobs-housing balance and ESC in multiscale units by income groups and explored the aggregate patterns of jobs-housing balance and ESC using LEHD data (2016). Using the Monte Carlo simulated individual data, we applied K-medoids to classify 4 aggregate patterns of jobs-housing balance and 2 aggregate patterns of ESC and resulted in 7 joint-patterns of jobs-housing balance and ESC by each income group. The patterns reveal the jobs-housing relationship and ESC and unveil spatial distribution of the jobs-housing matched and mismatched areas by different income groups. Comparing with previous studies, our study is improved in three ways. First, the aggregate patterns of jobs-housing relationship incorporated more information than previous literature. Previous studies analyzed JWR and ESC at a certain scale and analyzed the impact on ESC at each scale discretely (Zhou & Yeh, 2021; Zhou, Yeh, Li, & Yue, 2018). Our study analyzed aggregate patterns of jobs-housing relationship considering various incremental scales (1 km – 10 km searching radius) using K medoid clustering method. Instead of finding the 'perfect' searching radius or scale (Zhou & Yeh, 2021), we explored patterns across different scales and identify the overall patterns by integrating all proposed scales. Second, our study reveals the limitation of only considering one scale when analyzing jobs-housing relationships. The result shows the impact of JWR on ESC varies as the scale changes. The association increases from around 0.17 to 0.85 (Fig. 13). Focusing on one scale will neglect the variation on other scales and unveil partial information. Third, our study illustrates the differing impact of JWR on ESC by higher-income, medium-income, and lower-income groups. Our result confirmed the positive effect of JWR on ESC (Blumenberg & King, 2021; Ta, Chai, Zhang, & Sun, 2017; Zhou, Yeh, & Yue, 2018), but further discloses differences of that on income groups. We found that within a 4 km searching radius, JWR affected lower-income group more than that of higher-income group (Fig. 12). Whereas beyond the 4 km searching radius, JWR affected higher-income group more. JWR has least impact on medium-income group among the three groups.

The 4 aggregate patterns of JWR are *Job Rich, Balanced, Resident Rich-Suburb, and Residential Rich-Rural* clusters. The JWR of the 4 patterns varies across all the sizes of spatial units and forms distinct spatial distributions across the metropolitan area. The JWR curves of *Job Rich* and *Balanced* have turning points around radius from 5 km to 6 km (3–3.5 miles Euclidean distance), after which the JWR of these two patterns barely changes. The radius around the turning points corresponds to the suitable measuring buffer size in previous studies (Cervero, 1989; Deakin, 1989.; Peng, 1997; Wang, 2000; Stoker & Ewing, 2014). We explored the patterns of JWR at 6 km radius buffers and compared the result (Fig. 14) with the aggregate patterns of multi-scale JWR (Fig. 8. Overall). *Job Rich* (19.82%) and *Balanced* (31.50%) areas (Fig. 14) account for over 50% of all grids, which presents a more balanced jobs-housing than patterns of clusters formed by considering JWR at varying spatial units (29.59%, see Fig. 8 and Table 2 for more details). The difference in the patterns indicates the limitation of only considering the JWR measured within one scale. With regards to the aggregate patterns by different income groups, most of the *Job Rich* and *Balanced* area are concentrated in central cities for higher-income group, while there are more *Job Rich* and *Balanced* areas both in central cities and peripheral cities for lower-income group. This result is consistent with pervious study (Schleith et al., 2016) which revealed the lower-income group had relative better jobs-housing balance.

The 2 aggregate patterns of ESC are *High Matched* and *Low Matched*, and each pattern has varying values of ESC across the sizes of the spatial units. The values of ESC within small sizes of buffers (1 km and 2 km) are almost 0, implying workers barely work within 2 km radius buffers. This indicates the small sizes of spatial unit may have limitations in evaluating ESC in Cincinnati. Zhou and Yeh (2021) applied 2 km grid as a suitable unit to evaluate ESC in a case study of Shenzhen. Though 2 km grid is a relatively small size of spatial unit to measure ESC for Cincinnati, it may be a suitable unit to measure Shenzhen. The discrepancy may be due to the difference in travel mode, job density, and residential density of the two cities. Zhou and Yeh (2021) also pointed out that the hypothetical spatial analysis unit of the turning point of ESC in U.S. cities may be larger than that of Chinese cities, and our result echoed with their conclusion. The ESC increases as the size of the spatial units

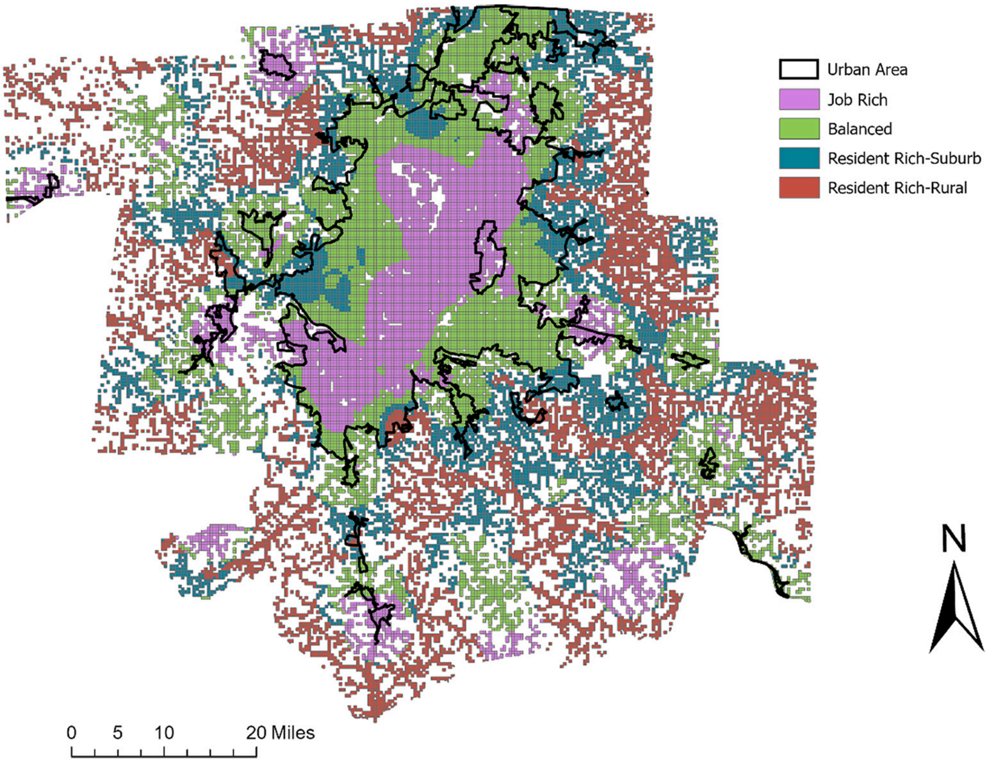

Fig. 14. The pattern of JWR at 6 km radius buffers.

increases, but there is no apparent turning point when the radius increases from 1 km to 10 km, indicating there may not be a 'suitable' scale to measure ESC in Cincinnati. Only considering ESC on one scale may overlook information at other scales and aggregate patterns incorporating ESC at all scales are needed. Looking at ESC patterns (*High Matched* and *Low Matched*) by different income groups, the lower-income group has a highest percentage of *High Matched* grids, suggesting the lower-income workers are more likely to have jobs close to them, which is consistent with previous studies(Hu & Schneider, 2017). However, the inclinations of having jobs nearby for lower-income workers may not imply they choose to find these low-paying jobs, but rather be limited by the inadequate transportation access, either lack of private cars or public transit service. In Cincinnati, the public transit service only covers the core area of the metro (such as Hamilton County), while most of suburban area are out of service, specifically *Resident Rich-Suburb, and Residential Rich-Rural* where jobs are deficient. Some lower-income workers, such as workers in agriculture and mining, may choose to work locally due to the industry nature, but some lower-income workers may have to choose jobs nearby due to the unaffordability of commuting cost and limit of access to transit.

The joint-patterns of JWR and ESC reveal the relationship between jobs-housing balance and ESC and possible commuting issues in Cincinnati. *Job Rich-High Matched*, *Balanced-High Matched*, and *Resident Rich-High Matched* are the matched areas, accounting for 25.58% of the metropolitan areas. The severe mismatched areas are *Job Rich-Low Matched* and *Balanced-Low Matched* areas, accounting for 7.34% of the metropolitan area. Most of the severe mismatched areas are in the suburban areas that surround central cities. Though there are relative high job opportunities within a reasonable commuting distance, residential workers need to travel long to work. Analyzing the distribution of these areas, we found most of the residential workers in the sever mismatched area are higher-income earners, who prefer the favorable environment in the suburb, but need to work in central cities (Hu et al., 2017). The joint-patterns of JWR and ESC by different income groups further illustrate this preference. The severe mismatched areas of higher-income group make up 16.95%, while those of lower-income group makes up around 9.53%.

This study has limitations for future research. First, the patterns of clusters revealed the relative distribution of jobs-housing balance and ESC for each income group but could not provide the absolute comparison across different income groups. The clustering method mostly depends on the input data, and thus the change of data would result in different absolute values and variance of each cluster. Though grids from higher-income group and lower-income group were both classified as *Job Rich* cluster, the absolute values and range of these groups were not same. Second, we applied K-medoids clustering method to explore the aggregated patterns of JWR and ESC but did not compare with other methods in depth. We did evaluate the performance of K-medoids and K means through Davies-Bouldin's index (a lower value means a better performance) (Martino, Rizzi, & Mascioli, 2018). The K-medoids clustering (1.26) outperformed the K means clustering (1.28) for overall JWR with four clusters, and the K-medoids clustering (0.21) also outperformed the K means clustering (0.22) for overall ESC with two clusters. But we did not compare the clusters classified by these two methods as it is not the focus of this research.

7. Conclusion

This study examined and visualized the aggregate patterns of jobs-housing balance and ESC by different income groups within multiscale units and identified the jobs-housing matched areas and mismatched areas. The result shows though the balanced jobs-housing is more likely to associated with a high ESC, there are severe mismatched areas in the metropolitan area which may lead to longer commuting distance and severe peak-hour congestion. Higher-income group has relatively more jobs mismatch areas than that of lower-income group especially in suburban areas. Jobs-housing balance affects lower-income group more when the searching radius <4 km, while has more effect on higher-income group when searching radius is beyond 4 km. Our result can provide insights to planners and policy makers to understand the jobs-housing relationship and its spatial variation of income groups across the metropolitan area. If jobs-housing policies focus on increasing local jobs to address congestion and commuting issues, it needs to pay attention to those mismatch areas and locate jobs that correspond to the needs of different income groups, such as locating low-paying jobs within 4 km of lower-income workers. Though lower-income group has relatively less mismatch areas, considering the lack of transportation access, such as in adequate public transit service in suburban areas, and the relative less options of residential choices, the mismatch area may affect lower-income workers more than higher-income workers. Policies that increase the matched jobs for lower-income workers and the access to transportation help reduce inequality of job access.

CRediT authorship contribution statement

Zhiyuan Yao: Conceptualization, Methodology, Writing – original draft. **Changjoo Kim:** Conceptualization, Writing – review & editing.

References

Abowd, J. J., Haltiwanger, J., & Lane, J. (2004). Integrated longitudinal employer-employee data for the United States. *American Economic Review, 94*(2), 224–229.

Antipova, A., Wang, F., & Wilmot, C. (2011). Urban land uses, socio-demographic attributes and commuting: A multilevel modeling approach. *Applied Geography, 31*(3), 1010–1018.

Biermann, S., & Martinus, K. (2020). Reducing the need to travel: the challenge of employment self-containment. In *Handbook of Sustainable Transport* (pp. 248–257). Edward Elgar Publishing.

Blumenberg, E., & King, H. (2021). Jobs–housing balance re-re-visited. *Journal of the American Planning Association, 87*(4), 484–496.

Cervero, R. (1989). Jobs-housing balancing and regional mobility. *Journal of the American Planning Association, 55*(2), 136–150.

Cervero, R. (1996). Jobs-housing balance revisited: Trends and impacts in the San Francisco Bay Area. *Journal of the American Planning Association, 62*(4), 492–511.

Curtis, C., & Olaru, D. (2010). The relevance of traditional town planning concepts for travel minimization. *Planning Practice and Research, 25*(1), 49–75.

Deakin, E. (1989). Land use and transportation planning in response to congestion problems: A review and critique (Vol. 1237, no. 1237-1244, p. 77). In *Transportation Research Board, Commission on Sociotechnical Systems, National Research Council, National Academy of Sciences*.

Debenham, J., Stillwell, J., & Clarke, G. (2003). The estimation of self-containment and catchment size indicators for use in small area classification. *International Journal of Population Geography, 9*(3), 253–271.

Giuliano, G. (1991). *Is jobs-housing balance a transportation issue?* University of California Transportation Center.

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

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

Horner, M. W., & Schleith, D. (2012). Analyzing temporal changes in land-use–transportation relationships: A LEHD-based approach. *Applied Geography, 35*(1–2), 491–498.

Hu, L., & Schneider, R. J. (2017). Different ways to get to the same workplace: How does workplace location relate to commuting by different income groups? *Transport Policy, 59*, 106–115.

Hu, Y., & Wang, F. (2015). Decomposing excess commuting: A Monte Carlo simulation approach. *Journal of Transport Geography, 44*, 43–52.

Hu, Y., & Wang, F. (2016). Temporal trends of intraurban commuting in Baton Rouge, 1990–2010. *Annals of the American Association of Geographers, 106*(2), 470–479.

Hu, Y., Wang, F., & Wilmot, C. G. (2017). Commuting variability by wage groups in Baton Rouge, 1990–2010. *Papers in Applied Geography, 3*(1), 14–29.

Huang, J., Hu, Y., Wang, J., & Li, X. (2021). A tale of two cities: Jobs–housing balance and urban spatial structures from the perspective of transit commuters. *Environment and Planning B: Urban Analytics and City Science, 48*(6), 1543–1557.

Kaufman, L., & Rousseeuw, P. J. (1990). Partitioning around medoids (program pam). In *, 344. Finding Groups in Data: An Introduction to Cluster Analysis* (pp. 68–125).

Kelobonye, K., Mao, F., Xia, J. C., Swapan, M. S. H., & McCarney, G. (2019). The impact of employment self-sufficiency measures on commuting time: Case study of Perth, Australia. *Sustainability, 11*(5), 1488.

Levinson, D. M., & Kumar, A. (1994). The rational locator: Why travel times have remained stable. *Journal of the American Planning Association, 60*(3), 319–332.

Lin, D., Allan, A., & Cui, J. (2016). Exploring differences in commuting behaviour among various income groups during polycentric urban development in China: New evidence and its implications. *Sustainability, 8*(11), 1188.

Ma, X., Liu, C., Wen, H., Wang, Y., & Wu, Y. J. (2017). Understanding commuting patterns using transit smart card data. *Journal of Transport Geography, 58*, 135–145.

Martino, A., Rizzi, A., & Mascioli, F. M. F. (2018, July). Distance matrix pre-caching and distributed computation of internal validation indices in k-medoids clustering. In *2018 International Joint Conference on Neural Networks (IJCNN)* (pp. 1–8). IEEE.

Martinus, K., & Biermann, S. (2018). Strategic planning for employment self-containment in metropolitan sub-regions. *Urban Policy and Research, 36*(1), 35–47.

Niedzielski, M. A., Horner, M. W., & Xiao, N. (2013). Analyzing scale independence in jobs-housing and commute efficiency metrics. *Transportation Research Part A: Policy and Practice, 58*, 129–143.

Office of Management and Budget. (2010). 2010 standards for delineating metropolitan and micropolitan statistical areas. *Federal Register, 75*(123), 37246–37252.

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

Schleith, D., & Horner, M. W. (2014). Commuting, job clusters, and travel burdens: Analysis of spatially and socioeconomically disaggregated longitudinal employer–household dynamics data. *Transportation Research Record, 2452*(1), 19–27.

Schleith, D., Widener, M., & Kim, C. (2016). An examination of the jobs-housing balance of different categories of workers across 26 metropolitan regions. *Journal of Transport Geography, 57*, 145–160.

Small, K. A., & Song, S. (1994). Population and employment densities: structure and change. *Journal of urban economics, 36*(3), 292–313.

Spear, B. D. (2011). *NCHRP 08–36, Task 098 improving employment data for transportation planning.* American Association of State Highway and Transportation Officials.

Stevens, D. W. (2002). *Employment that is not covered by state unemployment (No. 2002–16)*. US Census Bureau: Center for Economic Studies.

Stoker, P., & Ewing, R. (2014). Job–worker balance and income match in the United States. *Housing Policy Debate, 24*(2), 485–497.

Suárez, M., Murata, M., & Delgado Campos, J. (2016). Why do the poor travel less? Urban structure, commuting and economic informality in Mexico City. *Urban Studies, 53*(12), 2548–2566.

Sultana, S. (2002). Job/housing imbalance and commuting time in the Atlanta metropolitan area: Exploration of causes of longer commuting time. *Urban Geography, 23*(8), 728–749.

Ta, N., Chai, Y., Zhang, Y., & Sun, D. (2017). Understanding job-housing relationship and commuting pattern in Chinese cities: Past, present, and future. *Transportation Research Part D: Transport and Environment, 52*, 562–573.

Thomas, R. (1969). *London's new towns: A study of self-contained and balanced communities*. London: PEP.

Vichi, M., & Saporta, G. (2009). Clustering and disjoint principal component analysis. *Computational Statistics & Data Analysis, 53*(8), 3194–3208.

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

Wang, F. (2000). Modeling commuting patterns in Chicago in a GIS environment: A job accessibility perspective. *The Professional Geographer, 52*(1), 120–133.

Yan, L., Wang, D., Zhang, S., & Xie, D. (2019). 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.

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

Zhao, P., Lü, B., & De Roo, G. (2011). Impact of the jobs-housing balance on urban commuting in Beijing in the transformation era. *Journal of Transport Geography, 19*(1), 59–69.

Zheng, M., Liu, F., Guo, X., & Lei, X. (2019). Assessing the distribution of commuting trips and jobs-housing balance using smart card data: A case study of Nanjing, China. *Sustainability, 11*(19), 5346.

Zhou, X., & Yeh, A. G. (2021). Understanding the modifiable areal unit problem and identifying appropriate spatial unit in jobs–housing balance and employment self-containment using big data. *Transportation, 48*(3), 1267–1283.

Zhou, X., Yeh, A. G., Li, W., & Yue, Y. (2018). A commuting spectrum analysis of the jobs–housing balance and self-containment of employment with mobile phone location big data. *Environment and Planning B: Urban Analytics and City Science, 45*(3), 434–451.

Zhou, X., Yeh, A. G., & Yue, Y. (2018). Spatial variation of self-containment and jobs-housing balance in Shenzhen using cellphone big data. *Journal of Transport Geography, 68*, 102–108.