

Article

# Exploring Urban Spatial Feature with Dasymetric Mapping Based on Mobile Phone Data and LUR-2SFCAe Method

Lingbo Liu <sup>1</sup> , Zhenghong Peng <sup>2</sup>, Hao Wu <sup>2,\*</sup> , Hongzan Jiao <sup>2</sup> and Yang Yu <sup>1</sup> 

<sup>1</sup> Department of Urban Planning, School of Urban Design, Wuhan University, Wuhan 430072, China; lingbo.liu@whu.edu.cn (L.L.); yuyang1@whu.edu.cn (Y.Y.)

<sup>2</sup> Department of Graphics and Digital Technology, School of Urban Design, Wuhan University, Wuhan 430072, China; pengzhenghong@whu.edu.cn (Z.P.); Jiaohongzan@whu.edu.cn (H.J.)

\* Correspondence: wh79@whu.edu.cn; Tel.: +86-27-6877-3062

Received: 22 May 2018; Accepted: 10 July 2018; Published: 12 July 2018



**Abstract:** Dasymetric mapping of high-resolution population facilitates the exploration of urban spatial feature. While most relevant studies are still challenged by weak spatial heterogeneity of ancillary data and quality of traditional census data, usually outdated, costly and inaccurate, this paper focuses on mobile phone data, which can be real-time and precise, and also strengthens spatial heterogeneity by its massive mobile phone base stations. However, user population recorded by mobile phone base stations have no fixed spatial boundary, and base stations often disperse in extremely uneven spatial distribution, this study defines a distance-decay supply–demand relation between mobile phone user population of gridded base station and its surrounding land patches, and outlines a dasymetric mapping method integrating two-step floating catchment area method (2SFCAe) and land use regression (LUR). The results indicate that LUR-2SFCAe method shows a high fitness of regression, provides population mapping at a finer scale and helps identify urban centrality and employment subcenters with detailed worktime and non-worktime populations. The work involving studies of dasymetric mapping based on LUR-2SFCAe method and mobile phone data proves to be encouraging, sheds light on the relationship between mobile phone users and nearby land use, brings about an integrated exploration of 2SFCAe in LUR with distance-decay effect and enhances spatial heterogeneity.

**Keywords:** dasymetric mapping; urban spatial feature; land use regression (LUR); two-step floating catchment area (2SFCA); mobile phone data; urban centrality

## 1. Background

Population mapping at a finer scale and higher resolution can play an important role in understanding urban spatial features, especially in the measurement of urban centrality [1,2] and identification of employment centers [3–7]. Therein, dasymetric mapping is considered as an effective method [8], which helps allocate population data to finer spatial units with ancillary data.

Progress has been made in studies on dasymetric mapping with the employment of various methods such as Weighted areal interpolation [9], Binary filtered areal weighting [10], Three-class and limiting variable [11,12], and Image texture [13]. However, two main challenges are found in previous literature on dasymetric mapping: (1) the ignorance of spatial heterogeneity inside every geographic unit, which may contain inhabitable and uninhabitable parts, or parts of different population density; and (2) the quality of source population data.

Most studies try to address spatial heterogeneity by fining geographic scale of ancillary data, ranging from Land use/land cover data (LULC) [14], soil sealing degree [15], nighttime lights, transportation network, elevation, and slope data extracted from satellite maps [16] to data with more classifications, such as cadastral information [17], tax parcel [18], buildings [19], Points of Interests (POIs) [20], and Volunteer Geographic Information (VGI) [21]. Although such methods have proven effective in improving the resolution of population mapping by enhancing spatial variance with increased amount of data categories, the accuracy of ancillary data still affects the results. Moreover, the census data used in most studies usually lag behind in timeliness, and are costly and inaccurate, which could jeopardize the result of population distribution as well.

Meanwhile, statistical linear regression has often been employed to examine the correlation result between population density and ancillary data to verify the result of dasymetric mapping [17,22]. Especially, in many related air pollution studies [23–25], dasymetric mapping for regression between land use and observed air quality data, known as land use regression (LUR), is used to identify the connection between variables and the associated land uses.

This article proposes that taking mobile phone data as a source of population distribution would highly improve the accuracy of dasymetric mapping. With the rapid development of Information and Communication Technology (ICT), mobile phone data have become an important source for studies of population distribution and behaviors of urban residents, such as identifying the commuting of residents [26,27]. Even so, few studies take mobile phone data as a solution to address those two problems, even though they provide relatively accurate user population and precise spatial location. Meanwhile, the massive amount of mobile phone base stations with corresponding data of user distribution all over the cities could offer information of spatial heterogeneity which has always been lacking in previous dasymetric mapping methods.

However, compared with the clear boundary of traditional census data, the spatial distribution of population data associated with mobile phone base stations has no stable boundary of service area. Generating Thiessen polygons with base stations could help define a border, but the highly uneven spatial distribution of base stations would bring more uncertainty into the mapping result.

The paper further presents an integrated methodology of LUR-2SFCAe with two-step floating catchment area (2SFCA) and LUR to address the problem. As the mobile phone user population recorded by base stations are related to surrounding land use distribution and influenced by the distance-decay pattern [28], the mathematical relation between user population and neighbor land uses is similar to the classic distance-decay demand–supply model in accessibility studies on public service, wherein the 2SFCA method has proven effective. That is to say, the population of nearby land is determined by not only user population of nearby base stations and various land uses, but also the distances to them.

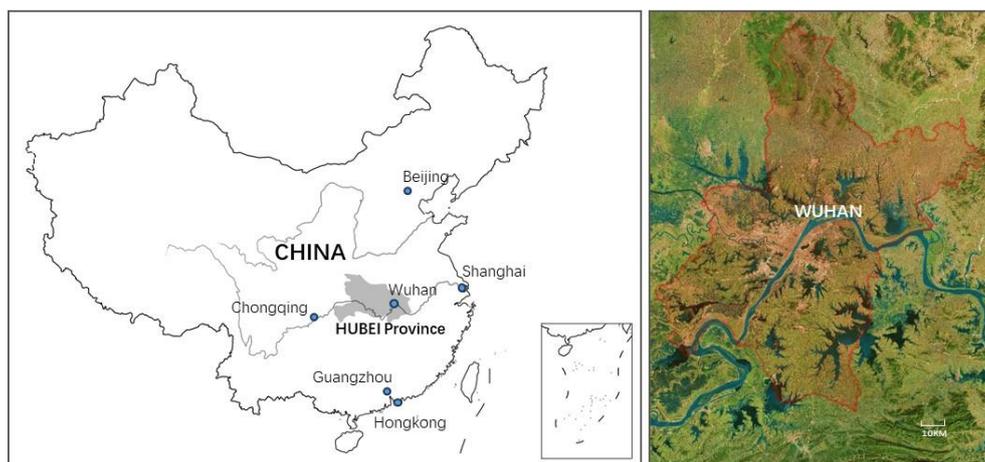
This paper begins with an assumption that the original mobile phone data and land use data would result in unacceptable correlation due to an extremely uneven distribution of base stations which may locate in close proximity but with largely different user population. Therefore, a 1 km grid and its centroids are applied to statistically transform original data into gridded mobile phone data, which show relatively high correlation with land use data by the LUR-2SFCAe method. With a population mapping during work time and non-worktime, the spatial feature of Wuhan is explored.

The result of this study indicates that LUR-2SFCAe can be utilized in mapping population of high resolution with land use data and mobile phone data in forms of grids. It can also be used to identify spatial feature of population at different time. The methodology of LUR-2SFCAe in the present article is also applicable to related LUR studies.

## 2. Data Preparation

### 2.1. Study Area: Wuhan, China

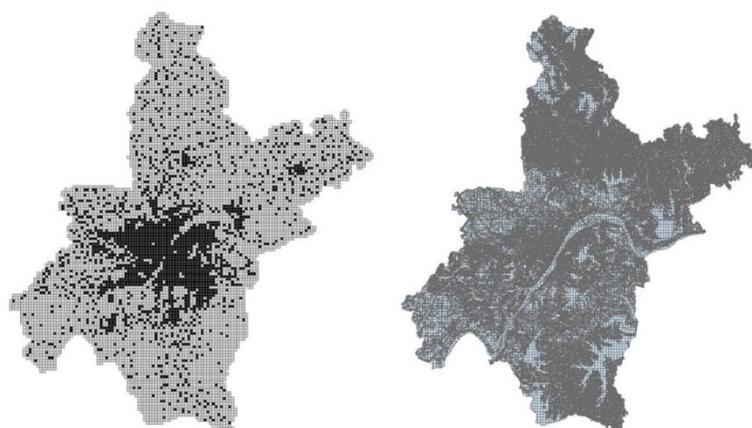
The case study area is Wuhan city, located in central China (Figure 1). As the capital of Hubei Province and one of the nine National Central Cities of China, Wuhan is the most populous city in Central China. The city boasts abundant mountain and water resources and is divided into “Three Towns” as Wuchang, Hankow, and Hanyang by the Yangtze River and Han River. The complicated geographic condition demands dasymetric mapping rather than simple interpolating tools such as Kriging, Kernel density or Voronoi polygons for population mapping.



**Figure 1.** Location of Wuhan, Hubei Province in China.

### 2.2. Data and Preprocessing

The present study utilizes two kinds of data in Wuhan metropolitan area: mobile phone data as the source population and existing land use as ancillary data; both were overlaid on the 1 km grids (Figure 2). The gridded data not only reflect the output resolution of population mapping but also improve the efficiency of model computing. Most importantly, the gridded mobile phone data can reduce the negative influence of uneven spatial distribution of original base stations in regression.



**Figure 2.** The 1 km × 1 km grid applied for mobile phone base stations and land use data.

Land use data for 2013, provided by Wuhan Planning Bureau, were sorted in codes as A (Administration and public services), B (Commercial and business facilities), R (Residential),

MW (Industrial, logistics and warehouse), S (Road, street and transportation), U (Municipal utilities), G (Green space and square) in urban built area, H (Development land, including town and country land as H1, regional infrastructure as H2, etc.) and E (Non-development land, including water as E1, farmland as E2, etc.) in the regional background.

After the grid was applied, each gridded land parcel and its centroid were assigned with the various types land uses.

Data of phone call records used in the present study were provided by a partner telecommunication operator whose market share is about 60%, verified for representing whole population distribution proportionally in Wuhan [29]. Mobile phone data of 7,300,000 users in November 2015 in Wuhan City were used in the study. As most urban studies of mobile phone data have discussed, the sample data can represent population distribution on a global scale, while users of other service providers and multiple data for the same user area have not been considered in the macro scope of whole city. Data were pre-processed, eliminating all privacy-related information. The basic format is a multi-field table tagged with the user ID. Data from the busiest base stations during one month were categorized into three time periods: work-time (from 7:00 a.m. to 7:00 p.m., Monday to Friday), and on-worktime (7:00 p.m. to 7:00 a.m., Monday to Friday; Saturdays and Sundays), and all-time.

The distance based on the length statistics of TIN of all base stations in ArcGIS (Figure 3) shows that the minimum distance is 0.095877 m, the average value is 726.931858 m, and the standard deviation is 1873.130695 m. The spatial distribution of base station is extremely uneven, wherein base stations located nearby but with different user population would create significant error for land use regression.

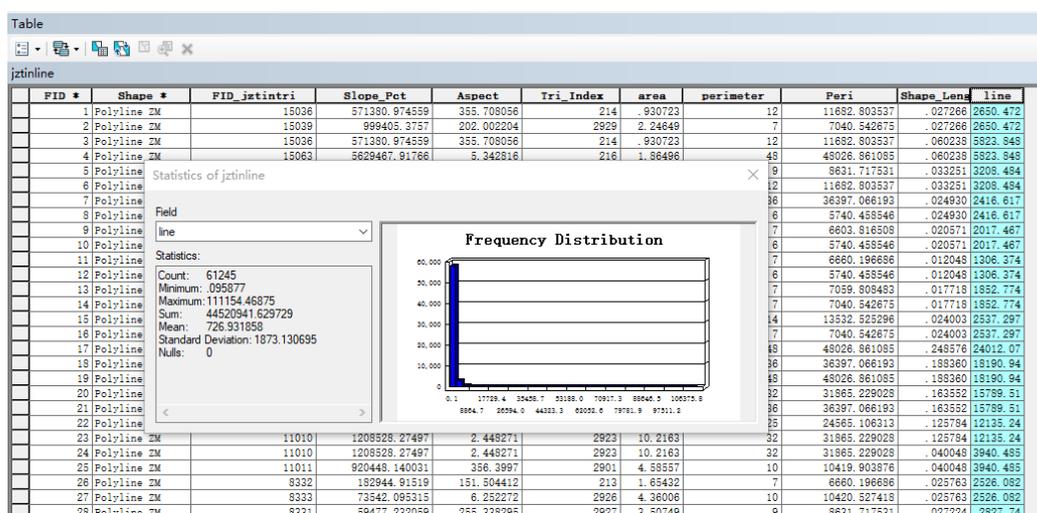


Figure 3. Statistics of the TIN length of mobile phone base stations.

Therefore, a 1 km grid was also applied to allocate the mobile phone data, generating a simulative grid of mobile phone data with centroids, evenly distributed spatially. The original 33,587 points of base stations are aggregated into 2235 points, which also could help accelerate the calculation of the model.

### 3. Methods

The major workflow of the study is presented in Figure 4. The major steps are: (1) Create the 1 km × 1 Km grid and centroids in study area to represent users population of mobile phone base stations inside and the land area of various kinds of land use. (2) Apply the 2SFCA method to assign grid points with land use value. (3) Implement LUR with population and land use value to verify the regression fit by substituting coefficients to verification data. (4) Substitute coefficients back to grid

points with land use variables to acquire the population distribution in worktime and non-worktime, based on which urban spatial structure was explored. The main procedure is the 2SFCAe model.

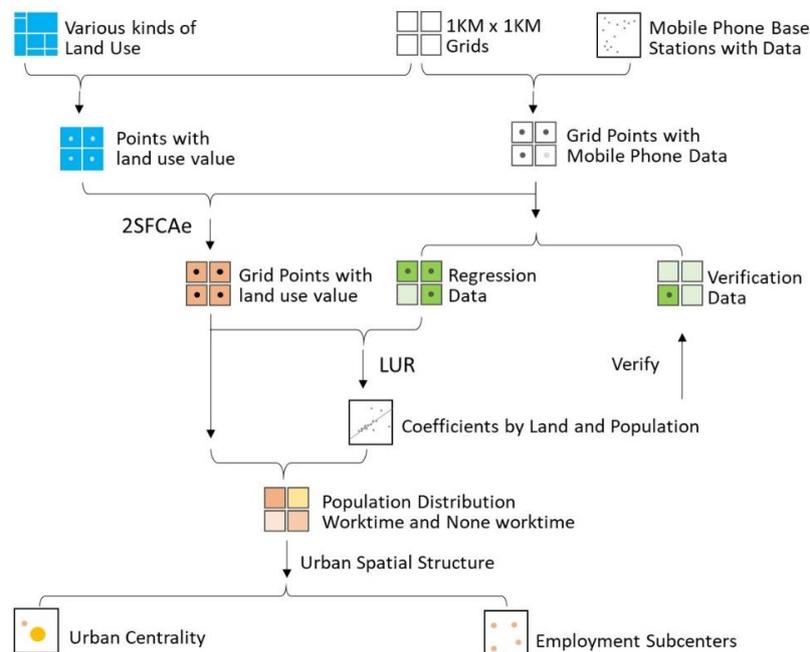


Figure 4. Workflow of LUR-2SFCAe.

### 3.1. Land Use Regression Model

Land use regression model is a multi-variable linear regression model, prevalently utilized for urban air pollution modeling with four elements: observed data, geographic predictors, model development and validation [24], offering pollution mapping of higher resolution compared to traditional interpolation methods. With a connection and regression between geographic elements and part of monitoring data, LUR sets a linear equation with coefficients for corresponding geographic variables, verifies equations with the rest monitoring data by substituting coefficients back, and then assigns the value to space around the observation spots.

People commonly rest in residential areas, and work in offices, commercial and industrial zones; land uses surrounding base station directly relate to the size of mobile phone user's population. LUR model can interpolate the size of mobile phone population to the surrounding land uses, which has the same equation as the traditional form in air pollution studies:

$$Y_i = k_0 + k_1x_1 + k_2x_2 + \dots \dots + k_nx_n + \varepsilon_i \quad (1)$$

where  $Y_i$  is the population number provided in mobile phone data of base station  $i$ ,  $K_0$  is a constant as intercept,  $x_n$  is the value of type  $n$  land use surrounding base station  $i$  with a corresponding coefficient  $K_n$ , and  $\varepsilon_i$  is the random error.

In contrast,  $Y_i$  is a summarized value of adjacent area other than density value applied in LUR models of air pollution, constituting a similar demand–supply relation between population value of base station and nearby lands, which make 2SFCA method employed in accessibility studies appropriate for combination with LUR.

### 3.2. Two-Step Floating Catchment Area with Entropy Gravity Method

As the prototype of 2SFCA, Floating Catchment Area method (FCA) merely considers land use effects around observation point as most of traditional LUR models have done. Both models consider

the distance threshold and decay effect, yet ignore spatial heterogeneity as base stations with the same land use pattern may differentiate in population for their different spatial location such as CBD and subcenters. 2SFCA takes land influence into account with a weighted approach of population as a reflection of spatial heterogeneity, which means that the larger the population of the base station is, the larger that of surrounding land is as well, though the corresponding land value may be the same.

The basic ideas underlying 2SFCA is that the supply and demand points are used as centers to perform floating catchment searches, respectively. For the first search, public facility ( $j$ ) is used as the center point, all settlements ( $k$ ) within a threshold distance ( $d_0$ ) are searched. Thus, the ratio ( $R_j$ ) between the service capacity of the facility and the served population within the corresponding area is calculated. For the second search, each settlement ( $i$ ) is used as the center point, all locations of public facilities within a threshold distance ( $d_0$ ) are searched. The aggregated services ( $R_j$ ) from various public facilities are summarized to acquire the service capacity of a public facility (accessibility) at point ( $i$ ).

Despite its relative popularity, the 2SFCA method still has its limitation for its dichotomy, determining accessibility value merely according to a fixed threshold distance (or time). Some studies have constructed a distance decay effect model using a kernel density function, gravity model or Gaussian function. As a summarization to which, Wang presented G2SFCA model with an integrated equation about decay function  $f(d)$  and illustrated the different pattern as follow: (a) Gravity function; (b) Gaussian function; (c) 2SFCA; (d) E2SFCA; (e) Kernel; and (f) three-zone hybrid approach [30].

$$A_i = \sum_{j=1}^n \left[ S_j f(d_{ij}) / \sum_{k=1}^m P_k f(d_{kj}) \right] \tag{2}$$

As new physics holds that gravity is a kind of entropic force [31], this article takes Wilson’s entropy gravity model as decay effect pattern in 2SFCAe. Wilson deduced an interactive model based on the principle of Maximum Entropy rather than a metaphor of Newton gravity model [32], changing the interaction equation from power function to exponential function:

$$T_{ij} = A_i O_i B_j D_j \exp(-\beta C_{ij}) \tag{3}$$

The difference between them is that the Newton power function is more sensitive to the reduction of distance. As the distance becomes greater, the difference becomes smaller.

With a simplification to Wilson Entropy Gravity model, this article defined the gravity value between  $O_i$  and  $D_j$  as  $T_{ij} = O_i D_j \exp(-2d_{ij})$ , giving the coefficient  $\beta$  value of 2 as most traditional gravity function does.

### 3.3. LUR-2SFCAe Method

As mobile phone base station  $j$  with population  $p_j$ , land point  $L_i$  with  $k$  different types of land use of corresponding area value ( $M_{i1}, M_{i2}, \dots, M_{ik}$ ) and the distance between them as  $d_{ij}$ , the procedures of LUR-2SFCAe are as follows:

(1) Calculate the population entropy gravitational value for each piece of land  $L_i$ , and the surrounding  $j$  base stations (population  $p_j$ ) within the threshold range  $d_0$ :

$$f_{ij} = p_j \exp(-2d_j) \forall d_j \leq d_0 \tag{4}$$

(2) Aggregate population gravitational entropy for each land point  $L_i$ :  $\sum f_{ij}$ ;

(3) Assign land value back of  $L_i$  ( $M_{i1}, M_{i2}, \dots, M_{ik}$ ) to surrounding base stations proportionally:

$$l_{ik} = M_{ik} \cdot f_{ij} / \sum f_{ij} \tag{5}$$

(4) Summarize the total land value of base station  $j$  assigned by surrounding land points within the threshold range  $d_0$ :

$$X_{jk} = \sum l_{ik} \forall d_j \leq d_0 \tag{6}$$

(5) Explore a linear regression function with all population  $p_j$  and its land value set  $(X_{j1}, X_{j2}, \dots, X_{jk})$  of base stations as  $P = f(X_k)$ :

$$P = a_0 + a_1X_1 + a_2X_2 + a_3X_3 + \dots + a_kX_k + \epsilon_k \tag{7}$$

where  $X_k$  is a set of land value of different types,  $a_k$  is the corresponding coefficient,  $a_0$  is the constant and  $\epsilon_k$  is standard error.

(6) Verify the reliability of the regression model with verification data of base stations, substitute the regression coefficient back to calculate the total population of land point  $L_i$ .

The illustration of the regression model for LUR-2SFCAe and a comparison with FCA method is as follows (Figure 5):

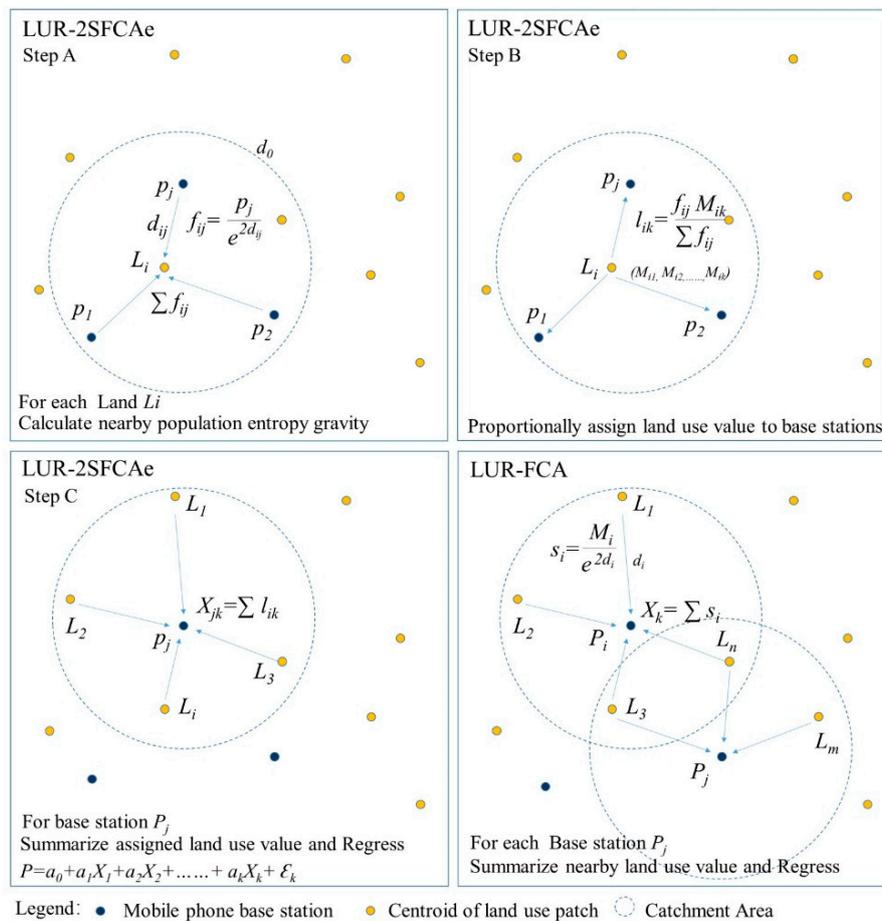


Figure 5. Illustration of LUR-2SFCAe and LUR-FCAe.

## 4. Results

### 4.1. Regression Result of LUR-2SFCA

This study defined the search threshold as 5000 m, larger than most of buffering zone of LUR in air pollution studies, taking most of Land use types into account as R (Residential), A (Administration and public services), B (Commercial and business facilities), MW (Industrial, logistics and warehouse), U (Municipal utilities), G (Green space and square), S (Road, street and transportation), H1 (Town and country), H2 (Regional infrastructure), E1 (water) and E2 (Farmland) and all-time population. While some of the base stations and land point overlay in same grids, between which the interaction distance is defined as 200 m.

Performed 2SFCAe with previous land use data and all-time population, and then a stepwise regression in SPSS, the regress retains the following six parameters, R, A, B, MW, H1, and E with a final adjusted  $R^2$  of 0.792, showing a high fitness of LUR-2SFCAe model (Table 1). It could also be concluded that residential zone played a dominant role in population distribution, then public service, commercial, industrial, logistics and warehouse in the main urban area, town and country and water in the regional background.

**Table 1.** Summarization of stepwise regression (all-time).

Model Summary					
Model	R	$R^2$	Adjusted $R^2$	Std. Error of the Estimate	Predictors
1	0.821	0.674	0.674	3208.54425171590500	(Constant), R
2	0.852	0.726	0.726	2942.90174974674660	(Constant), R, A
3	0.875	0.765	0.765	2725.59070375458400	(Constant), R, A, B
4	0.883	0.780	0.779	2641.12225592935500	(Constant), R, A, B, MW
5	0.887	0.786	0.786	2603.89655684452830	(Constant), R, A, B, MW, H1
6	0.890	0.792	0.792	2566.13576819830540	(Constant), R, A, B, MW, H1, E1

For exploring the urban spatial feature, this experiment implements regression with the population of worktime and non-worktime to get population distribution in different periods. As compared with all-time population, non-worktime displays a higher adjusted  $R^2$  of 0.82 while worktime shows the same (Table 2).

**Table 2.** Summarization of regression (worktime and non-worktime).

SUMMARY OUTPUT					
Regression Statistics (Worktime)			Regression Statistics (Non-Worktime)		
Multiple R		0.890264	Multiple R		0.906303421
R Square		0.792571	R Square		0.82138589
Adjusted R Square		0.791919	Adjusted R Square		0.820904883
Standard error		2564.744	Standard error		2370.731631
Observation		2235	Observation		2235
ANOVA			(worktime)		
	df	SS	MS	F	Significance F
Regression	6	$5.6 \times 10^{10}$	$9.33 \times 10^{10}$	1416.09	0
Residual	2228	$1.47 \times 10^{10}$	6,585,053		
Total	2234	$7.06 \times 10^{10}$			
ANOVA			(non-worktime)		
	df	SS	MS	F	Significance F
Regression	6	$5.76 \times 10^{10}$	$9.6 \times 10^{10}$	1707.636	0
Residual	2228	$1.25 \times 10^{10}$	5,620,368		
Total	2234	$7.01 \times 10^{10}$			

Comparing coefficients with non-worktime (Table 3), worktime has larger amounts in intercept, A, B, WM and H1 type land, which provides job opportunities, while it has a smaller amount in R land and shares the same amount in E.

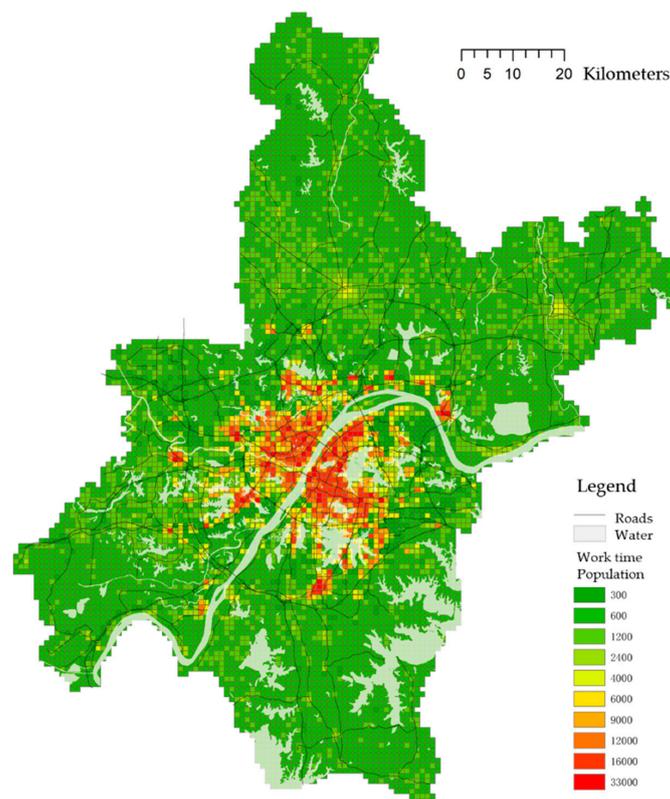
By substituting the coefficients back to LUR equation, the rest of the mobile phone data were verified with the corresponding population with a resulting  $R^2$  of 0.897, which confirmed the applicability of LUR model, then the population of every land points could be computed as well in worktime and non-worktime.

**Table 3.** Regression Coefficients (worktime and non-worktime).

Worktime	Coefficients	Standard Error	t Stat	p-Value
Intercept	477.8392	80.59048	5.929226	$3.52 \times 10^{-9}$
R	0.018144	0.000499	36.33294	$2.1 \times 10^{-227}$
A	0.017012	0.000819	20.77093	$9.65 \times 10^{-88}$
B	0.039378	0.001912	20.59004	$2.24 \times 10^{-86}$
MW	-0.00431	0.00036	-11.9656	$5.04 \times 10^{-32}$
H1	0.002634	0.000228	11.56714	$4.3 \times 10^{-30}$
E	-0.00016	$1.97 \times 10^{-5}$	-8.19038	$4.34 \times 10^{-16}$
Non-Worktime	Coefficients	Standard Error	t Stat	p-Value
Intercept	454.0856	74.45374	6.098896	$1.26 \times 10^{-9}$
R	0.02138	0.000461	46.34211	0
A	0.0149	0.000757	19.6916	$1.04 \times 10^{-79}$
B	0.029965	0.001767	16.95971	$9 \times 10^{-61}$
MW	-0.0051	0.000333	-15.3391	$1.47 \times 10^{-50}$
H1	0.00256	0.00021	12.16957	$4.92 \times 10^{-33}$
E	-0.00016	$1.82 \times 10^{-05}$	-8.51949	$2.9 \times 10^{-17}$

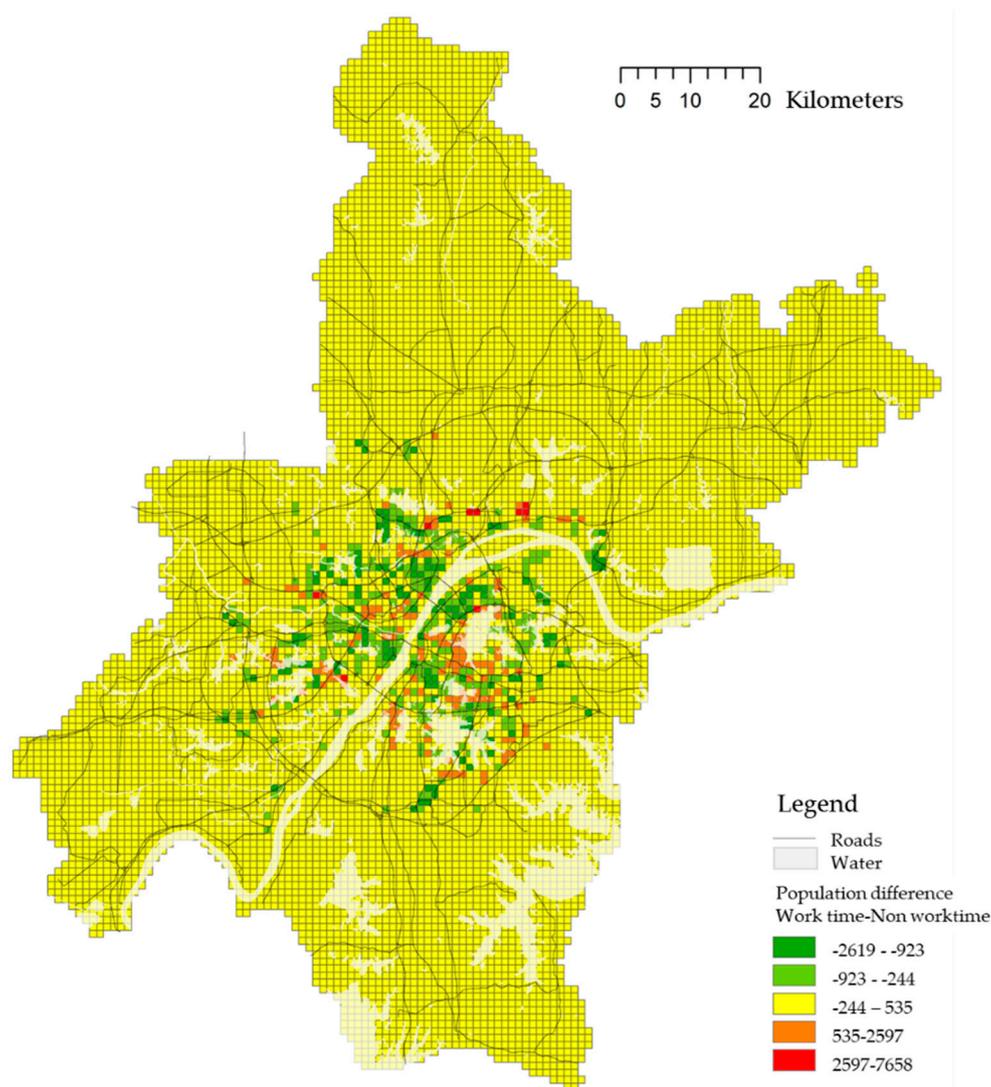
#### 4.2. Exploring Spatial Feature with Gridded Population

Resubstituting the coefficients to the equation can generate the population. As the population is computed by the regression equation with the land entropy gravity, the value is not a direct output but proportional to the actual population of the cells. Plotted with the population of the whole 8999 cells of worktime and non-worktime in ten levels with natural breaks in ArcGIS (Figure 6), both population distribution maps with a similar hierarchical distribution, clearly distinguish the waters, especially in the central area, showing higher accuracy as compared to other interpolation methods such as Kernel, Kriging, etc.

**Figure 6.** Population distribution in worktime.

From the density pattern, Wuhan shows a multi-nuclei model in spatial structure with an obvious dominant and compact core in the main city and several sub-cores around its periphery mostly located across the waters. Such natural environment as rivers and lakes impose strong constraints on the structural development of Wuhan which expanded mostly along the waters and roads. In general, urban spatial form and pattern of Wuhan are affected and limited by natural environments.

The population difference between worktime and non-worktime can represent the characteristics of each land distinguished as employment center or residential area, in that, if the worktime population is greater than non-worktime, it will be the employment center, otherwise, it is a residential area. Mapping the population variance in ArcGIS with five groups and Jenkin's breaks, most of the cells outside the main city were in middle level, reflecting a natural metropolis boundary of Wuhan, wherein the larger positive numbers are the employment centers and smaller negative numbers are the residential areas (Figure 7).



**Figure 7.** Population difference between worktime and non-worktime.

With a relatively compact employment cluster in the central area, the urban spatial structure shows a trend of sprawling to the outskirts, especially in the north along the Yangtze River and in the east. The phenomenon that employment center grows on the periphery reflects the characteristics of top-down urbanization driven by the governments in China. The planning led by the governments

also has a greater impact on an intentional move out of the industrial park, universities, colleges, and commercial wholesale market from downtown to periphery under the policy of land finance [33].

Since land types A, B, and WM play important roles in worktime population, the conclusion is drawn from the regression above that the percentage of different types of land uses in units can be analyzed to show the distribution of different categories of employment centers.

When mapped in ArcGIS (Figure 8), most of the employment centers of A land are located in Wuchang District which has many universities and research institutes, those of B land situate in Hankow District which is a historical commercial district, and those of WM land disperse around the outskirts. The universities and markets around the outskirts are a result of planning by the government which relocated them to the suburbs intentionally for a density evacuation in the central city, a motivation for urban sprawling and a result of the land economy pursuit.

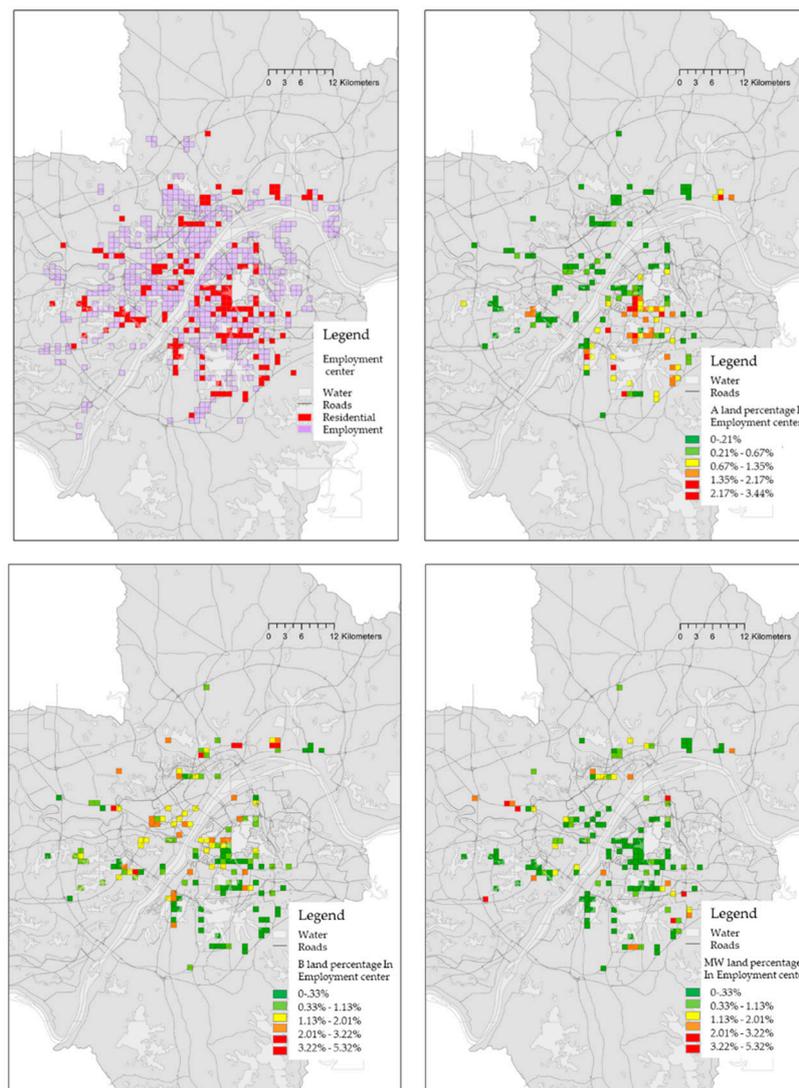


Figure 8. Employment centers of different kinds of land use.

As compared with the traditional concentric model, sector model and multi-nuclei model, Wuhan shows a mixed pattern: a dominant center and several minor sub-centers of different functions, expanding along the river or traffic corridors.

## 5. Discussion

These results suggest that LUR-2SFCAe is fit for dasymetric mapping with mobile phone data and to present land uses, wherein the distance decay setting of 2SFCAe tackles the problem of the undefined boundary of base station service area. These findings are understandable because there exists certain connection between user population of mobile phone base stations and surrounding land uses.

These results agree with Anto's findings [26] that mobile phone data reflect surrounding population density. The regression  $R^2$  in this experiment is higher than that reported in Image Texture method [13], and similar to building population mapping with PopShape GIS [34], which further verify the conclusion of Bakillah [20] that finer scale ancillary data provide more accurate dasymetric mapping. Furthermore, 2SFCAe method explores the distance-decay function and weighted population effect in interpolation which can be seen as an improvement of kernel density surface method with population-weighted census centroids [35], and is better than FCA method, as in public service studies [36].

As a comparison of the assumption of gridded mobile phone data and the uncertainty in boundary issue, this study also applied the same LUR-2SFCA method on original mobile phone data and explored LUR with Thiessen polygon boundary of base stations. Both regressions showed a lower  $R^2$ . While LUR-2SFCAe was tested and calculated with all-time population of original mobile phone data and land use data, the regression result in SPSS (Table 4) retained only three parameters as R, H1 and MW, showing an unacceptable  $R^2$  which indicates the data or model need to be modified. Meanwhile, when Thiessen polygons were applied on original base stations, the regression with user population and the land uses inside the polygons retained similar parameters as previous models and a same low  $R^2$  (Table 5).

**Table 4.** Summarization of Regression with original mobile phone data.

Model Summary					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Predictors
1	0.080	0.006	0.006	323.201	(Constant), R
2	0.115	0.013	0.013	322.084	(Constant), R, H1
3	0.120	0.014	0.014	321.899	(Constant), R, H1, MW

**Table 5.** Summarization of Regression with Thiessen Polygons.

Model Summary					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Predictors
1	0.164	0.027	0.027	896.784	(Constant), R
2	0.182	0.033	0.033	894.020	(Constant), R, A
3	0.192	0.037	0.037	892.199	(Constant), R, A, H1
4	0.200	0.040	0.039	890.932	(Constant), R, A, H1, E1
5	0.206	0.042	0.042	889.737	(Constant), R, A, H1, E1, MW

The interesting thing is that the Thiessen polygons method does not considered spatial heterogeneity, unlike many traditional methods [21], since different polygons may contain the same land use components but vary in user population. However, the former has verified the unevenness in spatial heterogeneity as nearby station may have different population. Although the 1 km grid was applied in this study with a persuasive regression result, a balance of unevenness and heterogeneity remains to be explored for finer-grid mapping.

These results provide substantial evidence for the assumption that the combination of LUR and 2SFCA can map population with mobile phone data and land use, which addresses the spatial heterogeneity in most dasymetric mapping, and tackles the problem for the undefined boundary of

base stations in 2SFCA. Furthermore, the 2SFCAe defines a new demand–supply model comparing with traditional LUR model which often simply summarizes the land use of its buffering zone.

## 6. Conclusions

This paper develops a methodology of LUR-2SFCA with mobile phone data and land use data to alleviate the spatial heterogeneity issue which challenges most dasymetric mapping methods and explore urban spatial feature with population mapping of different time at finer scale. As a result of the experiments, it is concluded that mobile phone data reflect more temporal, detailed and accurate population than census data. Furthermore, the distance-decay model in 2SFCA can solve the uncertain service boundary issue of mobile phone base station and its user population. Additionally, the 2SFCA assigns a weighted population to nearby land patches which again strengthens the spatial heterogeneity of observed population, hence improving the fitness of regression. Finally, the work-time and non-worktime population distribution calculated by LUR-2SFCAe can help identify urban centrality and employment centers.

On the other hand, based on comparison of original mobile phone data and Thiessen polygon method, spatial heterogeneity and unevenness of mobile phone data are supposed to affect the outcome of dasymetric mapping, which needs to be further studied.

Although land use and mobile phone data utilized in this paper can provide population mapping of a relatively high resolution, both data types could only be acquired locally, which may limit their implementation on a global scale. On the other hand, this method might face the challenge of computation due to the sizes of grids and floating catchment area, when the grids are smaller and the spatial resolution finer. The calculation task will increase exponentially in the procedure of distance calculation within a floating catchment area, which might exceed the calculation capability of the computer or related software.

In general, the contribution of the present study lies in the methodology combination of LUR and 2SFCAe, which provide a distance-decaying demand–supply model in dasymetric mapping with mobile phone data and land use data. The methodology of LUR-2SFCAe is also applicable to related dasymetric mapping and LUR studies.

**Author Contributions:** L.L. and H.W. conceived and designed the experiments; L.L. performed the experiments; Z.P. acquired and analyzed the data; Y.Y. and H.J. contributed reagents/materials/analysis tools; and H.W. and L.L. wrote the paper.

**Funding:** The study was funded by China Postdoctoral Science Foundation (No. 2016M600609); and China Postdoctoral Science Foundation (No. 2016M602357).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Taylor, P.J.; Pain, K. Polycentric mega-city regions: Exploratory research from Western Europe. In Proceedings of the Healdsburg Research Seminar on Megaregions, Healdsburg, CA, USA, 4–6 April 2007; Lincoln Institute of Land Policy and Regional Plan Association: Cambridge, MA, USA, 2007.
2. Chou, T.-L. The transformation of spatial structure: From a monocentric to a polycentric city. In *Globalizing Taipei: The Political Economy of Spatial Development*; Wiley: Hoboken, NJ, USA, 2005; pp. 55–77.
3. Wu, F.; Yeh, G.O. Urban spatial structure in a transitional economy the case of Guangzhou, China. *J. Am. Plan. Assoc.* **1999**, *65*, 377–394.
4. Krehl, A.; Siedentop, S.; Taubenböck, H.; Wurm, M. A comprehensive view on urban spatial structure: Urban density patterns of German city regions. *ISPRS Int. J. Geo-Inf.* **2016**, *5*, 76. [[CrossRef](#)]
5. Jun, M.J.; Choi, S.; Wen, F.; Kwon, K.H. Effects of urban spatial structure on level of excess commutes: A comparison between Seoul and Los Angeles. *Urban Stud.* **2018**, *55*, 195–211. [[CrossRef](#)]
6. Goswami, A.G.; Lall, S.V. Jobs in the city: Explaining urban spatial structure in Kampala. In *Policy Research Working Paper*; World Bank: Washington, DC, USA, 2016.

7. Bento, A.M.; Cropper, M.; Mobarak, A.M.; Vinha, K. The impact of urban spatial structure on travel demand in the United States. *Rev. Econ. Stat.* **2016**, *87*, 466–478. [[CrossRef](#)]
8. Mennis, J. Generating surface models of population using dasymetric mapping. *Prof. Geogr.* **2003**, *55*, 31–42.
9. Holt, J.B.; Lo, C.P.; Hodler, T.W. Dasymetric estimation of population density and areal interpolation of census data. *Am. Cartogr.* **2004**, *31*, 103–121. [[CrossRef](#)]
10. Poulsen, E.; Kennedy, L.W. Using dasymetric mapping for spatially aggregated crime data. *J. Quant. Criminol.* **2004**, *20*, 243–262.
11. Eicher, C.L.; Brewer, C.A. Dasymetric mapping and areal interpolation: Implementation and evaluation. *Am. Cartogr.* **2001**, *28*, 125–138. [[CrossRef](#)]
12. Mennis, J.; Hultgren, T. Dasymetric mapping for disaggregating coarse resolution population data. In Proceedings of the 22nd Annual International Cartographic Conference, A Coruña, Spain, 9–16 July 2005.
13. Liu, X.H.; Herold, M.; Clarke, K. Population density and image texture: A comparison study. *Photogramm. Eng. Remote Sens.* **2006**, *72*, 187–196. [[CrossRef](#)]
14. Bielecka, E. A dasymetric population density map of Poland. In Proceedings of the 22nd International Cartographic Conference, A Coruña, Spain, 9–16 July 2005.
15. Krunić, N.; Bajat, B.; Kilibarda, M. *Dasymetric Mapping of Population Distribution in Serbia Based on Soil Sealing Degrees Layer*; Springer International Publishing: Berlin, Germany, 2015; pp. 137–149.
16. Wu, S.S.; Qiu, X.; Wang, L. Population estimation methods in GIS and remote sensing: A review. *Mapp. Sci. Remote Sens.* **2005**, *42*, 80–96. [[CrossRef](#)]
17. Maantay, J.A.; Maroko, A.R.; Herrmann, C. Mapping population distribution in the urban environment: The cadastral-based expert dasymetric system (CEDS). *Am. Cartogr.* **2007**, *34*, 77–102. [[CrossRef](#)]
18. Jia, P.; Qiu, Y.; Gaughan, A.E. A fine-scale spatial population distribution on the high-resolution gridded population surface and application in Alachua County, Florida. *Appl. Geogr.* **2014**, *50*, 99–107. [[CrossRef](#)]
19. Ural, S.; Hussain, E.; Shan, J. Building population mapping with aerial imagery and GIS data. *Int. J. Appl. Earth Obs. Geoinf.* **2011**, *13*, 841–852. [[CrossRef](#)]
20. Bakillah, M.; Liang, S.; Mobasheri, A.; Arsanjani, J.J.; Zipf, A. Fine-resolution population mapping using OpenStreetMap points-of-interest. *Int. J. Geogr. Inf. Sci.* **2014**, *28*, 1940–1963. [[CrossRef](#)]
21. Kunze, C.; Hecht, R. Semantic enrichment of building data with volunteered geographic information to improve mappings of dwelling units and population. *Comput. Environ. Urban Syst.* **2015**, *53*, 4–18. [[CrossRef](#)]
22. Liu, X.H.; Kyriakidis, P.C.; Goodchild, M.F. Populationata with volunteered geographic information to improve mappings of dwel. *Int. J. Geogr. Inf. Sci.* **2008**, *22*, 431–447. [[CrossRef](#)]
23. Ryan, P.H.; Lemasters, G.K. A review of land-use regression models for characterizing intraurban air pollution exposure. *Inhal. Toxicol.* **2007**, *19*, 127. [[CrossRef](#)] [[PubMed](#)]
24. Hoek, G.; Beelen, R.; Hoogh, K.D.; Vienneau, D.; Gulliver, J.; Fischer, P.; Briggs, D. A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmos. Environ.* **2008**, *42*, 7561–7578. [[CrossRef](#)]
25. Cervantes-Larios, A.; Hystad, P.; Setton, E.; Poplawski, K.; Deschenes, S.; Demers, P.A. Estimating Canadians' exposure to PM<sub>2.5</sub> and NO<sub>2</sub> using national land use regression models: Implications of scale and population location measures. *Epidemiology* **2011**, *22*, S106.
26. Aasa, A. Application of mobile phone location data in mapping of commuting patterns and functional regionalization: a pilot study of Estonia. *J. Maps* **2013**, *9*, 10–15.
27. Frias-Martinez, V.; Soguero, C.; Frias-Martinez, E. Estimation of urban commuting patterns using cellphone network data. In Proceedings of the ACM SIGKDD International Workshop on Urban Computing, Beijing, China, 12–16 August 2012.
28. Lulli, A.; Gabrielli, L.; Dazzi, P.; Dell'Amico, M.; Michiardi, P.; Nanni, M.; Ricci, L. Scalable and flexible clustering solutions for mobile phone-based population indicators. *Int. J. Data Sci. Anal.* **2017**, *4*, 285–299. [[CrossRef](#)]
29. Wu, H.; Liu, L.; Yu, Y.; Peng, Z. Evaluation and planning of urban green space distribution based on mobile phone data and two-step floating catchment area method. *Sustainability* **2018**, *10*, 214. [[CrossRef](#)]
30. Wang, F.H.; Liu, Y. Effects of mullite fiber content on friction and wear properties of ceramic-based friction material. *J. Mater. Eng.* **2012**, *12*, 61–65.
31. Verlinde, E. On the origin of gravity and the laws of Newton. *J. High Energy Phys.* **2010**, *2011*, 1–27. [[CrossRef](#)]
32. Wilson, A.G. Entropy in urban and regional modeling. *Econ. Geogr.* **2011**, *48*, 446–447.

33. Cao, G.; Feng, C.; Tao, R. Local “land finance” in China’s urban expansion: Challenges and solutions. *China World Econ.* **2008**, *16*, 19–30. [[CrossRef](#)]
34. Lwin, K.K.; Murayama, Y. Accuracy assessment of GIS based building population estimation algorithm. In *Spatial Analysis and Modeling in Geographical Transformation Process: GIS-Based Applications*; Murayama, Y., Thapa, R.B., Eds.; Springer Netherlands: Dordrecht, The Netherlands, 2011; pp. 99–112.
35. Martin, D. An assessment of surface and zonal models of population. *Int. J. Geogr. Inf. Syst.* **1996**, *10*, 973–989. [[CrossRef](#)]
36. Wang, F. *Quantitative Methods and Applications in GIS*; CRC Press: Boca Raton, FL, USA, 2008; Volume 60, pp. 434–435.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).