Article

Measurement of Potential Victims of Burglary at the Mesoscale: Comparison of Census, Phone Users, and Social Media Data

Zhuofang Zhang ^{1,2} , Lin Liu ^{3,4,*} and Sisun Cheng ¹

¹ School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China; zhangzhf27@mail2.sysu.edu.cn (Z.Z.); chengss@mail2.sysu.edu.cn (S.C.)

² Guangdong Provincial Engineering Research Center for Public Security and Disaster, Guangzhou 510275, China

³ Center of GeoInformatics for Public Security, School of Geography and Remote Sensing, Guangzhou University, Guangzhou 510006, China

⁴ Department of Geography and GIS, University of Cincinnati, Cincinnati, OH 45221, USA

* Correspondence: liulin1@gzhu.edu.cn

Abstract: Since the target of burglars is generally the property of the inhabitant, it is crucial to accurately measure potential victims when analyzing burglaries, especially in small areas. Previous studies on burglary are mostly based on large units such as census tracts or communities. One of the difficulties is the measurement of the potential victims of burglary at the mesoscale. We compare the measuring effects of census population, census households, nighttime mobile phone users, and nighttime social media, such as the Tencent regional heatmap (TRH), on potential victims of burglary on 150 m × 150 m grids. Based on the rational choice theory, and controlling for the potentially confounding effects of risks and cost, we show that the TRH performed best, followed by census households and census population, and phone users performed poorly. The best-performing time period for TRH data was 3:00–5:00 am on weekends. These findings could lead to an improved measurement of potential victims of burglary at the mesoscale, and could provide scientific insight for crime prevention.

Keywords: crime geography; Tencent regional heatmap; mobile phone users; residential population; potential victims

Citation: Zhang, Z.; Liu, L.; Cheng, S. Measurement of Potential Victims of Burglary at the Mesoscale: Comparison of Census, Phone Users, and Social Media Data. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 280. <https://doi.org/10.3390/ijgi10050280>

Academic Editor: Wolfgang Kainz

Received: 3 March 2021

Accepted: 26 April 2021

Published: 28 April 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The target of home burglary is a residential household. Data such as resident population [1], households [2], and building characteristics [3] in the census are often used to estimate the actual residential population targets. However, such data are not available in small geographical areas in China and many other countries.

Emerging data sources, such as phone users [4,5], the Tencent regional heatmap (TRH) [6], taxi ridership, and subway ridership [7], have been used to estimate ambient population, capturing the dynamic distribution of population. Among them, phone users are mostly applied in theft research. He et al. [8] found that the population indicators calculated from mobile phone big data significantly correlate with the spatial variation of thefts. Song et al. [7] compared the measurement effects of four types of data (phone users, residential population, subway ridership, and taxi ridership) on the population at risk of theft at different times on 1 km × 1 km grids, and concluded that phone users and taxi ridership outperformed residential population in the majority of periods of a day. Recently, researchers [9,10] discovered that TRH helps explain the distribution of drug crime hotspots, and found that it has a significant positive impact on street contact crime. The former [9] used dynamic data to represent the overall population, but did not make full use of the time-varying characteristics of dynamic data; the latter [10] divided a day into three time periods to assess the varying effects on street contact crime. So far, this new ambient population measure has not been applied to burglary studies.

With the development of crime analysis from the macro- to the meso- and micro-scales [11,12], burglary also needs more meso- and micro-research [13]. At the macro scale, the social disorganization theory is commonly used to explore the influencing factors of burglary [12,14]. In contrast, the rational choice theory and crime pattern theory have been applied at the meso- and micro-scales. Previous literature has considered the influences of bus stops [15], road density [2,16], neighborhood commercial density [17], population composition [2], and target concentration [16,18]. Crime prevention through environmental design theory (CPTED) pays more attention to housing accessibility [16,18], housing usage patterns, and traffic accessibility [19]. Similarly, in police practice, measurement, prevention, and intervention have been addressed in the literature [20,21]. Measurement and representation of potential burglary victims at the mesoscale have been a consistent challenge.

While studies of burglary are abundant, few exist that compared different measurements of potential victims of burglary. It is in this context that this paper explores alternative measurements using phone users and TRH data. Phone users are derived from the interactions of individual phones and the base stations. TRH is based on the location information collected by mobile apps. Both cell phone data and TRH data have a very high temporal resolution, but the latter has a much higher spatial resolution. Theoretically, cell phone and TRH data at nighttime, when people have returned home, can better capture the actual residential population. In reality, in large cities in southern China, even at night, residents engage in many activities, such as working overtime, having midnight snacks, partying at entertainment facilities, etc. It is worth noting that there may be a significant difference between the regularity of these activities during weekday nights and weekend nights. All of these will affect the measurement of residential population by dynamic population data [10]. However, exactly what time periods are best suited for burglary studies has not been tested empirically.

This study chose the rational choice theory to control for crime risks, costs, and rewards, and aims to identify the most appropriate measure of residential population as a surrogate measure of potential burglary targets. Four measures are compared in this study, including census population, census households, mobile phone users, and THR.

2. Theories

As argued in the introduction, it is very important to choose a theoretical framework based on the actual social situation of the study area. The rational choice theory is considered suitable for crime research at the mesoscale, and it has been proved to be able to explain China's burglary effectively [3,22]. It emphasizes offenders' perception and judgment of the target, and comprehensively considers the potential benefits, risks, and cost of approaching the target. Obviously, these three elements, which are outlined below, are closely related to the state of potential victims.

- (1) Potential benefits. Previous studies [23,24] considered the wealth and exposure of targets in residential areas to represent their attractiveness to criminals. The proceeds of burglaries come from the residence of the victim, so the number of potential victims and their possible wealth have become the primary conditions for the offender to measure the benefits. In this study, census population, census households, nighttime mobile phone users, and the Tencent regional heatmap were used to represent potential targets. Housing types also imply the offender's estimation of the possible wealth of the household. Three housing types were used in this study: residential district buildings, apartments and dormitory buildings, and commercial-residential buildings [17,25].
- (2) Cost of traveling to the target. Traffic convenience affects the cost of burglaries. Following prior research [3], bus stops and road density were selected to represent traffic accessibility. The more convenient the transportation in the target area, the lower the travel cost.

- (3) Risks of getting caught. Crime risks stem from civil defense, physical defense, and technical defense. The risk of surveillance and standards of physical security are primary deterrents for burglars [26]. In China, surveillance cameras installed by the police are a direct supervision force, which increases the risk of burglars being discovered.

Figure 1 outlines the application of rational choice theory in this study. Specifically the study considered the potential benefits, risks, and costs that may be brought by contacting potentially victimized targets; sorted out the specific corresponding indicators in the built environment and population environment; and analyzed the factors that affect the occurrence of burglaries. From the perspective of the built environment, the risk of exposure and the physical condition of criminal travel are described by surveillance-camera distribution, road density, and various housing-type POIs. From the perspective of the social environment, residential population (census population, census household, mobile phone users, Tencent regional heatmap) was used to represent potential victims.

Figure 1. Representation of theoretical elements.

3. Study Area, Data, and Methods

3.1. Study Area

The study area is located in XT JieDao, ZG City, South China. ZG City, as the “South Gate” of China, is a metropolis with a developed economy, a large population, and a very high degree of openness. There are very different urbanization processes, with complex population structure and diverse land use. The southern part of XT JieDao is an urban area with a high degree of urbanization; the northern part is dominated by mountainous areas and sparsely populated. We choose the south as the study area, with a total area of 4.4 square kilometers, with a nonlocal population accounting for nearly 70%. As of November 2010, it had a nonlocal population of 118,900 and a local permanent population of 54,500. The rate of crime, especially burglary, of the study area is significantly higher than in the rest of the city. The study area is composed of 15 communities, including 3 urban villages (Figure 2). The urban village is a unique urban phenomenon in the process of urbanization in China (it is completely different from the “urban village” in the UK, an idealized model for sustainable development and planning). Urban villages are essentially islands surrounded by newly developed areas. There are major differences in the appearance and internal composition between the urban villages and other communities. The former is composed of disorderly low-rent housing, mixed with various business for basic services, small restaurants, retail shops, and street vendors along narrow alleys. The crowded back-to-back buildings may not follow the proper building codes. In general, urban villages are not well maintained. The latter is composed of well-planned high-rise buildings and green spaces [27].

3.2. Data and Methods

We constructed a negative binomial regression model by taking the grid as the sample unit to compare the effects of different indicators on the measurement of the potential victims (the actual resident population). The research data included burglary cases, census data, mobile phone users, the Tencent regional heatmap, surveillance cameras, POIs of housing types, traffic networks, and other basic geographic data. Since the nighttime population can represent the residential population distribution more effectively, mobile phone users and TRH data from 18:00 to 08:00 on both weekdays and weekends were selected for the study. Following a previous study [7], the data were divided in 5 3-h time periods. Descriptive statistics of all variables are shown in Table 1.

Table 1. Summary of the statistics of the variables (N = 192 sample grid cells).

Variables		Mean	Variance	Minimum	Maximum
Number of burglaries(cases)		2.85	28.98	0.00	24.00
Census population (/100 people)		3.99	2.23	1.46	7.75
Census households (/100 households)		1.76	0.93	0.57	4.03
Mobile phone users (weekday)	18:00–20:00	4.51	16.57	0.03	26.55
	21:00–23:00	4.41	16.45	0.02	26.82
	00:00–02:00	3.33	12.21	0.01	26.47
	03:00–05:00	2.60	7.99	0.00	21.76
	06:00–08:00	3.18	9.49	0.01	22.40
Mobile phone users (weekend)	18:00–20:00	4.75	17.09	0.04	23.77
	21:00–23:00	4.64	18.26	0.04	28.70
	00:00–02:00	3.74	14.75	0.03	28.74
	03:00–05:00	2.82	9.23	0.02	24.16
	06:00–08:00	3.16	10.16	0.01	24.61
Tencent regional heatmap (weekday)	18:00–20:00	4.69	10.97	0.03	14.54
	21:00–23:00	4.55	13.72	0.02	16.31
	00:00–02:00	2.30	4.47	0.07	9.70
	03:00–05:00	1.05	0.86	0.00	3.73
	06:00–08:00	2.31	2.95	0.05	7.48
Tencent regional heatmap (weekend)	18:00–20:00	4.60	11.19	0.09	13.94
	21:00–23:00	4.06	9.94	0.09	14.41
	00:00–02:00	2.51	4.03	0.07	8.58
	03:00–05:00	1.11	0.88	0.00	3.65
	06:00–08:00	2.19	3.26	0.00	7.75
Number of surveillance cameras		1.66	7.20	0.00	19.00
Road density		0.27	0.03	0.00	1.00
Number of bus stops		0.23	0.35	0.00	3.00
Number of apartment dormitories		0.15	0.19	0.00	3.00
Number of commercial–residential buildings		0.64	3.60	0.00	16.00
Number of residential district buildings		0.81	1.77	0.00	8.00

3.2.1. Spatial Units of Analysis

In order to not be restricted by administrative boundaries, a grid treatment was applied. Referring to the spatial grid division methods in prior research [28,29], the optimal grid size formula of Griffith et al. [30] was applied, and the optimal grid side length turned out to be about 130 m. Considering the actual situation of the study area, the grid side length was determined to be 150 m [10]. Along the edge of the study area, cells with less than 50% inside the study area were removed from the analysis. A total of 192 cells made up the sample for subsequent analysis.

Figure 2. Distribution map of communities, villages, and roads in XT JieDao.

3.2.2. Data and Processing

The data sources and relevant information are shown in Table 2. For burglary cases, we first adopted a procedure to automatically settle the case points, and then manually adjusted the case points one by one to ensure accuracy (number of cases: 284 in 2017, 153 in 2018, and 110 in 2019). For surveillance cameras, abnormal points such as unclear positioning were eliminated first, and the remaining errors were manually adjusted (616 points were in XT JieDao, and 9 points were eliminated, leaving 607 points; among these, 318 cameras were distributed in the southern part of XT JieDao). For road-network data, prior to processing, manual updates were completed by comparing remote-sensing images and field surveys. Referring to prior research [7], after processing, all data were summarized into the grid by intersecting to estimate the grid value of the corresponding variable.

Table 2. Data overview and processing methods.

Data	Source	Acquisition Time	Attributes	Processing Method
Burglaries cases	Municipal Public Security Bureau	2017–2019	Table: 547 cases	Drop point, intersection, summarize
Surveillance cameras	Municipal Public Security Bureau	2018	Point: 318 cameras	Intersect, summarize
Bus stops	“Daodaotong” electronic map	2016	Point: bus-stop location	Intersect, summarize
Housing types	“Daodaotong” electronic map	2016	Point: apartment dormitory, commercial–residential building, residential district building	Intersect, summarize
Road density	“Daodaotong” electronic map	2016	Line: community roads, low-grade roads but for traffic, village and town-level roads, county roads, provincial roads, domestic roads, national roads	7 levels, weighted from 1 to 7; weighted kernel density analysis by level (radius 150 m); intersect
Census data	The sixth national census	November 2010	Plane: community units	Intersect, weighted by area
Tencent regional heatmap	Tencent Company	9–15 April 2018	Point: 25 m interval sampling	Intersect, summarize
Mobile phone users	China Unicom	12–18 May 2016	Point: base station units	Create Thiessen polygons, intersect, weight by area

1. Mobile phone users

Mobile phone user data were derived from the signal data based on base stations provided by one of the three major carriers in China, including the total number of anonymous mobile users counted hourly in each base station during the whole week of 12–18 May 2016. In south XT JieDao, the adjacent cellular towers are within 500 m of each other. As long as a mobile phone is powered on, the mobile phone antenna will automatically look for base stations to obtain service, and its signal searching is recorded in the mobile phone user data [31]. Since phones often try to connect to the nearest base station, it made sense to use the Thiessen polygon approach to aggregate mobile user groups into grids. To ensure the integrity and accuracy of each Thiessen polygon in the study area, all base stations within 1 km of the periphery of the study area also were collected. It is worth noting that the number of mobile phone users was relative and cannot represent the absolute number of users. Considering the difference between the activities of residents on weekdays and weekends [32], the 114 h from 00:00 on Monday to 17:00 on Friday were considered weekdays, and the remaining 54 h from 17:01 on Friday to 23:59 on Sunday were considered weekends.

2. Tencent Regional Heatmap

The Tencent regional heatmap (TRH) was a capture of current relative population data in the study area based on the map. It can be viewed in the mini-program “Yichuxing” (based on the social media software WeChat), which shows the dynamic spatial distribution of the app users. Compared with the traditional static census data, commuter survey data, and other dynamic population data, the TRH has the advantages of strong real-time performance, high accuracy, wide coverage, and easy access [33]. Tencent apps have a large user base, and as long as the app is in use or running in the background, the real-time location of users is collected [34]. Through the big data acquisition method with a Python program, the dynamic population data of the whole week of 9–15 April 2018 was obtained, with a sampling spatial interval of 25 m × 25 m and a time interval of 1 h. Although TRH data can only collect the population data of Tencent-related products on smart phones, considering the general extremely high utilization rate of Tencent products in major cities of China, it had a good coverage rate in the research area [34]. Due to its higher accuracy than that of the grids, to make data interpretation more practical, prior studies [33] were referred to convert data to obtain the population at the sampling point, and the corresponding sampling point data in each grid was combined in the overlay analysis as the value of each grid population.

3.2.3. Negative Binomial Regression Model

As is typical in crime studies, our dependent variable, the number of burglary cases, was overdispersed. Therefore, the negative binomial model was more appropriate than the Poisson model [7,35]. The marginal effect of the independent variable of the model is called the incidence rate ratio (IRR), which means for every increase unit of the independent variable, the incidence rate of a case will increase by a factor of the IRR.

For model comparison, Akaike’s information criterion (AIC) was adopted to reflect the effectiveness of the model; the smaller the value is, the better the model performed [7]. Variance inflation factors (VIF) were used to test the multicollinearity between independent variables prior to the estimation. In principle, the tolerable range of VIF was less than 10.

4. Analysis and Results

4.1. Spatial Distribution of Burglary

The burglaries in south XT JieDao showed obvious clustering characteristics (Figure 3), and the case volume was significantly high in hotspot A (located in the central and western area) and hotspot B (located in the eastern area). Compared with hotspot A, hotspot B had a weaker concentration and a wider range. There also were a few cases in the far east and west.

Figure 3. Distribution map of burglary cases (A, B, C, and D indicate hotspots; N = 192 sample grid cells).

The investigation found that hotspot A was a typical urban village, and it was bounded by three main roads in the periphery. Hotspot B was composed of two urban villages juxtaposed from east to west that were separated by a zigzag path with little morphological differences. What hotspot A and hotspot B had in common was that the land space inside the village city was mixed and disordered, including a large number of high-density cheap rental buildings, mixed with commercial and residential buildings and schools; while the surrounding area close to the main road was another well-constructed scene, composed of office buildings, commercial buildings, and government agencies, with distinct functions. Obviously, burglary cases were concentrated in the inner urban village area. In addition, scattered burglary cases also occurred in regions C and D. There were two main roads on the west side of region C and on the east side of region D, and there were many well-constructed high-rise residential buildings in each region.

In terms of time distribution, the cases were mainly concentrated in the morning and evening (Figure 4). Most people become aware of a burglary when they got up in the morning or returned home from work during the evening. It should be noted that the exact time of a burglary was uncertain, because the burglary occurred in the absence of the owner's awareness [36]. The "time of incident" in the crime data is often the best guess.

Figure 4. The number of burglary cases in each time period.

4.2. Measurement Result Comparison

Census population, census households, mobile phone users, and TRH data were used to represent the residential population to capture the potential victims of burglary. The first

two were static, while the second two were dynamic data that changed in real time. Due to the mobility of residents between their working and living spaces, the nighttime data may be significant for the representation of the residential population. However, before the morning rush hour, the residents may be generally in their homes. Meanwhile, residents' daily routines on weekdays were affected by work and overtime, while on weekends, they were affected by various recreational activities. In order to measure the performance of dynamic data at different time periods at night accurately, 15 h, from 18:00 to 8:00, were selected separately on weekdays and weekends.

As the VIF shows (Table 3), the collinearity of independent variables was weak in all models (mean VIFs were all less than 1.2, and the Max VIFs were all less than 2).

Table 3. The VIF values of the independent variables in each model (22 models in total).

Variables Representing Potential Victims in Each Model		Mean VIF	Max VIF
Census population (/100 people)		1.17	1.50
Census households (/100 households)		1.18	1.52
Mobile phone users (weekday)	18:00–20:00	1.14	1.37
	21:00–23:00	1.14	1.39
	00:00–02:00	1.14	1.38
	03:00–05:00	1.14	1.38
	06:00–08:00	1.14	1.37
Mobile phone users (weekend)	18:00–20:00	1.14	1.38
	21:00–23:00	1.14	1.39
	00:00–02:00	1.14	1.38
	03:00–05:00	1.14	1.38
	06:00–08:00	1.14	1.37
Tencent regional heatmap (weekday)	18:00–20:00	1.18	1.50
	21:00–23:00	1.17	1.46
	00:00–02:00	1.17	1.44
	03:00–05:00	1.18	1.45
	06:00–08:00	1.17	1.45
Tencent regional heatmap (weekend)	18:00–20:00	1.18	1.51
	21:00–23:00	1.18	1.47
	00:00–02:00	1.17	1.43
	03:00–05:00	1.18	1.44
	06:00–08:00	1.17	1.42

All models were established for the 5 periods with a corresponding potential victim representation variable, and their AIC values are shown in Figure 5. The AIC value was used to measure the fitting of the statistical model. The smaller the AIC value was, the more accurate the independent variable value was in capturing the real situation. According to Figure 5, TRH data performed best, followed by census households and census population, and mobile phone users performed poorly. Both mobile phone users and the TRH data performed better in the late night and predawn hours. However, the performance of the TRH data was more stable throughout the night, and much better fitted than that of the cell phone data.

4.3. Influencing Factors Analysis

According to the minimum AIC principle, for mobile phone users and TRH data, the time periods with the best performances (03:00–05:00 on weekdays or weekends) were selected. Four types of different data were used to construct four negative binomial regression models to analyze the influencing factors of burglary, and the results are shown in Table 4.

Figure 5. AIC values of various residential population indicators in the negative binomial regression model.

Table 4. Results of the negative binomial regression model.

Variables	Model I		Model II		Model III Weekday 03:00–05:00 Phone Users		Model IV Weekend 03:00–05:00 TRH	
	Census Population		Census Households					
	IRR	Coefficient	IRR	Coefficient	IRR	Coefficient	IRR	Coefficient
Residential population	1.99 **	0.69 **	2.82 **	1.04 **	1.11 **	0.10 **	1.67 **	0.51 **
Surveillance cameras	1.02	0.02	1.01	0.01	1.09 *	0.08 *	0.98	−0.02
Road density	0.89	−0.11	1.26	0.23	0.21	−1.58	1.69	0.53
Bus stops	0.37 **	−0.98 **	0.41 **	−0.88 **	0.50 **	−0.70 **	0.81	−0.21
Apartment dormitories	1.77	0.57	1.73	0.55	1.50	0.41	1.08	0.08
Commercial–residential buildings	1.02	0.02	1.09	0.08	1.04	0.04	1.12 **	0.12 **
Residential district building	1.10	0.10	1.13	0.12	1.06	0.06	1.19 **	0.17 **
Constant	0.11 **	−2.25 **	0.22 **	−1.52 **	1.14	0.14	0.12 **	−2.15 **
/lnα		0.79		0.71		1.01		−1.15
α		2.20		2.04		2.74		0.32
AIC		693.22		687.20		715.32		564.69
Pseudo R ²		0.09		0.09		0.06		0.26

Note: AIC = Akaike's information criterion; IRR = incidence rate ratio; * $p < 0.05$, ** $p < 0.01$ (two-tailed, based on robust standard errors).

The AIC value of model IV was the smallest, and the Pseudo R^2 value of model IV also was the largest. Both suggest that model IV had a better goodness of fit than the other three models. The clustering of residuals (Figure 6) shows that the residuals were highest in the urban villages and their surrounding areas. The Moran's I indicated that models I, II, and III all had significant but low spatial autocorrelation (Moran's I was 0.22, 0.16, and 0.09, respectively), while model IV had no significant spatial autocorrelation (Moran's I was 0), further underscoring that the TRH data from 03:00–05:00 was the best measure for potential targets.

The results of the models indicated that the four types of residential population data played an integral role in measuring the distribution of burglary victims. The IRR values of the four types of residential population were significantly larger than other independent variables, which showed that it had a positive effect on the occurrence of burglaries. The incidence rate ratio (IRR) suggested that the burglaries in a grid increased by approximately 67% (IRR = 1.67) with every single unit increase in the TRH on weekends from 03:00–05:00. Interestingly, although the TRH data could reflect the population distribution more accurately, it was not the absolute real population on the numerical magnitude, and only the census population could be equivalent to the real population value. According to

models I and II, the burglaries in a grid increased by approximately 99% (IRR = 1.99) with every 100-person increase and 182% (IRR = 2.82) with every 100 households.

Figure 6. Residual clusters of various residential population indicators in the negative binomial regression model (N = 192 sample grid cells).

Theoretically, surveillance cameras represent the power of supervision and guardianship, but only showed weak or no effects in our models. This was consistent with the literature, stating that there was limited evidence of the effectiveness of CCTV for reducing burglaries, as it is easy to mask/hide a face from the camera.

Road density was not significantly correlated with burglaries in all models. This was probably related to the reality of the study area: the roads outside the villages in the city were long, straight, and simple, while the roads inside the villages in the city are mostly narrow alleys, many of which were not passable by cars. The complexity in the composition of different types of roads may have contributed to this result.

Bus stops showed weak negative effects in models I, II, and III. One possible explanation here is that offenders arrived at a locality by bus, and then moved away from the stop to a high-density residential area to find suitable targets, thus bus stops were negatively related to burglaries. The inclusion of the TRH data offset the role of bus stops, making it insignificant in model IV.

In terms of residential types, only the residential district buildings and commercial-residential buildings showed significant positive effects in model IV. In fact, in south XT JieDao, apartments and dormitory buildings were often mixed with commercial-residential buildings, with residential district buildings embedded into them in slices. Apartments and dormitories can reflect some short-term residents who are not registered. These populations also existed in the TRH data, but not in the census population. In model IV, the IRR suggested that burglaries increased by 12% (IRR = 1.12) with every single unit increase in commercial-residential buildings, and by 19% (IRR = 1.19) with every single unit increase in residential district buildings. This suggested that burglars preferred residential areas to commercial districts, likely due to a greater number of suitable and more rewarding targets. Unexpectedly, in the first three models, every residential type was not significantly correlated with burglaries. This may be due to the low data accuracy leading

to mixed information. It also proved the superiority of TRH data in the measurement of the distribution of potential victims.

5. Discussion

As Favarin said, the analysis of specific types of crime in small areas is better for understanding where and how crime happens in urban areas [37], and we confirmed that TRH data with high spatial and temporal resolution had a significant advantage in measuring the distribution of burglary targets at the mesoscale, and census population also played an important role in quantitative measurement. Compared with prior research [38], which analyzed potential victims by using only a certain kind of spatiotemporal sampling data, our research combined the precise population distribution (TRH data) with the absolute value of the population (census data), which was more realistic.

Although the TRH data was a relative population value, it reflected the real residential population at the micro- to mesoscale due to its high spatial resolution. The TRH data during night hours were viable measures of burglary targets. More potential targets attracted more burglars [39,40]. This was consistent with the rational choice theory that the number of targets and their spatial distribution strongly influence a criminal's choice [16].

As we know, most burglaries occur when a person is not in their home. There may be confusion regarding the logic of using ambient population for burglary targets, as people may argue that the ambient population can deter burglars from committing a crime. Our study actually helped answer this concern. We discovered that the active ambient population at daytime did not explain burglary well, and that only the TRH data during late night and predawn hours, when people may have gone to sleep, were viable surrogate measures of potential burglary targets. The derived variable for potential targets was then applied to the entire 24 h of a day.

The performance of mobile phone users during weekdays was similar to that during weekends. This was consistent with the findings of Song et al. [7]. However, the accuracy of mobile phone data had limitations. The service areas of a cell tower were estimated through Thiessen polygons [35], which were significantly larger than the size of individual cells. In addition to the mismatch in spatial resolution, the cell phone data was from 2014, a few years prior to the year of the TRH and the crime data. The mismatch in time may have contributed to the poor performance. Further, the cell phone data was from one of the three largest service providers, and it may not have been fully representative of all cell phone users, especially for a small study area.

The performance of the TRH during weekdays was slightly different from that of weekends [41]. However, the best performing time period was consistently 03:00–05:00, whether on weekdays or weekends. This was in line with the regularity of people's daily activities observed by Rizwan et al. based on location-based social network (LBSN) data sets. They found that the activities of people were the weakest from 03:00 to 05:00 in Shanghai, China [42]. This meant that residents most likely were asleep during this period. Therefore, the TRH data for 03:00–05:00 was a good indicator of the occupied apartments, leading to a viable measure of potential targets for burglars.

One major advantage of TRH is its high resolution of 25 m × 25 m. The other is its superior representation of social media. TRH data is captured by all apps in the big family of Tencent, the most dominant social media platform. These two factors combined help explain why the TRH outperforms the others.

Census population and census households were consistent in many aspects. However, it seemed that household is more important than population, given that the IRR was much higher for the household counts than the population counts. This result was logical, as burglars target households, not individual people.

The deterrent effect of surveillance cameras was not significant. In theory, cameras may discourage street crime [43]. However, their deterrence of burglary has not been universally supported in the literature. It is not really a surprise that this study did not reveal any significant effects of the cameras.

The results for road density were not significant, but the bus stops mainly had a negative effect. Bus stops were mainly located along main and arterial roads. The areas immediately adjacent to these roads were fairly affluent. The inner area further away from the main roads were mostly mid- to low-rise buildings, most of which were cheap rental apartments. These properties, characteristic of the urban village, greatly promoted the risk of victimization in urban villages [19,44,45]. As a result, bus stops showed negative or insignificant effects, which was consistent with the findings of Liu et al. [15].

The rapid drop of burglary cases from 2017 to 2019 might have affected the performance of the model. This drop was coupled with a rapid increase in fraud cases. The convenience and increasing popularity of mobile payments, such as WeChat Pay and Alipay, might have facilitated such a switch.

6. Conclusions

Based on the rational choice theory, taking south XT JieDao in ZG city, a megacity in South China, as the research area, we compared census population, census households, mobile phone users, and the Tencent regional heatmap (TRH) as measures of potential victims of burglary at the meso-scale of 150 m × 150 m grids. The following were the main conclusions:

- (1) The ranking of the performance from high to low was: TRH data, census households, census population, and mobile phone users.
- (2) The performance of TRH data varied in time. There existed minor differences in the performances between weekdays and weekends. The best time period for TRH data was 03:00–05:00 on weekends.

There were also limitations in our research. The housing type was determined by the POI data, which was viable but may not have been completely accurate. The study area was relatively small, so community-based census variables such as education level and population makeup were omitted. In addition, it cannot be ignored that both mobile phone data and TRH data were relative values of the actual population distribution. They may have underrepresented the location where cell phone coverage was limited and the population who were less inclined to use cell phones and apps. Another limitation was that the exact time of burglaries tends to be unknown. Further, the census data was from 2010, and the cell phone data was from 2016. All of these contributed to the limitations of applying the findings of this research for preventing burglaries.

In sum, our main innovation was that we introduced dynamic population data with a high spatial and temporal resolution to effectively measure burglary targets at the mesoscale. While we believe the findings are generally applicable, their robustness must be tested in future studies.

Author Contributions: Conceptualization, Lin Liu and Zhuofang Zhang; data curation, Zhuofang Zhang and Lin Liu; funding acquisition, Lin Liu; investigation, Zhuofang Zhang and Lin Liu; methodology, Zhuofang Zhang and Lin Liu; project administration, Lin Liu; resources, Lin Liu; supervision, Lin Liu; visualization, Zhuofang Zhang; writing—original draft, Zhuofang Zhang, Lin Liu, and Sisun Cheng; writing—review and editing, Zhuofang Zhang, Lin Liu, and Sisun Cheng. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Key Project of Guangzhou Scientific Research Plan (No. 201804020016) and the National Key R&D Program of China (Nos. 2018YFB0505500, 2018YFB0505503).

Data Availability Statement: Data of Burglary are not publicly available due to data protection regulation. The 2010 census data are openly available on request from the National Bureau of Statistics of China (<http://www.stats.gov.cn/> (accessed on 1 December 2020)). Mobile phone user data, TRH data, and basic geographic information data presented in this study can be purchased at the corresponding data provider company (the corresponding company has been specified in the article). Due to the confidentiality of crime data and purchasing agreement with the other data providers, we cannot share relevant data.

Acknowledgments: The authors would like to thank the editor and the anonymous reviewers.

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

References

1. Roth, J.; Lee, S.; Joo, J. The effect of community-level alarm ownership on burglary rates. *J. Appl. Secur. Res.* **2018**, *13*, 160–171. [\[CrossRef\]](#)
2. Xiao, L.; Liu, L.; Song, G.; Ruiter, S.; Zhou, S. Journey-to-crime distances of residential burglars in china disentangled: Origin and destination effects. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 325. [\[CrossRef\]](#)
3. Xiao, L.; Liu, L.; Song, G.; Zhou, S.; Long, D.; Feng, J. Impacts of community environment on residential burglary based on rational choice theory. *Geogr. Res.* **2017**, *36*, 2479–2491.
4. Zhong, W.; Wang, D.; Xie, D.; Yan, L. Dynamic characteristics of shanghai’s population distribution using cell phone signaling data. *Geogr. Res.* **2017**, *36*, 972–984.
5. Malleson, N.; Andresen, M.A. Exploring the impact of ambient population measures on London crime hotspots. *J. Crim. Justice* **2016**, *46*, 52–63. [\[CrossRef\]](#)
6. Duan, Y.; Liu, Y.; Liu, X.; He, D. Measuring polycentric urban structure using Easygo big data: A case study of Chongqing metropolitan area. *Prog. Geogr.* **2019**, *12*, 1957–1967. [\[CrossRef\]](#)
7. Song, G.; Liu, L.; Bernasco, W.; Xiao, L.; Zhou, S.; Liao, W. Testing Indicators of Risk Populations for Theft from the Person Across Space and Time: The Significance of Mobility and Outdoor Activity. *Ann. Am. Assoc. Geogr.* **2018**, *108*, 1370–1388. [\[CrossRef\]](#)
8. He, L.; Páez, A.; Jiao, J.; An, P.; Lu, C.; Mao, W.; Long, D. Ambient population and larceny-theft: A spatial analysis using mobile phone data. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 342. [\[CrossRef\]](#)
9. Liu, Y.; Li, W.; Wang, X. Spatiotemporal distribution of drug crimes at micro-scale: Taking NH and DM residential communities in SZ City as an example. *Prog. Geogr.* **2020**, *39*, 841–852. [\[CrossRef\]](#)
10. Liu, L.; Liang, S.; Song, G. Explaining street contact crime based on dynamic spatio-temporal distribution of potential targets. *Int. J. Geogr. Inf. Sci.* **2020**, *22*, 887–897.
11. Wang, Z.; Liu, X.; Lu, J. Multiscale geographic analysis of burglary. *Acta Geogr. Sin.* **2017**, *72*, 329–340.
12. Roh, S.; Kim, E.; Yun, M. Criminal victimization in South Korea: A multilevel approach. *J. Crim. Justice* **2010**, *38*, 301–310. [\[CrossRef\]](#)
13. Liu, D.; Song, W.; Xiu, C. Spatial analysis on robbery, forcible seizure, vehicle theft and burglary in Changchun. *Sci. Geogr. Sin.* **2014**, *34*, 1344–1352.
14. Zhang, L.; Messner, S.F.; Liu, J. A multilevel analysis of the risk of household burglary in the city of Tianjin, China. *Brit. J. Criminol.* **2007**, *47*, 918–937. [\[CrossRef\]](#)
15. Liu, L.; Jiang, C.; Zhou, S.; Liu, K.; Du, F. Impact of public bus system on spatial burglary patterns in a Chinese urban context. *Appl. Geogr.* **2017**, *89*, 142–149. [\[CrossRef\]](#)
16. Xu, J.; Chen, P.; Chen, J. Research on spatial-temporal distributions of burglary based on the environmental criminology: Based on the analysis of crime in Beijing. *Hum. Geogr.* **2018**, *33*, 43–50.
17. Sohn, D. Do all commercial land uses deteriorate neighborhood safety? Examining the relationship between commercial land-use mix and residential burglary. *Habitat Int.* **2016**, *55*, 148–158. [\[CrossRef\]](#)
18. Vandeviver, C.; Bernasco, W. “Location, Location, Location”: Effects of Neighborhood and House Attributes on Burglars’ Target Selection. *J. Quant. Criminol.* **2019**, *36*, 1–43. [\[CrossRef\]](#)
19. Chang, D. Social crime or spatial crime? Exploring the effects of social, economical, and spatial factors on burglary rates. *Environ. Behav.* **2009**, *43*, 26–52. [\[CrossRef\]](#)
20. Stokes, N.; Clare, J. Preventing near-repeat residential burglary through cocooning: Post hoc evaluation of a targeted police-led pilot intervention. *Secur. J.* **2019**, *32*, 45–62. [\[CrossRef\]](#)
21. Jones, P.A.; Brantingham, P.J.; Chayes, L.R. Statistical models of criminal behavior: The effects of law enforcement actions. *Math. Models Methods Appl. Sci.* **2011**, *20*, 1397–1423. [\[CrossRef\]](#)
22. Wang, Z.; Liu, L.; Haberman, C.; Lan, M.; Yang, B.; Zhou, H. Burglaries and entry controls in gated communities. *Urban Stud.* **2021**. [\[CrossRef\]](#)
23. Long, D.; Liu, L.; Feng, J.; Song, G.; He, Z.; Cao, J. Comparisons of the community environment effects on burglary and outdoor-theft: A case study of ZH peninsula in ZG city. *Acta Geogr. Sin.* **2017**, *72*, 341–355.
24. Song, G.; Xiao, L.; Zhou, S.; Long, D.; Zhou, S.; Liu, K. Impact of residents’ routine activities on the spatial-temporal pattern of theft from person. *Acta Geogr. Sin.* **2017**, *72*, 356–367.
25. Montoya, L.; Junger, M.; Ongena, Y. The relation between residential property and its surroundings and day- and night-time residential burglary. *Environ. Behav.* **2014**, *48*, 515–549. [\[CrossRef\]](#)
26. Armitage, R. Domestic burglary: Burglar responses to target attractiveness. In *Reducing Burglary*; Springer International Publishing: Cham, Switzerland, 2018; pp. 45–75.
27. Li, L.; Lin, J.; Li, X.; Wu, F. Redevelopment of urban village in China—A step towards an effective urban policy? A case study of Liede village in Guangzhou. *Habitat Int.* **2014**, *43*, 299–308. [\[CrossRef\]](#)
28. Xu, C.; Liu, L.; Zhou, S.; Jiang, C. Spatial heterogeneity of micro-spatial factors’ effects on street robberies: A case study of DP peninsula. *Geogr. Res.* **2017**, *12*, 2492–2504.

29. Lyu, X. The characteristics of the criminal intelligence predictive analytical technology in the united states: An analysis from rand's report. *J. Intell.* **2016**, *35*, 7–12.
30. Griffith, D.A.; Amrhein, C.G.; Desloges, J.R. Statistical analysis for geographers. *J. Am. Stat. Assoc.* **1999**, *94*, 654.
31. Niu, X.; Ding, L.; Song, X. Understanding urban spatial structure of Shanghai central city based on mobile phone data. *Urban Plan. Forum.* **2014**, *6*, 71–77.
32. Wu, Z.; Ye, Z. Research on urban spatial structure based on Baidu heat map: A case study on the central city of Shanghai. *City Plan. Rev.* **2016**, *40*, 33–40.
33. Shen, L.; Zhang, C.; Li, H.; Wang, Y.; Wang, Z. Relationship between urban rail transit commuting and jobs-housing balance: An empirical analysis from Beijing based on big data methods. *Prog. Geogr.* **2019**, *38*, 791–806.
34. Chen, W.; He, X. Social work driven by big-data: Prospects and challenges. *J. Soc. Sci.* **2017**, *7*, 74–81.
35. Feng, J.; Liu, L.; Long, D.; Liao, W. An examination of spatial differences between migrant and native offenders in committing violent crimes in a large Chinese city. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 119.
36. Ratcliffe, J.H. Aoristic signatures and the temporal analysis of high volume crime patterns. *J. Quant. Criminol.* **2002**, *18*, 23–43. [[CrossRef](#)]
37. Favarin, S. This must be the place (to commit a crime). Testing the law of crime concentration in Milan, Italy. *Eur. J. Criminol.* **2018**, *15*, 702–729. [[CrossRef](#)]
38. Hanaoka, K. New insights on relationships between street crimes and ambient population: Use of hourly population data estimated from mobile phone users' locations. *Environ. Plan. B Urban Anal. City Sci.* **2016**, *45*, 295–311. [[CrossRef](#)]
39. Mburu, L.W.; Helbich, M. Crime risk estimation with a commuter- harmonized ambient population. *Ann. Am. Assoc. Geogr.* **2016**, *106*, 804–818. [[CrossRef](#)]
40. Andresen, M.A. Crime measures and the spatial analysis of criminal activity. *Br. J. Criminol.* **2006**, *46*, 258–285. [[CrossRef](#)]
41. Muhammad, R.; Zhao, Y.; Liu, F. Spatiotemporal analysis to observe gender based check-in behavior by using social media big data: A case study of Guangzhou, China. *Sustainability* **2019**, *11*, 2822. [[CrossRef](#)]
42. Rizwan, M.; Wan, W.; Gwiazdzinski, L. Visualization, spatiotemporal patterns, and directional analysis of urban activities using geolocation data extracted from LBSN. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 137. [[CrossRef](#)]
43. Liu, L.; Jiang, C.; Li, L. The effects of police CCTV camera on crime: A case study from Gusu district in Suzhou, China. *Sci. Geogr. Sin.* **2019**, *39*, 61–69.
44. Davies, T.; Johnson, S.D. Examining the relationship between road structure and burglary risk via quantitative network analysis. *J. Quant. Criminol.* **2015**, *31*, 481–507. [[CrossRef](#)]
45. Zawisza, T.; Garza, R. Using an eye tracking device to assess vulnerabilities to burglary. *J. Police Crim. Psychol.* **2017**, *32*, 203–213. [[CrossRef](#)]