Cities 166 (2025) 106187

Contents lists available at ScienceDirect

Cities

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

Housing vacancy rate estimation in high-rise residential communities: An experiment utilizing multi-sourced data in Beijing

Huimin Zhao [a], Changcheng Kan [b], Ying Long [a,c,*]

[a] School of Architecture, Tsinghua University, China
[b] School of Urban Planning and Design, Peking University Shenzhen Graduate School, China
[c] Hang Lung Center for Real Estate, Key Laboratory of Eco Planning & Green Building, Ministry of Education, Tsinghua University, China

ARTICLE INFO

Keywords:
Housing vacancy rate
Building resource wastage
Location-based services data
Mobile signaling data
Nighttime light data
Wi-fi probe

ABSTRACT

Amidst rapid urbanization, high-rise residential buildings have emerged as the predominant housing typology, a trend projected to persist and intensify. However, a significant proportion of these residential buildings exhibit high vacancy rates, resulting in significant inefficiencies and waste. Accurately ascertaining the housing vacancy rate (HVR) at the residential community scale is therefore essential for improving resource allocation. Traditional methods for evaluating HVR often face challenges related to limited spatial coverage and insufficient resolution, particularly in high-rise urban settings. This study compares several methods for estimating HVR at the residential community level, including the use of location-based services (LBS) data, mobile signaling data, and three types of nighttime light data. These methods were applied to 5737 residential communities within Beijing's Fifth Ring Road. To validate the accuracy of these approaches, field surveys were conducted in 103 representative communities, utilizing Wi-Fi probes to estimate household counts. The results indicate that LBS data yields the highest accuracy (correlation coefficient: 0.76), underscoring its effectiveness and potential for large-scale HVR assessment. Furthermore, the global accessibility of LBS data suggests its applicability for HVR estimation.

1. Introduction

Amidst the global trend of urbanization, high-rise urban environments have become a defining feature of contemporary cities, characterized by concentrated populations and high building density (Wang & Shaw, 2018). For example, in 2019, the United States constructed approximately 352,000 high-rise residential buildings (Mosey & Deal, 2021). In major cities such as New York and Chicago, high-rise residential buildings are prevalent, including apartment buildings and multi-family dwellings (Scheu & Garascia, 2016). Similarly, in metropolitan areas like Tokyo and Osaka, high-rise apartments constitute the dominant housing form (Peng & Wang, 2022; Sano, Filipović, & Radović, 2020). In China, large cities such as Beijing, Shanghai, Guangzhou, and Shenzhen also feature predominantly high-rise residential buildings (Zhang, Zhao, & Long, 2025). Driven by accelerated urbanization and limited land availability in central urban areas, the proliferation of high-rise housing has become increasingly prominent (Sun, 2020).

The occupancy and utilization of high-rise residential buildings play a crucial role in global energy consumption and resource use (Nie, Kemp, & Fan, 2023; Olaya, Vásquez, & Müller, 2017; Yang, Hu, Zhang, & Steubing, 2022; Zhang, Li, Chen, Yan, & Liu, 2023). Within this context, the issue of housing vacancy, defined as residential units that remain unoccupied for extended periods, has emerged as a key indicator of resource inefficiency, attracting growing attention from both academic researchers and policymakers (Flas, Halleux, Cools, & Teller, 2022).

Beyond its association with resource waste, housing vacancy generates significant negative externalities. High housing vacancy rate (HVR) may indicate speculative real estate bubbles and can lead to declines in both residential and nearby property values (Jang, Ahn, Kim, & Song, 2018; Leskinen, Vimpari, & Junnila, 2020). Additionally, elevated vacancy levels are linked with increased psychological stress, reduced subjective well-being, and heightened feelings of fear, anxiety, and distrust, which can ultimately contribute to adverse health outcomes (Newman, Li, & Park, 2022; Snedker & Herting, 2016). Vacant housing is also frequently associated with higher incidences of violent crime (Arnio, Baumer, & Wolff, 2012; Yonas, O'Campo, Burke, & Gielen, 2007). Moreover, long-term vacancies can trigger a downward spiral of neighborhood decline, resulting in further property abandonment.

* Corresponding author at: School of Architecture, Tsinghua University, Haidian District, Beijing 100084, China.
E-mail addresses: zhaohm21@mails.tsinghua.edu.cn (H. Zhao), ylong@tsinghua.edu.cn (Y. Long).

https://doi.org/10.1016/j.cities.2025.106187
Received 21 March 2024; Received in revised form 29 March 2025; Accepted 17 June 2025

Therefore, accurately assessing HVR at the residential community level is essential for mitigating these negative externalities, enhancing resource efficiency, and supporting broader goals such as sustainable urban development and carbon neutrality.

Unfortunately, housing vacancy is a widespread issue observed across both developed and developing countries. Recent studies and reports have documented its growing prevalence globally. In the United States, the number of vacant housing units increased by 44 % between 2000 and 2010 (US Government Accountability Office, 2011). A similar trend is evident in Asia; for instance, Japan's HVR rose from 2.5 % in 1963 to 13.6 % by 2018 (Kubo & Mashita, 2020). In China, a 2015 survey reported HVR ranging from 22 % to 26 % in major cities (Pan & Dong, 2021). These statistics highlight the global relevance of housing vacancy and underscore its importance as a critical issue that transcends economic development status.

While countries such as the United States and Japan have incorporated HVR into their national statistical systems, considerable variation exists in the methods and standards used to calculate and measure HVR, often shaped by differing national contexts (Flas et al., 2022). Additionally, the HVR data is frequently reported at aggregated administrative levels, which constrains the ability to conduct detailed analyses of regional housing conditions (U.S. Government Accountability Office, 2011). Notably, in some countries or regions, including China, HVR is not included in official statistical reporting, resulting in a lack of government-released data on this metric. With the advent of big data, researchers have increasingly explored using alternative data sources, such as nighttime light data, to estimate HVR. This approach enables a more granular analysis of spatial heterogeneity within urban areas (Chen et al., 2015). However, due to the relatively coarse resolution of spatial grids of the data, existing studies are typically constrained to analyses conducted at scales ranging from 130 to 1000 m. (Liu et al., 2023; Yang & Pan, 2022). Although there is a growing interest in analyzing residential blocks (Shi et al., 2022), current data limitations, along with the prevalence of high-rise housing, hinder the ability to examine housing vacancy at the finer scale of individual residential communities (Fig. 1).

Accordingly, this study develops and evaluates multiple methodologies for accurately estimating HVR at the residential community level in high-rise urban environments, using central Beijing as a case study. The findings aim to establish a foundational framework for future research in this and related domains, offering new insights and methodological approaches for large-scale, residential community-level HVR estimation.

2. Literature review

2.1. Estimating HVR with traditional data

Traditional methods for estimating HVR include the use of governmental surveys, taxation records, field investigations, and electricity consumption data.

Governmental data sources primarily comprise census surveys and taxation records. Surveys, typically conducted through official census efforts, provide demographic and housing information, often accessible via digital government archives. For instance, administrative data from the United States Postal Service (USPS) and the American Community Survey (ACS) offer insights into vacant housing units at the census tract level (Immergluck, 2016; Wegmann, 2020). However, such data often lack the spatial granularity required for detailed vacancy analysis. While taxation records can be used to identify vacant housing (Wyatt, 2008), their availability and utility vary significantly across countries and regions.

To address these limitations, researchers have employed structured interviews and questionnaires to gather vacancy-related data within specific regions (Fahmi & Sutton, 2008; Ma, Zhou, & Fang, 2003). However, as sampling-based methods, these surveys often lack comprehensive coverage. Concurrently, field investigations, such as nighttime photography to observe lighting as a proxy for occupancy, or inspections of property exteriors and yards, have been used to assess housing status (Akiyama et al., 2020; Yao, Teng, & Li, 2013). While these approaches yield detailed household-level data, they are highly labor-intensive, time-consuming, and difficult to scale. Furthermore, they are typically based on short-term observations and rely heavily on subjective judgments, which may compromise the accuracy of vacancy assessments.

Electricity consumption data has also been utilized to identify vacant housing by analyzing usage patterns over time (Jiang, Cheng, Xiang, Yin, & Wang, 2020; Wang et al., 2023). Although this method is effective in detecting occupancy status, access to such data is often restricted due to privacy concerns and requires collaboration with utility providers, thereby limiting its applicability in broader research contexts.

2.2. Estimating HVR with new data

The emergence of new data sources and technological advancements has significantly improved the capacity to estimate HVR.

Among these, nighttime light data has become an increasingly popular data due to its broad spatial coverage and ability to capture illumination patterns indicative of human activity. Three commonly used nighttime light datasets in HVR research include DMSP/OLS (Yao & Li, 2011), NPP/VIIRS (Chen et al., 2015; Liu et al., 2023), and Luojia-01 (Tan, Wei, & Yin, 2020; Yang & Pan, 2022), with their spatial resolutions of approximately 1000 m, 500 m, and 130 m, respectively. While these datasets are publicly accessible, they face limitations in spatial granularity and challenges in filtering non-residential light sources, such as street lights. In 2022, the Changguang Satellite Technology Company launched the Jilin-01 satellite, capable of capturing nighttime light data at a remarkable spatial resolution of 0.92 m (Du, Wang, Zou, & Shi,

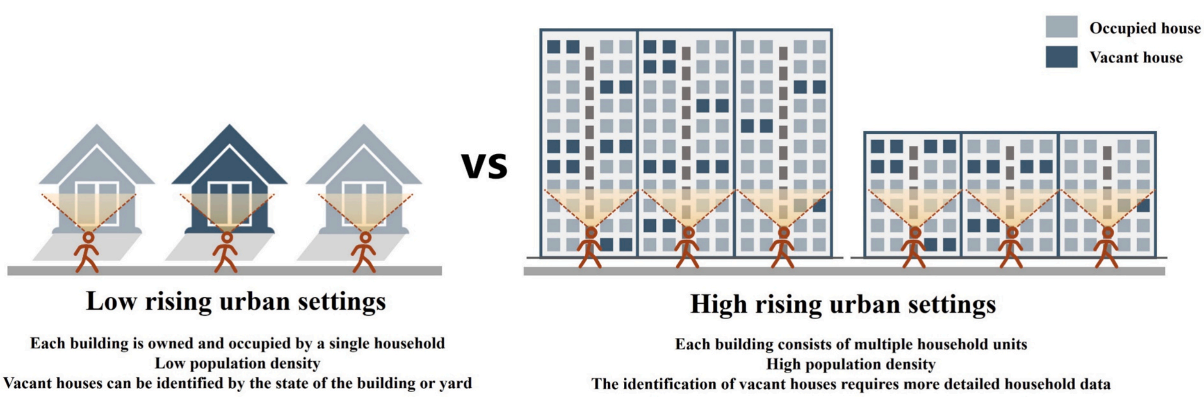


Fig. 1. Differences in the identification of vacant houses between low and high-rise urban settings.

2018). Although this advancement presents new opportunities for detailed HVR analysis, the data is not publicly accessible and typically requires customized satellite imaging services, which limits scalability and widespread application.

In addition to nighttime light data, some researchers have explored the use of high-resolution remote sensing imagery (Zou & Wang, 2021b) and street view images (Zou & Wang, 2021a) to identify vacant housing, using visual indicators such as signs of habitation or prolonged lack of maintenance as proxies for occupancy status. While these image-based data sources are generally accessible, their applicability is largely confined to single-family housing contexts and remains limited in high-rise urban environments, where multiple households occupy a single building.

With the widespread adoption of mobile devices, location-based services (LBS) data and mobile signaling data have emerged as valuable resources for analyzing urban human activity patterns. Previous studies have employed these data sources to estimate population density and mobility (Jin et al., 2017; Yao et al., 2017). However, their use in HVR estimation is still nascent, in part due to data accessibility challenges, as they are typically obtained through commercial purchase from internet companies or telecommunications providers.

In summary, while previous research has made significant progress in identifying vacant housing and estimating HVR through innovative data sources and methodologies (Table 1), challenges persist (Fig. 2). These include variations in spatial resolution across datasets, their contextual suitability for different urban environments, and the uncertainty of HVR estimates at fine spatial scales. In response, this study employs LBS data, mobile signaling data, and nighttime light data to estimate HVR and evaluate their comparative accuracy, with the goal of refining residential community-level HVR estimation methods in high-rise urban environments.

3. Research design

3.1. Study area and data collection

(1) Study area

This study uses Beijing as its case study area. As China's capital, Beijing serves as a major hub for politics, culture, education, technology, and international exchange. High-rise buildings dominate the architectural landscape in central Beijing. However, publicly available government data on HVR is currently lacking, underscoring the need to better understand the city's vacancy dynamics. This study focuses on the urban area within Beijing's Fifth Ring Road, treating all residential communities in this zone as research objects and basic units of analysis.

(2) Data description and preprocessing

The paper utilizes multi-source big data (Table 2), with the core datasets comprising five types: Baidu Map LBS data, Unicom mobile signaling data, Jilin-01 nighttime light data, Luojia-01 nighttime light data, and NPP/VIIRS nighttime light data. All data processing procedures were anonymized to ensure the protection of individual privacy.

3.1.1. Baidu Map LBS data

The LBS data used in this study is derived from statistical information obtained through the Baidu Maps Location Software Development Kit (SDK). Baidu Maps processes over 120 billion location service requests daily and supports more than 1.2 billion active smart devices per month, covering all administrative regions in China. The study extracted smartphone location features and their temporal distribution from internet-based location data between February and June 2022. Based on nighttime location patterns and other indicators, the spatial locations of residential dwellings were identified with high accuracy (See Appendix 1 for details). It is acknowledged that Baidu LBS data may not represent the entire population within the study area. Consequently, the number of residents in each community was proportionally adjusted based on official demographic statistics. All relevant data processing steps were anonymized, and no individual privacy was compromised at any stage or in any output.

3.1.2. Unicom mobile signaling data

China Unicom, one of the major telecommunication operators in China, estimates residential locations nationwide using the most frequently observed nighttime positions after user account registration. To account for potential discrepancies between the number of Unicom users and the total population, adjustments were made accordingly. In this study, Unicom mobile signaling data from July 2022 was used to estimate the total population of each residential community.

3.1.3. Nighttime light data

Three sources of nighttime light data were employed: Jilin-01, Luojia-01, and NPP/VIIRS. The Jilin-01 satellite offers a high spatial resolution of 0.92 m. Jilin-01 nighttime light data was acquired for the study area, specifically between October and December 2022, from 20:00 to 21:00 on weekdays.[1] Furthermore, Luojia-01 nighttime light data from September 2018, with a spatial resolution of 130 m, was selected due to its cloud-free quality. The NPP/VIIRS nighttime light data from September 2022, characterized by a 500 m resolution, was also utilized. In accordance with established research practices (Tang, Du, Lin, Guo, & Qie, 2020), all three datasets underwent preprocessing steps including coordinate projection, geographic registration, radiometric calibration, and the removal of anomalous values.

3.2. Methodology

This paper follows a five-step process. First, the boundaries of residential communities are delineated, followed by the calculation of household capacity and the actual household count. The performance of various datasets is validated to identify the highest-performing high-resolution data. The workflow is illustrated in Fig. 3.

(1) Identification of community boundaries

The boundary data of residential communities was extracted using AOI data and land use data. In addition, POI data and remote sensing imagery were incorporated to refine certain boundaries. This study focuses exclusively on residential communities completed prior to 2022, excluding those under construction in 2022. Moreover, given that some residential areas in Beijing still consist of self-built houses, such as courtyard residences, which typically feature low building heights and do not align with the high-rise urban context, the study calculated the average building height for each residential area and excluded those with fewer than two stories. After this filtering process, 5737 valid residential communities were identified, covering a total area of 140.1km^2. To minimize the influence of non-residential areas, such as schools and commercial buildings, on the results, non-residential zones were manually excluded from these boundaries. The final dataset consists of 5737 adjusted residential community boundaries, covering an area of 139.7 km^2 (Fig. 4). The size of each residential community ranges from 0.054 ha to 61.70 ha, with an average area of 2.44 ha and a standard deviation of 3.19 ha. Within the study area, the average building density, defined as the ratio of the total building area to the area of the residential community, is 2.62, and the average number of building floors is 9.06.

(2) Calculation of household capacity

[1] This timeframe is strategically chosen to precede most urban residents' bedtime, as suggested by Feng et al. (2020).

Table 1
Data used for the evaluation of HVR.

Data source		Spatial resolution	Data collection method	Research unit	Scalability in high-rise urban settings	Reference
Government survey data		Low	Open access	Administrative unit	Low	Immergluck, 2016; Wegmann, 2020
Taxation record data		High	Government requests	House	Low	Wyatt, 2008
Field survey data		High	Self-collection	House	Low	Akiyama et al., 2020; Yao et al., 2013
Interviews and questionnaires		High	Self-collection	House	Low	Ma et al., 2003;
Electricity consumption data		High	Government requests	House	Low	Jiang et al., 2020; Wang et al., 2023
Nighttime light data	DMSP /OLS	Low (1 km)	Open access	Grid(1 km)/ City	Low	Yao & Li, 2011
	NPP /VIIRS	Moderate (500 m)	Open access	Grid(500 m)/ City	High	Chen et al., 2015; Liu et al., 2023
	Luojia −01	High (130 m)	Open access	Grid(130 m)/ Residential blocks	High	Shi et al., 2022; Tan et al., 2020
	Jilin-01	High (0.92 m)	Purchase	Grid (0.92 m)	High	Du et al., 2018
Remote sensing image data		High	Open access	House	Low	Zou & Wang, 2021b
Street view image data		High	Open access	House	Low	Zou & Wang, 2021a
LBS data		High	Purchase	Residential community	High	Jin et al., 2017
Mobile signaling data		High	Purchase	Residential community	High	Yao et al., 2017

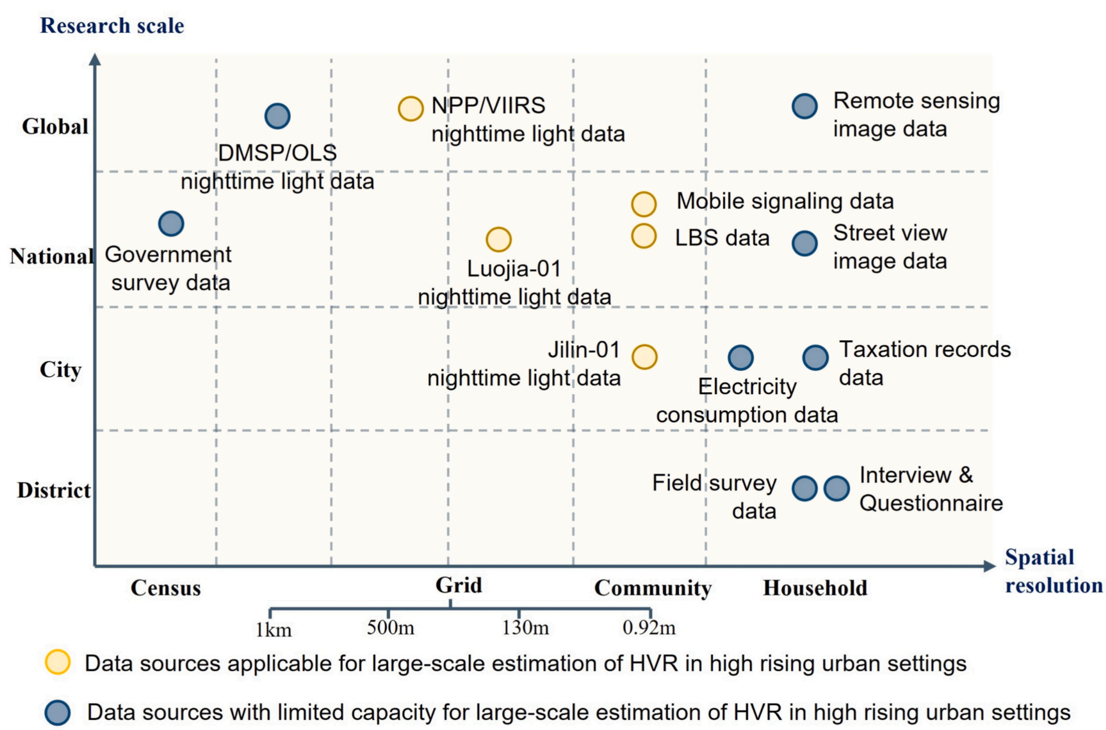


Fig. 2. Spatial resolution of various data sources and their corresponding appropriate scales.

The building data used in this study includes both the building footprints and their respective heights. The footprint area of each building was calculated, and the number of floors for each building was determined, assuming a residential building height of 3 m per floor. The total area of each building was then calculated by multiplying its footprint area by the number of floors. By integrating the area designated for each residential unit, an estimation of the household capacity per building can be derived. This paper includes a comprehensive classification of buildings to more accurately estimate household capacity.[2] In cases where data on unit areas are unavailable for specific residential communities, the average unit area for that building category across all communities is used as a proxy. By summing up the total household capacity of all buildings, the aggregate household capacity for each residential community was calculated (Eq. 1).

[2] Specifically, the identified building types include villas and up to 6 floors, 7–18 floors, 19–33 floors and >33 floors residential buildings. These classifications are primarily based on the prevalent distribution of floor numbers observed in residential structures throughout China (Li et al., 2023).

Table 2
Data description and application.

Application	Data	Spatial resolution	Time	Source	Data information
Identification of residential community boundary	Point of interest (POI) data	Point	2022	Baidu Maps	Latitude, longitude and category of POIs
	Area of interest (AOI) data	Parcel	2022	Baidu Maps	Location and category of AOIs
	Land use data	Parcel	2014	Government	Residential land boundary in Beijing
	Remote sensing image data	0.3 m	2022	Google Maps	Remote sensing images for the area within Beijing's Fifth Ring Road
Calculation of household capacity	3D building data	Building	2023	Baidu Maps	Building footprint and building height
	Residential unit area data	Unit	2022	Anjuke.com	Area, type and location of unit
Calculation of housing count	LBS data	Point	September 2022	Baidu	Latitude and longitude of each user's residence
	Mobile signaling data	Point	July 2022	Unicom	Latitude and longitude of each user's residence
	Jilin-01 nighttime light data	0.92 m	October to December 2022	Changguang Satellite Technology Company	Nighttime light remote sensing imagery for the area within Beijing Fifth Ring Road
	Luojia-01 nighttime light data	130 m	September 2022	Wuhan University et al.	Nighttime light remote sensing imagery for the area within Beijing's Fifth Ring Road
	NPP/VIIRS nighttime light data	500 m	September 2022	NASA and NOAA	Nighttime light remote sensing imagery for the area within Beijing Fifth Ring Road
	Household size data	District	2020	Seventh National population survey	Household size of the district
	Permanent population data	Sub-district	2020	China's seventh national population census	The number of permanent residents in each sub-district

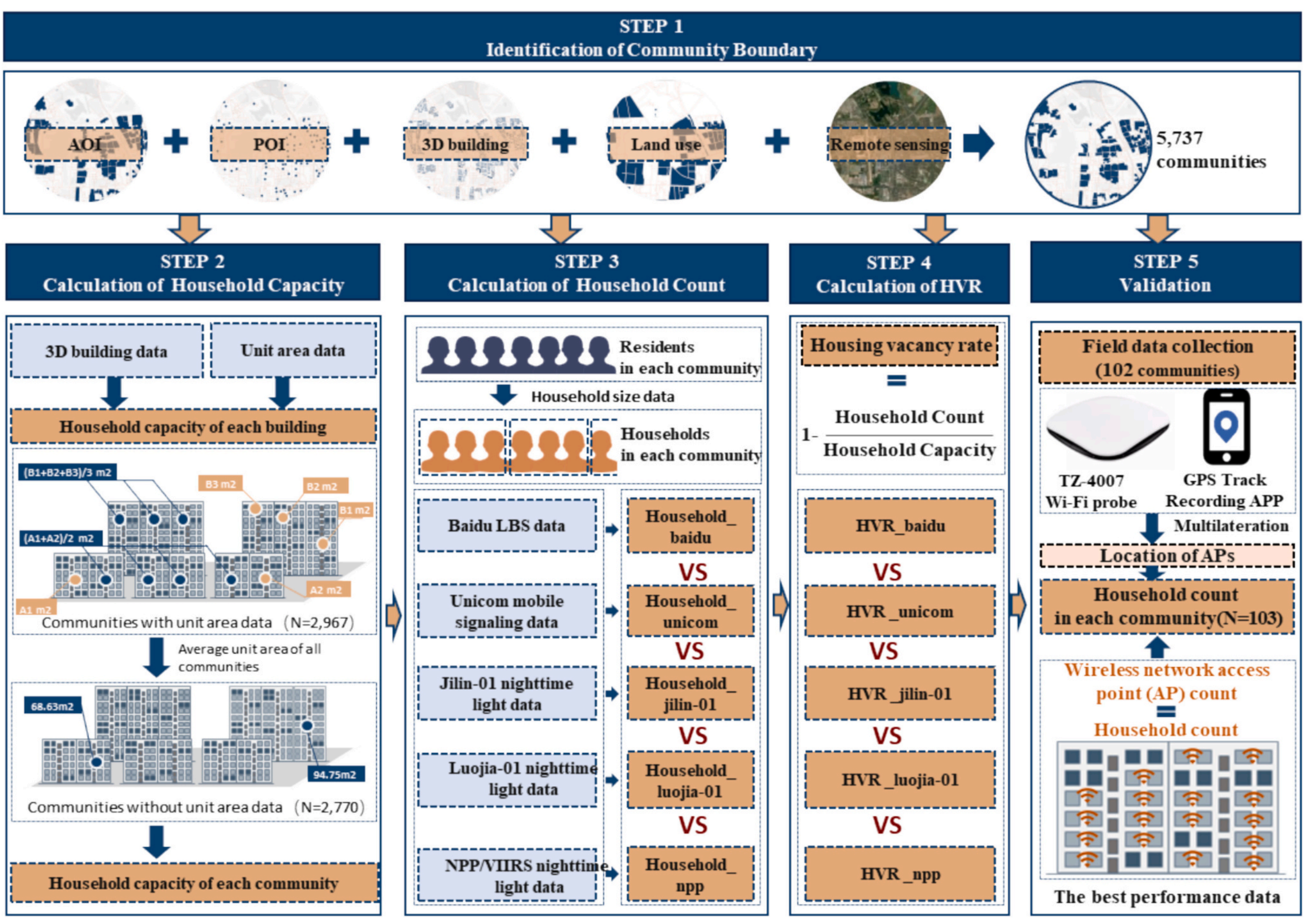


Fig. 3. Methodology flowchart

$$\text{Capacity}_j = \sum_i \frac{A_{base\_i} \times H_i}{A_{avg\_unit\_type,i} \times 3} \quad (1)$$

Where Capacity_j represents the household capacity of community j; i denotes each building within community j; A_{base_i} refers to the base area of building i; H_i indicates the height of building i; $A_{avg\_unit\_type,i}$ represents the average unit area for the specific building type; the constant value of 3 represents the height of each floor in residential buildings.

(3) Calculation of household count

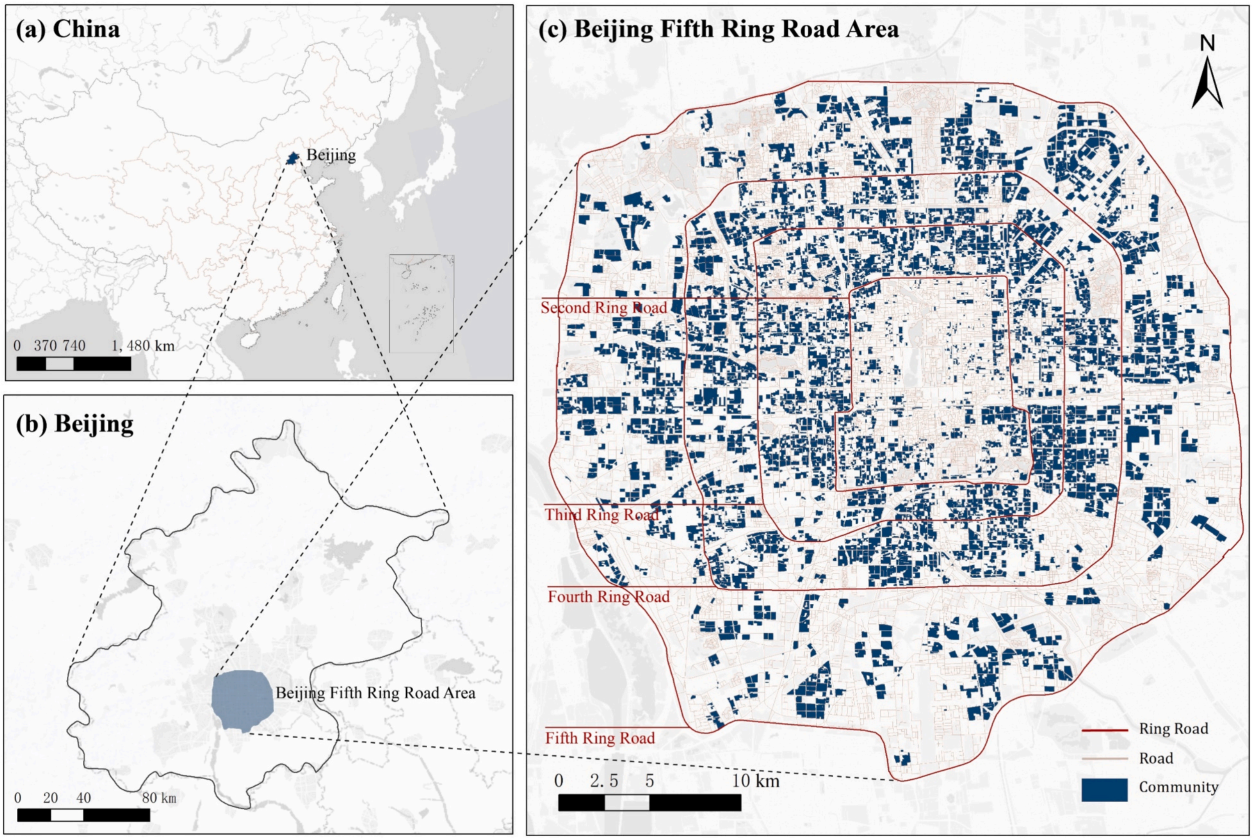


Fig. 4. Study area and the residential community boundaries.

The actual household count of various residential communities was estimated using five distinct datasets. The estimation methodologies applied to Baidu Map LBS data and Unicom mobile signaling data are similar. The specific formula applied to calculate the actual household count with Baidu Map LBS data is represented as Eq. 2, while the formula of Unicom mobile signaling data is given in Eq. 3.

$$Household_{baidu,j} = Pop_{baidu,j} \Big/ House\_size_j \tag{2}$$

$$Household_{unicom,j} = Pop_{unicom,j} \Big/ House\_size_j \tag{3}$$

Where $Household_{baidu,j}$ represents the household count of community j as derived from Baidu LBS data; $Pop_{baidu,j}$ denotes the population count of community j according to Baidu LBS data; $Household_{unicom,j}$ represents the household count of community j as derived from Unicom mobile signaling data; $Pop_{unicom,j}$ denotes the population count of community j according to Unicom mobile signaling data;$House_size_j$corresponds to the household size of community j (that is, the household size of the district where the community is located).

The study also employed three sources of nighttime light data to estimate the actual household count within residential communities. Building on a hypothesis established in prior research, it was postulated that a linear relationship exists between the number of households and the nocturnal light emissions from residential buildings, characterized by a constant linear coefficient (Shi et al., 2022). To determine the N value for non-vacant household units, the study drew on previous research (Huang and Zhao, 2023; Wang, Fan, & Wang, 2019) and assumed that residential communities located in the city center are fully occupied. A total of 382 residential community samples located in the city center were selected for analysis. These residential communities were considered fully occupied, meaning the housing capacity equaled the number of households. Based on this assumption, the study calculated the average household N value of these communities and applied it as the estimated N value for non-vacant housing units. The actual number of households in a residential community was determined by taking the ratio of the community's N value to the N value of non-vacant housing units (Eq. 4–5) (See Appendix 2 for details).

$$N_{avg,j} = N_{Rj} \Big/ Capacity_j \tag{4}$$

$$Household_{night\ time\ light\ data,j} = N_{Rj} / N_{avg\_median} \tag{5}$$

Where $N_{avg,j}$ refers to the average N value per household within the community j; N_{Rj} denotes the DN value in the community j; $Capacity_j$represents the household capacity of the community j; N_{avg_median} is the median of $N_{avg,j}$ across all communities. $Household_{night\ time\ light\ data,j}$ represents the household count of community j as derived from different nighttime light data.

(4) Estimation of HVR

The calculation formula (Eq. 6) for the HVR used by all five datasets is consistent, defined as the difference between 1 and the ratio of the actual household count to the household capacity within a residential community.

$$HVR_j = 1 - Household_j \Big/ Capacity_j \tag{6}$$

Where HVR_j refers to the HVR of the community j; $Household_j$ represents the household count of community j;

$Capacity_j$represents the household capacity of the community j.

(5) Validation of HVR

To validate the accuracy of different datasets, the study selected 103 representative residential communities for field-based Wi-Fi probe data collection. The HVR derived from the field data was used as the ground truth, against which the HVR values from the various datasets were compared. The detailed validation process is described in the following Results section.

4. Results

4.1. Household capacity in Beijing

Within the Beijing Fifth Ring Road, residential communities collectively accommodate 3.71 million households, with an average household capacity of 587 households. The maximum capacity is 9834 households, while the minimum is 12 households, with a standard deviation of 721.

4.2. Household count in Beijing obtained from various big data

Distinct variations are observed in household count estimates derived from different data sources (Table 3). Specifically, the figures from Baidu LBS, Unicom mobile signaling, and Jilin-01 nighttime light data all exceed the total household capacity within the study area. Baidu LBS data recorded the highest actual household count, with 3,614,545 households, averaging 630 per community. Unicom mobile signaling data reported a slightly lower total of 3,461,831 households, with an average of 603 per community. Jilin-01 nighttime light data estimated 3,035,028 households, averaging 529 per community. In contrast, both Luojia-01 and NPP/VIIRS nighttime light data reported household count below the area's capacity. Luojia-01 data recorded 2,237,763 households (average of 390), while NPP/VIIRS data recorded 1,140,533 households (average of 199). The distribution map (Fig. 5 a-e) further highlights these disparities.

Analysis of pairwise correlations among the five data estimates (Fig. 5f) reveals that, aside from the strong correlation of 0.80 between the Baidu and Unicom data, correlations between other data pairs are below 0.60. These findings highlight the need for validation of performance of different data.

4.3. Validation of HVR

(1) Field data collection

This study conducted field surveys across 103 representative residential communities (Fig. 6a), selected for their geographical diversity and the variety of architectural typologies they encompass (See Appendix 3 for details). The average number of floors in the buildings across these communities was 9.02, ranging from 3 to 34 floors. To validate the household count estimations from the five datasets, the study assumed that each occupied house in these communities was equipped with a wireless network access point (AP), such as Wi-Fi routers (See Appendix 4 for details).

Between August 15 and 23, 2023, four volunteers equipped with Wi-Fi probes, GoPro cameras and smartphones running the 2bulu application gathered data by cycling through every road within the residential communities (Fig. 6b). During data collection, the volunteers maintained a speed of no more than 7 km/h, and the Wi-Fi probes were not obstructed by metal objects or other coverings. The Wi-Fi probe used was the Zhongke Aoxun TZ-4007, with a detection range of up to 100 m. It recorded data every 5, s, including source and destination MAC address, frame type, frame subtype, signal strength, and timestamp. Concurrently, the 2bulu application tracked GPS trajectories, while the GoPro camera captured visual images of the community. The combined data collection efforts resulted in a comprehensive dataset, which included 10,913,632 entries of probe data, 103 GPS trajectories, and 345.91 GB of image data.

(2) Data processing

The study focused on extracting data characterized by a frame type of 0 and a frame subtype of 8, as defined in the 802.11 protocol, which is typically used to identify AP devices. A total of 1,263,568 records, corresponding to 168,436 APs, were extracted. The location of the probe at the time of AP signal detection was determined by synchronizing the data collection timestamp with the corresponding GPS trajectory (Fig. 6c and Fig. 6d). Assuming an inverse correlation between signal strength and the distance from the probe to the AP, multilateration techniques were used to estimate the likely positions of each AP (Fig. 6e). This approach was crucial in excluding APs located outside the communities.

(3) Correlation analysis

The number of routers within a residential community, extracted from the Wi-Fi probe data, was used as the define $Household_{Wi-Fi}$, and the HVR_{Wi-Fi} for the 103 communities was calculated. A bivariate correlation analysis was performed on a sample of these 103 typical residential communities, with HVR_{Wi-Fi} as the dependent variable and the results from the five datasets as independent variables, to assess the accuracy of the datasets. When compared to the Wi-Fi probe data, the Baidu LBS data exhibited a strong correlation coefficient of 0.76 ($p < 0.001$), followed by the Unicom mobile signaling data at 0.69 ($p < 0.001$) and the Jilin-01 nighttime light data at 0.66 ($p < 0.001$). In contrast, the LuoJia-01 and NPP/VIIRS nighttime light datasets showed comparatively weaker correlations. The LuoJia-01 dataset had a correlation coefficient of 0.46 ($p < 0.001$), while the NPP/VIIRS dataset demonstrated a minimal correlation, with a coefficient of only 0.23($p < 0.001$).

4.4. HVR at residential community level in Beijing

In conclusion, the Baidu LBS data demonstrates the best performance among the five datasets. The results indicate that the average HVR of residential communities within Beijing's Fifth Ring Road is 13.31 %. It was observed that 3932 residential communities had an HVR of <10 %, while 592 residential communities exhibited an HVR exceeding 50 %. The spatial distribution of HVR across these communities is illustrated in Fig. 7. Residential communities with higher HVRs are primarily concentrated on the outskirts of the study area.

The study selected residential communities with varying HVRs and collected on-site images. Residential communities 1 and 2, characterized by relatively lower vacancy rates, showed a noticeable aggregation of both motorized and non-motorized vehicles on their roads, along with visible resident activity. Residential community 3, with a moderate vacancy rate, exhibited some signs of residential life; however, the presence of vehicles was less evident compared to the first two communities. Residential communities 4 and 5, both exhibiting higher vacancy rates, displayed distinct characteristics. The former was undergoing partial building renovations, while the latter was a newly constructed villa area completed in 2021. Overall, the Baidu LBS data was found to closely align with the Wi-Fi data findings.

Table 3
Descriptive statistics of household capacity and household counts ($N = 5737$).

	Total	Mean	Min	Max	Std.
Capacity	3,708,109	587	12	9834	721
$Household_{baidu}$	3,614,545	630	0	13,042	838
$Household_{unicom}$	3,461,831	603	0	12,339	934
$Household_{jilin-01}$	3,035,028	529	8	27,330	864
$Household_{luojia-01}$	2,237,763	390	20	11,793	425
$Household_{npp}$	1,140,533	199	0	2286	216

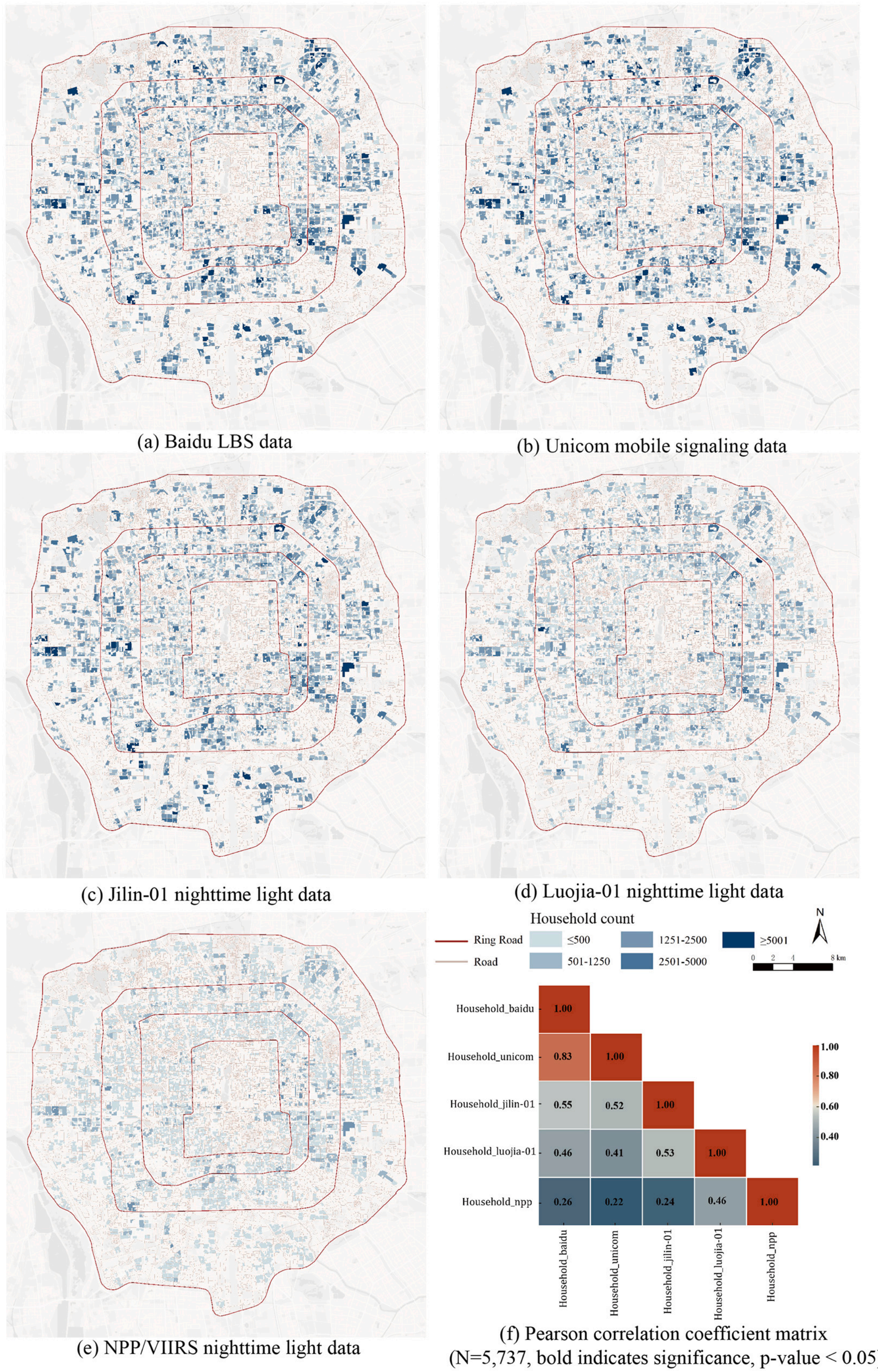


Fig. 5. The distribution and Pearson correlation analysis of diverse household count outcomes (N = 5737).

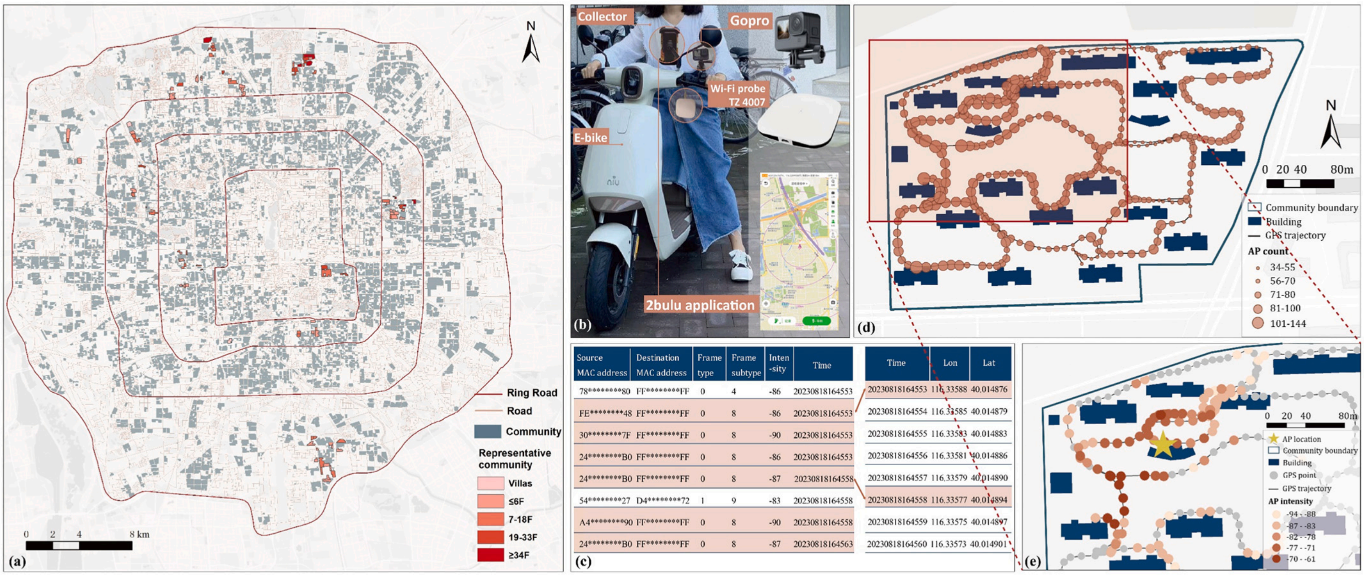


Fig. 6. Diagrammatic representation of Wi-Fi probe data collection and processing workflow
(a) Distribution of 103 representative residential communities; (b) Data collection; (c) Extraction of AP data and geospatial coordinate alignment; (d) Quantification of AP densities at varied locations; (e) Implementation of multilateral positioning for precise AP localization (Illustrated with a singular AP example).

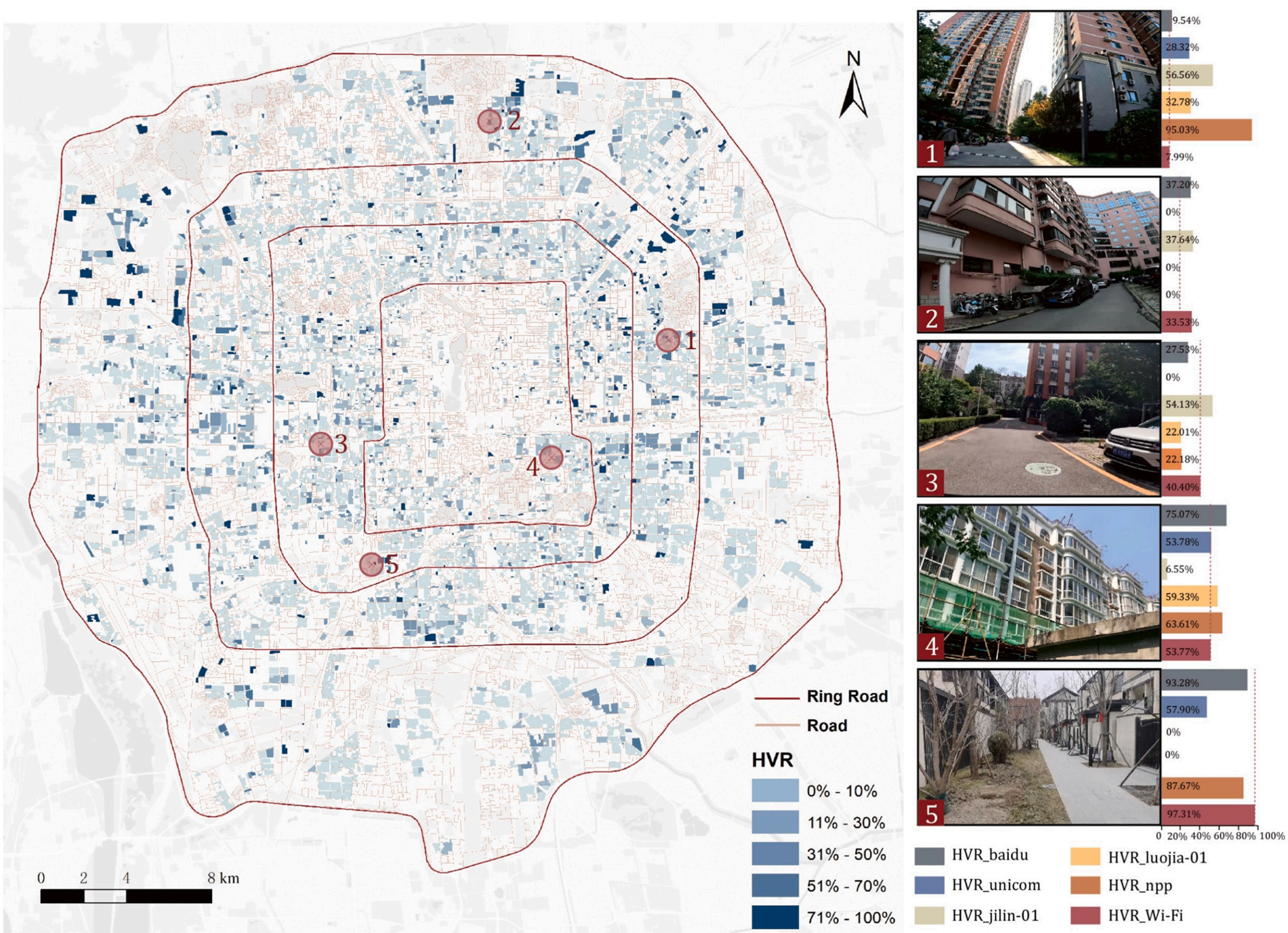


Fig. 7. HVR estimation results utilizing Baidu LBS data.

5. Discussion

5.1. Academic contribution compared to previous studies

High-rising buildings are a prominent feature of major urban landscapes. These areas often contain a substantial number of vacant housing units, indicating a significant waste of construction resources and potentially posing negative consequences, including adverse effects on the health of nearby residents. Consequently, accurately identifying the housing vacancy rate (HVR) at a fine-grained residential community level is essential for the efficient repurposing of vacant housing, reducing the consumption of building materials, and minimizing the generation of demolition waste. However, the distinct architectural characteristics of high-rise urban environments present considerable challenges for the detection of vacant housing. This paper proposes a methodological framework for large-scale identification of residential vacancy rates in high-rise urban areas.

Accurately calculating housing vacancy rates (HVR) requires precise estimation of both the total household capacity and the actual household count within a given area. Previous research has frequently relied on land use data, specifically, the area of residential communities, to estimate housing capacity (Lu, Zhang, Liu, Ye, & Miao, 2018; Yang & Pan, 2022). However, such approaches may introduce substantial errors in high-rise urban settings. Other studies have utilized three-dimensional (3D) building data to estimate capacity, yet often overlook important residential building attributes, such as building type and number of floors, which are correlated with unit area (Hu, Yan, Guo, Cui, & Dong, 2017). Some scholars have used household capacity data obtained from online real estate platforms. However, coverage remains limited. In our study, 1731 (26.7 %) residential communities lacked available capacity data. Specifically, 1595 residential communities were not listed on the Anjuke platform, and 136 were listed, but did not publish housing capacity information. Furthermore, Anjuke provides only geographic coordinates, which may not align precisely with the actual spatial boundaries used in our analysis. To address these challenges, our study integrates unit area data from various residential building types, collected from real estate information service platforms, with 3D building data. Additionally, we conducted a linear regression analysis comparing the available housing capacity data from real estate platforms with our estimations. The resulting R^2 value of 0.78 supports the reliability and validity of our proposed methodology.

This study utilizes five distinct data sources, including Baidu Location-Based Services (LBS) data, Unicom mobile signaling data, Jilin-01, Luojia-01 and NPP/VIIRS nighttime light data, to estimate the actual household count at the residential community level. These estimations are then employed in calculating the HVR, offering innovative methodologies for fine-grained, large-scale assessments of housing vacancy. To the best of our knowledge, this is the first study to systematically compare multiple methods for HVR estimation at the residential community level, and it also marks the inaugural application of LBS and mobile signaling data for these estimations. A key finding of the study is the substantial variation in HVR estimates produced by the five datasets. Notably, the results derived from Baidu LBS and Unicom mobile signaling data exhibit a strong correlation (correlation coefficient: 0.83), whereas the other datasets show weak inter-correlations, with coefficients consistently below 0.55. This outcome highlights the critical importance of validating the accuracy and reliability of data sources used in vacancy estimation.

As previously mentioned, accurately determining household count in high-rise urban settings presents significant challenges. The external appearance of residential units provides no reliable indication of occupancy, making traditional field surveys particularly difficult. Furthermore, in cities such as Beijing, access to electricity consumption data, which could potentially assist in occupancy estimation, is not available due to data restrictions. To address these challenges, this study introduces a novel method for collecting actual household count data in high-rise urban settings, which is used to validate the estimates derived from the five datasets. This method utilizes Wi-Fi probes to detect and count wireless network access points (APs) within the residential communities, using AP count as a proxy for the actual household count. Compared to traditional field surveys, this method significantly reduces labor and time costs. According to collection experience, a single data collector can survey 7 residential communities, covering approximately 9000–11,000 households, within an average of 6 h. Moreover, this approach minimizes potential biases resulting from inconsistent auditing standards among surveyors and offers scalability for application in other high-rise urban areas.

5.2. Performance of different datasets

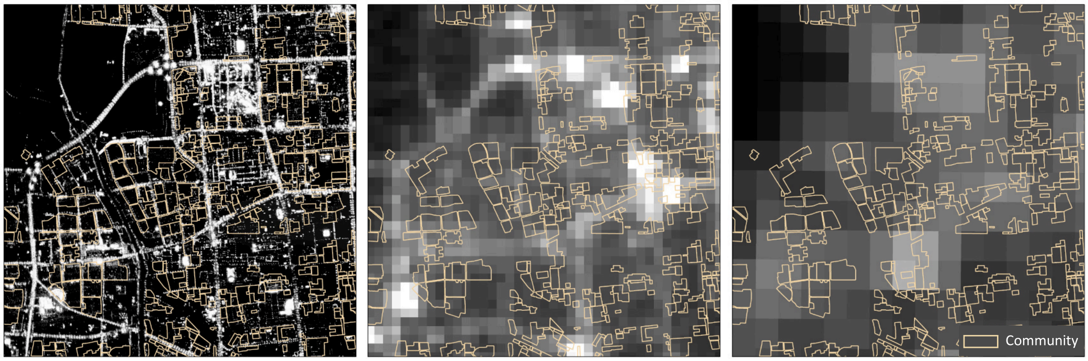
The paper collected the AP counts in 103 residential communities to validate HVR estimates derived from the five datasets. Among the datasets evaluated, Baidu Map LBS data demonstrated the best performance, showing a strong correlation with Wi-Fi probe results (correlation coefficient: 0.76). This finding is consistent with prior research, which indicates that LBS data significantly outperforms land use/cover data and nighttime light data in mapping population distributions (Xu, Song, Cai, & Zhu, 2021). Therefore, LBS data can be considered a reliable source for broader estimations of HVRs at the residential community level.

In contrast, datasets based on Unicom mobile signaling data and the Jilin-01 nighttime light data exhibited weaker correlations, with coefficients of 0.69 and 0.66, respectively. The relatively lower accuracy of Unicom mobile signaling data may be attributed to its dependence on the locations of connected base stations to estimate user locations (Zhai, Wu, Fan, & Wang, 2018), which lacks the spatial precision of GPS-enabled LBS data. Furthermore, Unicom's resident identification algorithm, based on mobile card registration, may not accurately capture the current population residing within residential communities.

The Jilin-01 nighttime light data also has notable limitations. Although the current analysis is based solely on luminosity values within a 2-m buffer zone surrounding residential buildings, it is not possible to fully eliminate the influence of non-residential light sources, such as streetlights and vehicle headlights. Additionally, the high spatial resolution of the data introduces inefficiencies in image acquisition. As a result, the nighttime light imagery used in this study is synthetic, meaning that different regions were not captured simultaneously. Moreover, unlike the Baidu Map LBS and mobile signaling data, the nighttime light data does not allow for extended temporal observation, thereby introducing a higher degree of randomness into the HVR estimates. The relatively low correlation with LBS data (correlation coefficient: 0.55) further underscores the limitations of nighttime light data for this application.

Although previous studies have demonstrated the feasibility of using Luojia-01 and NPP/VIIRS nighttime light data to estimate HVR at block or grid scale (Shi et al., 2022; Wang et al., 2019), our findings suggest that these datasets are less suitable for HVR estimation at the residential community scale. Specifically, they show only weak correlations with the Wi-Fi data, with correlation coefficients of 0.46 and 0.23, respectively. This is likely due to the low spatial resolution of these datasets (Fig. 8). Statistical analysis reveals that 3534 (61.60 %) residential communities are smaller than a single grid cell in the Luojia-01 dataset, and 5695 (99.27 %) residential communities are smaller than a single grid cell in the NPP/VIIRS dataset. Additionally, the Luojia-01 dataset ceased updates after 2019. This introduces a significant temporal mismatch between the Luojia-01 data and the other datasets used in this study. Consequently, the use of Luojia-01 nighttime light data to assess housing vacancy status beyond 2020 is not feasible.

It is important to note that Baidu Map LBS data, China Unicom mobile signaling data, and Jilin-1 nighttime light data are not publicly accessible (Table 4). Access to these datasets requires either direct purchase or customized services from the respective operators, which



(a) Jilin-01 nighttime light data (b) Luojia-01 nighttime light data (c) NPP/VIIRS nighttime light data

Fig. 8. Comparative of spatial resolution among three variants of nighttime light data in the same area.

Table 4
The comprehensive comparison of five datasets.

Dataset	Accuracy in HVR estimation	Spatial resolution	Time span	Collection method	Based on long-term observation or not
Baidu Map LBS data	High	Point	From 2017 to the present	Purchase	Yes
China Unicom mobile signaling data	Moderate	Point	2019–2025	Purchase	Yes
Jilin-01 nighttime light data	Moderate	0.92 m	From 2021 to the present	Purchase	No
Luojia-01 nighttime light data	Low	130 m	2018–2019	Open access	No
NPP/VIIRS nighttime light data	Low	500 m	From 2013 to the present	Open access	Yes

may constrain very large-scale HVR, particularly at the national level, due to the high costs of data acquisition. In contrast, the Luojia-01 and NPP/VIIRS nightlight datasets are publicly available. Previous studies have demonstrated their utility in estimating HVR at both parcel and urban scales (Shi et al., 2022; Wang et al., 2019), suggesting that these datasets can partially mitigate the issue of limited data accessibility.

5.3. Robustness testing

To ensure the robustness of the findings, three robustness checks were conducted. The first involved calculating the HVR directly from the residential community boundaries obtained from the AOI and other datasets, without manually verifying to exclude public buildings or other non-residential structures. The second robustness check adjusted the method for estimating the number of households in residential communities based on nighttime light data. Specifically, it assumed a linear relationship between nighttime light intensity and population size within residential communities. By combining nighttime light datasets with census data, population estimates were generated for each residential community. These estimates were then divided by average household size to derive the number of households, followed by HVR calculation. The third robustness check used housing capacity data directly obtained from a real estate platform (Anjuke.com) for each residential community. Results from these robustness checks consistently showed that the Baidu Map LBS data outperformed all other datasets in HVR estimation (Fig. 9).

5.4. Practical implications and limitations

In summary, this study presents a comprehensive methodological framework for large-scale and fine-grained analysis of housing vacancy rates (HVR) using LBS data. The application of this dataset enabled accurate estimation of HVR across a wide range of residential communities within Beijing's Fifth Ring Road, revealing an average vacancy rate of 13.31 %. However, despite the presence of vacant units in some communities, the actual household count exceeded the available housing capacity in others, indicating an uneven spatial distribution of residents. These findings carry important policy implications, offering government agencies a deeper understanding of the dynamics of Beijing's housing market and supporting the development of targeted strategies to rebalance resident and housing distribution. To address disparities, policies could be introduced to incentivize resource reallocation within the housing market. For communities with higher HVR, interventions such as rental or purchase subsidies, along with increased investment in public facilities, including transportation, commerce, education, and healthcare, could help attract residents. Such measures would ease population pressure in overcrowded areas, promote a more balanced distribution of residents across the city, and support equitable urban development. Additionally, the establishment of a dynamic monitoring system based on LBS data is recommended. This system would enable continuous tracking of residential patterns, early identification of vacancy trends, and timely prediction of potential housing-related issues, thereby facilitating proactive policymaking and intervention.

While Baidu LBS data is exclusive to China, comparable international platforms, such as Google, Facebook, and Foursquare, offer similar opportunities for housing vacancy analysis. Furthermore, while this study focuses on high-rise urban environments, the proposed methodology is also applicable to low-rise urban areas, which typically feature simpler architectural structures and may facilitate more straightforward analysis.

Nonetheless, this study is not without its limitations. It primarily utilizes HVR as an indicator to assess the vacancy patterns in Beijing but it does not delve into unit-level vacancy analysis, which warrants further exploration. Additionally, residential boundary determination in this study required a degree of manual correction to exclude non-residential buildings. In fact, several recent studies have demonstrated the potential of integrating POI data (Deng et al., 2022), remote sensing imagery (Lin et al., 2021), street view imagery (Zhang, Fukuda, & Yabuki, 2021), and social media imagery (Hoffmann, Abdulahhad, & Zhu, 2023), in combination with deep learning models, to infer building functions rapidly and accurately. Future research could build upon these approaches to enhance and automate the boundary identification process. Moreover,

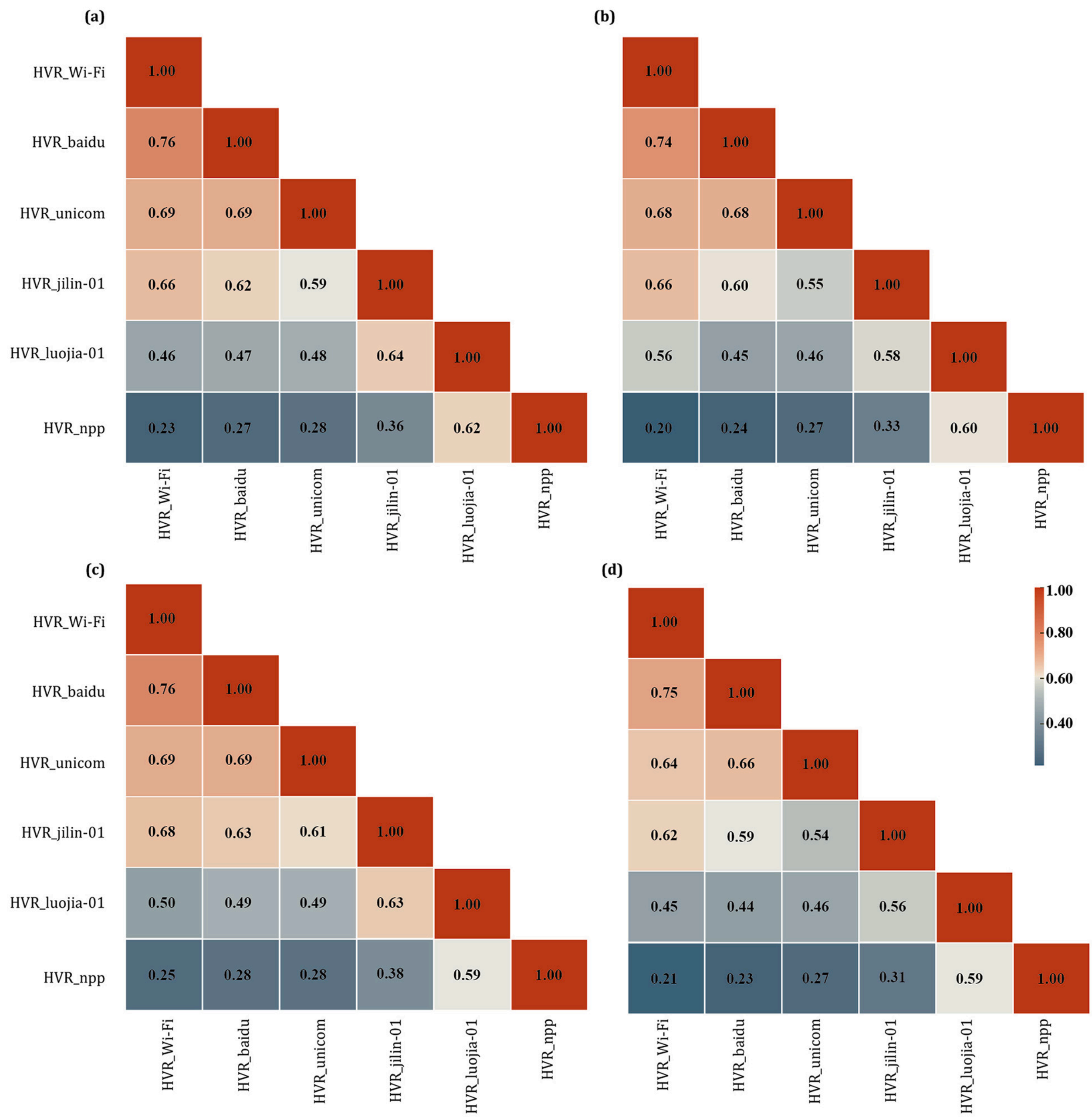


Fig. 9. Robustness test of the HVR correlation computed with five datasets and the Wi-Fi probe dataset: (a) HVR correlation analysis results without any adjustments; (b) HVR correlation analysis results using unmodified residential area boundaries; (c) HVR correlation analysis results with adjustments based on nightlight data to estimate household count; (d) HVR correlation analysis results using housing capacity data from real estate platform.

the current methodology is limited in its ability to distinguish seasonal vacancies, highlighting another area for potential refinement and methodological advancement.

6. Conclusions

In this study, we conducted a comprehensive analysis of 5737 residential communities within Beijing's Fifth Ring Road and compared five data sources to develop an accurate and effective methodology for estimating housing vacancy rates (HVR) at the residential community level in high -rise urban environments. This proposed framework, characterized by its precision and efficiency, demonstrates strong potential for application in other regions. Our main findings can be summarized as follows:

1) **Significant variation exists among datasets in estimating HVR.** Comparative analysis of five high-resolution datasets, along with their correlation with Wi-Fi probe data, revealed notable discrepancies in estimation performance.
2) **Location-Based Services (LBS) data proved to be the most reliable for HVR estimation.** The estimated HVR derived from LBS data exhibited the highest correlation with Wi-Fi probe data, with a coefficient of 0.76.
3) **The effectiveness of Luojia-01 and NPP/VIIRS nighttime light data at the residential community level is limited.** Their correlation coefficients with Wi-Fi probe data were only 0.46 and 0.23, respectively, indicating relatively low reliability for fine-scale vacancy analysis.
4) **Using LBS data, the HVR within Beijing's Fifth Ring Road was estimated to be 13.31 %.** This finding holds substantial policy

implications, offering valuable guidance for urban housing policy formulation, the revitalization of vacant housing stock, and the reduction of resource inefficiencies.

CRediT authorship contribution statement

Huimin Zhao: Writing – review & editing, Writing – original draft, Visualization, Validation, Project administration, Methodology, Investigation, Formal analysis, Data curation. **Changcheng Kan:** Writing – review & editing, Visualization, Resources, Project administration, Methodology. **Ying Long:** Writing – review & editing, Writing – original draft, Resources, Project administration, Methodology, Funding acquisition, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by three projects from the National Natural Science Foundation of China (52178044, 62394331 and 62394335). The authors extend their sincere appreciation to Nanxi Su, Qiyuan Hong, and Tangqi Tu for their invaluable contribution in the collection of Wi-Fi probe data.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cities.2025.106187.

Data availability

The authors do not have permission to share data.

References

Akiyama, Y., Ueda, A., Ouchi, K., Ito, N., Ono, Y., Takaoka, H., & Hisadomi, K. (2020). Estimating the spatial distribution of vacant houses using public municipal data. In *, 22. In geospatial Technologies for Local and Regional Development: Proceedings of the 22nd AGILE conference on geographic information science* (pp. 165–183). Springer International Publishing. pp.

Arnio, A. N., Baumer, E. P., & Wolff, K. T. (2012). The contemporary foreclosure crisis and US crime rates. *Social Science Research, 41*(6), 1598–1614. https://doi.org/10.1016/j.ssresearch.2012.06.002

Chen, Z., Yu, B., Hu, Y., Huang, C., Shi, K., & Wu, J. (2015). Estimating house vacancy rate in metropolitan areas using NPP-VIIRS nighttime light composite data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 8*(5), 2188–2197. https://doi.org/10.1109/JSTARS.2015.2418201

Deng, Y., Chen, R., Yang, J., Li, Y., Jiang, H., Liao, W., & Sun, M. (2022). Identify urban building functions with multisource data: A case study in Guangzhou, China. *International Journal of Geographical Information Science, 36*(10), 2060–2085. https://doi.org/10.1080/13658816.2022.2046756

Du, M., Wang, L., Zou, S., & Shi, C. (2018). Modeling the census tract level housing vacancy rate with the Jilin1-03 satellite and other geospatial data. *Remote Sensing, 10*(12), 1920. https://doi.org/10.3390/rs10121920

Fahmi, W., & Sutton, K. (2008). Greater Cairo's housing crisis: Contested spaces from inner city areas to new communities. *Cities, 25*(5), 277–297. https://doi.org/10.1016/j.cities.2008.05.003

Flas, M., Halleux, J. M., Cools, M., & Teller, J. (2022). Identifying housing vacancy using data on registered addresses and domestic consumption. *Housing Studies, 1-23*. https://doi.org/10.1080/02673037.2022.2031943

Hoffmann, E. J., Abdulahhad, K., & Zhu, X. X. (2023). Using social media images for building function classification. *Cities, 133*, Article 104107. https://doi.org/10.1016/j.cities.2022.104107

Hu, S., Yan, D., Guo, S., Cui, Y., & Dong, B. (2017). A survey on energy consumption and energy usage behavior of households and residential building in urban China. *Energy and Buildings, 148*, 366–378. https://doi.org/10.1016/j.enbuild.2017.03.064

Huang, H., & Zhao, H. (2023). Research on high-precision housing vacancy rate measurement and influencing factors based on Jilin-01 nightlight data. *Shanghai Urban Planning Review, 02*, 101–108. https://doi.org/10.11982/j.supr.20230215

Immergluck, D. (2016). Examining changes in long-term neighborhood housing vacancy during the 2011 to 2014 US national recovery. *Journal of Urban Affairs, 38*(5), 607–622. https://doi.org/10.1111/juaf.12255

Jang, H., Ahn, K., Kim, D., & Song, Y. (2018). *Detection and prediction of house price bubbles: Evidence from a new city. In computational science–ICCS 2018: 18th international conference, Wuxi, China, June 11–13, 2018 proceedings*. part III 18 (pp. 782–795). Springer international publishing. https://doi.org/10.1007/978-3-319-93701-4_58.

Jiang, J., Cheng, X., Xiang, F., Yin, X., & Wang, J. (2020). Research on vacancy rate and idle rate of residential houses based on power data. In *, Vol. 9. In 2020 IEEE 9th joint international information technology and artificial intelligence conference (ITAIC)* (pp. 2017–2019). IEEE. https://doi.org/10.1109/ITAIC49862.2020.9339006. , December.

Jin, X., Long, Y., Sun, W., Lu, Y., Yang, X., & Tang, J. (2017). Evaluating cities' vitality and identifying ghost cities in China with emerging geographical data. *Cities, 63*, 98–109. https://doi.org/10.1016/j.cities.2016.12.016

Kubo, T., & Mashita, M. (2020). Why the rise in urban housing vacancies occurred and matters in Japan. *The Rise in Vacant Housing in Post-growth Japan: Housing Market, Urban Policy, and Revitalizing Aging Cities, 3-22*. https://doi.org/10.1007/978-981-15-4165-1_1

Leskinen, N., Vimpari, J., & Junnila, S. (2020). Using real estate market fundamentals to determine the correct discount rate for decentralised energy investments. *Sustainable Cities and Society, 53*, Article 101953. https://doi.org/10.1016/j.scs.2019.101953

Li, Y., Lin, D., Zhang, Y., Song, Z., Sha, X., Zhou, S., & Yu, Z. (2023). Quantifying tree canopy coverage threshold of typical residential quarters considering human thermal comfort and heat dynamics under extreme heat. *Building and Environment, 233*, Article 110100. https://doi.org/10.1016/j.buildenv.2022.110100

Lin, A., Sun, X., Wu, H., Luo, W., Wang, D., Zhong, D., & Zhu, J. (2021). Identifying urban building function by integrating remote sensing imagery and POI data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14*, 8864–8875. https://doi.org/10.1109/JSTARS.2021.3107543

Liu, X., Li, Z., Fu, X., Yin, Z., Liu, M., Yin, L., & Zheng, W. (2023). Monitoring house vacancy dynamics in the pearl river delta region: A method based on NPP-viirs night-time light remote sensing images. *Land, 12*(4), 831. https://doi.org/10.3390/land12040831

Lu, H., Zhang, C., Liu, G., Ye, X., & Miao, C. (2018). Mapping China's ghost cities through the combination of nighttime satellite data and daytime satellite data. *Remote Sensing, 10*(7), 1037. https://doi.org/10.3390/rs10071037

Ma, Y. D., Zhou, C. S., & Fang, Y. P. (2003). Analysis on the present situation and contributing factors of vacant commercial housing in Guangzhou urban district. *Chinese Geographical Science, 13*, 182–189. https://doi.org/10.1007/s11769-003-0014-7

Mosey, G., & Deal, B. (2021). Optimizing multi-family building massing for affordability and envelope performance: An investigation of the trade-offs implicit in low-rise residential buildings. *Buildings, 11*(3), 99. https://doi.org/10.3390/buildings11030099

Newman, G., Li, D., & Park, Y. (2022). The relationships between neighbourhood vacancy, probable PTSD, and health-related quality of life in flood-disaster-impacted communities. *Urban Studies, 59*(15), 3077–3097. https://doi.org/10.1177/00420980209627779

Nie, H., Kemp, R., & Fan, Y. (2023). Investigating the adoption of energy-saving measures in residential sector: The contribution to carbon neutrality of China and Europe. *Resources, Conservation and Recycling, 190*, Article 106791. https://doi.org/10.1016/j.resconrec.2022.106791

Olaya, Y., Vásquez, F., & Müller, D. B. (2017). Dwelling stock dynamics for addressing housing deficit. *Resources, Conservation and Recycling, 123*, 187–199. https://doi.org/10.1016/j.resconrec.2016.09.028

Pan, J., & Dong, L. (2021). Spatial identification of housing vacancy in China. *Chinese Geographical Science, 31*, 359–375. https://doi.org/10.1007/s11769-021-1198-8

Peng, T. C., & Wang, C. C. (2022). The application of machine learning approaches on real-time apartment prices in the Tokyo metropolitan area. *Social Science Japan Journal, 25*(1), 3–28. https://doi.org/10.1093/ssjj/jyab029

Sano, S., Filipović, I., & Radović, D. (2020). Public-private interaction in low-rise, high-density Tokyo. A morphological and functional study of contemporary residential row-houses. *The Journal of Public Space, 5*(2), 63–88. https://doi.org/10.32891/jps.v5i2.1285

Scheu, R., & Garascia, M. (2016). Chicago multifamily market characterization: Developing a comprehensive picture of the multifamily housing landscape. *Cityscape, 18*(3), 265–276.

Shi, L., Wurm, M., Huang, X., Zhong, T., Leichtle, T., & Taubenböck, H. (2022). Estimating housing vacancy rates at block level: The example of Guiyang. *China. Landscape and Urban Planning, 224*, Article 104431. https://doi.org/10.1016/j.landurbplan.2021.104431

Snedker, K. A., & Herting, J. R. (2016). Adolescent mental health: Neighborhood stress and emotional distress. *Youth & Society, 48*(5), 695–719. https://doi.org/10.1177/0044118X14544432

Sun, L. (2020). *Housing affordability in Chinese cities*. Lincoln Institute of Land Policy.

Tan, Z., Wei, D., & Yin, Z. (2020). Housing vacancy rate in major cities in China: Perspectives from night time light data. *Complexity, 2020*, 1–12. https://doi.org/10.1155/2020/8822924

Tang, P., Du, P., Lin, C., Guo, S., & Qie, L. (2020). A novel sample selection method for impervious surface area mapping using JL1-3B night time light and Sentinel-2 imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13*, 3931–3941. https://doi.org/10.1109/JSTARS.2020.3002316

US Government Accountability Office. (2011). *Vacant properties: Growing number increases communities' costs and challenges*.

Wang, L., Fan, H., & Wang, Y. (2019). An estimation of housing vacancy rate using NPP-VIIRS night-time light data and OpenStreetMap data. *International Journal of Remote Sensing, 40*(22), 8566–8588. https://doi.org/10.1080/01431161.2019.1615655

Wang, Y., & Shaw, D. (2018). The complexity of high-density neighbourhood development in China: Intensification, deregulation and social sustainability challenges. *Sustainable Cities and Society, 43*, 578–586. https://doi.org/10.1016/j.scs.2018.09.014

Wang, Z., Ma, J., Zhang, B., Yang, Y., Wang, B., & Zhao, W. (2023). Does high speed railway alleviate housing vacancy rates? Evidence from smart meter data of household electricity consumption. *Transportation Research Part A: Policy and Practice, 176*, Article 103787. https://doi.org/10.1016/j.tra.2023.103787

Wegmann, J. (2020). Residences without residents: Assessing the geography of ghost dwellings in big US cities. *Journal of Urban Affairs, 42*(8), 1103–1124. https://doi.org/10.1080/07352166.2019.1669956

Wyatt, P. (2008). Empty dwellings: The use of council-tax records in identifying and monitoring vacant private housing in England. *Environment and Planning A, 40*(5), 1171–1184. https://doi.org/10.1068/a39176

Xu, Y., Song, Y., Cai, J., & Zhu, H. (2021). Population mapping in China with Tencent social user and remote sensing data. *Applied Geography, 130*, Article 102450. https://doi.org/10.1016/j.apgeog.2021.102450

Yang, P., & Pan, J. (2022). Estimating housing vacancy rate using nightlight and POI: A case study of Main urban area of xi ‘an City. *China. Applied Sciences, 12*(23), 12328. https://doi.org/10.3390/app122312328

Yang, X., Hu, M., Zhang, C., & Steubing, B. (2022). Key strategies for decarbonizing the residential building stock: Results from a spatiotemporal model for Leiden, the Netherlands. *Resources, Conservation and Recycling, 184*, Article 106388. https://doi.org/10.1016/j.resconrec.2022.106388

Yao, W., Teng, G., & Li, H. (2013, December). Recognition methods of housing vacancy based on digital image processing. In *In 2013 international conference on information science and cloud computing companion* (pp. 675–680). IEEE.

Yao, Y., & Li, Y. (2011). House vacancy at urban areas in China with nocturnal light data of DMSP-OLS. In *Proceedings 2011 IEEE international conference on spatial data mining and geographical knowledge services* (pp. 457–462). IEEE. https://doi.org/10.1109/ICSDM.2011.5969087. , June.

Yao, Y., Liu, X., Li, X., Zhang, J., Liang, Z., Mai, K., & Zhang, Y. (2017). Mapping fine-scale population distributions at the building level by integrating multisource geospatial big data. *International Journal of Geographical Information Science, 31*(6), 1220–1244. https://doi.org/10.1080/13658816.2017.1282615

Yonas, M. A., O'Campo, P., Burke, J. G., & Gielen, A. C. (2007). Neighborhood-level factors and youth violence: Giving voice to the perceptions of prominent neighborhood individuals. *Health Education & Behavior, 34*(4), 669–685. https://doi.org/10.1177/1090198106293942

Zhai, Y., Wu, H., Fan, H., & Wang, D. (2018). Using mobile signaling data to exam urban park service radius in Shanghai: Methods and limitations. *Computers, Environment and Urban Systems, 71*, 27–40. https://doi.org/10.1016/j.compenvurbsys.2018.03.011

Zhang, J., Fukuda, T., & Yabuki, N. (2021). Development of a city-scale approach for façade color measurement with building functional classification using deep learning and street view images. *ISPRS International Journal of Geo-Information, 10*(8), 551. https://doi.org/10.3390/ijgi10080551

Zhang, X., Li, Y., Chen, H., Yan, X., & Liu, K. (2023). Characteristics of embodied carbon emissions for high-rise building construction: A statistical study on 403 residential buildings in China. *Resources, Conservation and Recycling, 198*, Article 107200. https://doi.org/10.1016/j.resconrec.2023.107200

Zhang, Y., Zhao, H., & Long, Y. (2025). CMAB: A multi-attribute building dataset of China[J]. *Scientific Data, 12*(1), 430. https://doi.org/10.1038/s41597-025-04730-5

Zou, S., & Wang, L. (2021a). Detecting individual abandoned houses from google street view: A hierarchical deep learning approach. *ISPRS Journal of Photogrammetry and Remote Sensing, 175*, 298–310. https://doi.org/10.1016/j.isprsjprs.2021.03.020

Zou, S., & Wang, L. (2021b). Individual vacant house detection in very-high-resolution remote sensing images. In *Smart spaces and places* (pp. 117–129). Routledge.