



Full length article

Uncovering the hidden resource cost of urbanization: A globally scalable open-source framework for urban housing vacancy monitoring

Huimin Zhao^{a,b,1}, Yecheng Zhang^{a,1} , Fangming Liu^c, Ke Wang^{c,*},
Ying Long^{a,d,e,**}

^a School of Architecture, Tsinghua University, Beijing, China

^b School of Architecture and Urban Planning, Beijing University of Technology, Beijing, China

^c School of Ecology & Environment, Renmin University of China, Beijing, China

^d Hang Lung Center for Real Estate, Tsinghua University, Beijing, China

^e Key Laboratory of Ecological Planning & Green Building, Ministry of Education, Tsinghua University, Beijing, China

ARTICLE INFO

Keywords:

Housing vacancy rate

SDG 11

Open-source data

Embodied carbon

Urbanization

China

ABSTRACT

Rapid urbanization often outpaces population growth, creating massive resource inefficiencies. Quantifying this structural mismatch between housing supply and actual occupancy remains a global challenge due to costly censuses and imprecise proxy data. Here, we present a globally scalable, low-cost framework integrating open-source 3D building data with gridded population data to monitor block-level urban housing vacancy rates (HVR). Applying this framework to China, we produced the first high-resolution national vacancy map across 2813 spatial cities, revealing an average HVR of 23.46%. This equates to 77 million vacant units and 10.8 billion square meters of unutilized floor space. Our block-level analysis further reveals substantial intra-urban heterogeneity, including systematic core-periphery occupancy gradients and localized mismatches between housing supply and resident population. Evaluated using complementary agreement, classification, and external-consistency assessments, this open-source tool empowers nations to monitor housing resources (SDG 11), mitigate embodied carbon waste, and support evidence-based sustainable development.

1. Introduction

Amid the accelerating climate crisis, urban expansion and the building sector have emerged as a critical battleground (Zhang et al., 2026; Zhang et al., 2026), accounting for nearly 40% of global energy-related carbon emissions (Ürge-Vorsatz et al., 2020; UNEP, 2022). Within this sector, housing vacancy presents a profound "resource paradox": while approximately 1.1 billion people worldwide live in slums (UN-Habitat, 2022), vast quantities of housing stock remain uninhabited (Aalbers, 2016; Çelik, 2024; Molloy, 2016). This phenomenon transcends economic and social concerns, representing a multidimensional environmental failure. Economically, vacant housing represents trillions in underutilized capital (Baker, 2011; Aalbers, 2019) and lost property tax revenue, and it distorts market dynamics by presenting a supply-demand imbalance (Glaeser and Gyourko, 2005; Glaeser et al., 2008). Socially, vacancy is linked to neighborhood decline

and increased crime (Spader et al., 2016; Chen and Rafail, 2020; Morrison, 2003; Branas et al., 2018), a phenomenon often explained by the "broken windows" theory (Windows, 1982). Most critically, from an environmental perspective, every vacant unit represents a significant "sunk cost" of embodied energy and material consumption (Röck et al., 2023), specifically carbon emissions released for structures that serve no function (Langevin et al., 2023). This reality reveals a stark fracture in the UN Sustainable Development Goals (SDGs): the coexistence of surplus stock and housing poverty directly undermines the mandate for Sustainable Cities and Communities (SDG 11), while the resulting embodied carbon waste violates the principles of Responsible Consumption (SDG 12) and Climate Action (SDG 13). Therefore, from a resource-efficiency perspective, the underlying ownership status (e.g., investment, seasonal homes) is irrelevant; any physically vacant housing is an environmental burden.

However, our ability to mitigate this "concrete burden" is severely

* Corresponding author.

** Corresponding author at: School of Architecture, Tsinghua University, Beijing, China.

E-mail addresses: wangkert@ruc.edu.cn (K. Wang), yulong@tsinghua.edu.cn (Y. Long).

¹ These authors contributed equally to this work.

constrained by a global data divide (Mirza et al., 2021). Effective governance requires precision, yet housing vacancy remains a blind spot in planetary monitoring. Our analysis of 195 nations indicates that 73% lack any official census data on housing vacancy (Fig. 1, Supplementary Table S1). The deficit is most acute in the rapidly urbanizing Global South, where the speed of development often outpaces the capacity of statistical institutions (Wardrop et al., 2018). Even among the few nations that do collect such data, significant methodological and accessibility challenges persist. Most of these nations still rely on conventional methods that are not only costly and labor-intensive but also suffer from extremely low update frequencies, with census cycles typically spanning five to ten years. Such approaches are incapable of capturing the dynamics of a rapidly changing housing market. More critically, the utility of existing data is fundamentally limited by its coarse granularity. Among those that do publish, the information is typically aggregated at broad spatial resolutions, such as the city or census-tract level. This coarse data is insufficient for the intra-urban insights required by policymakers for targeted interventions (Brelsford et al., 2017). The extreme scarcity of high-resolution data impedes effective urban governance and fundamentally limits the potential for evidence-based policymaking (Espey et al., 2025).

To bypass these administrative bottlenecks, researchers have explored various proxy data sources, yet each faces inherent limitations (Fig. 2, Supplementary Table S2). For instance, night-time light (NTL) data (Li et al., 2019; Wang et al., 2019; Liu et al., 2023; Chen et al., 2015;

Tan et al., 2020; Shi et al., 2022; Zhao et al., 2025; Du et al., 2018), despite its global coverage (Kyba et al., 2017), is fundamentally constrained by its inability to distinguish residential from commercial or public lighting (Chen et al., 2015; Ma et al., 2014). Utility consumption data (e.g., electricity, water), while a direct indicator of activity (Akiyama et al., 2020; Zhao et al., 2025; Zou and Wang, 2020; Kumagai et al., 2016), is systematically inaccessible for large-scale research due to stringent privacy and commercial confidentiality barriers (Zhao et al., 2025). Similarly, online real estate platforms suffer from a severe sample selection bias (Zheng et al., 2025; Zou and Wang, 2022), as they only capture explicitly marketed properties while ignoring the vast stock of "implicit" vacancies, leading to systematic underestimation. Other visual sources like remote sensing (Zou and Wang, 2020) or street-level imagery (Zou and Wang, 2022; Zou and Wang, 2021; Yue et al., 2022) are ineffective for high-rise building settings where individual unit occupancy is externally invisible. Finally, emerging sources like mobile signaling data and location-based service (LBS) data, while precise, face substantial procurement costs and complex commercial partnerships that limit their scalability and reproducibility (Zhao et al., 2025). In short, current urban monitoring frameworks still lacks a scalable, low-cost, and high-resolution solution to audit its housing vacancy.

To bridge this gap, we introduce a globally scalable, low-cost and highly accurate framework that democratizes access to high-precision urban diagnostics. Moving beyond the limitations of traditional proxies, our method synergizes two categories of globally available open-

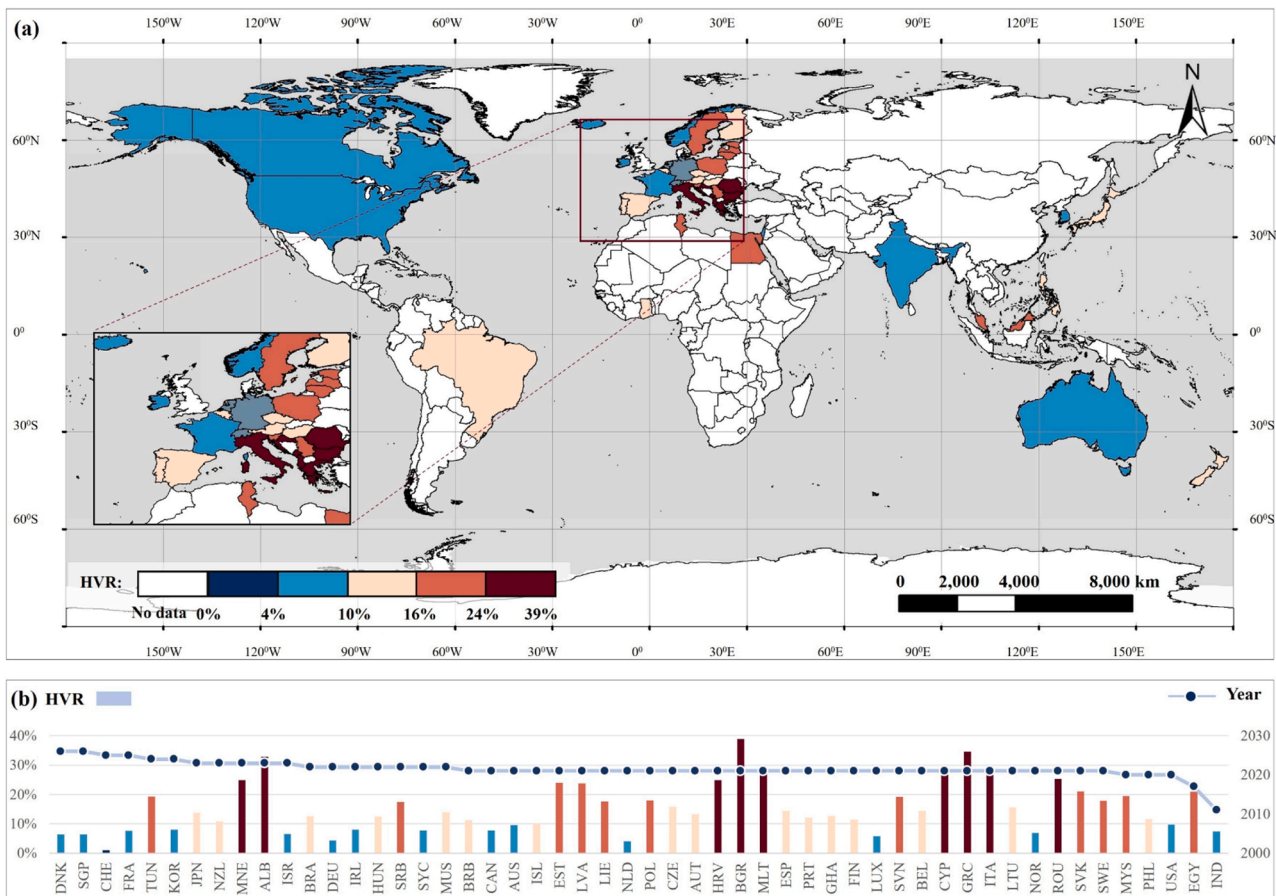


Fig. 1. Global landscape of housing vacancy rate (HVR). (a) An analysis of 195 countries and regions reveals that the vast majority (73%, $N = 143$) lack official census data on housing vacancy. Among the 52 jurisdictions that conduct such censuses, 22 report an HVR exceeding 15%. Spain distinguishes itself as one of the few instances utilizing national electricity big data, whereas the remaining 51 rely on traditional methods such as surveys. Furthermore, 50 of these 52 jurisdictions provide data only at a coarse granularity (e.g., city or census tract level). Cyprus (block level) and Germany (100-m grid level) are the notable exceptions offering fine-grained data. (b) The line chart indicates the year of the most recent housing vacancy survey for each country, while the bar chart displays the corresponding HVR values. Data source: Authors' compilation of housing vacancy statistics published by national statistical agencies, census authorities, and other relevant governmental institutions. Country-specific data sources and links are provided in Table S1.

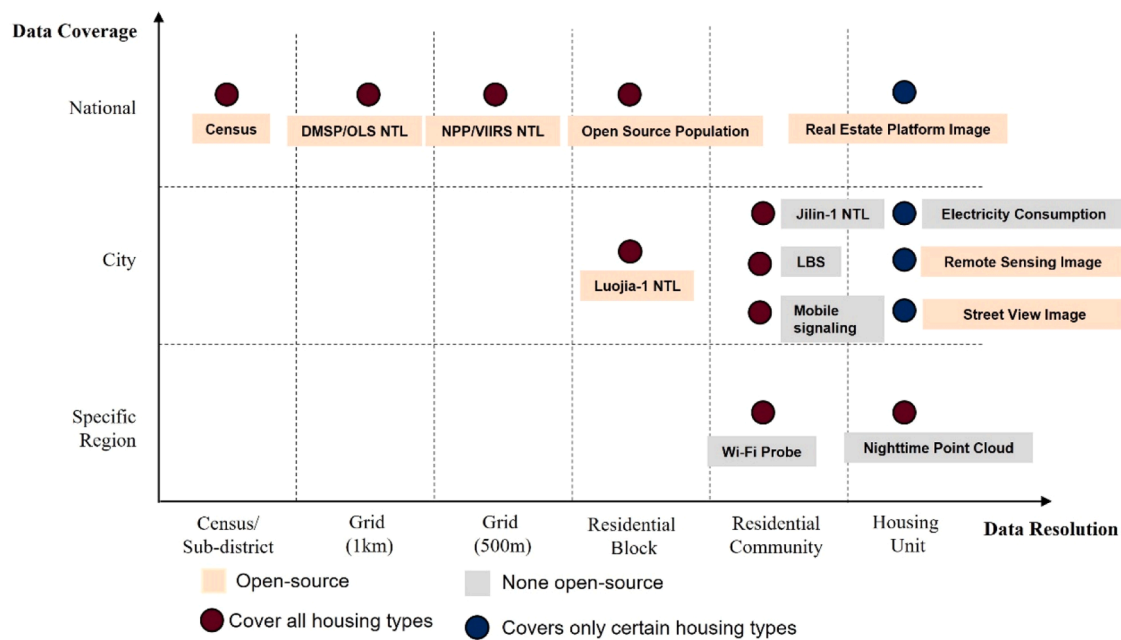


Fig. 2. Comparison of data resolution and coverage of different data sources. The chart classifies current housing vacancy monitoring data based on data resolution (x-axis) and data coverage extent (y-axis). This visualization underscores the critical gap in datasets that simultaneously achieve national scale, high resolution, and universal accessibility.

source data: comprehensive 3D building models (to quantify housing supply) and high-resolution gridded population data (to map housing demand). By integrating these datasets at the "natural block" level, we achieve a granular alignment of supply and demand that was previously attainable only through labor-intensive field surveys.

We apply this framework across urban China, encompassing urban settings that range from low-rise settlements to dense high-rise developments (Chen et al., 2018; Sun et al., 2020; Bai et al., 2014). Under fiscal decentralization, local governments' reliance on land-conveyance revenues has encouraged land supply, infrastructure investment, and new housing development, sometimes ahead of effective residential demand (Liu, 2023; Tong et al., 2023; Yu and Zhou, 2024). This supply-led and presale-oriented development model, combined with slowing population growth and uneven inter-city migration, can produce a persistent spatial mismatch between housing construction and household formation. Recent evidence also shows that unused housing is more prevalent where housing completions outpace population growth and where land availability and fiscal pressure create stronger incentives for residential land sales. China's high HVR may therefore partly reflect the legacy of land-financed urban expansion rather than household-level market decisions alone.

By mapping 2,813 spatial cities, we produced the first national-scale, block-level housing vacancy map for the country. This high-resolution lens allows us to look beyond aggregate statistics and unmask the latent laws of urban morphological evolution, distinguishing between market-driven density gradients and planning induced spatial mismatches. Validated against multiple independent datasets, including field-collected Wi-Fi probe data, ground-truth video data, the number of for-rent or for-sale listings on real estate platforms and the news reports from search engines, this study provides complementary evidence while quantifying the remaining disagreement and bias. Ultimately, this work provides a standardized, affordable toolkit for nations to monitor housing resources (SDG 11) and quantify the hidden environmental costs of urbanization, supporting evidence-based sustainable development policies worldwide.

2. Method

2.1. Study area

To accurately capture housing vacancy across China's diverse urban landscapes, this study covers all urban regions nationwide. Because China's administrative city boundaries often extend far beyond the physical urbanized extent to include vast rural territories, their use can severely distort urban statistics. Therefore, this study focuses on a spatial city, which is defined as a continuous built-up area characterized by significant population density, high-density urban infrastructure and non-agricultural land use and economic activities, and thereby more authentically represents a functional urban economy. We used the 2022 updated version of the spatial city boundaries developed by Ma and Long (2020), identifying 2813 cities ≥ 5 km². These spatial cities have a total area of 136,600 km² and span various administrative levels and climate zones, excluding the regions of Hong Kong, Macau, and Taiwan (Fig. 3).

2.2. Data sources

This study synergistically integrates multiple open-source or publicly available geospatial datasets. The core datasets are summarized in Fig. 4 (Supplementary Table S3).

2.3. A scalable framework for HVR monitoring

This study employs a fine-grained, block-centric methodology (Fig. 5) that integrates multi-source open geospatial data to enable precise, high-resolution measurement of the HVR at a national scale.

2.3.1. Calculation of housing stock

The precise calculation of housing stock is a cornerstone of this study, for which we designed a three-step workflow:

Building Data Selection and Processing: This study primarily utilized the comprehensive China Multi Attribute Building (CMAB) dataset (Zhang et al., 2025). This dataset, generated by integrating

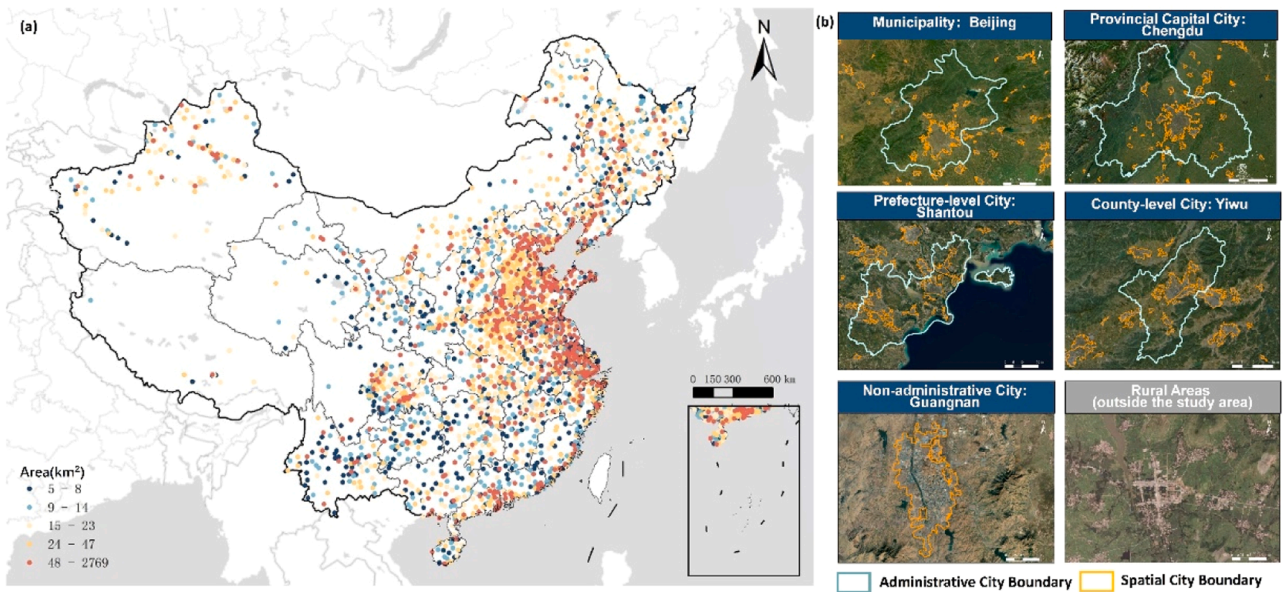


Fig. 3. Study area. (a) Distribution of spatial cities and their corresponding areas. (b) Contrasting physical and administrative city definitions. Visual inspection of high-resolution remote sensing data demonstrates that physical city boundaries better reflect the spatial extent of human-built environments than administrative demarcations.

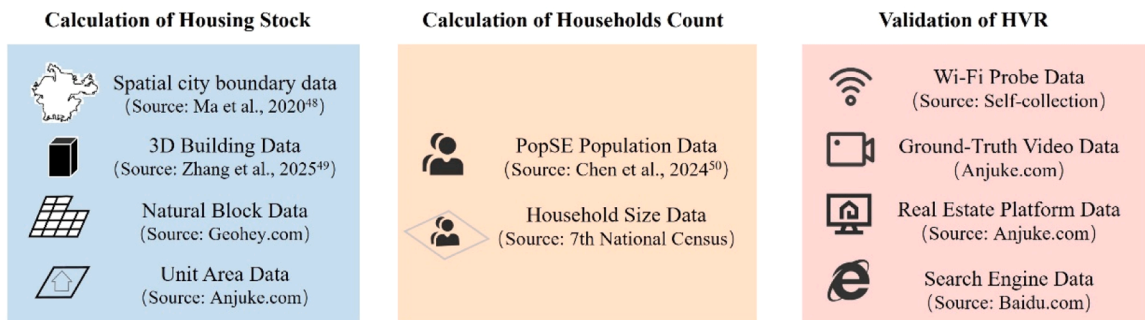


Fig. 4. Publicly available datasets utilized in this study. The estimation of housing stock was based on the Spatial city boundary data, 3D Building data, natural block, and unit area data. The calculation of household count utilized PopSE population data and household size data. For validation purposes, Wi-Fi probe data, ground-truth video data, real estate platform data, and search engine data were employed.

multi-source data via deep learning models, provides vectorized building footprints, floor heights, and functional classifications (e.g., residential, commercial, industrial) nationwide. Its comprehensiveness and accuracy far exceed those of other open-source counterparts (Supplementary Table S4).

Residential Block Identification: We used the "natural block" (an urban block naturally demarcated by the road network) as the basic analytical unit. First, all buildings classified as "residential" were extracted from the CMAB dataset. For each natural block, we calculated the residential-building-area share as the gross floor area of buildings classified as residential divided by the gross floor area of all buildings within the block. Blocks with a residential-building-area share of at least 40% were retained for HVR estimation. Throughout this article, the term "residential block" is used as shorthand for a block meeting this operational criterion and therefore denotes a block with a substantial residential component rather than an exclusively residential block. We assessed the sensitivity of the results to this classification choice by repeating the national estimation using minimum residential-building-area-share thresholds of 30%, 40%, 50%, and 60% (Supplementary Fig. S1).

Unit Area Identification and Housing Stock Calculation: To convert gross building area into a number of dwelling units, we

needed to determine the "area per household" for the residences within each block. We web-scraped 3.8 million second-hand housing listings, which included detailed unit areas and geolocations, from a mainstream real estate platform (Anjuke). This measure is distinct from both in-unit construction area and usable floor area. Under Chinese property-area measurement conventions, dwelling gross floor area comprises the in-unit construction area, including usable floor area, wall area, and balcony area, together with the allocated share of common building space, such as elevator shafts, stairwells, public entrance halls, and corridors. Therefore, the area reported in the listings is not merely the floor space available for private use; it already includes walls and allocated common areas. These listings were matched to residential blocks to calculate a typical area per household for different building typologies. We categorized residential structures into urban self-built housing and collective housing, with the latter further subdivided by floor count: 1–3 floors, 4–6 floors, 7–18 floors, 19–33 floors, and >33 floors. For blocks with no matching housing listings, we employed a robust, hierarchical imputation method to fill missing values. We preferentially used the mean value for the same building type within the sub-district. If unavailable, we sequentially fell back to the mean of the host spatial city, administrative city, and province. This ensured that every

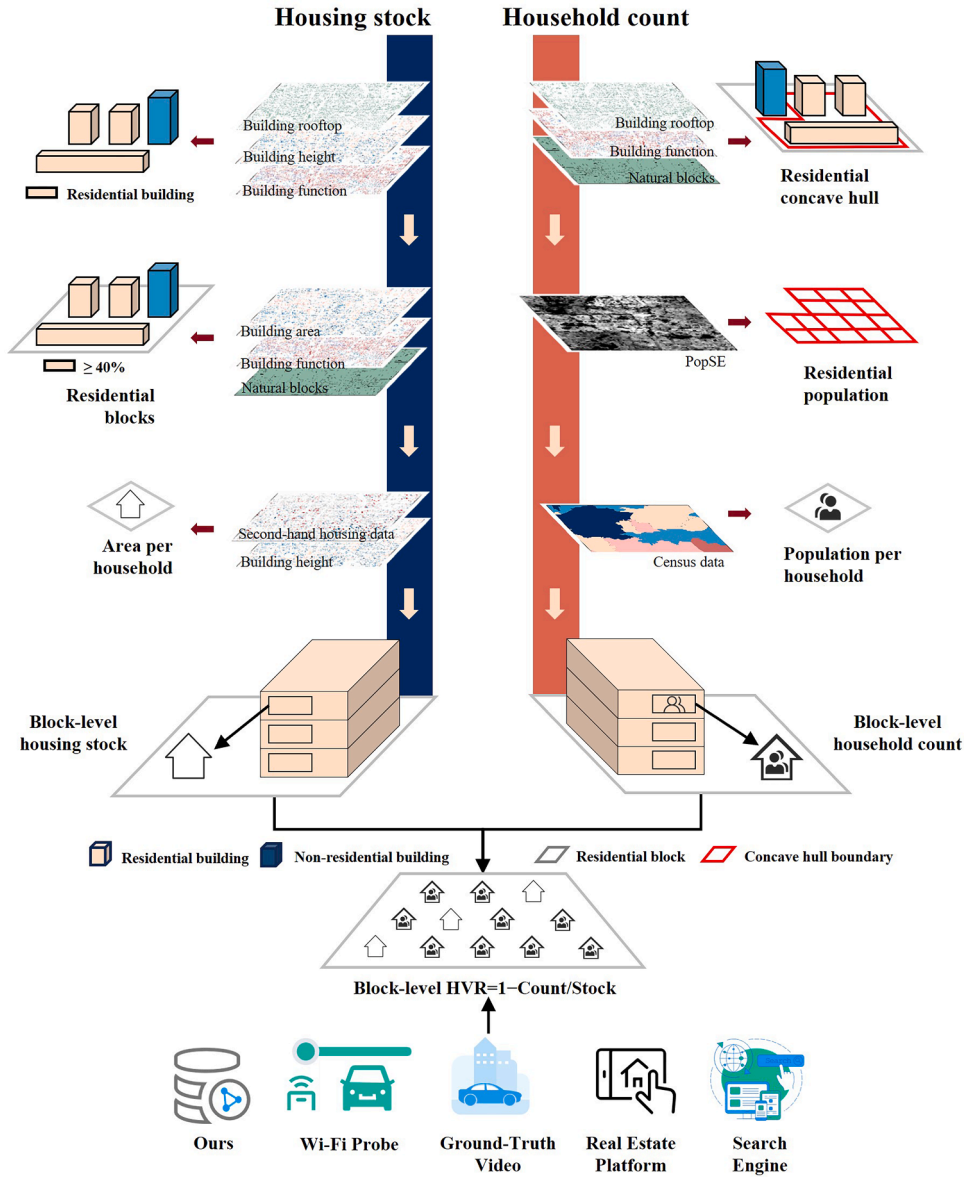


Fig. 5. Methodological framework for the estimation of the national-scale, block-level HVR. The overall logic of this framework comprises three core steps executed at the fine-grained "residential block" unit: quantifying housing stock, assessing household count, and comparing this supply-demand differential.

residential building was assigned a reasonable and representative area per household (Supplementary Table S5).

Following these steps, the housing stock for an individual residential building is its total gross floor area (footprint area($A_{base,i}$) $\times$ number of floors(F_i)) divided by the corresponding area per household for its type ($A_{avg_unit_type,i}$). The housing stock of a residential block ($Stock_b$) is the sum of the stock of all residential buildings within it (Eq. 1). The total housing stock of a spatial city ($Stock_s$) is the sum of the stock of all its residential blocks (Eq. 2).

$$Stock_b = \sum_i \frac{A_{base,i} \times F_i}{A_{avg_unit_type,i}} \quad (1)$$

$$Stock_s = \sum_i Stock_b \quad (2)$$

2.3.2. Calculation of household count

The estimation of households was also conducted in three steps, centered on the fine-grained spatial allocation of high-precision popu-

lation data:

Population Data Selection: This study utilized the PopSE-China2020 dataset, a 100-meter resolution gridded population product generated from multi-source data fusion and machine learning models. Compared to other global population datasets such as WorldPop and LandScan, PopSE has undergone more refined local calibration within China, demonstrating superior spatial resolution and consistency with actual population distribution characteristics. This makes it the ideal choice for this study (Supplementary Table S6).

Generation of Residential Zone Concave Hull Boundaries: To precisely allocate the gridded population data to residential buildings (and to avoid incorrectly including population from non-residential areas like associated commercial spaces, office buildings, or large green spaces), we generated a tightly-fitting 'concave hull' boundary for the residential building cluster within each residential block. The algorithm proceeds as follows (Supplementary Fig. S2): first, we extracted the vertex coordinates of all residential

building footprints within a block. We then applied a k-nearest neighbors-based concave hull algorithm to generate a minimum polygon that tightly envelops all residential buildings. Finally, we applied a 10-meter buffer to this concave hull boundary to smooth its edges and account for potential spatial positioning errors in the population grid, thereby ensuring a comprehensive capture of the population residing in and immediately adjacent to these residential buildings.

Calculation of Household Count: The number of households in a residential block ($Household_b$) was calculated by dividing its total population (Pop_b) by the average household size of its district/county ($House_{size_b}$) (3 and 4). Pop_b is defined as the sum of the population from all PopSE (Chen et al., 2024) grid cells whose centroids fall within the block's concave hull boundary. $House_{size_b}$ was obtained from the 7th National Population Census bulletins. The total number of households in a spatial city ($Household_s$) is the sum of the households in all of its constituent residential blocks.

$$Household_b = \frac{Pop_b}{House_{size_b}} \quad (3)$$

$$Household_s = \sum_i Household_b \quad (4)$$

2.3.3. Calculation of HVR

The HVR for a residential block (HVR_b) is calculated based on its housing stock and household count (Eq. 5). Subsequently, the absolute number of vacant housing units (VH_b) is calculated as the product of the block-level housing vacancy rate and the housing stock (Eq. 6), while the total vacant housing area (VH_Area_b) is derived by multiplying the vacant housing units by the block's average unit area ($House_area_b$) (Eq. 7).

$$HVR_b = 1 - \frac{Household_b}{Stock_b}, \text{ when } Household_b < Stock_b \quad (5)$$

$$HVR_b = 0, \text{ when } Household_b \geq Stock_b$$

$$VH_b = HVR_b \times Stock_b \quad (6)$$

$$VH_Area_b = VH_b \times House_area_b \quad (7)$$

For a spatial city, the overall housing vacancy rate (HVR_s) is the total number of vacant units in all its blocks divided by the city's total housing stock (Eq. 8). The total number of vacant units (VH_s) and total vacant housing area (VH_Area_s) for the city are the sums from all its blocks (9 and 10).

$$HVR_s = \frac{\sum VH_b}{Stock_s} \quad (8)$$

$$VH_s = \sum VH_b \quad (9)$$

$$VH_Area_s = \sum VH_Area_b \quad (10)$$

2.3.4. Modeling of spatial structure

To move beyond static quantification and decode the underlying spatial structure, we developed a standardized analytical framework based on the Equivalent Physical Radius. At the macro scale, we constructed a Stock-Weighted Inverse Vacancy Gradient Model to quantify the center-hinterland interaction. By calculating the Core-Periphery Occupancy Gap, defined as the difference in stock-weighted occupancy between the mature urban core and the expansion periphery, and combining this spatial metric with temporal demographic trends (shrinkage versus growth), we established a four-quadrant typology to characterize the spatial and demographic conditions of cities. At the micro scale, we dissected the internal continuity of urban structure using

an Inverse S-shape Function to model the radial distance-decay of occupancy levels. These radial profiles were subsequently categorized via K-Means clustering to identify distinct morphological archetypes (e.g., Compact vs. Sprawling), thereby systematically revealing the diverse structural footprints of housing vacancy across China.

While the HVR serves as a direct indicator of resource inefficiency, establishing a theoretical dialogue with the classical Monocentric City Model and density gradient theories requires a metric that adheres to distance-decay laws (Clark, 1951). Therefore, to examine the spatial conformity of vacancy patterns, we adopted the inverse vacancy metric (1-HVR). Given that classical urban theories posit a decay in density and intensity from the core (Bertaud and Malpezzi, 2003), analyzing this inverse metric allows us to directly employ the standard Inverse S-shape function to quantify the structural compactness and physical boundaries of the housing market.

We establish a standardized spatial benchmark by converting the physical land area ($Area_i$) of each entity into an Equivalent Physical Radius (R_i) (Eq. 11). Based on this benchmark, we construct a dynamic core-periphery structure: the Urban Core (D_{core}) covers the mature built-up area within a distance $d \leq R_i$ from the centroid, while the Urban Periphery (D_{outer}) represents the expansion hinterland extending from R_i to $m \times R_i$. The equivalent radius is defined as:

$$R_i = \sqrt{\frac{Area_i}{\pi}} \quad (11)$$

Macro-Scale: Quantification of the Core-Periphery Occupancy Gap. At the inter-urban scale, we quantify the center-hinterland interaction using a Stock-Weighted Inverse Vacancy Gradient Model. The HVR is first converted into an inverse vacancy metric (1 - HVR). To account for differences in housing-stock size among residential blocks, we compute the Core-Periphery Occupancy Gap (Gap) as the difference between the stock-weighted average inverse HVR of the core and that of the periphery (Eq. 12).

$$Gap_i = \underbrace{\frac{\sum_{j \in D_{core}} [(1 - HVR_{ij}) \times N_{ij}]}{\sum_{j \in D_{core}} N_{ij}}}_{\text{Weighted Core Inverse Vacancy}} - \underbrace{\frac{\sum_{k \in D_{outer}} [(1 - HVR_{ik}) \times N_{ik}]}{\sum_{k \in D_{outer}} N_{ik}}}_{\text{Weighted Outer Inverse Vacancy}} \quad (12)$$

where HVR_{ij} and N_{ij} denote the HVR and total housing stock (units) of parcel j in the core of city i , respectively; and similarly for parcel k in the periphery. A positive Gap indicates that stock-weighted occupancy is higher in the urban core than in the periphery, whereas a negative Gap indicates the opposite pattern.

To capture the demographic trajectory of each spatial city, we further define the Shrinkage Rate (S_i) using gridded population data (WorldPop 2015–2020) (Eq. 13):

$$S_i = \frac{Pop_{2015,i} - Pop_{2020,i}}{Pop_{2015,i}} \quad (13)$$

where a positive rate ($S_i > 0$) signifies population loss (Shrinking), while a negative rate signifies growth. By combining these two dimensions—the Core-Periphery Occupancy Gap and the Shrinkage Rate—we construct a four-quadrant urban typology comprising Core-Dominant Growth, Periphery-Dominant Growth, Core-Dominant Shrinkage, and Periphery-Dominant Shrinkage.

Micro-Scale: Internal Distance-Decay Modeling. To dissect the continuous internal structure of spatial cities, we model the radial attenuation of the inverse vacancy metric (1-HVR) using the Inverse S-shape Function. We calculate the average weighted inverse vacancy in concentric rings (r) from the city center to the fringe. The distance-decay behavior is fitted using the following Eq. (14):

$$f(r) = \frac{1 - c}{1 + e^{\alpha \left(\frac{2r}{D} - 1 \right)}} + c \quad (14)$$

where $f(r)$ is the inverse vacancy value at distance r ; c represents the baseline inverse vacancy level of the periphery; α is the slope parameter indicating the compactness of the urban form (a higher α implies a sharper drop in the inverse metric from core to edge); and D represents the estimated spatial extent of the urban entity.

Model goodness of fit was evaluated using the coefficient of determination, calculated as $R^2 = 1 - SSE/SST$, where $SSE = \sum_i (y_i - \hat{y}_i)^2$ and $SST = \sum_i (y_i - \bar{y})^2$. Because the Inverse S-shape Function is a constrained nonlinear model, its residual sum of squares is not guaranteed to be smaller than that of the mean-only benchmark. We therefore did not truncate R^2 at zero; a negative value indicates that the fitted curve performs worse than the mean-only benchmark rather than representing negative explained variance.

Morphological Typology Based on the extracted radial profiles, we employ K-Means Clustering to categorize spatial cities into distinct morphological archetypes (e.g., Standard, Compact, Sprawling and Anomalous).

Spatial Autocorrelation of Block-Level HVR: To quantify the intra-urban spatial clustering of block-level HVR, we calculated Global Moran's I separately for each spatial city containing at least 30 residential blocks and exhibiting non-zero within-city variation in HVR. Spatial weights were constructed using the eight nearest neighboring block centroids and were row-standardized. Statistical significance was assessed using 999 random permutations, and the Benjamini-Hochberg procedure was applied to control the false discovery rate across cities. We further used Spearman's rank correlation to examine the association between Global Moran's I and aggregate city-level HVR. Robustness was evaluated using alternative spatial-weight specifications with four and twelve nearest neighbors.

3. Results

3.1. A macro-scale panorama of national HVR

We applied the framework to urban areas across China. The study area encompasses substantial variation in urban scale and settlement form, ranging from low-rise villages and small towns to dense high-rise urban areas. Crucially, to ensure our metrics reflect functional urban economies rather than administrative abstractions, we restricted our analysis to 2813 "spatial cities" (continuous built-up areas ≥ 5 km²), thereby filtering out dilution from vast rural hinterlands.

The application of our framework generated the first panoramic, block-level assessment of this massive system (Fig. 6). The analysis reveals an aggregate HVR of 23.46% across China's urban areas. This percentage translates to a substantial inventory: 77.14 million vacant units, encompassing 10.8 billion square meters of unutilized gross floor area. When recast as a resource metric, this dormant stock exposes a massive "hidden capacity": theoretically sufficient to accommodate approximately 202 million people. This implies that China could elevate its urbanization rate by 14.3%, from the 2020 baseline of 63.9% (Xu et al., 2023) to 78.2%, without laying a single new brick.

Nevertheless, city-level estimates at the upper tail of the distribution require cautious interpretation. Several small spatial cities, including Yumen, exhibit estimated HVRs exceeding 75%. In these cities, relatively small population and housing-stock denominators may magnify the proportional effects of uncertainties in residential-building classification, gridded-population allocation, and spatial-boundary delineation. At the same time, severe population loss or housing development that substantially preceded effective demand may also produce genuinely high levels of underutilization. These extreme values should therefore be interpreted primarily as indicators of a severe housing

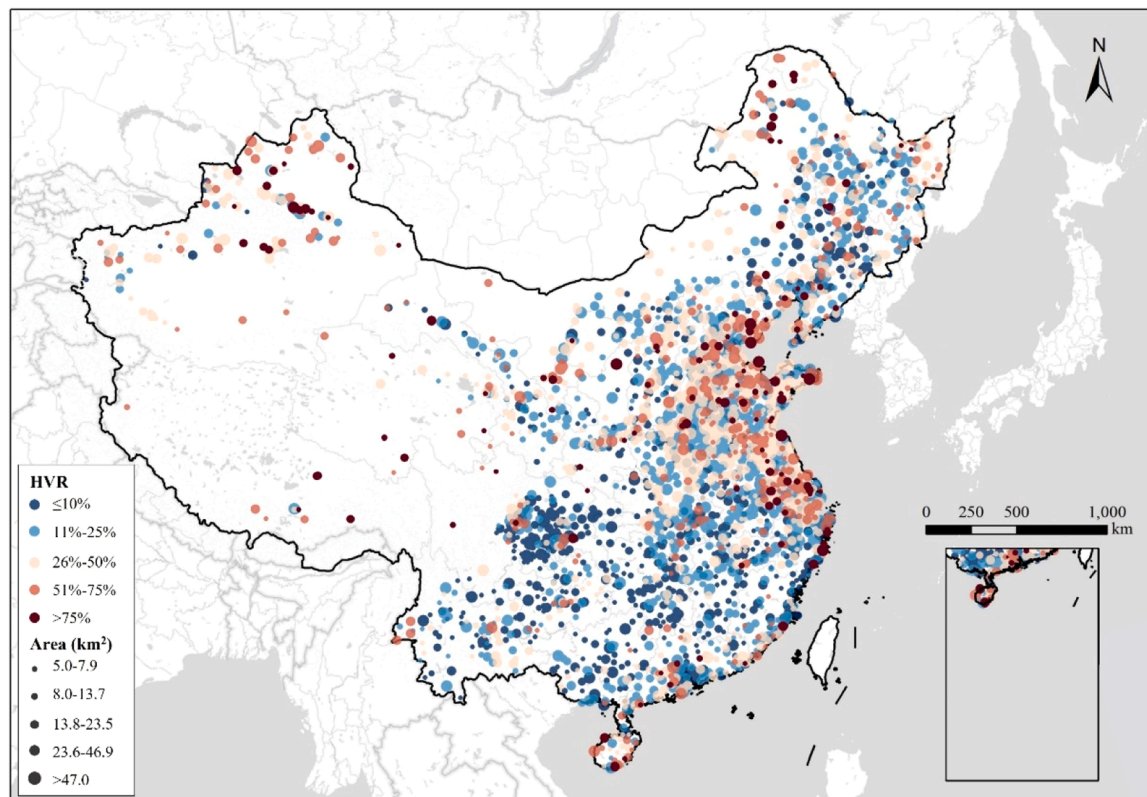


Fig. 6. Spatial distribution map of the HVR in China's spatial city regions. The color of each circle indicates the HVR, ranging from low (blue) to high (red). The size of the circle is proportional to the spatial urban area. City-level HVR estimates exceeding 75%, particularly those for small spatial cities, should be interpreted cautiously because scale-related uncertainty may magnify proportional estimation errors. These values are intended to indicate potentially severe housing underutilization rather than provide precise rankings among cities.

supply–demand mismatch rather than as precise rankings of city-level vacancy.

3.2. Inter-city core–periphery occupancy differentiation

At the macro scale, we identify substantial differences in core–periphery occupancy patterns among spatial cities (Fig. 7). By combining the Core–Periphery Occupancy Gap (Gap) with the Shrinkage Rate (S), we classify spatial cities into four categories: Core-Dominant Growth, Periphery-Dominant Growth, Core-Dominant Shrinkage, and Periphery-Dominant Shrinkage. Core-Dominant Growth describes growing cities in which stock-weighted occupancy is higher in the core than in the periphery, whereas Periphery-Dominant Growth describes growing cities with the opposite spatial pattern. Core-Dominant Shrinkage and Periphery-Dominant Shrinkage distinguish shrinking cities according to the direction of their core–periphery occupancy contrast. The relationship between the Core–Periphery Occupancy Gap and urban size is nonlinear: the Gap initially tends to increase with urban size but shifts toward negative values after a certain size threshold.

3.3. Intra-city distribution patterns

Within individual cities, our block-level estimates enable a fine-grained characterization of the intra-urban spatial organization of housing vacancy. The results show that housing vacancy is spatially clustered rather than randomly distributed and exhibits systematic associations with urban structural characteristics.

- (1) **Spatial Clustering of Block-Level HVR:** The block-level resolution of this study reveals substantial heterogeneity in the intra-urban distribution of housing vacancy (Fig. 8). To quantify these patterns rather than relying solely on visual interpretation, we calculated Global Moran's I for 1,625 spatial cities. The median Global Moran's I was 0.230 (interquartile range: 0.110–0.381). Overall, 81.8% of the analyzed cities exhibited significant positive spatial autocorrelation at the 0.05 level, and 80.6% remained significant after false-discovery-rate correction. Global Moran's I was also positively associated with aggregate city-level HVR (Spearman's $\rho = 0.495$, $p < 0.001$). This association remained consistent under alternative spatial-weight specifications

($\rho = 0.496$ for $k=4$ and $\rho = 0.476$ for $k=12$, both $p < 0.001$). These results indicate that higher overall vacancy tends to be associated with greater spatial clustering of block-level HVR.

- (2) **Scale Dependence of Radial Occupancy Gradients:** The goodness of fit of the radial distance-decay model varied systematically with city size. Among the analyzed spatial cities, 406 had an equivalent physical radius exceeding 50 km and together accounted for 74% of the total urbanized area in China (Supplementary Fig. S3). Smaller spatial cities frequently showed low or negative R^2 values, indicating that a monocentric radial function provided only a limited description of their occupancy distributions (Supplementary Fig. S4). Model fit generally increased with city size, and many larger spatial cities had R^2 values between 0.6 and 0.9 (Supplementary Fig. S5). These results indicate that the radial distance-decay model describes occupancy patterns more consistently in larger spatial cities than in smaller ones. We interpret this finding as scale-dependent variation in model performance rather than evidence of a self-organizing process, a phase transition, or a causal size threshold.
- (3) **Morphological Archetypes of Inverse Vacancy:** Beyond the general gradient, K-means clustering of radial inverse vacancy profiles (1 - HVR) uncovers four distinct morphological archetypes that reflect divergent developmental logics (Fig. 9). The 'Standard Decay' cluster (green line) typifies the urban norm, exhibiting a moderate, smooth logistic attenuation consistent with canonical bid-rent theory, thereby representing the baseline equilibrium of general urban development. In contrast, the 'Compact Steep Decay' cluster (red line) is characterized by high inverse HVR in the urban core followed by a sharp decline with increasing distance. Prevalent in major metropolises, this pattern indicates a pronounced core–periphery occupancy contrast, with substantially higher stock-weighted occupancy in the center than at greater distances. The 'Polycentric Sprawl' pattern (blue line) maintains a flattened trajectory at consistently high levels, indicating low global vacancy. Characteristic of regionally integrated megacities, this morphology reflects the blurring of urban–rural boundaries driven by the intense spatial spillover effects inherent to metropolitan amalgamation. Finally, the 'Anomalous New Town' typology (orange line) exhibits a significant structural deviation characterized by an anomalous inverted U-shaped curve. This trajectory indicates a profound spatial mismatch

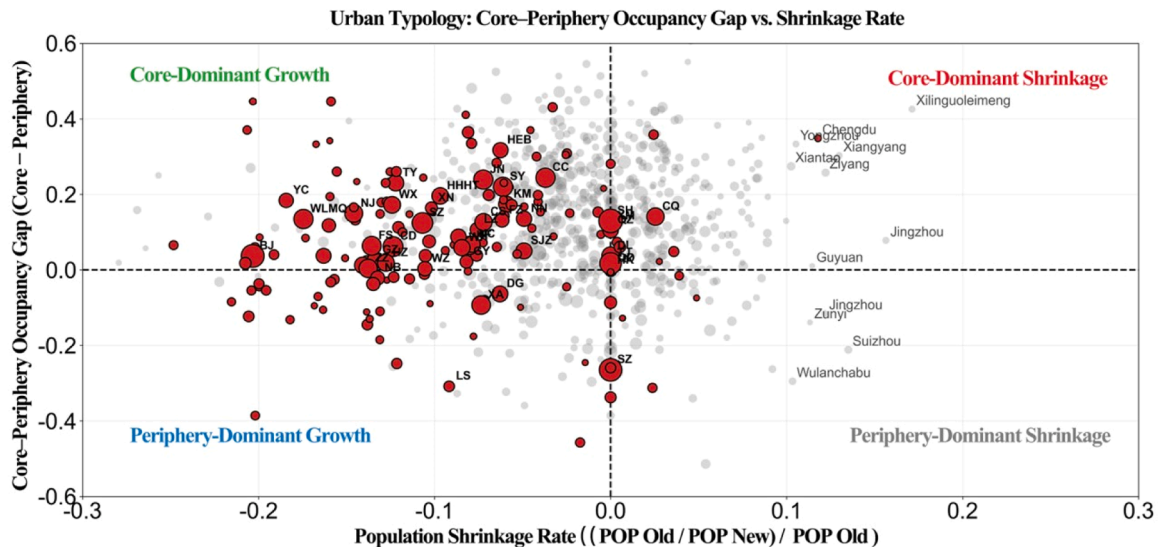
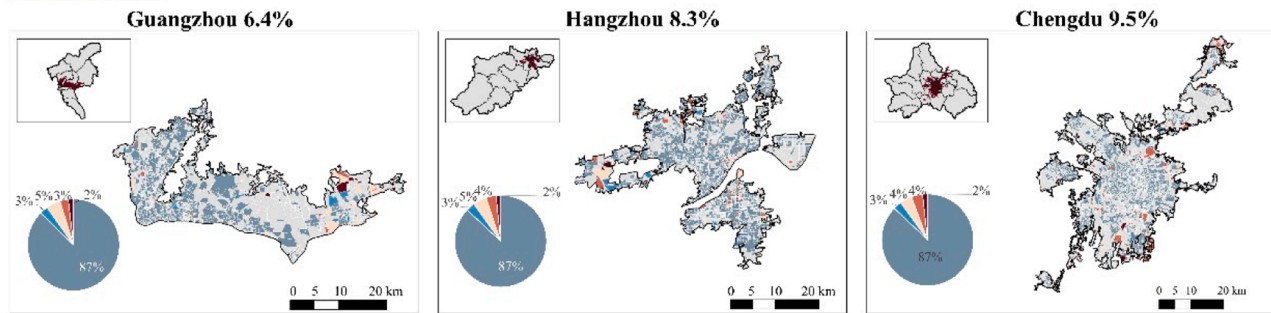
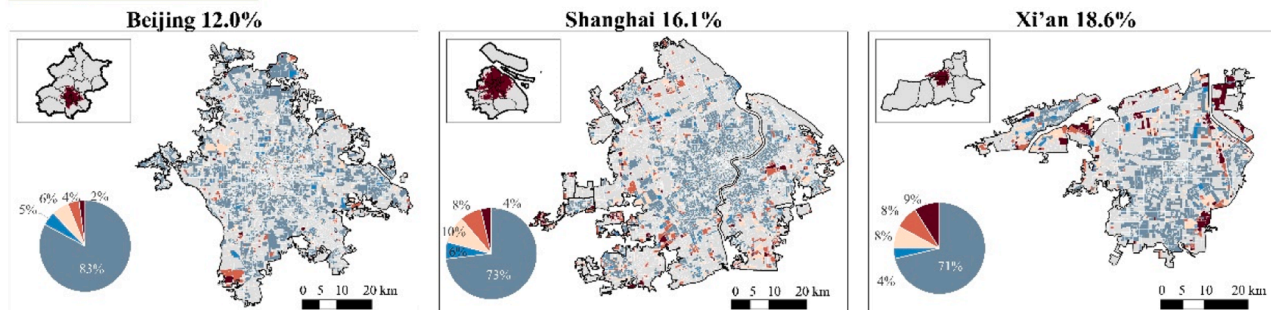
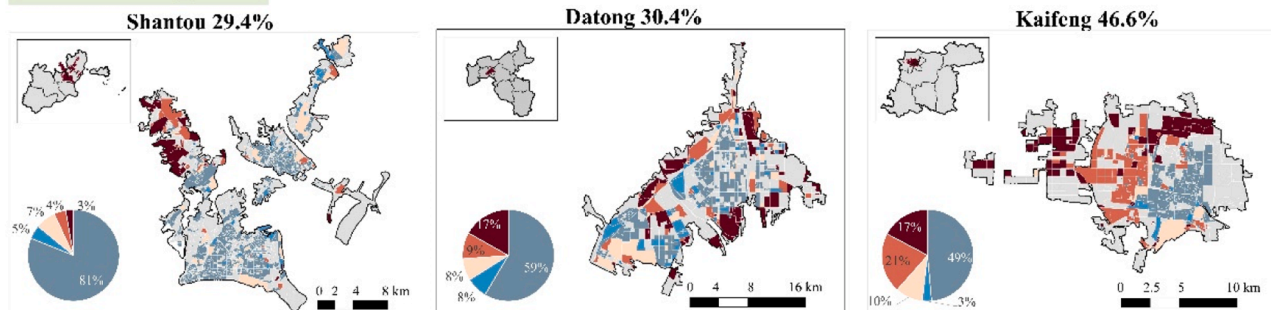
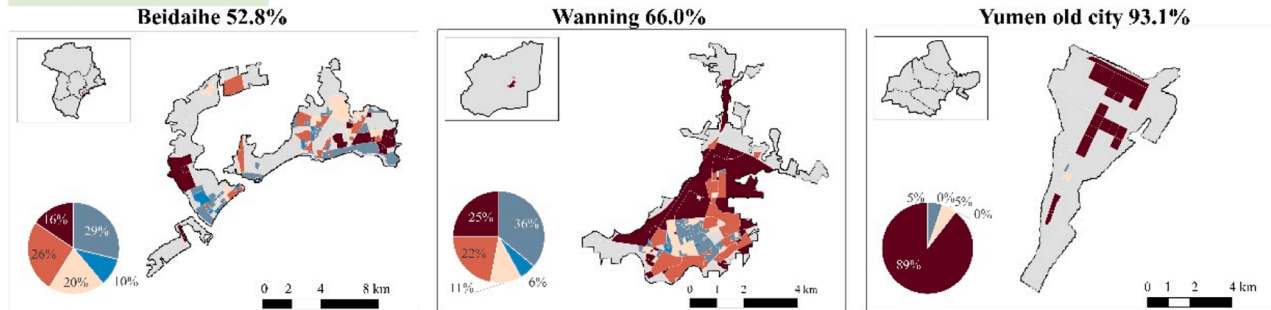


Fig. 7. Typology of urban spatial and demographic conditions based on the Core–Periphery Occupancy Gap and Shrinkage Rate. The four quadrants represent Core-Dominant Growth, Periphery-Dominant Growth, Core-Dominant Shrinkage, and Periphery-Dominant Shrinkage. Bubble size indicates the physical area of each spatial entity.

HVR ≤ 10%**10% < HVR ≤ 25%****25% < HVR ≤ 50%****50% < HVR ≤ 100%****Location of spatial city:**

Administrative city Spatial city

HVR:

Non-residential block 0% 10% 25% 50% 75% 100%



Fig. 8. Intra-urban housing vacancy patterns at residential block scale in representative cities. This figure presents a panel of 12 cities, categorized into four groups by their overall HVR (%) from low to high. Within each city, individual blocks are colored according to their vacancy range (0–10%, gray blue; 11–25%, light blue; 26–50%, orange; 51–75%, light red; 76–100%, dark red). For each spatial city, the top-left inset shows its location within the larger administrative area, and the bottom-left inset illustrates the percentage of blocks within each vacancy category.

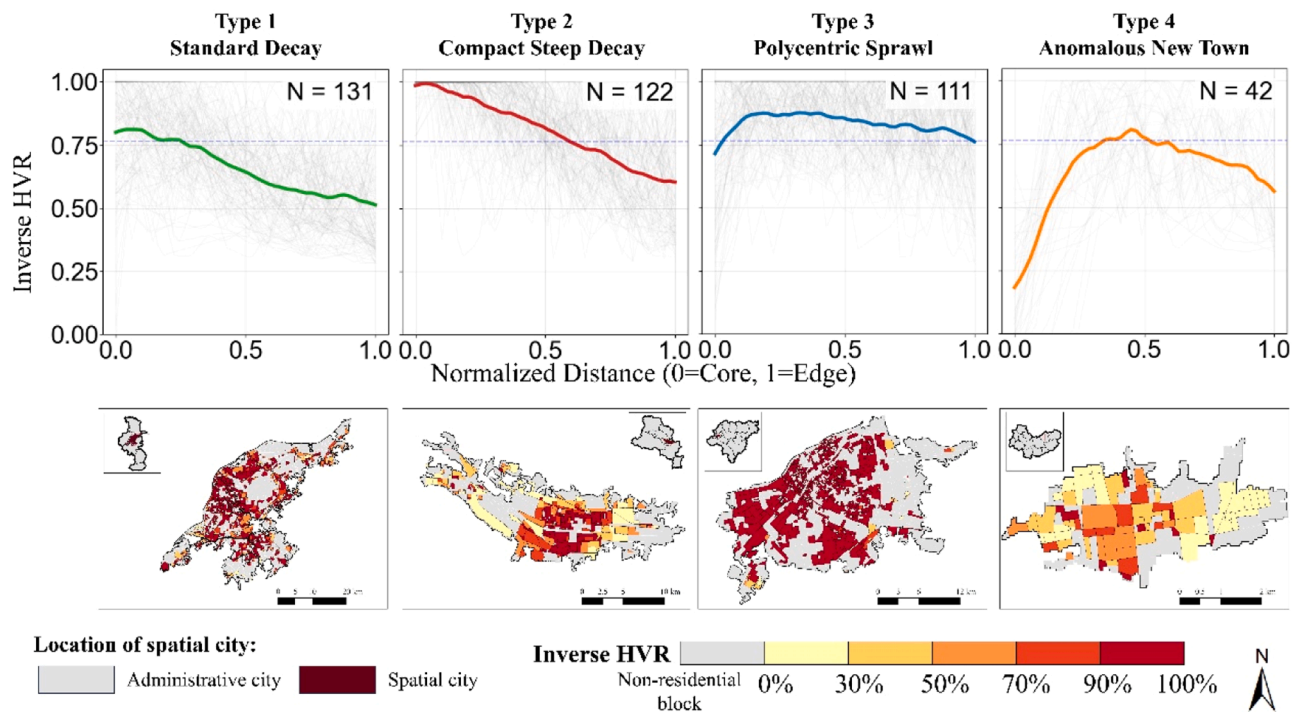


Fig. 9. Morphological archetypes of intra-urban structure identified via K-means clustering of radial inverse vacancy profiles. The four types correspond to Standard Decay, Compact Steep Decay, Polycentric Sprawl, and Anomalous New Town forms.

rooted in rapid leapfrog development or inner-city decay, which fundamentally disrupts natural density gradients.

3.4. Multi-source validation and external-consistency assessment

To ensure the scientific validity and reliability of all aforementioned findings, we performed a series of rigorous, independent validations of our methodological framework against data from diverse sources (Fig. 10).

Validation against Field-Collected Wi-Fi Probe Data: This method operates on the assumption that an occupied household is highly likely to operate an active Wi-Fi router (Access Point, AP). Therefore, the number of active APs within a residential block can serve as an effective proxy for the number of resident households. We conducted on-site data collection using Wi-Fi probes in 53 residential blocks, selected to represent diverse locations, construction scales, housing price levels, and completion years. During collection, volunteers traversed all roads within the community on bicycles while a probe device continuously scanned for and recorded the MAC address, signal strength, and frame type of all Wi-Fi signals within its detection range (up to 100 meters) at 5-second frequency. A mobile app simultaneously recorded the GPS trajectory. Finally, a series of algorithms was applied to these data to estimate the number of resident households and calculate a field-derived HVR (See Supplementary Materials 7.1 for details). The comparison yielded an MAE of 7.92 percentage points, an RMSE of 9.96 percentage points, and a mean bias of -1.25 percentage points.

Validation against Ground-Truth Video Data: We selected 2,700 residential blocks nationwide using a stratified sampling approach. The stratification was based on two criteria: (1) China's official urban administrative hierarchy, comprising five distinct levels (centrally-administered municipalities, provincial capitals, prefecture-level cities, county-level cities, and other functionally defined urban areas), and (2) three calculated HVR categories ($<25\%$, $25\%–50\%$, $>50\%$). For these blocks, we collected 36,985 on-site videos from a major real estate platform (Anjuke). An auditor with a background in urban planning individually reviewed this visual evidence and manually annotated the

vacancy level of each block into the same three tiers (See Supplementary Materials 7.2 for details). Comparison with the video annotations yielded an overall classification accuracy of 84.5%, a Cohen's κ of 0.767, and a macro-F1 score of 0.845 across the 2,700 residential blocks.

Validation against Real Estate Platform Data: Logically, cities with a higher number of vacant housing units should also have more properties listed for rent or sale on online real estate platforms. We scraped for-sale and for-rent listings (3.77 million total) from the Anjuke platform (<https://www.anjuke.com/>) and aggregated them to the spatial city level. To ensure the robustness of the analysis and mitigate potential biases from sparse data coverage in smaller cities, we included only the 289 cities with over 500 total listings in this validation. A comparison between our estimated number of vacant units and the number of listed properties in these cities revealed a significant positive correlation (correlation coefficient = 0.63).

Validation against Search Engine Data: We queried the top 20 search results for the HVR of each city on China's most popular search engine (Baidu), yielding 7,032 pieces of information for 356 cities. Using the GPT-4 large language model API, we systematically retrieved, parsed, and structured this information to extract the reported city-level HVR, year, and source. Results from credible sources, such as commercial institution reports and government bulletins, were manually curated to serve as a validation benchmark, resulting in a dataset for 43 cities (See Supplementary Materials 7.3 for details). Despite variations in the methodologies and years of these media-reported estimates, they exhibited a consistent trend with our calculated results across the vast majority of cities, with an MAE of 7.43 percentage points, an RMSE of 9.46 percentage points, and a mean difference of $+1.19$ percentage points.

Together, these four analyses provide complementary but distinct evidence: panels a and d quantify continuous error and signed bias, panel b evaluates categorical agreement, and panel c assesses external consistency only. The results support the utility of the framework.

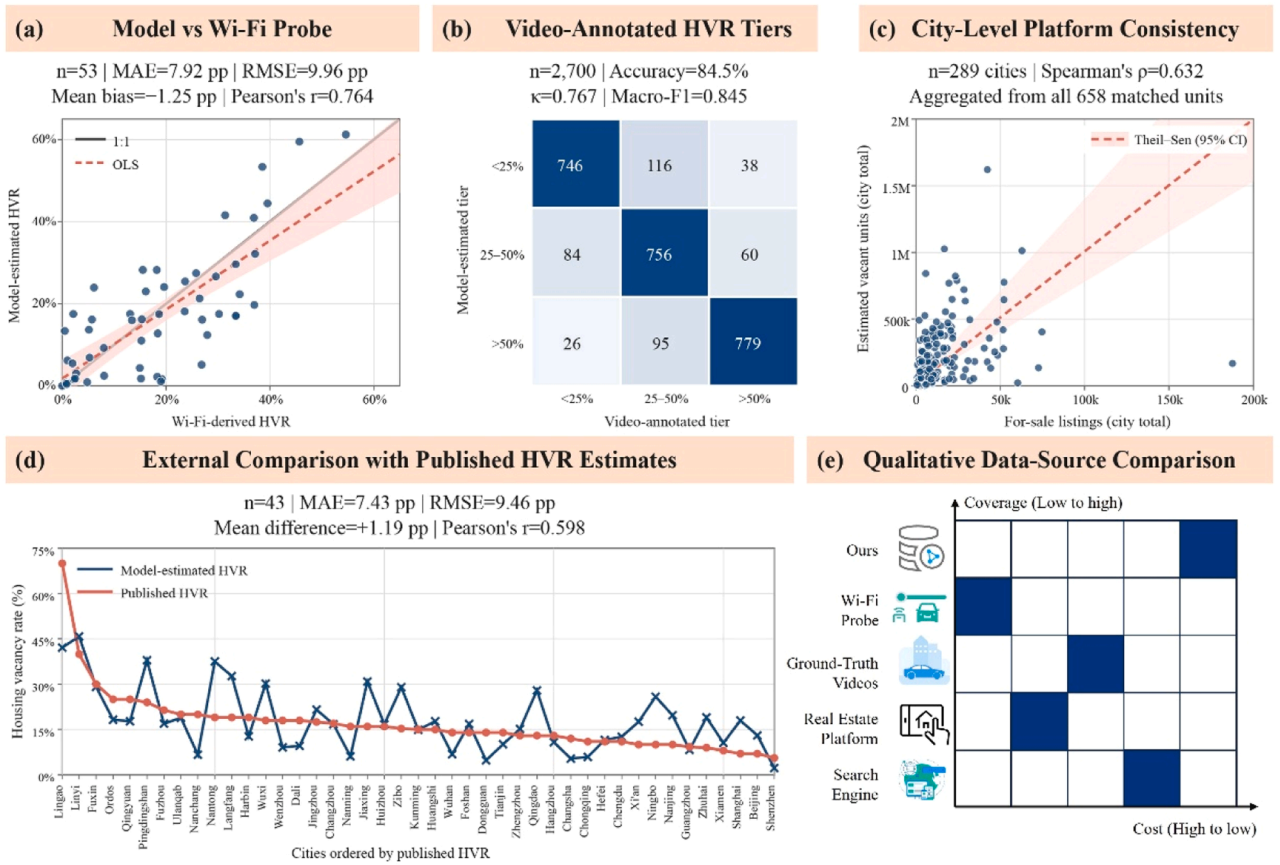


Fig. 10. Multi-source evaluation of the HVR estimation framework. (a) Agreement between model-estimated and Wi-Fi-probe-derived block-level HVRs for 53 paired residential blocks. The panel reports MAE, RMSE, mean bias, and Pearson's r and displays the 1:1 reference line, OLS trend, and 95% confidence interval. (b) Confusion matrix comparing model-estimated HVR tiers with video-annotated tiers for 2,700 residential blocks. Overall accuracy, Cohen's κ , and macro-F1 are reported. (c) City-level external-consistency assessment comparing the total number of for-sale and for-rent listings with the estimated total number of vacant housing units in 289 cities. Spearman's ρ and a Theil-Sen robust trend with a 95% bootstrap confidence band are shown. (d) Comparison between model-estimated and published HVRs for 43 cities, reporting MAE, RMSE, mean difference, and Pearson's r . (e) Qualitative comparison of data-source coverage and acquisition cost.

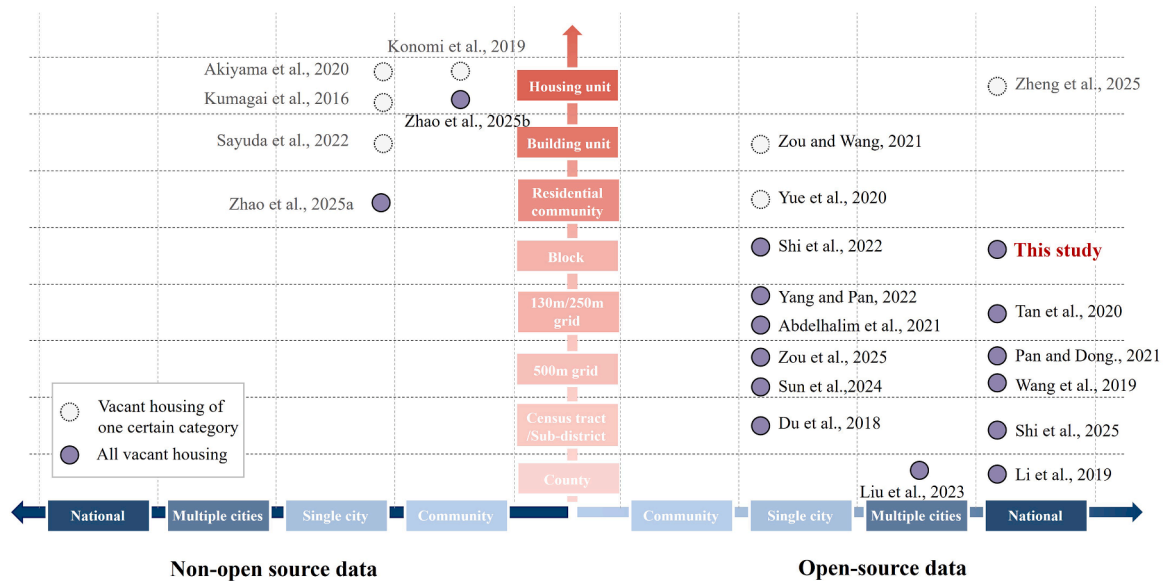


Fig. 11. Comparison of our methodology against existing studies. The comparison is structured along two key methodological dimensions: research scope (horizontal axis) and the spatial unit of analysis (vertical axis). The chart distinguishes between studies based on non-open data sources (left panel) and those utilizing open data (right panel). Our work ('This study') is positioned in the upper right, demonstrating its unique capability to achieve both national scope and fine-grained, block-level granularity using an open-data framework.

4. Discussion

This study, through a framework built entirely on open-source data, presents the first comprehensive, high-resolution assessment of China's housing vacancy. It reveals a national average urban housing vacancy rate (HVR) of 23.46%, corresponding to over 77 million vacant housing units in China. This finding not only quantifies the profound mismatch between housing supply and actual residential demand but, more importantly, provides an unprecedentedly granular empirical foundation for housing policy, urban planning, and sustainable development research in China and globally.

4.1. Bridging the global data divide: a scalable, high-precision open-source framework

Accurately estimating the HVR, particularly in nations lacking a systematic official census, has long been a global challenge. While previous studies have explored various avenues, these efforts have been persistently constrained by the fundamental limitations of available data and methods. Our work constitutes a key step forward by systematically overcoming these long-standing barriers through a series of methodological innovations. Fig. 11 below provides a detailed comparison of our methodology against existing studies, visually summarizing our contributions (Supplementary Table S7).

Unifying Block-Level Granularity with National Coverage. A central challenge in large-scale urban studies has been overcoming the inherent methodological trade-off between spatial resolution and geographical extent. This study represents a significant methodological breakthrough by reconciling these two critical dimensions: it is the first to advance housing vacancy monitoring to the fine-grained "block" scale while simultaneously achieving comprehensive national coverage across 2,813 cities. A review of existing literature identifies six pioneering studies that established the foundation for national-scale housing vacancy measurement in China. While instrumental in shaping the field, these foundational works inevitably navigated trade-offs dictated by the data environments of their time. For instance, studies effectively leveraging night-time light (NTL) data (e.g., DMSP/OLS (Li et al., 2019; Pan and Dong, 2021), NPP/VIIRS (Wang et al., 2019), LuoJia-01 (Shi et al., 2022)) successfully achieved broad national coverage. However, due to the inherent sensor resolutions available, these analyses prioritized scale over detail, operating primarily at coarser grid (130 m–1 km) or city levels. Conversely, innovative approaches using interior real estate imagery provided unprecedented, high-fidelity insights into specific vacancy typologies (e.g., sold-but-unrenovated units). Yet, this precision inherently necessitated a focus on specific market segments rather than the complete vacancy landscape (Zheng et al., 2025). Similarly, census-based research offered unmatched demographic depth at the sub-district level but prioritized broader housing utilization efficiency over explicit vacancy rates (Shi et al., 2025). By transcending these historical data constraints, our approach provides a unified solution that does not compromise granularity for scale, offering a new baseline for precision urban governance.

Global Scalability via Open-Source Data. The methodology possesses high scalability, a feature directly resulting from a low-cost framework built on open-source and public data. While the vast majority of prior research has been confined to single-city analyses (Yang and Pan, 2022; Abdelhalim et al., 2021; Niu, 2018; Sun et al., 2024; Zou et al., 2025) due to the prohibitive financial and temporal costs of data acquisition, our approach transcends these geographical constraints. By relying on globally available resources—such as 3D building models and high-resolution population grids—our framework ensures broad applicability without the need for expensive commercial partnerships. Although this study utilized a China-specific dataset (PopSE), global alternatives like WorldPop (Tatem, 2017) and LandScan (Lebakula et al., 2025) ensure the framework's portability. The nationwide application allowed us to evaluate whether the framework could

be implemented consistently across differences in regional development, urban scale, and housing typology within a large study area. This scalability stands in stark contrast to alternative methods where cost directly limits scale; for instance, approaches based on electricity consumption are constrained by commercial data access (Akiyama et al., 2020; Zou and Wang, 2020), while field surveys (Konomi et al., 2019) and mobile signaling data or LBS data (Zhao et al., 2025) are rendered unscalable by prohibitive labor and partnership expenses. Consequently, our methodology offers a practical and affordable tool for other nations to monitor housing resources and advance sustainable development goals.

Methodological Improvements in Housing Supply and Demand Estimation. Our method achieves high accuracy by fundamentally improving the assessment of both housing supply and demand. On the housing supply side, we used a comprehensive 3D building dataset to estimate total housing stock, thereby overcoming the flaws of relying on real estate platform data. Such platform data suffer from severe sample selection bias (Zhao et al., 2025), as they only cover actively listed properties. This leads to a systematic underestimation of total stock by missing unlisted vacant properties (e.g., developer-held new homes, private investment properties) and by having poor coverage in smaller cities and for non-commodity housing types (Chen et al., 2015; Zheng et al., 2025; Zou and Wang, 2022). Our approach covers all building types, providing a far more comprehensive and unbiased stock estimate. On the residential demand (population) side, we utilized a 100-meter resolution population grid (PopSE) combined with the concave hull technique to estimate the number of resident households, achieving precision far exceeding coarse estimates from NTL data. The use of NTL data as a population proxy is constrained by several factors, including its inability to distinguish residential from commercial illumination and distortions from blooming effects and saturation (Du et al., 2018). Our method directly uses calibrated population distribution data and precisely allocates it to residential building spaces, significantly enhancing the accuracy of occupancy assessment. To quantitatively verify this accuracy, we conducted a rigorous uncertainty propagation analysis incorporating errors from remote sensing extraction, height prediction, and population downscaling (see Supplementary Materials 9). The results demonstrate the robustness of our bottom-up estimation: 74% of residential blocks nationwide exhibit an HVR estimation error margin of less than 0.1.

4.2. Validating methodological necessity through the case of China

Applying the framework across China allowed us to examine the information gained from block-level analysis relative to coarser spatial units. The results reveal substantial within-city variation in HVR across urban areas with different settlement forms and housing typologies. These patterns demonstrate that block-level resolution provides information on intra-urban vacancy distributions that may be obscured by spatial aggregation.

Implications for Urban Shrinkage and Regeneration. Combining the Core-Periphery Occupancy Gap with the Shrinkage Rate (Fig. 7) provides a diagnostic framework for urban policy. Cities classified as Core-Dominant Shrinkage simultaneously exhibit demographic decline and lower stock-weighted occupancy in their peripheral areas than in their cores. For these cities, conventional growth-oriented planning may be inappropriate; policy responses should instead emphasize 'smart shrinkage', consolidation of existing urban functions, and the planned management of underused housing stock in peripheral areas.

Mechanisms of Anomalous Urban Forms: The identification of Type 2 (Compact Steep Decay) and Type 4 (Anomalous New Town) highlights the dual influence of state planning and geographic constraints (Fig. 9). The "hollow core" or "leapfrog" patterns observed in Type 4 cities, manifesting as a mid-range depression in the inverse vacancy curve are direct consequences of top-down planning interventions, such as the establishment of remote industrial parks or

administrative new towns, which forcibly decouple jobs from housing. Conversely, the steep core-to-periphery decline in inverse HVR observed in Type 2 indicates a pronounced spatial concentration of housing occupancy in the urban core. These non-conforming patterns necessitate a revised urban theory that accounts for the "visible hand" of the state and the "invisible hand" of geography alongside market forces.

These institutional mechanisms provide a plausible interpretation of the spatial patterns identified in this study. In particular, the concentration of high HVRs in peripheral new developments and shrinking cities indicates a temporal and spatial decoupling between housing construction and household formation. Land-conveyance revenues and land-based financing enabled local governments to fund infrastructure and open new development areas, while real-estate construction supported short-term fiscal and growth objectives. Where the anticipated population inflows subsequently failed to materialize—particularly in lower-tier and shrinking cities—the newly created housing stock remained spatially fixed and persistently underutilized. This interpretation is consistent with recent evidence that unused housing is more prevalent where housing completions outpace population growth, developable land is more abundant, and fiscal deficits strengthen local incentives to sell residential land. It also corresponds with the documented decline in housing utilization efficiency across Chinese cities.

These findings confirm that a high-resolution lens representing a methodological necessity that extends far beyond simple technical refinement. Without the block-level precision provided by our framework, these critical mechanisms of urban evolution, and the resulting resource waste, would remain hidden behind average statistics.

4.3. Implications for sustainability: from "concrete burden" to evidence-based policy

The deployment of this framework unmask the profound environmental scale of housing vacancy, revealing that the 77 million identified vacant units are not merely an economic inefficiency but a massive "concrete burden" of embodied carbon that conflicts directly with Climate Action (SDG 13) and Responsible Consumption (SDG 12). This phenomenon signifies a critical failure in global resource allocation, where vast quantities of emissions are released for structures that serve no function. However, by quantifying this stock, our framework transforms this environmental liability into a diagnostic asset for achieving Sustainable Cities and Communities (SDG 11), facilitating a pivotal shift in urban governance from "growth-oriented" expansion to "stock optimization".

The precise diagnostics enabled by this method allow for the operationalization of this shift through differentiated policy responses. For instance, in cities classified as Core-Dominant Shrinkage, the results support restricting additional construction and adopting 'smart shrinkage' strategies that consolidate existing urban functions and manage underused housing stock in peripheral areas. In contrast, the explicit localization of vacancy belts in Anomalous New Towns can support targeted functional repurposing. Ultimately, as we approach the 2030 SDG deadline, this tool provides the essential evidence base to curb the hidden environmental costs of urbanization and transform vacant housing from a wasted resource into a sustainable opportunity for meeting future housing needs.

4.4. Limitations and future directions

Despite its contributions, this study has several limitations. First, as a static snapshot of 2020–2021, the analysis cannot capture the temporal evolution of housing vacancy or causally identify the relative contributions of land finance, real-estate development, and demographic change. Second, while our framework captures total physical vacancy—an appropriate and necessary aggregation for assessing environmental costs and embodied carbon—it lacks the capacity to classify the underlying subtypes (e.g., frictional vacancies, seasonal second

homes, and investment holdings). Disentangling these specific categories remains crucial, as different drivers of vacancy necessitate distinct socioeconomic policy interventions. Third, the framework's precision is inherently constrained by the spatial heterogeneity of open-source data availability. In regions with extreme data scarcity or low digitization, the estimated HVR may exhibit biases, necessitating the integration of self-collected field data for targeted, localized calibration. Fourth, extreme city-level HVR estimates may be subject to scale-dependent uncertainty. In small spatial cities, relatively small population and housing-stock denominators increase the sensitivity of HVR estimates to uncertainties in building-use classification, population allocation, and spatial-boundary delineation. Future research should integrate multi-source data (e.g., consumer-behavior and platform data) to distinguish among vacancy types, develop longitudinal HVR estimates, and link these estimates to city-level land-conveyance revenues, residential land supply, housing completions, developer financing, and demographic change. Such longitudinal, multi-source data would enable the dynamic monitoring of housing vacancy and support more rigorous causal assessment of its institutional and socioeconomic drivers.

5. Conclusions

By integrating open-source 3D building morphologies with high-resolution population grids, this study establishes a globally scalable framework that bridges the persistent data divide in housing vacancy monitoring, enabling widespread access to granular urban diagnostics without the constraints of traditional census limitations or commercial data silos. Applying this framework to China's massive urban system as a rigorous stress test uncovers a pronounced national housing vacancy rate of 23.46%, corresponding to approximately 77 million vacant units and over 10 billion square meters of unutilized floor space—a "phantom inventory" theoretically capable of housing over 200 million people. These findings reframe vacancy beyond a purely economic inefficiency to a critical environmental liability, exposing a profound "resource mismatch" and "embodied carbon burden" that directly conflicts with the UN Sustainable Development Goals for Responsible Consumption (SDG 12) and Climate Action (SDG 13). Ultimately, as the 2030 deadline approaches, this work provides the essential empirical baseline to pivot global urban governance from growth-oriented expansion to stock optimization, ensuring that future development leverages existing dormant assets to meet housing needs sustainably.

CRedit authorship contribution statement

Huimin Zhao: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Project administration, Methodology, Investigation, Formal analysis, Data curation. **Yecheng Zhang:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Formal analysis, Data curation. **Fangming Liu:** Writing – review & editing, Writing – original draft. **Ke Wang:** Writing – review & editing, Writing – original draft, Methodology. **Ying Long:** Writing – review & editing, Resources, Methodology, Funding acquisition, Data curation.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Ying Long reports financial support was provided by National Natural Science Foundation of China. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.resconrec.2026.109185](https://doi.org/10.1016/j.resconrec.2026.109185).

Data availability

Data will be made available on request.

References

- Aalbers, M.B., 2016. The financialization of home and the mortgage market crisis. *The Financialization of Housing*. Routledge, pp. 40–63.
- Aalbers, M.B., 2019. Financial geography II: financial geographies of housing and real estate. *Prog. Hum. Geogr.* 43, 376–387. <https://doi.org/10.1177/0309132518819503>.
- Abdelhalim, Z., Matsunaka, R., Oba, T., 2021. Developing zonal house vacancy estimation models with a resolution of 250m: a case study of residential zones in Kyoto city. *Urban Reg. Plan. Rev.* 8, 103–122. <https://doi.org/10.14398/urpr.8.103>.
- Akiyama, Y., et al., 2020. Estimating the spatial distribution of vacant houses using public municipal data. In: *Geospatial Technologies for Local and Regional Development: Proceedings of the 22nd AGILE Conference on Geographic Information Science*, 22, pp. 165–183. https://doi.org/10.1007/978-3-030-14745-7_10.
- Bai, X., Shi, P., Liu, Y., 2014. Society: realizing China's urban dream. *Nature* 509, 158–160. <https://doi.org/10.1038/509158a>.
- Baker, D. (2011). False profits: recovering from the bubble economy.
- Bertaud, A., Malpezzi, S., 2003. The spatial distribution of population in 48 world cities: implications for economies in transition. *Cent. Urban Land Econ. Res.* 32, 54–55. University of Wisconsin.
- Branas, C.C., et al., 2018. Citywide cluster randomized trial to restore blighted vacant land and its effects on violence, crime, and fear. *Proc. Natl. Acad. Sci. USA* 115, 2946–2951. <https://doi.org/10.1073/pnas.1718503115>.
- Brelford, C., Lobo, J., Hand, J., Bettencourt, L.M., 2017. Heterogeneity and scale of sustainable development in cities. *Proc. Natl. Acad. Sci. USA* 114, 8963–8968. <https://doi.org/10.1073/pnas.1606033114>.
- Çelik, Ö., 2024. Cracking the housing crisis: financialization, the state, struggles, and rights. *Hous. Stud.* 39, 1385–1394. <https://doi.org/10.1080/02673037.2024.2333569>.
- Chen, M., Liu, W., Lu, D., Chen, H., Ye, C., 2018. Progress of China's new-type urbanization construction since 2014: a preliminary assessment. *Cities* 78, 180–193. <https://doi.org/10.1016/j.cities.2018.02.012>.
- Chen, X., Rafail, P., 2020. Do housing vacancies induce more crime? A spatiotemporal regression analysis. *Crime Delinq.* 66, 1579–1605. <https://doi.org/10.1177/0011128719854347>.
- Chen, Y., et al., 2024. A 100 m gridded population dataset of China's seventh census using ensemble learning and big geospatial data. *Earth Syst. Sci. Data* 16, 3705–3718. <https://doi.org/10.5194/essd-16-3705-2024>.
- Chen, Z., et al., 2015. Estimating house vacancy rate in metropolitan areas using NPP-VIIRS nighttime light composite data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 8, 2188–2197. <https://doi.org/10.1109/JSTARS.2015.2418201>.
- Clark, C., 1951. Urban population densities. *J. R. Stat. Soc. Ser. A (Gen.)* 114, 490–496. <https://doi.org/10.2307/2981088>.
- 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 Sens.* 10, 1920. <https://doi.org/10.3390/rs10121920>.
- Espey, J.M., Tatem, A.J., Thomson, D.R., 2025. Disappearing people: a global demographic data crisis threatens public policy. *Science* 388, 1277–1280. <https://doi.org/10.1126/science.adx8683>.
- Glaeser, E.L., Gyourko, J., 2005. Urban decline and durable housing. *J. Polit. Econ.* 113, 345–375. <https://doi.org/10.1086/427465>.
- Glaeser, E.L., Gyourko, J., Saiz, A., 2008. Housing supply and housing bubbles. *J. Urban Econ.* 64, 198–217. <https://doi.org/10.1016/j.jue.2008.07.007>.
- Konomi, S., Sasao, T., Hosio, S., Sezaki, K., 2019. Using ambient WiFi signals to find occupied and vacant houses in local communities. *J. Ambient Intell. Humaniz. Comput.* 10, 779–789. <https://doi.org/10.1007/s12652-018-0899-8>.
- Kumagai, K., Matsuda, Y., Ono, Y., 2016. Estimation of housing vacancy distributions: basic Bayesian approach using utility data. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* 41, 709–713. <https://doi.org/10.5194/isprs-archives-XLI-B2-709-2016>.
- Kyba, C.C., et al., 2017. Artificially lit surface of earth at night increasing in radiance and extent. *Sci. Adv.* 3, e1701528. <https://doi.org/10.1126/sciadv.1701528>.
- Langevin, J., et al., 2023. Demand-side solutions in the US building sector could achieve deep emissions reductions and avoid over \$100 billion in power sector costs. *One Earth* 6, 1005–1031. <https://doi.org/10.1016/j.oneear.2023.07.008>.
- Lebakula, V., Sims, K., Reith, A., et al., 2025. LandScan Global 30 arcsecond annual global gridded population datasets from 2000 to 2022. *Sci. Data* 12, 495. <https://doi.org/10.1038/s41597-025-04817-z>.
- Li, H., 2019. A remote sensing-based vacancy area index for estimating housing vacancy and ghost cities in China. In: *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*, pp. 7518–7521. <https://doi.org/10.1109/IGARSS.2019.8898197>.
- Liu, T., 2023. Explaining China's housing vacancies: a theory based on the incentives of local government officials. *J. Gov. Econ.* 10, 100077. <https://doi.org/10.1016/j.jge.2023.100077>.
- Liu, X., et al., 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, 831. <https://doi.org/10.3390/land12040831>.
- Ma, S., Long, Y., 2020. Functional urban area delineations of cities on the Chinese mainland using massive Didi ride-hailing records. *Cities* 97, 102532. <https://doi.org/10.1016/j.cities.2019.102532>.
- Ma, T., Zhou, C., Pei, T., Haynie, S., Fan, J., 2014. Responses of Suomi-NPP VIIRS-derived nighttime lights to socioeconomic activity in China's cities. *Remote Sens. Lett.* 5, 165–174. <https://doi.org/10.1080/2150704X.2014.890758>.
- Mirza, M.U., et al., 2021. Global inequality remotely sensed. *Proc. Natl. Acad. Sci. USA* 118, e1919913118. <https://doi.org/10.1073/pnas.1919913118>.
- Molloy, R., 2016. Long-term vacant housing in the United States. *Reg. Sci. Urban Econ.* 59, 118–129. <https://doi.org/10.1016/j.regsciurbeco.2016.06.002>.
- Morrison, N., 2003. Neighbourhoods and social cohesion: experiences from Europe. *Int. Plan. Stud.* 8, 115–138. <https://doi.org/10.1080/13563470305154>.
- Niu, X., 2018. Estimating housing vacancy rate in Qingdao city with NPP-VIIRS nighttime light and geographical national conditions monitoring data. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* 42, 1319–1326. <https://doi.org/10.5194/isprs-archives-XLII-3-1319-2018>.
- Pan, J., Dong, L., 2021. Spatial identification of housing vacancy in China. *Chin. Geogr. Sci.* 31, 359–375. <https://doi.org/10.1007/s11769-020-1171-7>.
- Röck, M., Balouktsi, M., Saade, M.R.M., 2023. Embodied carbon emissions of buildings and how to tame them. *One Earth* 6, 1458–1464. <https://doi.org/10.1016/j.oneear.2023.10.018>.
- Shi, L., et al., 2022. Estimating housing vacancy rates at block level: the example of Guiyang, China. *Landsc. Urban Plan.* 224, 104431. <https://doi.org/10.1016/j.landurbplan.2022.104431>.
- Shi, L., Leichte, T., Huang, X., Wurm, M., Taubenböck, H., 2025. The decreasing housing utilization efficiency in China's cities. *Nat. Cities* 1–11. <https://doi.org/10.1038/s44284-024-00177-8>.
- Spader, J., Schuetz, J., Cortes, A., 2016. Fewer vacants, fewer crimes? Impacts of neighborhood revitalization policies on crime. *Reg. Sci. Urban Econ.* 60, 73–84. <https://doi.org/10.1016/j.regsciurbeco.2016.07.005>.
- Sun, H., Zhou, G., Liu, Y., Fu, H., Jin, Y., 2024. Housing vacancy identification in shrinking cities based on multi-source data: a case study of Fushun city in Northeast China. *J. Geogr. Sci.* 34, 89–111. <https://doi.org/10.1007/s11442-024-2196-0>.
- Sun, L., Chen, J., Li, Q., Huang, D., 2020. Dramatic uneven urbanization of large cities throughout the world in recent decades. *Nat. Commun.* 11, 5366. <https://doi.org/10.1038/s41467-020-19158-1>.
- Tan, Z., Wei, D., Yin, Z., 2020. Housing vacancy rate in major cities in China: perspectives from nighttime light data. *Complexity* 2020, 5104578. <https://doi.org/10.1155/2020/5104578>.
- Tatem, A.J., 2017. WorldPop, open data for spatial demography. *Sci. Data* 4, 1–4. <https://doi.org/10.1038/sdata.2017.4>.
- Tong, D., Chu, J., MacLachlan, I., Qiu, J., Shi, T., 2023. Modelling the impacts of land finance on urban expansion: evidence from Chinese cities. *Appl. Geogr.* 153, 102896. <https://doi.org/10.1016/j.apgeog.2023.102896>.
- UN-Habitat. (2022). World cities report 2022: envisaging the future of cities.
- UNEP. (2022). 2022 Global status report for buildings and construction: towards a zero-emission, efficient and resilient buildings and construction sector.
- Ürge-Vorsatz, D., et al., 2020. Advances toward a net-zero global building sector. *Annu. Rev. Environ. Resour.* 45, 227–269. <https://doi.org/10.1146/annurev-environ-012420-054843>.
- Wang, L., Fan, H., Wang, Y., 2019. An estimation of housing vacancy rate using NPP-VIIRS night-time light data and OpenStreetMap data. *Int. J. Remote Sens.* 40, 8566–8588. <https://doi.org/10.1080/01431161.2019.1615655>.
- Wardrop, N.A., et al., 2018. Spatially disaggregated population estimates in the absence of national population and housing census data. *Proc. Natl. Acad. Sci. USA* 115, 3529–3537. <https://doi.org/10.1073/pnas.1715305115>.
- Windows, B., 1982. The police and neighborhood safety. *Atl. Mon.* 249, 29–38.
- Xu, X., Tan, M., Liu, X., Wang, X., Xin, L., 2023. Stability and changes in the spatial distribution of China's population in the past 30 years based on census data spatialization. *Remote Sens.* 15, 1674. <https://doi.org/10.3390/rs15061674>.
- 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. *Appl. Sci.* 12, 12328. <https://doi.org/10.3390/app122312328>.
- Yu, B., Zhou, X., 2024. Land finance and urban sprawl: evidence from prefecture-level cities in China. *Habitat Int.*, 103074. <https://doi.org/10.1016/j.habitatint.2024.103074>.
- Yue, X., Wang, Y., Zhao, Y., Zhang, H., 2022. Estimation of urban housing vacancy based on daytime housing exterior images—a case study of Guangzhou in China. *ISPRS Int. J. Geo-Inf.* 11, 349. <https://doi.org/10.3390/ijgi11060349>.
- Zhang, C., Yang, L., Wiedenhofer, D., Guo, J., Chen, Z., Li, S., Wang, Q., 2026. Building material stock drives embodied carbon emissions and risks future climate goals in China. *Nat. Clim. Change* 64–171.
- Zhang, Y., Hua, Y., Ma, W., Luo, Z., Long, Y., Li, Y., 2026. The computational microscope: 28,000 pathway tests reveal how urban expansion triggers the hidden carbon-heat-haze cascade. *Resour. Conserv. Recycl.* 232, 108944. <https://doi.org/10.1016/j.resconrec.2026.108944>.
- Zhang, Y., Zhao, H., Long, Y., 2025. CMAB: a multi-attribute building dataset of China. *Sci. Data* 12, 430. <https://doi.org/10.1038/s41597-025-04730-5>.

- Zhao, H., et al., 2025. Vacancy or occupancy? A comparative analysis of methods for identifying short-term residential unit vacancy in multi-family housing. *Build. Environ.* 285, 113608. <https://doi.org/10.1016/j.buildenv.2025.113608>.
- Zhao, H., Kan, C., Long, Y., 2025. Housing vacancy rate estimation in high-rise residential communities: an experiment utilizing multi-sourced data in Beijing. *Cities* 166, 106187. <https://doi.org/10.1016/j.cities.2025.106187>.
- Zheng, H., Zhang, R., Yin, X., Wu, J., 2025. Unused housing in urban China and its carbon emission impact. *Nat. Commun.* 16, 1985. <https://doi.org/10.1038/s41467-025-57217-7>.
- Zou, S., Wang, L., 2021. Detecting individual abandoned houses from google street view: a hierarchical deep learning approach. *ISPRS J. Photogramm. Remote Sens.* 175, 298–310. <https://doi.org/10.1016/j.isprsjprs.2021.03.020>.
- Zou, S., Wang, L., 2021. Individual vacant house detection in very-high-resolution remote sensing images. *Ann. Am. Assoc. Geogr.* 110 (2), 449–461. <https://doi.org/10.1080/24694452.2019.1665492>.
- Zou, S., Wang, L., 2022. Mapping individual abandoned houses across cities by integrating VHR remote sensing and street view imagery. *Int. J. Appl. Earth Obs. Geoinf.* 113, 103018. <https://doi.org/10.1016/j.jag.2022.103018>.
- Zou, Y., et al., 2025. A study on residential community-level housing vacancy rate based on multi-source data: a case study of Longquanyi District in Chengdu city. *Appl. Sci.* 15, 3357. <https://doi.org/10.3390/app15063357>.