An Optimized Approach to Job-Housing Spaces Identification in Urban Areas Using Location-Based Service Data: A Case Study in Haidian District of Beijing, China

Bin Liu[1,2], Lei Yang[1,2], Liding Chen[1,2,3*], Sike Ma[3,4,5]

1. State Key Laboratory of Urban and Regional Ecology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing 100085, China

2. University of Chinese Academy of Sciences, Beijing 100049, China

3. College of ecology and environment, Yunnan University, Kunming 650091, China

4. Institute of International Rivers and Eco-security, Yunnan University, Kunming 650500, China

5. College of Applied Arts and Science of Beijing Union University, Beijing 100191, China

Abstract: Rapid urbanization promotes socio-economic development but also poses challenges for urban management, particularly in achieving a balanced job-housing relationship. Such imbalances can aggravate traffic congestion, increase energy consumption, and reduce commuting efficiency. Addressing these urban issues requires accurate job-housing space identification (JHSI). The emergence of spatiotemporal big data in geography has popularized location-based service (LBS) data, especially mobile signaling data, for JHSI applications. However, employing mobile signaling data for JHSI presents challenges stemming from both dataset limitations and methodological complexities, including data accessibility constraints due to privacy concerns. This study develops an optimized JHSI approach using a novel LBS dataset and changing the identification lens into base stations. The newly adopted dynamic population data features simplified, privacy-sensitive fields. By establishing time thresholds for working and living hours based on local daily routines and applying straightforward statistical processing to these defined base station fields, we can derive estimated job-housing spaces. This approach not only achieves concise, high-precision identification with readily available data but also enables lightweight dataset applications with enhanced feasibility and broader applicability. We implemented this optimized approach in Haidian District, Beijing, using five days of 2023 data to evaluate method's applicability and quantify job-housing imbalances at subdistrict and town scales. Results demonstrate the approach's accuracy and multi-scale utility in assessing job-housing relationships. We contend that this optimized method advances JHSI-related perspectives in macro-level daily research, facilitates further LBS-driven urban applications, and contributes to improving human livability and quality of life in urban areas.

Keywords: Urban areas; Population Dynamic Data; Job-Housing Space Identification (JHSI); Location-Based Service (LBS) Data; Haidian District of Beijing

1 Introduction

With the development of urbanization, the function of urban areas has transformed from basic living and production spaces into complex, multifunctional landscapes. As human-designed and managed environments, urban areas demonstrate substantial differences in spatial organization compared to natural ecosystems. As increasing numbers of people migrate to cities in pursuit of better opportunities, high-quality job-housing spaces have emerged as a crucial metric for assessing urban planning effectiveness and living standards. The job-housing spaces concept rose to prominence following British sociologist Ebenezer Howard's 19th-century "Garden City" proposal (Howard, 1902). This model conceptualized urban environments where employment and residential spaces were intentionally co-located to promote a balanced lifestyle. However, urbanization-related issues continue to disrupt the job-housing balance in urban planning, posing growing challenges to sustainable development and the balance between human activities and land use.

Rapid urbanization, characterized by large-scale population mobility and extensive urban sprawl, has become the primary driver of spatial job-housing imbalance. This phenomenon aligns with the spatial mismatch hypothesis (Kain, 1968) observed during American suburbanization. Originally developed to analyze and quantify the growing spatial disparity between urban residential locations and suburban employment opportunities (E. Wang et al., 2011), this hypothesis has emerged as a critical concern in contemporary urban management (Schleith et al., 2016). Consequently, numerous urban challenges have become prevalent, including severe traffic congestion (Sultana, 2002; Li & Liu, 2016), reduced resident satisfaction, altered commuting behaviors (Lin et al., 2015), prolonged commute times, and increased commuting costs (Yan et al., 2019; Wang et al., 2021). Furthermore, research demonstrates that this imbalance exacerbates environmental degradation (D. Wang, 2017). Particularly, reliance on private vehicles for daily work-living travel has contributed to air pollution and other environmental problems in both developed and developing nations (Guo et al., 2021; Zhou et al., 2016; Sun et al., 2015). China, as the world's second-largest economy (Lin et al., 2015), has undergone exceptionally rapid urbanization since its 1978 economic reforms (Dong & Yan, 2021; Guan et al., 2018), with urbanization rates increasing by nearly 50 percentage points between 1978 and 2023. This transformation has made job-housing imbalance increasingly noticeable. Historically, before the 1990s, Chinese housing was primarily employer- or government-allocated and typically proximate to workplaces. However, post-reform housing policies transitioned from state provision to market-oriented systems (Guan et al., 2018; E. Wang et al., 2011; Zhou et al., 2016), leading to the current prevalence of spatial separation between workplaces and residences. This shift has generated substantial academic interest in China's job-housing dynamics (Ta et al., 2017; E. Wang et al., 2011; Zhou et al., 2016), particularly since the 1990s (Li & Liu, 2016; Ta et al., 2017). Additionally, China's commuting challenges are particularly severe and representative. The 2022 Annual Report on Commuting Time in Major Chinese Cities reveals that over 70% of Chinese cities exceed 60 minutes in average daily commute time, with this figure continuing to rise. Given these circumstances, understanding the relationship between job-housing spatial patterns and urban development becomes crucial for effective urban land use management, environmental planning, and accessibility enhancement transportation system design (Yao & Kim, 2022). Therefore, accurate job-housing space identification (JHSI) serves as a vital quantitative foundation for addressing these challenges and improving urban livability and quality of life.

The implementation of JHSI can be approached through two primary methodologies, distinguished by their data sources and operational frameworks. The first approach employs traditional statistical survey methods, including questionnaire surveys (Horner, 2002; Li & Liu, 2016; Long & Thill, 2015; Schleith et al., 2016) as well as population and economic censuses. However, these traditional methods are constrained by labor requirements, coupled with limited extrapolation capacity due to delayed and low-frequency data collection. The second methodology leverages the growing application of spatial big data in geographical research, which enables comprehensive, sophisticated, and multiscale urban analyses owing to its high precision and spatially embedded social attributes. Among contemporary geographical big data types, Location-Based Service (LBS) data have gained particular prominence. Comprising diverse specialized datasets within LBS data with strong capabilities for monitoring human activities through terminal positioning (Bi et al., 2023; Hadachi & Pourmoradnasseri, 2022; Mihaylova et al., 2007; Rousell & Zipf, 2017; Schmidtke, 2020; Zhao & Gao, 2023), LBS data applications have consequently attracted substantial research attention in JHSI and related urban studies. The accuracy and scope of these applications continue to expand alongside the increasing volume of available LBS-derived datasets. Common LBS datasets include: mobile signaling data from personal devices (Ahas et al., 2010; Alexander et al., 2015; Yang et al., 2021), cellular network data (Isaacman et al., 2011; Gundlegård et al., 2016; P. Zhang et al., 2017), vehicle GPS trajectory data (Mao et al., 2016; Bi et al., 2023; Liu et al., 2020), Points of Interest (POI) data (Jiang et al., 2015; Y. Zhang et al., 2021), internet-based positioning data (e.g., social media check-ins), public transportation tracking data (e.g., smart card data) (Huang et al., 2019; Long & Thill, 2015; Sari Aslam et al., 2019), and mobile phone trace data (Calabrese et al., 2013). Our study specifically focuses on comparative analysis with mobile signaling data-driven JHSI approaches, as this data type has recently become the most widely utilized LBS dataset in urban research. Beyond JHSI applications, mobility tracking through these datasets serves multiple purposes, including mobility network estimation (Louail et al., 2015), behavioral analysis (Calabrese et al., 2013; Ta et al., 2017; Yuan et al., 2012), commute pattern studies (Yan et al., 2019), and examination of mobility-socioeconomic relationships (Zhao & Gao, 2023). Nevertheless, JHSI implementation using mobile signaling data (Alexander et al., 2015; Calabrese et al., 2013; Yang et al., 2021; Wang et al., 2020; Pei et al., 2014) faces notable challenges and limitations, particularly concerning inherent dataset characteristics and methodological constraints of existing JHSI approaches. These include operational complexity, large sample size requirements, intricate processing demands, and limited accessibility due to privacy concerns surrounding personal information. Addressing these limitations represents both a significant research opportunity and necessity, as solutions would advance LBS-driven applications, refine JHSI methodologies for routine research use, and ultimately improve urban applications.

As the capital of China and a major international metropolis, Beijing has drawn considerable scholarly and planning attention due to its rapid urbanization and complex spatial patterns in job-housing relationships. Notably, the scale of the study area can significantly influence research outcomes in urban job-housing spatial analysis, as demonstrated by Horner & Murray (2002) and Small & Song (1992). In recent years, the intense pressures of urban life resulting from accelerated urbanization have led districts to evolve into increasingly distinct agglomerations of urban lifestyles, each with unique characteristics. However, prior research has predominantly focused on the city scale (Zhao et al., 2011), leaving the spatial relationship between job-housing spaces and their impact on daily life at the district or finer levels underexplored. Haidian District, one of Beijing's sixteen administrative divisions, stands out due to its distinctive demographic profile and its leading role in education and high-tech industries. The *Haidian District Planning (2017–2035)* underscores the district's strategic importance, outlining specific targets for population control (limiting the permanent population to 3.13 million by 2035), urban construction land use (capping urban and rural land use at 2,270 km² by 2035), and coordinated functional spatial patterns. Furthermore, as a critical component of the Beijing-Tianjin-Hebei (BTH) coordinated development strategy—which aims to establish a world-class urban cluster in the region by 2030—Haidian District not only serves as the primary hub for technological innovation to drive growth in Tianjin and Hebei but also functions as a key relocation zone for Beijing's non-capital functions. Given this strategic positioning, effective management of job-housing spaces in Haidian is crucial for achieving these multifaceted objectives.

In this study, we aim to accomplish two key objectives: (1) proposing an optimized JHSI approach that utilizes dynamic population data to address limitations in existing main LBS-driven JHSI frameworks, and (2) implementing this approach through a district-scale case study in Haidian District, Beijing, to analyze and map job-housing spaces while quantifying imbalance patterns across its 22 subdistricts and 7 towns. For the first objective, we introduce a novel LBS dataset—dynamic population data—which, like mobile signaling data, is derived from base station signals but differs in its representation of aggregated population clusters rather than individual trajectories. This dataset offers advantages in accessibility and privacy compliance, as it avoids sensitive personal information. Furthermore, we shift the analytical perspective from individual users to base stations in our optimized JHSI framework. Instead of relying on movement frequency during work and living periods, we classify base stations as either "working" or "living" stations based on their monitoring patterns. This allows JHSI to be assessed through two dimensions: (1) extracting geographical information and (2) estimating job-housing population distributions based on the categorized base stations. Overall, this optimized approach, leveraging dynamic population data and an adjusted analytical lens, enhances cost-efficiency, processing speed, feasibility, real-time updating capability, and broader applicability. For the second objective, we apply the optimized framework to Haidian District as a case study, generating quantitative insights into job-housing relationships at a finer urban scale. Using dynamic population data from April 10–14, 2023, we map the spatiotemporal distribution of job-housing spaces and employ the job-housing balance index (JHB) along with its standard deviation (SD) to measure imbalance patterns. The results reaffirm the method's lightweight implementation and high accuracy. Finally, we discuss its potential applications in urban planning, offering policy recommendations not only to address population and housing challenges but also to support its integration into the Beijing-Tianjin-Hebei (BTH) coordinated development strategy.

In summary, the job-housing relationship represents a crucial consideration in urban planning and policymaking for sustainable urban development. Expanding upon the aforementioned studies, our research seeks to develop optimized methods for monitoring job-housing relationships across diverse scenarios and scales, thereby supporting both daily commuting analysis and public policy applications. The study further enhances the understanding of job-housing dynamics in key urban zones while promoting the application of geospatial big data in urban planning and landscape design to mitigate housing-population conflicts. The remainder of the paper is structured as follows. Section 2 provides a comprehensive review of main existing LBS-driven JHSI methodologies, along with theoretical frameworks for data acquisition, including mobile terminal positioning technologies and the structural characteristics of dynamic population data. Section 3 elaborates on the methodological framework for the optimized JHSI approach. Section 4 presents an empirical case study conducted in Beijing's Haidian District from April 10-14, 2023. Sections 5 and 6 present the discussion and conclusions, respectively. The research flowchart is illustrated in Figure 1.

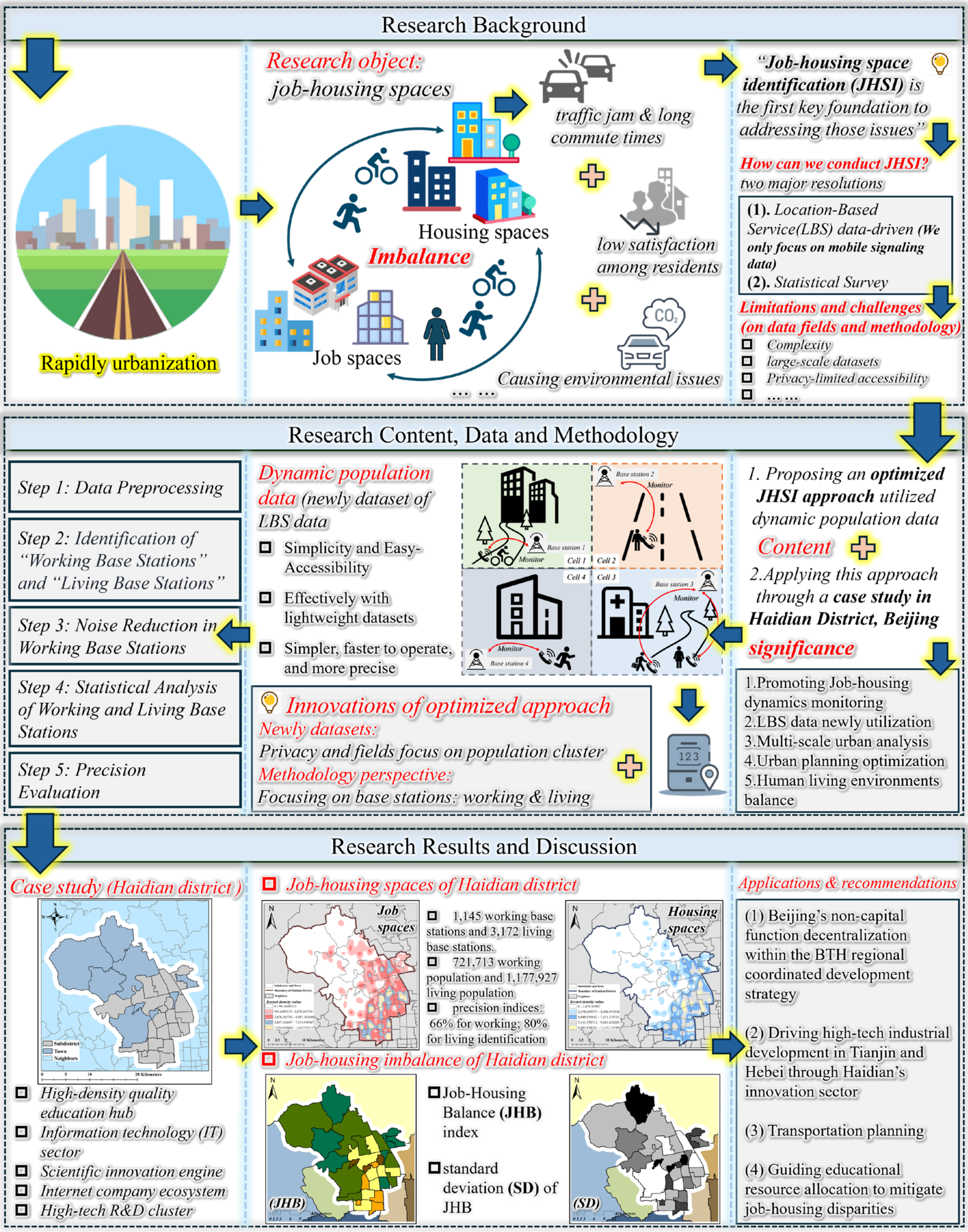

Figure 1: Flow chart of the research

2 Review of Existing JHSI Methodologies and Theories of Data Acquisition

2.1. Methodological Frameworks of Existing LBS Data-Driven JHSI

Among existing LBS data-based methods for JHSI identification, we focus on comparing approaches driven by mobile signaling data, which primarily follow two methodological frameworks: (1) The most widely used framework relies on individual mobile phone users' movement trajectories. Specifically, job-housing spaces are delineated by mining daily commuting patterns between residential and workplace locations. First, thresholds for working and living hours are established based on local daily routines. Next, the frequency of movement trajectories—including locations and time periods—is statistically summarized for each user. Finally, job-housing spaces are identified based on the prevalence of mobile users during working and living hours. High-frequency locations during workday working hours are classified as workspaces, while those dominant during living hours are designated as living spaces. The accuracy of this approach heavily depends on the predefined time thresholds for working/living periods and population mobility patterns (Alexander et al., 2015; Calabrese et al., 2013; Yang et al., 2021; Wang et al., 2020). (2) The second framework utilizes aggregated mobile signaling data, analyzing call volume records from base stations—including temporal distribution patterns and total volume—over short periods (e.g., one week). By assessing communication activity patterns and applying cluster analysis, this method infers land use types (e.g., residential and commercial zones, which encompass job-housing spaces) within each base station coverage area, typically delineated using Voronoi polygons and interpolated to grids (Pei et al., 2014).

While both frameworks are currently employed in mobile signaling data analysis, they still face unresolved challenges: (1) Complex operations on signaling data fields (e.g., DBSCAN and integrated complex cluster analysis), (2) limited scalability across different spatial or temporal resolutions, (3) data acquisition barriers and high costs due to privacy risks, (4) dependence on large-scale datasets, which impedes rapid JHSI implementation and updates, and (5) insufficient precision in identification. Consequently, optimizing both the dataset and the recognition logic within these frameworks remains the key bottleneck in advancing research on job-housing relationship.

2.2 Theories of Data Acquisition

The rapid development of wireless communication and information technologies has driven advances in geospatial and transportation-related research. Specifically, advancements in GIS, GPS, and RS (3S) technologies have enhanced LBS, enabling it to conduct positioning using multiple datasets from mobile terminal devices instead of traditional single-simplified spatial positioning devices. This progress has significantly promoted geospatial services in commercial use(Liao & Dong, 2017; Rousell & Zipf, 2017; Schmidtke, 2020; Weng et al., 2017). For example, Rousell and Zipf (2017) created a prototype navigation service that uses LBS data (sets of landmarks from OpenStreetMap) to generate pedestrian navigation instructions. Weng et al. (2017) proposed a method to extract urban landmarks rapidly from spatial databases, using LBS data (sets of check-ins and local accessibility) as weighted parameters. These examples demonstrate the tremendous potential of LBS technology in urban planning and navigation services. Until today, the widespread use of smartphones and advanced mobile positioning technologies has strengthened the importance of location services by providing alternatives to traditional GPS-based methods, and research on population mobility has expanded significantly (Hadachi & Pourmoradnasseri, 2022; Mihaylova et al., 2007; Yuan et al., 2012). In this study, dynamic population data, another type of LBS data, has not yet been widely used but has shown great potential. This data is obtained through base station positioning technology monitored on mobile phones. In this section, we will detail the principles of base station positioning technology and the structure of dynamic population data.

2.2.1 Base Station Positioning Technology

Base station positioning technology is based on the fundamental infrastructure of cellular networks. This network divides the service areas into a cellular structure with each cell having its own unique cell-ID. The cellular structure commonly has multitype shapes, Voronoi polygons are one of the widely methods employed to divide the whole cellular network (Perera et al., 2015; Sharifzadeh & Shahabi, 2006; Sharifzadeh & Shahabi, 2009; Pei et al., 2014). Base positioning technology offers a simple, economical, and highly available solution with wide space-coverages (both outdoors and indoors) without requiring any upgrades to handsets or network equipment, unlike GPS or WiFi positioning technologies(Perera et al., 2015; Trevisani & Vitaletti, 2004).

In this study, the dynamic population data has been fuzzy processed because the geographic locations of base stations are often irregularly distributed, and their data access is restricted in public research. Specifically, grids composed of square cells with a resolution of 200m × 200m were established based on the original base station data. The geographic coordinates of the grid cells were assigned to the base stations, allowing for spatial errors of 100 to 200 meters while maintaining only one base station per cell. Although hexagonal cells are typically used in cellular networks for more efficient coverage and reduced interference, we consider only square-cell networks where each cell contains a single base station to prevent monitoring interference (Fig. 2).

Based on the above, we can obtain the simplified general form of LBS data from the following principle: Specifically, when a visitor with a power-on mobile phone enters the serving area of a base station, his information will be recorded, thus LBS data on all visitors will be produced. This data is linked to the base stations by recording the visitor number within a certain period. The geographic information of the LBS data is the exact location of the base station, identified by a unique cell ID. The spatial precision of the data depends largely on the service area of the cellular network (Trevisani & Vitaletti, 2004). For example, the locations of Base Stations 1 to 4 (Fig. 2) are assigned to visitors with mobile phones when they enter the service area of the cell where the base station is located. Along with other information, a row of LBS data for each visitor will be generated. However, fast movement between base stations and staying time of the visitors will affect the precision of LBS data. For instance, repeat recordings of the moving drivers in different base stations will lead to error in LBS data. These effects often occur in certain LBS data, such as mobile signaling data, which may become extensive and noisy during specific periods. Data processing techniques are required to remove the noise, including ‘ping-pong switching’ analysis and clustering.

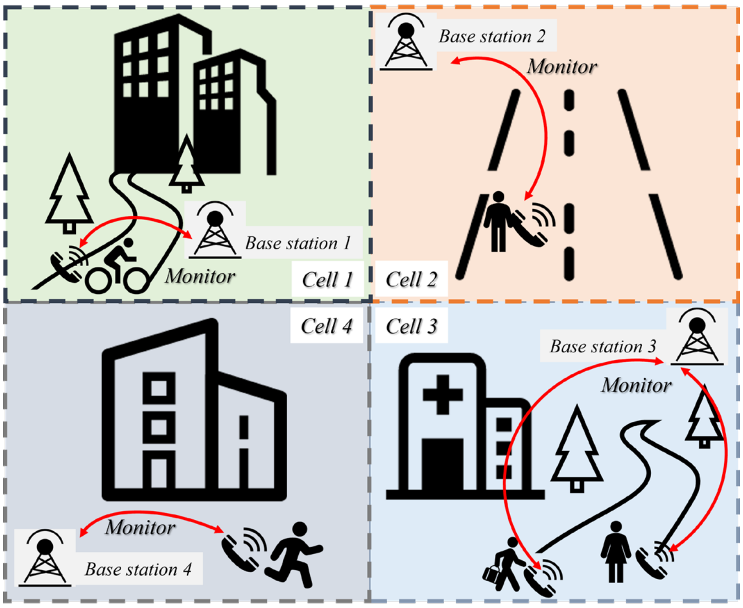

Figure 2: A Simplified View of the Cellular Network and four Examples of Base Station Monitoring

2.2.2 Overview of the Dynamic Population Data Structure

Dynamic population data is the sequence of population deduced from mobile phone signaling data. Mobile signaling data based on the monitoring functions of mobile base stations, is widely used in urban studies based on its diverse data attributes. These include unique IDs of each base station and the mobile phone users, personal information of mobile phone users (such as age), geographical location of base station, lasting time of the signal, and the types of mobile device (Okmi et al., 2023; Yang et al., 2021). In contrast, dynamic population data has more concise and general attribute about the base stations and mobile phone users. In this section, the overview of dynamic population data structure will be given in following 2 aspects:

1. Simplicity and Easy-Accessibility of Dynamic Population Data

Dynamic population data have simpler attributes than mobile signaling data. Only the key information of the mobile users is kept, such as the base station location (i.e., "longitude" and "latitude"), recording time (i.e., "Year-Day-Time"), the Administration Name, and the population number in each base station (i.e., "POPu"). In the dynamic population data, all information related to privacy is removed, making it more accessible to the publics compared to mobile signaling data.

2. More General and Simple Monitoring Principles of Dynamic Population Data

Compared with mobile signaling data, we utilized dynamic population data is based on general and simplified monitoring principles. It provides 24 recordings each day at one-hour intervals. Additionally, this principle avoids repeated recordings of the same visitor at the same base station within an hour and excludes useless short-period recordings. Only the recordings of the visitor who stayed the longest at a given base station within each hour will be retained. Table 1 gives a sample of dynamic population data from four base stations in Beijing's Haidian District at 0:00 on April 10, 2023. The first row of Table 1 indicates that "20**" visitors were in the base station of (116. *****°*E*, 39.*****°*N*) in Haidian District of Beijing at 00:00, April 10th, 2023. This data is simpler compared to mobile signaling data. In mobile signaling data, information about User A and his personal details such as age "M", location (Longitude, Latitude), and actions using the mobile from t_n to t_m are recorded in a field.

Table 1: Example of Dynamic Population Data Attribute Fields

Base station	Longitude	Latitude	Administration Name	Year-Day-Time	POPu
1	116.*****°*E*	39.*****°*N*	Haidian District, Beijing	2023-0410-00	20**
2	116.*****°*E*	39.*****°*N*	Haidian District, Beijing	2023-0410-00	15**
3	116.*****°*E*	39.*****°*N*	Haidian District, Beijing	2023-0410-00	15**
4	116.*****°*E*	39.*****°*N*	Haidian District, Beijing	2023-0410-00	13**

Note: This table presents a sample of dynamic population data monitored in the Haidian District, Beijing, with a spatial resolution of 200m x 200m. The geographic coordinates are based on the WGS84 coordinate system. The "***" is used for privacy protection.

This study employs dynamic population data that strictly adheres to China's Personal Information Protection Law (PIPL) and relevant regulations through comprehensive technical and administrative safeguards. The dataset contains no sensitive personal information, as all collected data were pre-aggregated and anonymized without including device IDs or any personally identifiable information. All analyses were conducted exclusively at the aggregate level, ensuring no individual behaviors could be traced. For spatial anonymization, raw base station positioning data were processed into 200m×200m grids, with system-generated grid centroids offset by at least 100m from actual base station locations, while temporal resolution was reduced to 1-hour intervals to further enhance privacy protection. The data processing protocol fully complies with the anonymization requirements specified in GB/T 35273-2020 *Information Security Technology -*

Personal Information Security Specification, ensuring complete regulatory compliance.

3 Methods

3.1 Optimized Framework for JHSI Using Dynamic Population Data

3.1.1 Framework of Processing Steps

Step 1: Data Preprocessing

The dynamic population data in Table 1 is originated from the monitoring records of all base stations. This step involves re-organizing the attribute fields of the data from the base stations (Fig. 3). It aims to ensure all the records of same base station being grouped into one dataset. Additionally, the fields "Day" and "Time" are separated from the "Year-Day-Time" field for further analysis. For example, Table 2 gives an example of organized data attribute fields by data preprocessing, the population number monitored by the base station 1 with (116.******°*E*, 40.******°*N*) in four periods: 2 AM, 3 AM, and 4 AM on April 10th, and 9 AM on April 11th was identified.

Table 2: Example of Organized Data Attribute Fields.

Base Station	Longitude	Latitude	Administration Name	Time	Day	POPu
1	116.******°*E*	40.******°*N*	Haidian District, Beijing	2	410	9**
1	116.******°*E*	40.******°*N*	Haidian District, Beijing	3	410	7**
1	116.******°*E*	40.******°*N*	Haidian District, Beijing	4	410	6**
1	116.******°*E*	40.******°*N*	Haidian District, Beijing	9	411	15**

Note: This table presents a sample of dynamic population data monitored in the Haidian District, Beijing, with a spatial resolution of 200m x 200m. The geographic coordinates are based on the WGS84. The "***" is used for privacy protection.

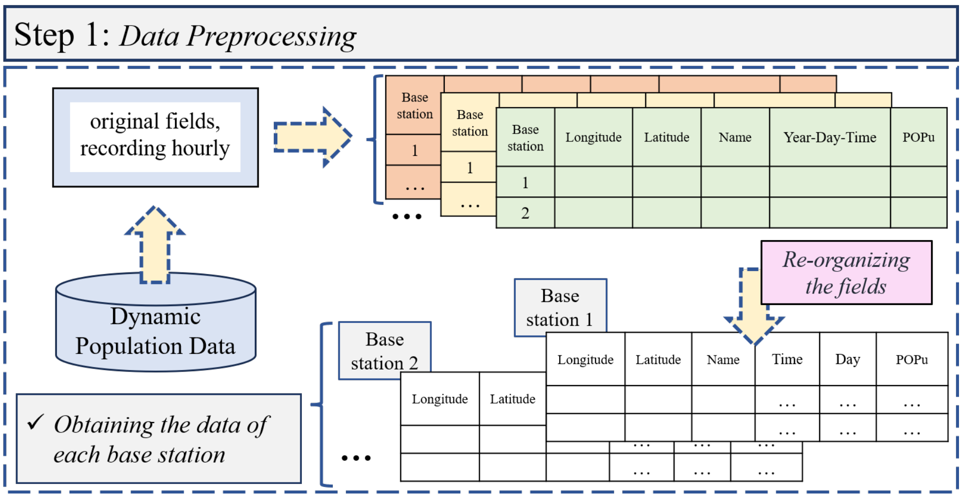

Figure 3: Data Preprocessing in the optimized JHSI Framework

Step 2: Identification of "Working Base Stations" and "Living Base Stations"

In this step, the thresholds on working and living time are set up based on the daily life patterns of local people. The base stations will be defined as working or living ones based on the thresholds through statistical frequency data processing (Fig. 4). After the working base stations and living base stations are defined, the spatial pattern of job-housing spaces can be derived. In practice, the location of the working and living base stations will used as symbols of the location information of job-housing spaces, respectively.

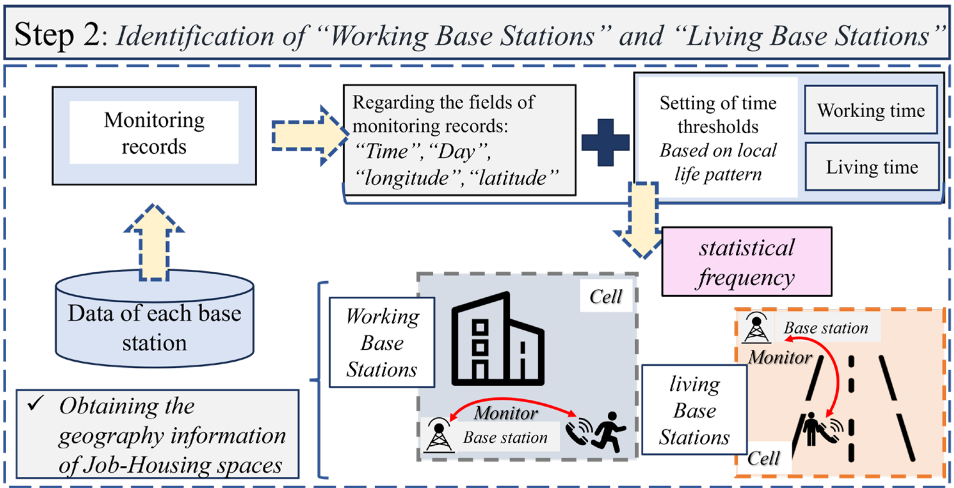

Figure 4: Identification of "Working Base Stations" and "Living Base Stations" in the Optimized JHSI Framework

Step 3: Noise Reduction in Working Base Stations

Before the statistical process, it is necessary to remove the noise information from the working base stations for better accuracy. This is because population mobility is more significant during working hours at working base stations compared to living base stations. Thus, we only performed noise reduction on working base stations. A group of all job base stations is organized by the "Day" field, as shown in Table 3. Detailed procedure is as follows (Fig. 5).

Firstly, based on all job base stations, we calculate the total "POPu" field for each "Day", referred to as "Total POPu". Then, the relationship between the "Time" field and the "Total POPu" is established to identify the general trends of this relationship. Secondly, one or more "Day" groups with significant fluctuation anomalies, which show trends obviously different from other "Day" groups, will be identified as noise. After that, we remove the entire field recordings of the noise in each working base station. Finally, in each base station, we calculate the average correlation coefficient for the remaining "Day" groups and remove the entire field recordings of the "Day" group with the lowest average correlation coefficient. The remaining fields in each working base station are then utilized for further analysis.

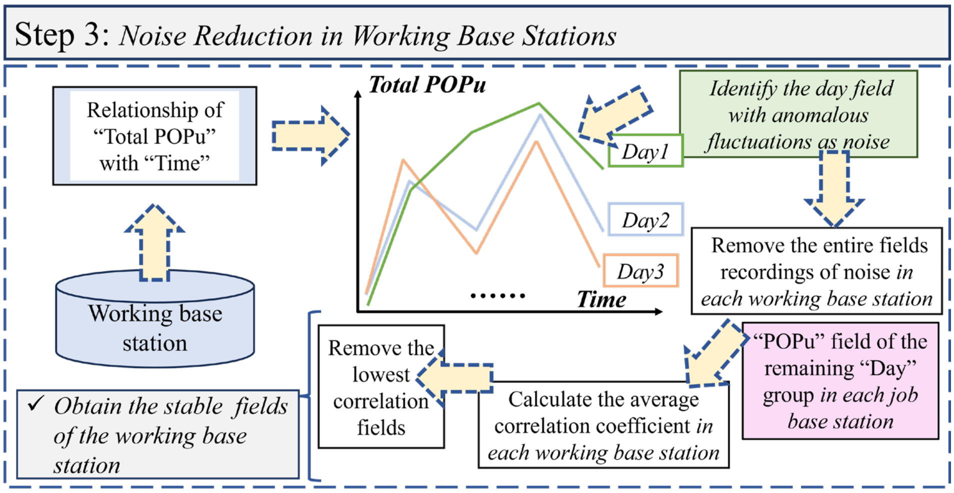

Figure 5: Noise Reduction in Working Base Stations in the optimized JHSI Framework

Step 4: Statistical Analysis of Working and Living Base Stations

Based on the previous step, the geographical spatial information of working-living spaces was obtained. In this step, the final population estimation for each working and living base station will be determined through data processing and statistical analysis (Fig. 6). Specifically, the average number of the remaining “POPu” field will be calculated based on the living time at each living base station. Additionally, due to the significant characteristics of population mobility during working hours, the average number for working base stations will be calculated during several designated working times, such as 8 am, 11 am, and 4 pm(Yang et al., 2021).

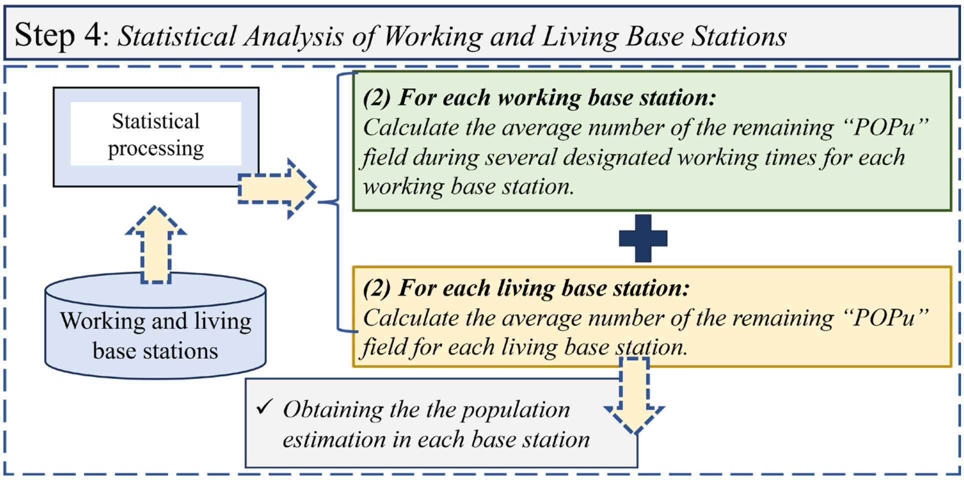

Figure 6: Statistical Analysis of Working and Living Base Stations in the optimized JHSI framework

Step 5: Precision Evaluation

Current research on JHSI often inadequately addresses precision evaluation (Zheng et al., 2023). Given the strong correlation between dynamic population data and mobile signaling data, this study adopts a precision evaluation framework for mobile signaling data-driven methods, emphasizing consistency between identified job-housing population distributions and the actual spatial logic of the study area (Fig. 7). Specifically, precision is verified by comparing regional living/working population statistics with identification results in terms of quantitative relationships (Wang et al., 2020). Two key considerations emerge: (1) Data representativeness: Dynamic population data primarily reflects mobile user monitoring, with coverage limited to specific carrier subscriber groups. Due to privacy constraints, obtaining complete datasets from all carriers remains challenging (not all provide well-protected dynamic population data). Thus, this study applies a scaling method using an “expansion coefficient” (estimated 60–70%) derived from mobile market statistics to align coverage with actual user proportions. (2) Statistical data preprocessing: Since job-housing relationships describe spatial connections between workplaces and residences of employed populations, non-working residents must be excluded from raw residential statistics before precision evaluation but retaining unprocessed employment data. Considering China’s compulsory education, age structure, and retirement policies (60 for men, 55 for women), we exclude populations aged 0–14 and over 65 from raw statistics before expansion coefficient adjustment to derive the final “comparative value” (Formula (1)). This standardized methodology ensures indicator comparability and rigorous precision evaluation.

Based on this framework, JHSI precision is evaluated by calculating a precision index using the identified populations and comparative values (Formula (2)), with the full process illustrated in Figure 7. Note that temporal mismatches between statistical surveys and identification results introduce unavoidable time-lag errors, which should be minimized where possible.

$$Comparative\ value_i = Expansion\ coefficient \times Processed\ Statistical\ data_i \qquad (1)$$

$$Precision\ index_i = \frac{Identified\ Population_i}{Comparative\ value_i} \tag{2}$$

In Formula (1) and (2), the subscript i, covers both working and living scenarios, which respectively correspond to the employment-related and residence-related population values processed in the aforementioned steps.

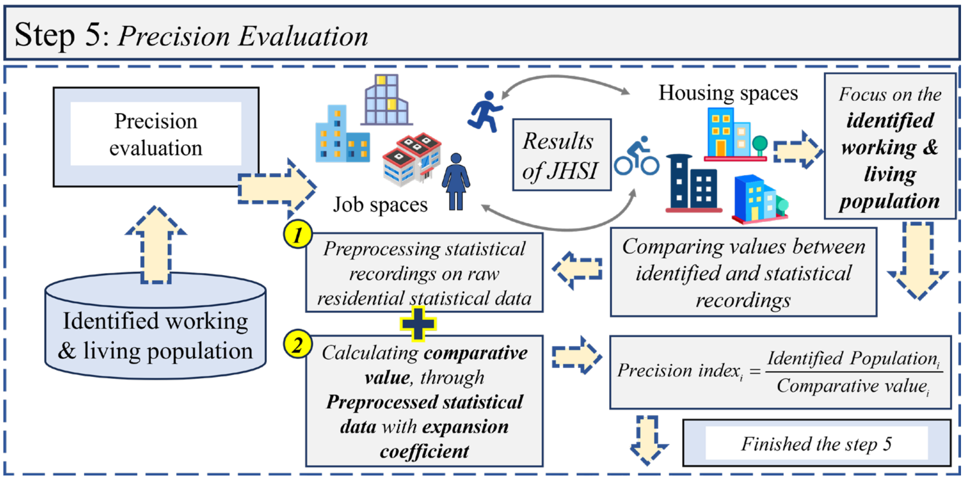

Figure 7: Precision Evaluation in the optimized JHSI framework

3.1.2 Improvement of the Optimized JHSI Approach with Dynamic Population Data

Compared to existing JHSI approaches based on LBS data, particularly those relying on mobile signaling data, our optimized method offers notable advancements. (1) The optimized JHSI approach is simpler, faster, and more precise. Dynamic population data employs straightforward fields (e.g., the "POPu" field for population mobility clustering), reducing data processing complexity. The method focuses on each base station—which have fixed locations—enabling rapid implementation and broader regional-scale applicability. Unlike mobile signaling data driven that require substantial computational resources and clustering algorithms (e.g., DBSCAN), the optimized JHSI framework is user-friendly and adaptable to diverse urban contexts. Furthermore, since dynamic population data contains no personally identifiable information, it is highly accessible, allowing the method to function as a simple, single-source-driven solution. Additionally, our approach more efficiently defines base stations from a functional perspective while using straightforward statistical processing to identify stable monitoring trends. This simplifies complex operations such as supervised classification and clustering while establishing a foundation for high-precision identification (particularly in contrast with the work of Pei et al., 2014). (2) The optimized JHSI framework performs effectively with smaller sample sizes. By shifting the analytical focus to spatially stable base stations—rather than individual mobility patterns—the method enables feasible implementation with lightweight datasets. In contrast, traditional approaches relying on mobile signaling data often demand larger samples (e.g., Wang et al., 2020, who used one month of mobile signaling data). The reduced data requirements also facilitate dynamic monitoring and significantly shorten update cycles. (3) The optimized JHSI approach demonstrates higher feasibility and broader applicability. Many LBS datasets are restricted due to corporate or institutional ownership and high access costs, limiting their availability to public and independent researchers. In contrast, dynamic population data is more accessible, cost-effective, and suitable for routine use by public and non-institutional organizations across various scenarios. Since it excludes sensitive user information or detailed activity records, the method enhances the practicality of JHSI applications. Moreover, it supports multi-scale adjustments based on research needs, enabling further investigations into job-housing imbalances and spatial functionality at finer scales.

3.2 Calculating of Job-Housing Imbalance

After completing the procedure, the urban job-housing spaces will be identified, including information on working and living locations, as well as the job-housing populations. Then, we can observe the overall distribution of job-housing spaces through visualization. Additionally, the more detailed imbalance in job-housing spaces at finer scales, such as subdistricts and towns, can be quantified by calculating indices of job-housing imbalance, commonly using the Job-Housing Balance (JHB) index and its standard deviation (SD) (formulas (3) and (4)) (Weitz et.al, 1997; Wang et.al, 2022).

$$JHB_{ij} = \frac{W_{ij} / W_i}{L_{ij} / L_i} \tag{3}$$

$$SD = \left| JHB_{ij} - 1 \right| \tag{4}$$

In formula (3) and (4), JHB_{ij} represents the JHB index for subdistricts or towns j in district i. W_{ij} denotes the number of employed people in subdistricts or towns j in district i, while W_i is the total number of employed people in district i. Similarly, L_{ij} indicates the number of residents in subdistricts or towns j in district i, and L_i is the total number of residents in district i.

If $JHB_{ij} = 1$, it indicates that the employment and residential functions are matched. If $JHB_{ij} > 1$, it means the proportion of employed people is higher than that of residents, suggesting that the employment function is stronger than the residential function. Conversely, if $JHB_{ij} < 1$, it indicates that the residential function is dominant. Additionally, the standard deviation (SD) of the JHB index can be used to measure the degree of job-housing spatial matching in the area. A smaller SD value indicates a better match between working and living spaces, while a larger SD value indicates a poorer match between them.

4 Case Study in Haidian District, Beijing, China

In this section, the optimized approach was employed to identify the Job-housing spaces in Haidian District by using dynamic population data from April 10th to April 14th, 2023 (five consecutive workdays, without including the effects of holidays or other factors).

4.1 Study area

Haidian District, one of Beijing's sixteen administrative districts, is situated in the northwestern and western parts of the city, covering an area of approximately 430.8 km². As of 2023, it had a permanent population of 3.125 million, accounting for about 2.6% of Beijing's total area. In the same year, Haidian's GDP reached 1,102.02 billion RMB, representing 25.2% of Beijing's total GDP (Haidian District Statistics Bureau 2024, 2024). The district is strategically positioned as a hub for high-quality education (hosting more than eighty universities), information technology, and scientific innovation. According to the Haidian District Planning (2017–2035), the area is slated to become a pivotal zone for China's political culture, technological advancement, and economic expansion. It has already emerged as a major center for internet companies and high-tech research and development (R&D). However, the rapid growth of education and internet technology has also accelerated urbanization, leading to associated urban challenges in Haidian District.

Administratively, Haidian consists of 22 subdistricts and 9 towns, each exhibiting distinct characteristics that contribute to a diversified development pattern. Data from the Haidian District Fourth National Economic Census Major Data Bulletin (2020) and the Haidian District Seventh National Population Census Bulletin (2021) highlight some key areas: Zhongguancun Subdistrict is renowned for its innovation-driven ecosystem, housing 19,250 legal entities in secondary and tertiary industries as of 2018. Haidian Subdistrict, home to prestigious institutions such as the Chinese Academy of Sciences, Peking University, and Tsinghua University, serves as an academic and research nucleus. Xueyuan Road Subdistrict had a permanent population of 226,315 in 2020 and is distinguished by its concentration of higher education institutions. Beixiaguan Subdistrict and Zizhuyuan Subdistrict, with populations of 146,366 and 129,367 respectively, attract residents and businesses due to their cultural and educational amenities. Ganjiakou Subdistrict stands out as the most economically dynamic, with total assets amounting to 2,297.73 billion RMB in 2018. Qinghe Subdistrict and Xisanqi Subdistrict have experienced rapid development, recording populations of 147,395 and 157,643, respectively. Meanwhile, Sijiqing Town and Xibeiwang Town exhibit high population densities, with 162,700 and 164,795 residents as of 2020. Figure 8 illustrates the study area, featuring a 90m digital elevation model (DEM) of Beijing and highlighting the spatial distribution of subdistricts and towns within Haidian District.

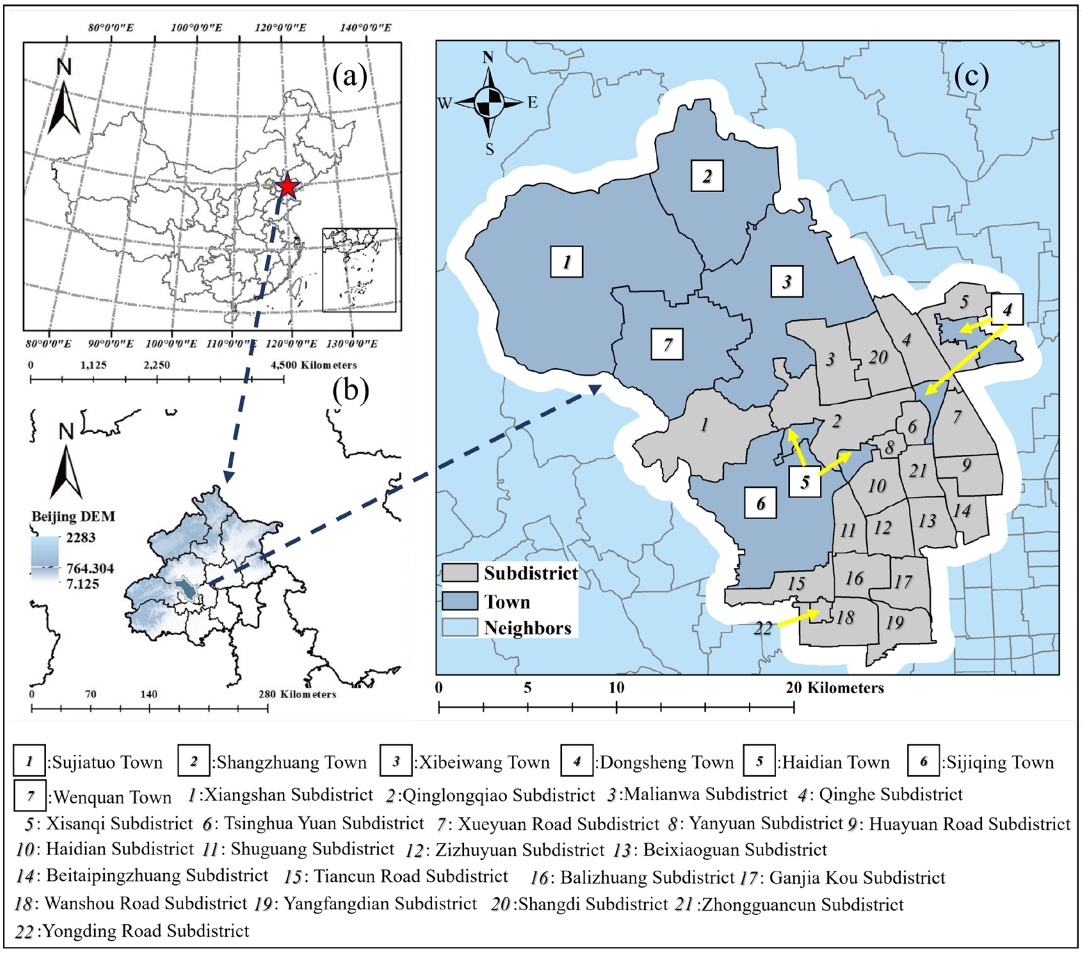

Figure 8: Study Area of Haidian District

Note: Figure 8: General information of the study area. (a) The location of Beijing in China; (b) The location of Haidian District in Beijing with 90m DEM; (c) The composition of Haidian District with detailed subdistrict information. All geographic data were collected from the Geospatial Data Cloud (http://www.gscloud.cn) and the National Platform for Common Geospatial Information Services (https://www.tianditu.gov.cn/?4).

4.2 Approach Utilization and Data sources

4.2.1 Data Sources

In this study, two types of datasets are used: the LBS datasets and the statistical datasets. Specifically, the description and sources of dynamic population data, permanent resident population, employee population, other population and aging population are all shown in Table 3.

Table 3 Data Sources in this Case Study

Data Type	Description	Data Source
Dynamic population data	Data is organized into 200m x 200m grids, from April 10th to April 14th, 2023, in Haidian District.	The dynamic population data was provided by a licensed telecommunications service provider in China under anonymized processing.
Permanent resident population	The 2023 permanent resident data for Haidian District (preprocessed with other and aging population) provided the reference for comparing of identified living population results.	Beijing Statistical Yearbook (2024): [https://nj.tjj.beijing.gov.cn/nj/main/2024-tjnj/zk/indexeh.htm]
Employed population	Haidian District's 2023 year-end urban non-private sector employment data served as the reference for comparing with identified working population results.	Beijing Regional Statistical Yearbook (2024): [https://nj.tjj.beijing.gov.cn/nj/qxnj/2024/zk/indexch.htm] (in Chinese)
Other population	Haidian District's 2023 year-end 0-14 age population data served as a proxy for non-working residents and was used as exclusion criteria when preprocessing the permanent resident dataset for precision evaluation.	Beijing Regional Statistical Yearbook (2024): [https://nj.tjj.beijing.gov.cn/nj/qxnj/2024/zk/indexch.htm] (in Chinese)
Aging population	Haidian District's 2023 elderly population (65+) statistics served as a proxy for non-working residents and were used as exclusion criteria when preprocessing the permanent resident dataset for precision evaluation.	Beijing Regional Statistical Yearbook (2024): [https://nj.tjj.beijing.gov.cn/nj/qxnj/2024/zk/indexch.htm] (in Chinese)

4.2.2 The Application of the Optimized JHSI Framework in Haidian District

Firstly, the thresholds and statistical time periods for defining working and living base stations are given. Specifically, according to the *Third National Time Use Survey Bulletin (No. 3)* released by *China's National Bureau of Statistics* in 2024 (National Bureau of Statistics of China, 2024), the survey analyzed residents' weekly time allocation across major life domains including paid work, unpaid labor, and transportation, with breakdowns by age group, gender, and household registration status (urban/rural). The data reveals that urban and rural residents averaged 6 hours 23 minutes and 6 hours 22 minutes of daily paid work respectively, with the working-age population (18-59 years) reaching 6 hours 32 minutes per day. For essential physiological activities, all demographic groups maintained a consistent daily average of approximately 12 hours 30 minutes, never dipping below 12 hours. Building upon these statistical recordings and incorporating the commute time thresholds established by Chinese researchers in JHSI-related studies (Table 4), we have accordingly set temporal parameters for job-housing patterns in Haidian District. In our study, the working time period was set from 7:00 am to 5:00 pm on workdays, and the living time period was set from 9:00 pm on the previous workday to 5:00 am on the next workday.

Table 4 Literature Parameters for Working and Living Time Thresholds

Time Segment Threshold	Threshold Value	Statistical Time Frame	Reference Source

Working time period	9:00-17:00, 8:00-18:00	Recurring for ≥10 days	(Wang et al., 2020), (Zheng et al., 2023)
Living time period	21:00-7:00(next day), 21:00-8:00(next day), 0:00-6:00, 20:00-5:00(next day)	Recurring for≥10 days (or≥15 days)	(Wang et al., 2020), (Zheng et al., 2023), (ZHANG et al., 2023)

Note: While the referenced studies in Table 4 analyzed monthly data, our study employs weekly observations to enable finer-grained temporal analysis.

Secondly, for identifying working and living base stations: When a base station conducts monitoring with full performance throughout the entire working time period on each research workday, it will be designated as a working base station. However, due to the complex regular trends of population mobility during living time, if a base station conducts monitoring that matches the living patterns of Haidian District—i.e., at six or more time points during each living time period on a full research workday—it will be designated as a living base station.

Finally, for noise reduction and statistical processing, we plotted a trend of the "Time" field against "Total POPu" for each "Day" in all working base stations (Fig. 9). After observing the regular trends in these relationships, an obvious fluctuation with a different trend was identified in the records of April 10, which was classified as noise (Fig. 9). Then, all the records of the noise were removed from each working base station. Additionally, in the final data processing for working base stations, based on Section 3.2.4 and table 4, the selected working time points to calculate the average number of the "POPu" field, as an estimate of the working population, are 8:00 am, 11:00 am, and 4:00 pm (Zheng et al., 2023).

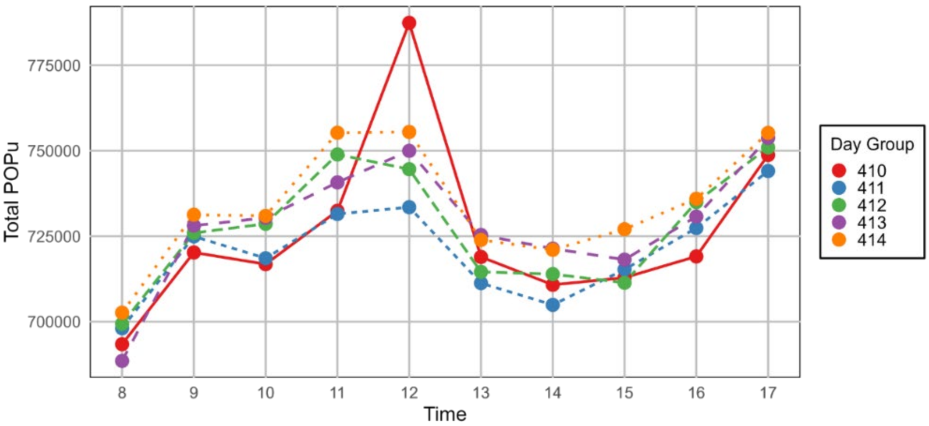

Figure 9: Trend Analysis for All Job Base Stations on "Total POPu" and "Day Group"

4.3 Results

4.3.1 The Results of JHSI in Haidian District

Based on the defined steps and time thresholds, the job-housing spaces in Haidian District, Beijing, were identified, and their spatial distribution is illustrated using kernel density maps in Fig. 10 (a) and (b). The analysis classified 1,145 base stations as working base stations and 3,172 as living base stations. In terms of population distribution, the estimated working population in Haidian District was approximately 721,713, while the living population was around 1,177,927. As shown in Fig. 10, the kernel density maps clearly reveal the spatial patterns of job-housing distribution, with concentrations primarily observed in the northern, eastern, and southeastern parts of Haidian District. The smaller spatial extent of residential areas compared to workplaces, combined with the kernel density analysis, suggests a notable imbalance between job and housing distributions.

To evaluate the precision of our estimates, we followed the methodology outlined in Step 5

(Section 3.1.1) and referenced the statistical data in Table 3. According to official statistics, Haidian District had a permanent resident population of 3.125 million in 2023, consisting of 1.684 million employed individuals, 0.484 million elderly residents, and 0.371 million other populations. To ensure comparability with the dynamic population dataset, we applied an expansion coefficient of 65% to adjust the statistical figures. After calibration, the working and living populations were estimated at approximately 1,094,600 and 1,475,500, respectively. When compared to our dynamic population-derived results, discrepancies of 372,887 (working population) and 297,573 (living population) were observed, yielding precision indices of 66% for working population identification and 80% for residential population identification. Despite these differences, the overall trends remain consistent, and the margin of error is within an acceptable range for job-housing space analysis from comparable studies (e.g., Wang et al., 2020). Potential sources of discrepancy include the incomplete exclusion of non-working residents and the absence of real-time employment records for private-sector workers in the current statistical indicators used to represent occupational populations (Table 3).

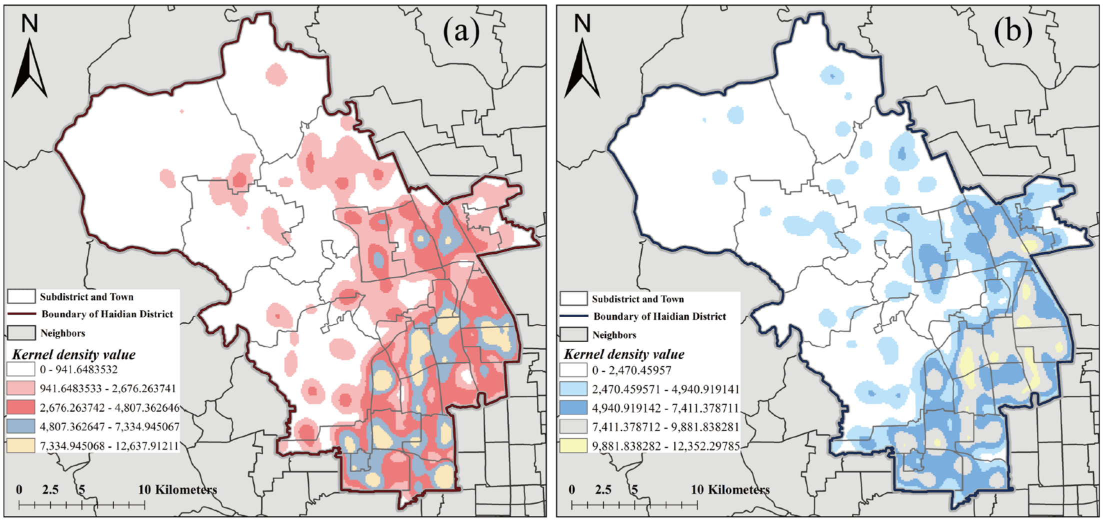

Figure 10: Kernel Density Map of Urban Working Spaces (a) and Living Spaces (b) in Haidian District

4.3.2 The Results of Job-Housing imbalance Calculation in Haidian District's 22 Subdistricts and 7 Towns

Based on the optimized JHSI framework results from Section 4.3.1 and Formulas (3)-(4), we computed the JHB and SD indexes for Haidian District, representing working-living functional intensity and job-housing spatial imbalance patterns respectively. Following the methodology in Section 3.2, we established five classification intervals to distinguish areas with JHB values below or above 1 (Figure 11a). The JHB values in Haidian District range from 0.237094 (Shangzhuang Town) to 1.987488 (Haidian Town), exhibiting relatively compact spatial patterns without fragmented distribution. Specifically, 8 subdistricts and 6 towns with JHB<1 occupy most of Haidian's periphery, particularly at district boundaries: northern/northeastern (adjacent to Changping District), western (bordering Shijingshan and Mentougou Districts), and eastern (neighboring Chaoyang District). This category includes subdistricts with values approaching 1, such as Xueyuan Road (0.985178), Malianwa (0.963796), and Tiancun Road (0.961074). Conversely, 14 subdistricts and 1 town with JHB>1 form a compact cluster in southern Haidian, predominantly showing values between 1-1.5. While Zizhuyuan Subdistrict (1.079088) approaches balance, areas like Zhongguancun (1.155604), Shangdi (1.194954), and Yongding Road (1.517918) demonstrate clear employment dominance. Overall, job-concentrated areas are orderly distributed in eastern/southeastern Haidian, proximate to Beijing's central districts (Dongcheng, Xicheng, and Chaoyang).

For SD analysis, we applied natural breaks classification to categorize the 29 values into five groups (Figure 11b). Haidian Town shows the maximum SD (0.987488), while Xueyuan Road Subdistrict has the minimum (0.014822). The SD distribution across 22 subdistricts and 7 towns reveals a highly fragmented spatial pattern, with most areas ranging 0.117052-0.245232. The most imbalanced areas include Haidian Town (0.987488), Tsinghua Yuan Subdistrict (0.970939), and Shangzhuang Town (0.762906) - the latter being residence-dominated while the former two are employment-centered. Yongding Road Subdistrict (0.517918) exhibits notably higher imbalance than Zhongguancun (0.155604) and Shangdi (0.194954). Furthermore, most boundary areas adjacent to Beijing's central districts (Chaoyang, Dongcheng, Xicheng, and Fengtai) display relatively high SD values, indicating pronounced job-housing imbalance.

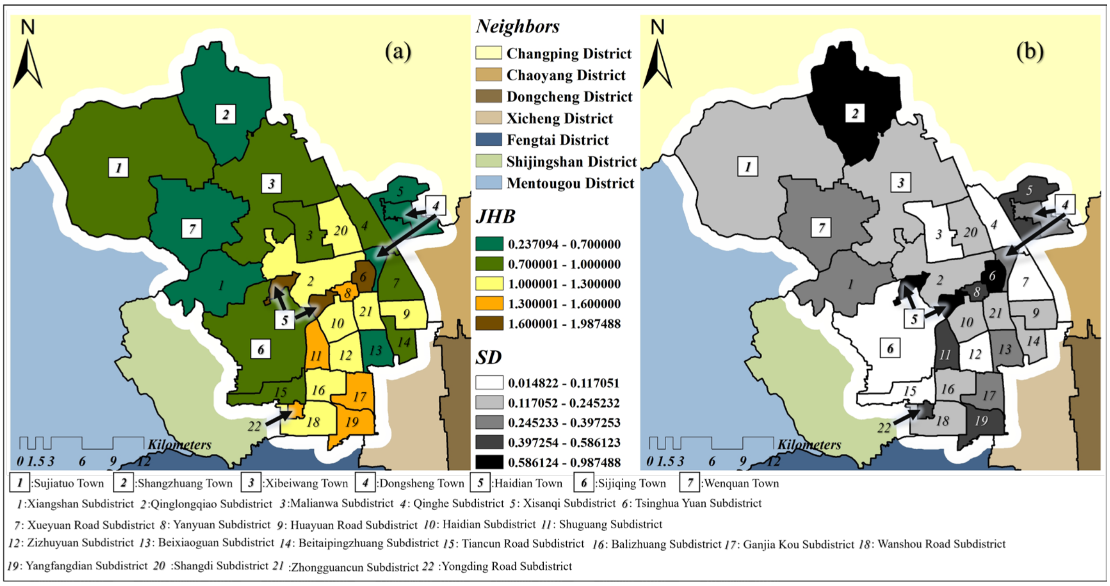

Figure 11: Distribution of the Job-Housing Balance (JHB) Index (a) and its Standard Deviation (SD) (b) at the Subdistrict and Town Scale in Haidian District

5 Discussion

5.1 Spatial Distribution of Job-Housing Relationships in Haidian District

The spatial distribution of job-housing relationships in Haidian District is illustrated through two maps: the distribution of job-housing spaces (Fig. 10) and imbalance patterns (Fig. 11). First, the analysis of job-housing spaces reveals distinct clustering patterns, with employment concentrations primarily located in two key regions: the northwest (e.g., Northwest Wang Town, Wenquan Town, and Sujiatuo Town), the southern and southeastern zones adjacent to Shijingshan, Chaoyang, Changping, and Beijing's core urban districts (Dongcheng and Xicheng) (Fig. 10). This spatial arrangement aligns with Haidian's industrial planning, as documented in the *Beijing Industrial Layout Map* (Beijing Municipal Development and Reform Commission, 2024). Specifically: (1) The Zhongguancun Future Science Park—specializing in next-generation IT, intelligent connected vehicles, tech services, and healthcare—is concentrated along Haidian's borders with Shijingshan, Chaoyang, Changping, and the urban core. (2) The Shangdi Cluster in northern Haidian (20.68 km²), spanning Northwest Wang, Wenquan, and Sujiatuo Towns, includes the Shangdi, Yongfeng, and Cuihu subdistricts and hosts R&D centers of major tech firms such as Baidu's headquarters and Huawei's Beijing Research Center. (3) The Zhongguancun Demonstration Zone follows a dual-core structure, with the northern core comprising the Shangdi-Xisanqi, Yongfeng, and Cuihu clusters, while the southern/southeastern core is anchored by the Zhongguancun Original Innovation Cluster. (4) High-tech parks (e.g., Zhongguancun No. 1, IC Design Park, Software Park, Dongsheng Tech Park, and JinYu Smart Factory) and specialized industrial zones (e.g., Shangdong Digital Valley and Digital TV Industrial Park) are situated near Haidian's boundaries with Changping and Chaoyang. These observations suggest that the primary drivers of job space formation in Haidian are its strong technology sector—concentrated along its periphery with central and neighboring districts—and its metro network (with key hubs such as Shangdi, Zhongguancun, Qinghe, and Xizhimen), which enhances connectivity to Changping, Dongcheng, and Xicheng, facilitating cross-regional industrial collaboration, labor mobility, and commuting flows. Consequently, this job concentration influences housing demand, living costs, and price disparities, resulting in a housing distribution that largely overlaps with employment spaces but covers a more limited area.

Second, to assess the job-housing imbalance exacerbated by Haidian's industrial layout, this study quantifies the dominant intensity and spatial patterns of employment-residence functions at the subdistrict and town level using joint measurements of JHB and SD (Fig. 11). Key findings include: (1) Job-dominant areas (JHB>1) exhibit strong spatial agglomeration, primarily in southern/southeastern core zones such as Shangdi and Zhongguancun subdistricts. This concentrated employment pressure corresponds with Haidian's high-tech industrial clusters but intensifies job-housing imbalances in core areas (higher SD values), reinforcing the link between job-housing dynamics and industrial spatial planning. Industrial agglomeration may also elevate housing prices, rents, and living costs, creating a filtering effect that displaces residential demand outward. (2) Housing-dominant areas (JHB<1) are clustered in northern and southern peripheral zones (e.g., Shangzhuang and Wenquan Towns). Notably, among the three areas with the highest SD values (Haidian Subdistrict, Qinghuayuan Subdistrict, and Shangzhuang Town), Shangzhuang exhibits a unique residence-dominated imbalance, contrasting with the employment-dominated imbalance in the other two. As a transitional zone between Haidian and Changping, Shangzhuang serves as a critical residential pressure absorber in the regional job-housing system. Additionally, China's school district policy—which ties school admission to residential location—likely exacerbates housing scarcity and contributes to intensified residential clustering in peripheral areas like Shangzhuang.

5.2 Spatiotemporal Evolution of Job-Housing Imbalance at the Subdistrict and Town Scale in Haidian District

Haidian District is a pivotal region in Beijing's development strategy. Using the optimized JHSI framework, we analyzed 2023 data, revealing that the spatial distribution of JHB exhibits a compact pattern, whereas SD appears fragmented. To assess the spatiotemporal evolution of job-housing imbalance at the subdistrict and town scale over the past five years, we compare our findings with a similar study by Wang et al. (2020) based on 2018 data. Although the overall JHB pattern remains largely unchanged since 2018, the fragmentation level of SD has increased in 2023. Notably, the value ranges for both JHB and SD in 2023 are significantly narrower than those in 2018, indicating substantial progress in job-housing balance across Haidian District. In 2018, a greater number of subdistricts and towns—particularly in southern Haidian, adjacent to Beijing's urban core—recorded JHB values exceeding 1.

Specifically, in 2018, employment functions were heavily concentrated in subdistricts such as Zhongguancun (JHB: 3.51–10.51), Shangdi (JHB: 2.51–3.5), and Yongding Road (JHB: 1.51–2.5), where JHB ranges were markedly wider than in surrounding areas. Similarly, SD values highlighted pronounced job-housing separation and elevated pressure in Zhongguancun, Shangdi, Qinghuayuan, Yanyuan, and Haidian subdistricts. By 2023, however, the overall SD range in Haidian District had contracted, reflecting a reduced concentration of employment functions and a more even distribution across the region. Furthermore, disparities in job and residential functions among subdistricts and towns have diminished. These improvements signify meaningful advances in mitigating urban land use pressures and enhancing living conditions, aligning with the population and land use targets set forth in the Haidian District Planning (2017–2035). The observed trends confirm that Haidian District's spatial development trajectory is consistent with its planning objectives.

5.3 From Methodology to Practice: Policy recommendations for Urban Planning

Our optimized JHSI methodology provides urban planners with an enhanced macro-perspective quantitative tool to address imbalanced job-housing relationships amid rapid urbanization. Using Haidian District as a case study, we examine both the technical scalability and practical applications of this approach to derive policy recommendations.

At the technical level, our geospatial big data-driven JHSI-optimized framework can be further integrated with AI to enable multi-scale regional data iteration and scenario upgrades, ultimately forming an intelligent job-housing interaction platform. For China's territorial planning, a public-private partnership model could effectively bridge internal assessments—such as identifying, evaluating, and optimizing key areas—with broader regional expansion strategies. A notable example is the Beijing Municipal Development and Reform Commission's interactive *Beijing Industrial Layout Map*, which integrates multi-objective industrial patterns with policy guidance for commercial and planning applications. These developments demonstrate that, with further technical refinement, our approach can support decision-making in residential location selection, corporate commercial planning, government transportation infrastructure, and built-environment development through spatially visualized solutions.

At the practical level, the method offers multidimensional value. First, it provides a scientific basis for regional industrial restructuring. Within the BTH coordinated development strategy, it enables Haidian District—a high-tech innovation hub—to amplify its impact by fostering a coordinated system where Beijing specializes in R&D while Tianjin and Hebei handle related production activities. Simultaneously, it facilitates the strategic relocation of Beijing's non-capital functions, such as manufacturing, service sectors, and government offices. Spatially visualized job-housing monitoring helps identify industrial clusters for relocation, evaluate spatial pressure on industries targeted for decentralization, and dynamically track balance adjustments. Second, the method can optimize public transit by targeting high-intensity or severely imbalanced job-housing zones through multi-period comparisons, thereby guiding station placement and route planning. For example, congestion in Haidian's Zhongguancun and Shangdi areas could be alleviated by adding subway stations or adjusting bus schedules. Additionally, the approach informs urban renewal projects, such as the redevelopment of older neighborhoods. Third, it can improve the balanced allocation of educational resources. China's school district system exacerbates spatial mismatches between jobs and housing, as concentrated elite schools inflate housing prices and induce long commutes. By embedding educational planning within industrial strategies—supported by our optimized JHSI tool—these systemic “job spaces-education spaces-housing spaces” conflicts can be mitigated, promoting sustainable urban integration.

5.4 Limitations

Although the approach optimized in this study is effective for identifying urban job-housing spaces, three key limitations remain when compared to existing methods based on mobile signaling data. First, a major constraint of our JHSI framework is its inability to capture daily commuting flow information. Given that commuting patterns are crucial for understanding urbanization—as they reflect the spatial distribution and mobility of urban populations—this omission represents a significant drawback. Consequently, our framework can only generate a broad spatial representation of job-housing distributions, which, though useful for macro-level analyses, lacks granularity. Second, the absence of up-to-date census data (due to delays in statistical monitoring) and representative indicators for working and residential populations (including occupational status) hinders further improvements in JHSI accuracy. Third, although age-based filtering of mobile device users has excluded most non-working residents, and our methodology ensures robustness through correlation coefficients and trend analysis, additional data refinement is still needed for two specific groups: (1) working-age non-working individuals (particularly university students), and (2) occupationally mobile populations with irregular residence patterns (e.g., frequent intercity commuters).

6 Conclusions

This study develops a base station-oriented JHSI framework, highlighting the practical utility of LBS data for urban planning applications. By replacing individual trajectories with aggregated population clusters and geospatially classified functional station data, we propose a lightweight, efficient, and privacy-preserving JHSI method. Implemented in Haidian District, the approach achieves precision indices of 66% for work identification and 80% for residence identification, demonstrating its effectiveness for urban planning and management—particularly in supporting the 2035 Plan objectives (e.g., housing allocation, population management, and spatial optimization)—while reinforcing Haidian’s pivotal role in the BTH coordinated development strategy. Furthermore, we examine its future application potential from both technical platform development and implementation perspectives, focusing on: (1) facilitating Beijing’s non-capital function decentralization under the BTH regional coordinated development strategy; (2) promoting high-tech industrial growth in Tianjin and Hebei through Haidian’s innovation sector; (3) optimizing transportation planning, particularly in station selection and micro-scale road network design; and (4) guiding educational resource allocation to alleviate job-housing imbalances.

Data availability

Dynamic population data used in this study will be made available upon request.

Declaration of competing interest

The authors declared that they have no conflicts of interest to this work.

Acknowledgments

This work was funded by the National Natural Science Foundation of China [42230718].

Reference

Ahas, R., Aasa, A., Silm, S., & Tiru, M. (2010). Daily rhythms of suburban commuters' movements in the Tallinn metropolitan area: Case study with mobile positioning data. *Transportation Research Part C: Emerging Technologies*, *18*(1), 45–54. https://doi.org/10.1016/j.trc.2009.04.011

AHAS R, SILM S, JÄRV O, et al. Using mobile positioning data to model locations meaningful to users of mobile phones[J]. Journal of Urban Technology, 2010, 17 (1): 3-27. https://doi.org/10.1080/10630731003597306

Alexander, L., Jiang, S., Murga, M., & González, M. C. (2015). Origin–destination trips by purpose and time of day inferred from mobile phone data. *Transportation Research Part C: Emerging Technologies*, *58*, 240–250. https://doi.org/10.1016/j.trc.2015.02.018.

Beijing Municipal Development and Reform Commission. (2024). Beijing Industrial Layout Map. Retrieved from https://project.fgw.beijing.gov.cn/zt/2024/cydt/

Beijing Municipal Commission of Planning and Natural Resources. Haidian District Planning (2017-2035). Beijing Municipal Commission of Planning and Natural Resources, https://ghzrzyw.beijing.gov.cn/zhengwuxinxi/ghcg/fqgh/202002/t20200213_1630027.html (in chinses)

Bi, S., Wang, L., Liu, S., Zhang, L., & Yuan, C. (2023). Identification of Urban Jobs–Housing Sites Based on Online Car-Hailing Data. *Sustainability*, *15*(2), 1712. https://doi.org/10.3390/su15021712

Calabrese, F., Diao, M., Di Lorenzo, G., Ferreira, J., & Ratti, C. (2013). Understanding individual mobility patterns from urban sensing data: A mobile phone trace example. *Transportation Research Part C: Emerging Technologies*, *26*, 301–313. https://doi.org/10.1016/j.trc.2012.09.009

Dong, R., & Yan, F. (2021). Revealing Characteristics of the Spatial Structure of Megacities at Multiple Scales with Jobs-Housing Big Data: A Case Study of Tianjin, China. *Land*, *10*(11), 1144. https://doi.org/10.3390/land10111144

Guan, X., Wei, H., Lu, S., Dai, Q., & Su, H. (2018). Assessment on the urbanization strategy in China: Achievements, challenges and reflections. *Habitat International*, *71*, 97–109. https://doi.org/10.1016/j.habitatint.2017.11.009

Gundlegård, D., Rydergren, C., Breyer, N., & Rajna, B. (2016). Travel demand estimation and network assignment based on cellular network data. *Computer Communications*, *95*, 29–42. https://doi.org/10.1016/j.comcom.2016.04.015

Guo, S., Pei, T., Xie, S., Song, C., Chen, J., Liu, Y., Shu, H., Wang, X., & Yin, L. (2021). Fractal dimension of job-housing flows: A comparison between Beijing and Shenzhen. *Cities*, *112*, 103120. https://doi.org/10.1016/j.cities.2021.103120

Howard E. Garden Cityes of Tomorrow[M] . London: Sonnenschein,1902:23 – 24.

Horner, M.W., 2002. Extensions to the concept of excess commuting. Environ. Plan. A 34, 543–566.

Horner M W, Murray A T. Excess commuting and the modifiable areal unit problem[J]. Urban Studies, 2002, 39(1): 131-139.

Haidian District Statistics Bureau. (2020). Haidian District Fourth National Economic Census Major Data Bulletin. Retrieved from https://zyk.bjhd.gov.cn/sjkf/tjgb/202004/t20200429_4401961.shtml (in chinses).

Haidian District Statistics Bureau. (2021). Haidian District Seventh National Population Census

Bulletin. Retrieved from https://zyk.bjhd.gov.cn/sjkf/tjgb/202106/P020210607318472808814.pdf (in chinses).

Hadachi, A., & Pourmoradnasseri, M. (2022). A Survey on the Advancement of Travel Time Estimation Using Mobile Phone Network Data. *IEEE Transactions on Intelligent Transportation Systems*, *23*(8), 11779–11788. https://doi.org/10.1109/TITS.2021.3107212

Huang, J., Levinson, D., Wang, J., & Jin, H. (2019). Job-worker spatial dynamics in Beijing: Insights from Smart Card Data. *Cities*, *86*, 83–93. https://doi.org/10.1016/j.cities.2018.11.021.

Haidian District Statistics Bureau. (2024). Beijing Haidian District Statistical Yearbook (2024). Retrieved from https://zyk.bjhd.gov.cn/sjkf/ndsj/202412/t20241230_4751623.shtml (in chinses).

Isaacman, S.; Becker, R.; Cáceres, R.; Kobourov, S.; Martonosi, M.; Rowland, J.; Varshavsky, A. Identifying Important Places in People's Lives from Cellular Network Data. In Proceedings of the Pervasive Computing, Heidelberg, Germany, 12–16 September 2011; pp. 133–151.

Jiang, S., Alves, A., Rodrigues, F., Ferreira, J., & Pereira, F. C. (2015). Mining point-of-interest data from social networks for urban land use classification and disaggregation. *Computers, Environment and Urban Systems*, *53*, 36–46. https://doi.org/10.1016/j.compenvurbsys.2014.12.001

Kain, J. (1968). Housing segregation, Negro unemployment and metropolitan segregation. Quarterly Journal of Economics, 82(2), 175e197.

Li, S., & Liu, Y. (2016). The jobs-housing relationship and commuting in Guangzhou, China: Hukou and dual structure. *Journal of Transport Geography*, *54*, 286–294. https://doi.org/10.1016/j.jtrangeo.2016.06.014

Liao, H., & Dong, W. (2017). An Exploratory Study Investigating Gender Effects on Using 3D Maps for Spatial Orientation in Wayfinding. *ISPRS International Journal of Geo-Information*, *6*(3), 60. https://doi.org/10.3390/ijgi6030060

Lin, D., Allan, A., & Cui, J. (2015). The impact of polycentric urban development on commuting behaviour in urban China: Evidence from four sub-centres of Beijing. *Habitat International*, *50*, 195–205. https://doi.org/10.1016/j.habitatint.2015.08.018

Liu, X., Tian, Y., Zhang, X., & Wan, Z. (2020). Identification of Urban Functional Regions in Chengdu Based on Taxi Trajectory Time Series Data. *ISPRS International Journal of Geo-Information*, *9*(3), 158. https://doi.org/10.3390/ijgi9030158

Long, Y., & Thill, J.-C. (2015). Combining smart card data and household travel survey to analyze jobs–housing relationships in Beijing. *Computers, Environment and Urban Systems*, *53*, 19–35. https://doi.org/10.1016/j.compenvurbsys.2015.02.005

Louail, T., Lenormand, M., Picornell, M., García Cantú, O., Herranz, R., Frias-Martinez, E., Ramasco, J. J., & Barthelemy, M. (2015). Uncovering the spatial structure of mobility networks. *Nature Communications*, *6*(1), 6007. https://doi.org/10.1038/ncomms7007

Mihaylova, L., Angelova, D., Honary, S., Bull, D., Canagarajah, C., & Ristic, B. (2007). Mobility Tracking in Cellular Networks Using Particle Filtering. *IEEE Transactions on Wireless Communications*, *6*(10), 3589–3599. https://doi.org/10.1109/TWC.2007.05912

Mao, F.; JI, M.; Liu, T. Mining spatiotemporal patterns of urban dwellers from taxi trajectory data. Front. Earth Sci. 2016, 10, 205–221.

National Bureau of Statistics of China. (2024). *The third national time use survey bulletin (No. 3)*. https://www.stats.gov.cn/sj/zxfb/202410/t20241031_1957215.html (in Chinese).

Okmi, M., Por, L. Y., Ang, T. F., & Ku, C. S. (2023). Mobile Phone Data: A Survey of Techniques, Features, and Applications. Sensors, 23(2), 908. https://doi.org/10.3390/s23020908.

Pei T, Sobolevsky S, Ratti C, Shaw S-L, Li T and Zhou CH, 2014. A new insight into land use classification based on aggregated mobile phone data, International Journal of Geographical Information Science, 28:9, 1988-2007, DOI: 10.1080/13658816.2014.913794

Perera K, Bhattacharya T, Kulik L, et al. Trajectory inference for mobile devices using connected cell towers [EB/OL] (2015-11-01) [2022-5-27] https://doi.org/10.1145/2820783.2820804.

Rousell, A., & Zipf, A. (2017). Towards a Landmark-Based Pedestrian Navigation Service Using OSM Data. *ISPRS International Journal of Geo-Information*, *6*(3), 64. https://doi.org/10.3390/ijgi6030064

Small K A, Song S. “Wasteful” commuting: a resolution[J]. Journal of political economy, 1992, 100(4): 888-898.

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

Sun, B., He, Z., Zhang, T., & Wang, R. (2015). Urban spatial structure and commute duration: An empirical study of China. International Journal of Sustainable Transportation, 10(7), 638–644. https://doi.org/10.1080/15568318.2015.1042175.

Sari Aslam, N., Cheng, T., & Cheshire, J. (2019). A high-precision heuristic model to detect home and work locations from smart card data. *Geo-Spatial Information Science*, *22*(1), 1–11. https://doi.org/10.1080/10095020.2018.1545884

Schleith, D., Widener, M., & Kim, C. (2016). An examination of the jobs-housing balance of different categories of workers across 26 metropolitan regions. *Journal of Transport Geography*, *57*, 145–160. https://doi.org/10.1016/j.jtrangeo.2016.10.008

Schmidtke, H. R. (2020). Location-aware systems or location-based services: A survey with applications to CoViD-19 contact tracking. *Journal of Reliable Intelligent Environments*, *6*(4), 191–214. https://doi.org/10.1007/s40860-020-00111-4

Sharifzadeh, M., & Shahabi, C. (2006). *Utilizing Voronoi Cells of Location Data Streams for Accurate Computation of Aggregate Functions in Sensor Networks*.

Sharifzadeh, M., & Shahabi, C. (2009). Approximate voronoi cell computation on spatial data streams. *The VLDB Journal*, *18*(1), 57–75. https://doi.org/10.1007/s00778-007-0081-y

Ta, N., Chai, Y., Zhang, Y., & Sun, D. (2017). Understanding job-housing relationship and commuting pattern in Chinese cities: Past, present and future. *Transportation Research Part D: Transport and Environment*, *52*, 562–573. https://doi.org/10.1016/j.trd.2016.11.011

Trevisani, E., & Vitaletti, A. (2004). Cell-ID Location Technique, Limits and Benefits: An Experimental Study. *Sixth IEEE Workshop on Mobile Computing Systems and Applications*, 51–60. https://doi.org/10.1109/MCSA.2004.9.

Weitz J, Schindler T. 1997. Are Oregon's communities balanced? A test of the jobs-housing balance policy and the impact of balance on mean commute times[R]. Portland, OR: Portland State University.

Wang, D. (2017). *The built environment and travel behavior in urban China: A literature review*.

Wang, E., Song, J., & Xu, T. (2011). From “spatial bond” to “spatial mismatch”: An assessment of changing jobs–housing relationship in Beijing. *Habitat International*, *35*(2), 398–409. https://doi.org/10.1016/j.habitatint.2010.11.008

Weng, M., Xiong, Q., & Kang, M. (2017). Salience Indicators for Landmark Extraction at Large

Spatial Scales Based on Spatial Analysis Methods. *ISPRS International Journal of Geo-Information*, *6*(3), 72. https://doi.org/10.3390/ijgi6030072

Wang H, Zeng W, Cao R. Simulation of the urban jobs–housing location selection and spatial relationship using a multi-agent approach[J]. ISPRS International Journal of Geo-Information, 2021, 10(1): 16.

Wang Bei, Wang Liang, Liu Yanhua, et al. Characteristics of jobs-housing spatial distribution in Beijing based on mobile phone signaling data. Progress in Geography, 2020, 39(12): 2028-2042. DOI: 10.18306/dlkxjz.2020.12.006.

Yan, L., Wang, D., Zhang, S., & Xie, D. (2019). Evaluating the multi-scale patterns of jobs-residence balance and commuting time–cost using cellular signaling data: A case study in Shanghai. *Transportation*, *46*(3), 777–792. https://doi.org/10.1007/s11116-018-9894-3

Yang, X., Fang, Z., Xu, Y., Yin, L., Li, J., & Zhao, Z. (2021). Revealing temporal stay patterns in human mobility using large-scale mobile phone location data. *Transactions in GIS*, *25*(4), 1927–1948. https://doi.org/10.1111/tgis.12750

Yao, Z., & Kim, C. (2022). Analyzing the multiscale patterns of jobs-housing balance and employment self-containment by different income groups using LEHD data: A case study in Cincinnati metropolitan area. *Computers, Environment and Urban Systems*, *96*, 101851. https://doi.org/10.1016/j.compenvurbsys.2022.101851

Yuan, Y., Raubal, M., & Liu, Y. (2012). Correlating mobile phone usage and travel behavior – A case study of Harbin, China. *Computers, Environment and Urban Systems*, *36*(2), 118–130. https://doi.org/10.1016/j.compenvurbsys.2011.07.003

Zhang, P., Zhou, J., & Zhang, T. (2017). Quantifying and visualizing jobs-housing balance with big data: A case study of Shanghai. *Cities*, *66*, 10–22. https://doi.org/10.1016/j.cities.2017.03.004

Zhang, Y., Liu, J., Wang, Y., Cao, Y., & Bai, Y. (2021). Research on the Method of Urban Jobs-Housing Space Recognition Combining Trajectory and POI Data. *ISPRS International Journal of Geo-Information*, *10*(2), 71. https://doi.org/10.3390/ijgi10020071

Zhao, P., & Gao, Y. (2023). Discovering the long-term effects of COVID-19 on jobs–housing relocation. *Humanities and Social Sciences Communications*, *10*(1), 633. https://doi.org/10.1057/s41599-023-02155-2

Zheng, J., Liu, M., Peng, K., & Zhang, Y. (2023). Job-housing relationships and commuting patterns in metropolitan fringe areas using mobile signaling data: A case study of Wuhan Huashan Eco-City. Housing and Real Estate Research, (9), 56-60.

ZHANG Yingna, WANG Yue, HU Haoyu, YUAN Chunlai. Analysis of Population Spatial-temporal Distribution and Mobility in Beijing-Tianjin-Hebei Urban Agglomeration Based on Mobile Phone Trajectory Big Data [J]. Area Research and Development, 2023, 42(03): 161-167+180. DOI: 10.3969/j.issn.1003-2363.2023.03.026

Zhao, P., Lü, B., & Roo, G. D. (2011). Impact of the jobs-housing balance on urban commuting in Beijing in the transformation era. *Journal of Transport Geography*, *19*(1), 59–69. https://doi.org/10.1016/j.jtrangeo.2009.09.008

Zhou, S., Liu, Y., & Kwan, M.-P. (2016). Spatial mismatch in post-reform urban China: A case study of a relocated state-owned enterprise in Guangzhou. *Habitat International*, *58*, 1–11. https://doi.org/10.1016/j.habitatint.2016.08.003