Manuscript

Urban Analytics and City Science

EPB: Urban Analytics and City Science
2025, Vol. 0(0) 1–16

DOI: 10.1177/23998083251346582
journals.sagepub.com/home/epb
Sage

Profile and theoretical advances in urban big data studies: A systematic review of 57 representative journals (2013–2023)

Tangqi Tu
Tsinghua University, China

Ningbo University, China

Enjia Zhang
Beijing University of Technology, China

Tsinghua University, China

Ying Long
Tsinghua University, China

Abstract

As big data becomes increasingly integral to urban studies, the evolving profile of its applications, including the trends in data, methodologies, and contributions to urban theories, remains unclear. This study systematically reviewed literature from 57 representative journals in the fields of Urban Studies, Geography, and Environmental Studies (JCR Q1) over the past decade (2013–2023), with a focus on quantitative research involving urban spatial big data. A total of 1,425 articles were identified through automated keyword filtering and manual screening, of which 82 were selected for further analysis due to their theoretical contributions. The results revealed key thematic areas, including urban transportation, spatial quality, vitality, and structure, with a predominant focus on developed countries and large cities. The applications of big data and novel methodologies in urban studies have increased from 2013 to 2023, with increasingly higher levels of spatiotemporal resolution. However, studies linking big data to urban theories remained limited, with most quantitative research centered on applying or proposing new methods or uncovering new phenomena. Theoretical contributions primarily explored the mechanisms underlying urban environment formation and their impacts on human activities and behaviors. This study provides a comprehensive understanding of big data applications and advances in theories in urban spatial studies and highlights key directions for future research.

Corresponding author:
Ying Long, School of Architecture and Hang Lung Center for Real Estate, Tsinghua University, Beijing 100084, China.
Email: ylong@tsinghua.edu.cn

Data Availability Statement included at the end of the article

Keywords
Big data, urban studies, systematic review, theoretical advances

Introduction

With the increasing applications of modern information and communication technologies (ICTs), the volume of related data has expanded substantially across diverse socioeconomic sectors. The term “big data” was first proposed by Cox and Ellsworth (1997) to describe the challenges associated with processing large-scale data in business applications. Since then, the academic community has been extensively discussing the definition and applications of big data. Unlike traditional “small” data—such as those derived from questionnaires and statistical censuses—big data includes new and open data sourced from ICTs embedded with electric sensors (Batty, 2013), which is typically characterized by its significantly higher volume, velocity, variety, veracity, and value.

The new data environment, encompassing both big data and open data, has brought new opportunities for urban research and planning practices. With the rapid rise in big data applications, the Fourth Paradigm—also known as data-intensive scientific research—is gradually becoming a new trend in the scientific paradigm of urban studies. Owing to the wider spatial coverage and higher spatiotemporal resolution, the application of big data shows advantages for urban studies, providing scientific foundations for policymakers to cope with emerging challenges in cities. Big data and related processing technologies have thus become important tools for promoting the Sustainable Development Goals (SDGs) in cities. However, merely possessing big data does not necessarily indicate advancement; the crucial value lies in deepening and broadening the understanding of urban phenomena.

In the field of urban big data studies, key questions remain: What is the current state of research progress? How do these findings contribute to the theoretical development of urban space issues? These questions need to be thoroughly explored to obtain meaningful insights. Thus, conducting a literature review on big data studies in urban spatial fields is crucial.

Several reviews of urban big data literature have already been conducted. Some reviews summarized the number of articles, countries, data sources, and research themes in big data-based urban sustainability research (Kong et al., 2020), highlighting progress in describing urban structure (Allam and Dhunny, 2019). Some research reviewed specific applications of urban big data, such as its combination with remote sensing or social sensing, machine learning applied to environmental images, urban digital twins, social media, urban structure identification, and transportation (Xing et al., 2024). Additionally, some scholars identified the limitations of big data in areas such as livability (Allam and Dhunny, 2019), addressing urban inequalities, and capturing temporal dynamics (Kandt and Batty, 2021). Furthermore, scholars have been increasingly recognizing the importance and challenges of volunteered or crowdsourced data in urban planning (Kong et al., 2020).

However, previous reviews have certain limitations that suggest opportunities for further enhancement. First, they were concentrated on static general characteristics, lacking a detailed analysis of temporal dynamics. Secondly, they have not adequately summarized and synthesized the spatiotemporal resolution and research scales employed in quantitative studies. Thirdly, while they described the scope of existing research, they have not comprehensively summarized the theoretical advances in the literature. As Batty (2013) highlighted the need for new theories to truly understand urban issues revealed by big data, the theoretical contributions in urban big data studies are essential for understanding contemporary cities and improving urban studies in the future.

To conduct a more systematic review, this study aims to address three key questions: What is the extent and proportion of quantitative studies in the field of urban big data? What are the profiles and limitations of these studies? What urban space theories have been advanced through these studies? The novelty of our literature review can be seen in three aspects. First, a detailed analysis of temporal dynamics was conducted to reveal the evolution and trends of urban big data studies. Secondly, spatiotemporal resolution and extent in urban big data studies were taken into consideration to present a more comprehensive picture of the existing research. Thirdly, related theoretical advances were analyzed to deepen the understanding of urban space based on big data analysis.

In line with these innovations, our literature review will provide valuable insights into the future development of urban big data studies and the advancement of socio-technical systems. First, we utilized a Large Language Model (LLM) to provide a new technical pathway for exploring urban big data studies. Secondly, we offered a comprehensive analysis of spatiotemporal scales in urban big data research, which can be a useful technical basis for selecting appropriate spatiotemporal analytical units for various urban spatial topics. Thirdly, our review focused on the comparison of methods and theories across research articles, demonstrating how big data methodologies contribute to enriching theoretical knowledge of urban space, which can inspire the effective development of socio-technical systems that optimize solutions for addressing urban planning issues.

The remainder of this paper is organized as follows. The materials and methodology section outlined the research framework and specific processing steps, including the criteria for filtering literature and methods for conducting categorized statistical analysis. In the result section, we first analyzed the statistics and general characteristics of the quantitative urban space literature and then explored the theoretical advances, and finally offered prospects for future research. Finally, the conclusions section provided a summary of the review.

Materials and methodology

General procedure

The general procedure for the literature review is illustrated in Figure 1. The process began with the selection of database sources to identify target journals. The names of all the journals at the first step with their categories and JCR quartiles are listed in Table S1 in the Supplementary Material (SM). Next, suitable keywords related to big data in urban space were compiled drawing from relevant review papers, as listed in Table 1. Screening criteria were then established, including time frame, language, access, study type, and content. Based on these screening criteria and listed keywords, quantitative studies about urban big data were identified and labeled. The attributes for labeling articles include general type, research field, study area, spatial analytic unit, acquired data, utilized methods, and study period. The labeling process was conducted aided by artificial intelligence (AI) automatic searching, and manual screening. Finally, quantitative studies related to urban space theories were identified and summarized.

Filtering steps of literature

In the first step, relevant literature reviews on urban spatial big data were examined (Allam and Dhunny, 2019; Batty, 2013; Kandt and Batty, 2021; Kitchin, 2013; Lim et al., 2018; Wang and Yin, 2023), which indicated that related studies are primarily published in journals categorized under Urban Studies, Geography, and Environmental Studies in the *Journal Citation Reports* (JCR). This focus is primarily because quantitative studies on urban big data center on analyzing the urban spatial environment and its complex dynamics. Specifically, studies addressing urban development,

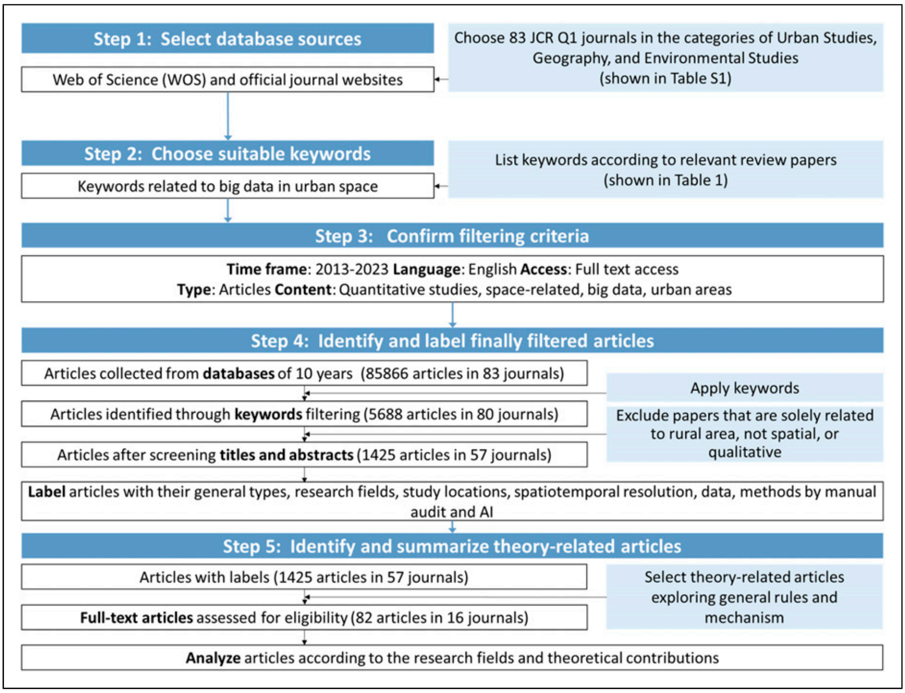

Figure 1. General procedure of the literature review.

Table 1. Keywords for automatic filtering literature.

Category	Big data related keywords
Basic concepts	Big data, multi-source data, massive data, emerging data
Street view	Street view image(s), street view picture(s), street view imagery
Open crowdsource map	POI (points of interest), AOI (areas of interest), OSM (OpenStreetMap), crowdsource data, VGI (volunteered geographic information)
Nightlight	Nighttime light, nightlight
Social media	Weibo (data), WeChat (data), Facebook (data), tweeter (data), social media data, TikTok (data), Instagram (data), Flickr (data), foursquare (data)
Sensor and locating	LBS (location-based service) data, GPS (global positioning system) data, IoT (internet of things), sensor data
Online service	Dianping (data), Yelp (data), online review (data), booking data, delivery data, website data, navigation data, Airbnb (data), Tencent map (data), Baidu map (data), Gaode map (data), Amap (data), Google Map (data)
Smart devices	Mobile phone data, smartphone data, phone signaling data, smart card data, camera (image/picture/data), wearable device (image/picture/data), WiFi (wifi) data, mobile application (data), cloud computing
Movement trajectory	Trajectory (data), mobility data, taxi data, ride-railing data, bicycle data, bike data, scooter data

governance, and spatial planning are closely aligned with journals in the Urban Studies category. Additionally, due to the methodological emphasis on spatial distribution patterns and geographic relationships, a significant portion of this research is published in Geography journals. Furthermore, since urban big data studies often explore the interaction between spatial characteristics, human activities, and socioeconomic impacts—such as healthcare accessibility and consumer behavior—many relevant articles are also found in Environmental Studies journals. Collectively, these three categories comprehensively cover the core themes and methodologies in urban spatial big data research.

To ensure efficiency and academic rigor, this review concentrated on high-impact journals, specifically those ranked in the first quartile (Q1) of JCR, a common approach in systematic literature reviews (Chang, 2022; Ennas and Guardo, 2015). Journals in the Q1 category are recognized for their significant influence within their respective academic fields and reflect the prevailing research trends. Therefore, all journals that were classified under JCR Q1 in any year from 2013 to 2023 from the aforementioned categories were included in the initial selection, excluding those unrelated to urban spatial quantitative research based on their journal scope. This process initially identified 85,866 articles across 83 journals.

Secondly, the literature from the 83 journals was filtered primarily through a combination of automatic keyword filtering and manual review. As shown in Table 1, the keywords were derived from terms related to urban spatial big data identified in relevant literature reviews. These literature reviews were searched on the *Web of Science* and *Google Scholar* using keywords such as "big data review," "open data review," and "new data review." Each article was manually reviewed to determine if it was a review of quantitative research on urban spatial big data. Based on these reviews (Allam and Dhunny, 2019; Batty, 2013; Kandt and Batty, 2021; Kitchin, 2013; Kong et al., 2020; Lim et al., 2018), keywords related to urban spatial big data were categorized into nine groups.

Automatic filtering was primarily performed on the literature from the *Web of Science* database using the predetermined keywords. However, since the database may not include all articles from the selected journals in 2023, literature from the official websites of these journals was also manually checked and filtered. After removing duplicate and irrelevant records, a total of 5688 articles were identified, and distributed across 80 journals.

Following the initial filtering, a finer manual review was conducted to ensure that only articles focusing on quantitative research on urban space were retained, while those covering unrelated topics were excluded. This manual filtering involved reading the titles and abstracts of the articles. If an article could not be definitively categorized based on its title and abstract alone, further reading was conducted in the order of methodology, data, conclusion, discussion, and, if necessary, the full text.

The exclusion criteria for article selection were as follows. First, articles were required to focus directly on urban issues; those concentrated solely on rural or national topics were excluded. Secondly, articles should be those analyzing geographic space, excluding those merely studying non-spatial city aspects like energy, society, economy, culture, or politics. Thirdly, articles needed to involve data analysis and quantitative research, excluding those only qualitatively discussing urban big data. These criteria were applied manually, resulting in 1,425 articles in 57 journals that were finally confirmed as being directly relevant to quantitative research on urban spatial big data. Articles from other journals that focused on unrelated topics, such as qualitative or non-spatial studies, were excluded.

Categorized labeling and statistical methods of literature

After a finer filtering process, key information from 1,425 articles was manually labeled. This labeling was comprised of two main dimensions: general characteristics and theoretical advances.

General characteristics included article types, research fields, study locations, spatiotemporal resolution (spatial analytic units, temporal units, and the number of data years), data sources, and methods utilized. The article types were divided into three aspects: proposing an approach, uncovering a phenomenon, and exploring a mechanism. The research fields were classified into nine categories: urban form and structure, land use location, urban spatial quality and health, urban perception, urban transport and mobility, urban vitality, spatial social equity, urban social safety, and urban history and culture. The categorization was based on related research themes and the SDGs for urban space (Kong et al., 2020; Nedungadi et al., 2024; UN, 2023), explained as follows. First, the urban form at the macro level and land use location at the micro level are typical research fields for urban spatial analysis, providing a foundation for understanding the urban physical environment (Barki et al., 2022; Hao et al., 2015; Kong et al., 2020). Secondly, the objective spatial quality conditions and their health effects are very crucial for achieving environmental sustainability and promoting healthy cities (Barki et al., 2022; Kousis and Tjortjis, 2023), and subjective perception of urban space also plays an important role in individual happiness and emotional well-being (Ito et al., 2024; Xing et al., 2024). Thirdly, research in urban transport mainly focuses on travel behaviors, which serve as a representative of flowing human activities in cities (Hao et al., 2015; Kousis and Tjortjis, 2023; Park and Lee, 2019), while urban vitality often reflects the dynamic flow of human activities and the economic sustainability of cities (Kim, 2025; Yang et al., 2020). Fourthly, addressing social issues of inequalities, such as gentrification and crimes, is vital for fostering equitable, inclusive, and safe cities as well as preserving history and culture, which is crucial for achieving social sustainability in SDGs (Kousis and Tjortjis, 2023; Pan et al., 2016; Yang et al., 2020). These categories collectively reflect the major research fields related to urban spatial big data, encompassing both environmental and social dimensions of urban life.

The labeling of general characteristics was conducted through manual reading with the assistance of AI tools. The AI tool, installed as the Copilot toolkit in the Microsoft Edge Browser, was used to open the PDF file of each article and provide answers to specific questions about its content. The questions included: (1) What is the geographic area of this study? (2) What is the spatial analytic unit in this study? (3) What data is used in this study? (4) What method is used in this study? And (5) What is the time period of this study? After obtaining key information automatically provided by the AI tool, we manually scanned each article to ensure the accuracy of the AI-generated responses. The labeled indicators were then summarized numerically to reveal the general characteristics of the quantitative articles across various aspects.

The dimensions of theoretical advances include whether an article was theory-related, and if so, the theoretical contribution type, theory category, and main findings were summarized and labeled for the article. In this study, theories were defined as well-substantiated explanations or frameworks grounded in empirical evidence and reasoning. Theoretical contributions were categorized into four aspects: proposing a new theory, testing existing theories (validation), testing existing theories (modification), and complementing existing theories. Articles related to theoretical advances were selected from those aimed at exploring a mechanism. However, the mere exploration of a mechanism did not necessarily indicate a theoretical contribution. Theoretical advances were categorized based on their relevance to either the physical built environment of urban space (e.g., urban image, spatial quality, urban form) or to human activities within cities (e.g., urban vitality, social safety, mobility behaviors). For each theory-related article, the main findings and theoretical achievements regarding urban space, as revealed by big data, were summarized.

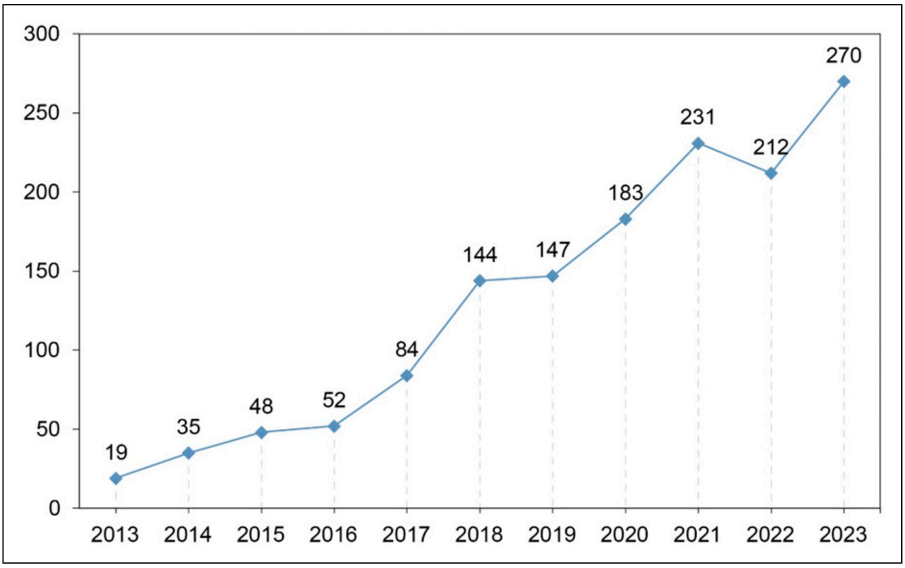

Figure 2. Total number of articles related to urban spatial big data quantitative research (2013–2023).

Results

Descriptive analysis

The numbers of articles counted by the filtering steps are presented in Table S2, Table S3, and Table S4 as in Figure 2. Table S2 provides basic statistics, including the total number of published articles and the number of articles left after the primary filtering by keywords. Table S3 shows the distribution of articles focusing on quantitative research related to urban space, totaling 1,425 articles across 57 journals. Regarding theoretical contributions, Table S3 further shows that 16 journals published 82 theory-related articles, which represent 5.75% of all the articles focused on quantitative research about urban spatial big data.

The journal with the highest number of articles on quantitative research of urban spatial big data is *Journal of Transport Geography*, accounting for 16.28% of the total number of articles, followed by *Computers Environment and Urban Systems* (11.79%), *Cities* (11.09%), and *International Journal of Geographical Information Science* (7.09%). As shown in Table S4, for most journals, the proportion of theory-related articles was below 10%. While some journals published a higher volume of articles on urban spatial big data, many of these focused on uncovering phenomena using big data, rather than exploring theoretical mechanisms, which results in fewer theory-related contributions. In terms of temporal trends, the number of articles on quantitative research of urban spatial big data has steadily increased from 2013 to 2023 (Figure 2). Table S3 indicates that journals such as *Cities* and *Applied Geography* have experienced particularly rapid growth in this area.

General characteristics

The general characteristics of research on urban spatial big data can be analyzed based on three key aspects. First, the distribution of articles was uneven across general types, research fields, and geographic areas, with the majority focusing on proposing new approaches or uncovering phenomena. Secondly, the data resolution in both temporal and spatial dimensions has steadily

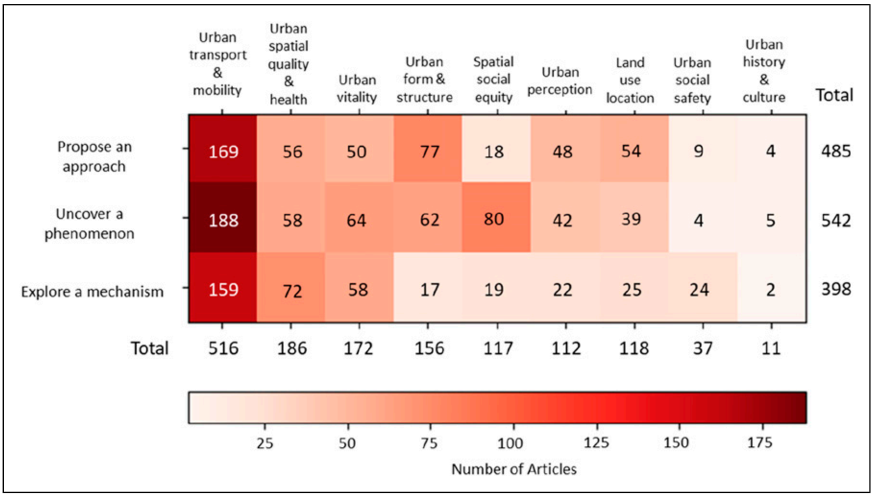

Figure 3. Numbers of articles for general types and research fields.

increased. Thirdly, the methods employed are becoming more diverse, with a growing number of studies using machine learning techniques to replace traditional statistical approaches. Specific findings are described as follows.

The results regarding general types and research fields are presented in Figure 3. Notably, the accumulation of over 150 articles was discernible in research domains such as urban transportation, urban spatial quality, urban vitality, and urban form. This trend underscores a significant emphasis on these topics within the academic discourse. Furthermore, an analysis of articles related to quantitative research of urban spatial big data revealed a predominant focus on proposing novel approaches and describing observable phenomena, while fewer explored the underlying mechanisms. The distribution of articles focused on mechanism exploration within each journal generally mirrors the overall article count. However, urban vitality and urban social safety exhibited a higher proportion of mechanism exploration, constituting 33.72% and 64.86% of their total numbers, respectively.

In terms of study locations, the labeled literature covered a total of 70 countries and 347 cities across six continents, with 10 countries being studied in 20 or more articles (Figure S1). At the country level, China emerged as the most frequently studied country, accounting for 45.82% of the articles. Regarding specific cities, 31 cities were studied in 10 or more articles, predominantly located in East Asia, as well as developed countries in Europe and North America. These cities are largely national capitals and major economic hubs.

The spatiotemporal resolution in the labeled literature has been highly diverse. Regarding spatial analytic units, the grid was the most widely used, accounting for 17.54% of all articles, followed by street or road (10.53%) and movement route (10.23%), as shown in Figure S2. The spatial units of the grid, movement route, and public transit stop often feature more temporally refined data (Figure S3). For example, in 64 articles, the spatial unit was the grid and the temporal unit was hourly, representing 4.49% of all articles. Over the years, there has been an increasing trend toward using more temporally refined data (Figure S4(a)). Regarding the number of data years, 87.09% of the articles were based on fewer than 4 years of data (Figure S4(b)).

In terms of data types used in all the articles (Figure S5(a)), the largest proportions were online map data (43.51%) and transport trajectory data (18.95%). From a temporal variation perspective, online map data, which reflected the urban built environment, showed obvious growth, with an average growth rate of 32.31%. For the big data reflecting human behaviors, cellphone signal data and transport trajectory data exhibited significant growth, with average growth rates of 27.10% and 47.58%, respectively.

Regarding the methods utilized in the articles (Figure S5(b)), non-spatial statistical analysis accounted for the highest proportion at 46.32%, followed by machine learning models (non-image) at 24.28%, and spatial correlation and regression models at 23.37%. From a temporal perspective, both image deep learning and other machine learning models showed rapid growth, with average growth rates of 46.00% and 30.20%, respectively. However, as of 2023, the proportion of image deep learning remained relatively low at 8.91%.

Theoretical advances

After reviewing and analyzing 1,425 articles on quantitative research in urban spatial big data, it was observed that only 82 articles contributed to theoretical development. These theory-related articles mainly focused on urban transport and mobility (34.15%), urban spatial environmental quality and health (21.95%), urban social safety (10.96%), and urban vitality (12.20%), as depicted in Figure S6, Table S5 and Table S6. Theory-related articles were more frequently found in research on human activities than on the built environment. From a temporal perspective, the number of theory-related articles has been increasing steadily, from just 1 article in 2015 to 21 articles in 2023, indicating continuous development in theoretical knowledge in the urban spatial field. Regarding the categories of theoretical contributions, these articles mainly aimed to complement existing theories (53.66%) or validate them (36.59%). In contrast, articles proposing theories or modifying existing ones were relatively scarce, accounting for 1.22% and 8.54%, respectively.

As shown in Figure S7, theoretical advances have progressed over the past decade, although their distribution across research fields remains uneven. Some fields have seen more progress than others, with an evolution from simplified, linear knowledge to more comprehensive, systematic, and non-linear insights into the research field, deepening and broadening the understanding of the spatial mechanism in cities. The most representative and productive research fields included urban vitality, urban mobility, and spatial quality.

Specifically, in the field of urban vitality, early studies found that mixed land use and higher development density were positively correlated with increased urban vitality (Long and Huang, 2019; Wu et al., 2018a). However, more recent research has suggested that development density has a non-linear impact on vitality, with certain threshold points, and that excessive land use diversity could diminish vitality due to shifting functional demands over time (Gan et al., 2021; Kang, 2020; Wang et al., 2022). Furthermore, these relationships have been found to be influenced by local socioeconomic status, and more detailed density indicators, such as plot ratio, have been tested to reveal their varying effects on vitality across different global contexts (Huang et al., 2023; Wang et al., 2023).

As for the research field of urban transport and mobility, early studies using big data verified the significant impacts of urban spatial attributes on travel behavior (Choi and Zhang, 2017; Wu et al., 2018b). Subsequent investigations have further explored these impacts, revealing their variation according to time, location, weather conditions, and national socioeconomic factors (Boakye et al., 2023; Chen et al., 2020; Wei, 2022; Zhang et al., 2019). Additionally, the emergence of new mobility modes, such as ride-hailing and bike-sharing, has been considered (Marquet, 2020; Yu and Peng, 2019), and the interplay between different travel modes, including their coopetition dynamics, has also been examined (Kim and Cho, 2021; Tang et al., 2021).

In the domain of urban spatial quality and health, researchers have traditionally viewed natural factors as indicators of good spatial quality, believed to improve public health, while urban development has generally been regarded as a threat to the ecological environment (Benita and Tuncer, 2019; Godwin et al., 2015; He et al., 2020; Li et al., 2018). However, as research has advanced, it has become apparent that natural factors, such as green space, may not always be beneficial if its expansion is unchecked, and the impact on health varies depending on geolocation (Kim et al., 2023; Yang et al., 2021; Zhang et al., 2023b). Moreover, the interaction between the ecological environment and urban development has been found to be complicated, with dynamic feedback mechanisms between the two (Wu et al., 2019). Meanwhile, independent dimensions for assessing spatial quality have been proposed (Li et al., 2022).

Other research fields exhibited similar patterns of how theoretical advances progress. That is, initially, simplified mechanisms were tested using big data sources, and subsequently, additional factors related to spatiotemporal heterogeneity were incorporated into the analysis. Newly emerging phenomena have been included in the analysis to assess their impact on classical topics. These studies have made significant contributions to the complementing, validating, and modifying of existing theories; however, research proposing entirely new theories remained rare, often due to a lack of sufficiently innovative perspectives.

Discussion and prospects

Big data technologies have played an increasingly important role in urban research and planning. Over the past decade, the volume of urban spatial big data quantitative research has grown from 19 articles in 2013 to 270 articles in 2023. However, when compared to the total number of articles published in JCR Q1 journals related to geography and urban studies (85,866), the body of work on urban spatial big data quantitative research remained relatively small (1,425), with theory-related articles constituting an even smaller subset (82). This suggests a significant need for further quantitative studies to expand the universal understanding of urban spatial variations and interactions. The research primarily focused on urban mobility, urban spatial quality, urban vitality, and urban form, while areas such as urban spatial equity, urban perception, and urban social safety were relatively less explored.

Profile of urban big data studies and shortcomings

The spatial distribution of study locations in urban spatial big data quantitative research exhibited significant global unevenness. Most studies focus on developed countries, with China being a notable exception. This indicates that many developing countries have not been adequately studied. Similarly, cities studied in 10 or more articles were predominantly located in national capitals and major economic hubs in developed countries and China, with secondary or smaller cities, especially in developing nations, receiving far less attention. These disparities are driven by multiple factors, such as the uneven distribution of ICT infrastructures and big data sources (Biljecki et al., 2023), even though these regions face numerous urban sustainability challenges that require attention and solutions.

The spatial resolution of existing research was mainly at the levels of grids, movement routes, and streets or roads. Finer spatial analytic units, such as buildings, land lots, urban places, and user locations, are still much less adopted in existing research. As for the temporal resolution, while the number of articles employing high-frequency temporal units has increased from 2013 to 2023, over half of all the articles still relyed on yearly or less frequent units. Studies with finer data in both spatial and temporal dimensions remain relatively rare. For example, in studies using streets or roads as their spatial analytic units, 102 articles used yearly or less frequent temporal units, while only 27 articles used hourly or more frequent units, accounting for less than a third of the former. Meanwhile, the study time periods of most articles were short, often limited to just 1 year.

However, it is crucial to note that the appropriate spatiotemporal resolution for big data research should be context-dependent, as the analysis accuracy required varies with the research purpose. Using finer data is not always better for all the studies. For research focused on built environments, such as urban spatial quality or land use location, finer spatial units (buildings or land lots) can provide more nuanced insights into these issues. In such cases, less frequent temporal data may suffice, as the built environment tends to be relatively stable within a short period. On the other hand, human activity-related research like urban mobility or vitality could use data in finer temporal units as human activities tend to vary rapidly for short periods within 1 day. Therefore, current studies should deepen their exploration of research topics by leveraging big data resources in ways that align with the unique attributes of their research objectives.

The application of big data has increased from 2013 to 2023, with online map data, cellphone signal data, and transport trajectory data becoming increasingly prominent. Notably, studies using online map data were much more prevalent than those using other categories of big data, likely due to the greater availability of such data. New methods for processing big data, such as image deep learning technologies, have also increased in recent years, though the total number of these articles remains relatively small compared to those employing traditional spatial correlation and regression models.

Theoretical advances in urban big data studies and research gaps

In contrast to 72.07% of articles focusing on describing phenomena or improving methods, fewer than 30% explored the underlying mechanisms, with a noticeable scarcity of theoretical exploration. It is anticipated that robust urban theories will gradually emerge as a result of extensive efforts to identify universal patterns. Theory-related articles primarily concentrated on fields such as urban transport and mobility, urban spatial environmental quality and health, and urban vitality. Regarding theoretical contributions, studies that complement and validate existing theories far outweighed those introducing new theoretical frameworks.

The theory-related articles mainly examined how the urban built environment has developed and how people have interacted with it. The integration of big data enables a more refined and granular depiction of urban space and human behavior. For example, online map data including POI and street view data can provide significantly more detailed information about socioeconomic function categories and ground objects in urban space than traditional remote sensing images, facilitating the analysis of more urban spatial differences at a much finer scale (Ito et al., 2024). This advantage can contribute to a better understanding of the urban spatial structure, spatial quality level, and evolutionary trends. In combination with big data reflecting human activities more precisely like movement trajectories, cellphone signals, social media, and data from various wearable or portable devices, the interaction mechanisms, associations, and causal relationships between the urban built environment and human activities can be more deeply explored. Thereby, researchers can make attempts to delineate where people are settled or traveling and how they are affected by various details of urban physical and functional space reflected in their thoughts, feelings, and behaviors from big data. This combination can provide pathways for significant theoretical explanations like the health effects of urban spatial quality conditions.

These theory-related findings can enhance the understanding of socio-technical systems by utilizing big data to enrich theoretical knowledge about urban space, thereby promoting more effective urban planning solutions. As cities around the world face the challenges of urban renewal, these findings can provide valuable guidance for more rational planning and policy formulation to improve spatial quality and residents' well-being. Moreover, these findings can also open new pathways for optimizing the allocation of public resources to foster social equity and enhance urban resilience, grounded in the accuracy of big data (Liu et al., 2024).

However, existing theory-related studies mainly investigated traditional urban space issues with big data, and studies about the impacts of new lifestyles and technologies in the digital era on traditional theories were relatively rare. For instance, although some studies have mentioned new lifestyles with big data technologies such as online ride-hailing, the research issues were still about traditional urban structures and travel needs. Some researchers have noticed the variation in the land use location of commercial functions in the context of the application of big data, such as massive online reviews and convenient delivery services (Zhang et al., 2023a). Nevertheless, how big data technologies reshape urban space and corresponding human activities was still expected for more mature and systematic theories. In addition, while existing theory-related research emphasized pedestrian accessibility and ecological factors in urban space, human behaviors and residents' needs are multifaceted and complicated. Therefore, a more integrated approach combining spatial and non-spatial elements is needed, through deeper analysis of big data, to establish a more intact framework for understanding the city organism.

Prospects and future research agendas

Building on the aforementioned advances, advantages, and limitations of urban spatial big data studies, a further research agenda is proposed as follows. First, existing research should deepen its focus on specific research topics by the attributes of the issues at hand, utilizing big data resources. For example, studies analyzing built environments should aim for a finer scale of analysis, such as buildings or land lots, to gain more nuanced insights, taking advantage of the improved spatial resolution of data aided by AI achievements (Pei et al., 2024).

Secondly, more urban big data studies are encouraged to shift from a primary focus on the applications of novel methods and descriptions of urban phenomena to the explorations of intricate mechanisms between different urban spatial factors and flows, and research issues with emerging challenges for urban sustainability. Future theoretical frameworks should consider multifaced benefits and trade-offs of different urban environmental conditions, from the perspectives of authorities, agencies, and individuals. A critical area for future theory-related studies lies in understanding how to make optimal planning and policies to save fiscal expenditure, boost economic prosperity, and improve residents' satisfaction for establishing the positive-feedback circulation within the urban space.

Thirdly, given that various disruptive technologies like ICTs, artificial general intelligence (AGI), artificial intelligence-generated content (AIGC), and robotic process automation (RPA) are increasingly reshaping people's behaviors and location preferences of socioeconomic activities in cities, more attention is expected to be paid to newly formed lifestyles and urban space for developing a systematic framework of new urban spatial theories in the digital era. On the other hand, traditional small data and conventional statistics are limited in analyzing cities as complex giant systems, while more theoretical knowledge about urban spatial mechanisms could be recognized with advanced AI methods and big data, transitioning from sensing intelligence to cognitive intelligence. For example, AIGC, when integrated into urban planning (Liu et al., 2024; Zheng et al., 2023), could use theoretical knowledge as an input in its planning process and subsequently adjusted or rediscovered by the output. In this way, spatial development strategies and design options could be more precisely predicted to witness the beginning of the new model of urban governance.

Limitations

While this literature review presents several innovative insights, it still has some limitations for improvement in the future. First, this literature review only focuses on JCR Q1 journals, which are relatively more highly reputable. As a result, it does not consider other journals that might also feature valuable quantitative analysis of urban big data. Secondly, although quantitative studies on urban spatial big data also involve abundant research issues and topics, this review concentrates on the most significant trends within the main research fields. Finally, given the rapid development of AI techniques, there is inherent uncertainty regarding changes in urban spatial conditions and research paradigms that may emerge in the future.

Conclusions

This literature review has analyzed the profile and theoretical advances of articles focusing on quantitative research about urban spatial big data across 57 representative journals. In terms of general characteristics, the predominant themes turned out to be urban transportation, urban spatial quality, and urban vitality. Studies were mainly concentrated in developed countries and large cities. The applications of big data and new methods like deep learning have increased from 2013 to 2023, while there is still a need for expansion compared with traditional methods. Some studies need to improve their spatiotemporal resolution of data based on research purposes. Meanwhile, this literature review also found a scarcity of theory-related studies exploring mechanisms, with most studies focusing on applying or proposing novel methods and uncovering new phenomena.

Theoretical contributions mainly focused on the formation mechanisms of the urban physical environment and its impacts on human activities, including people's locations and movements with their thoughts, feelings, and behaviors. The general trajectory of theoretical advances in each research field can be summarized as follows: initially, simplified mechanisms were tested using big data sources; subsequently, more factors of spatiotemporal heterogeneity were taken into consideration; besides, newly emerging phenomena were analyzed for their impacts on classical topics. These studies have significantly contributed to the complementing, validating, and modifying of theories, although the research proposing a theory was rare due to a lack of sufficiently innovative perspectives. Moreover, studies addressing the impacts of new lifestyles and technologies in the digital era on traditional theories were still relatively rare.

Future studies are encouraged to choose more appropriate spatiotemporal resolutions, enhance exploration of mechanisms of human-environment interactions in cities, and consider more about the impacts of smart technologies on urban space. The key issue is returning to the fundamental goals of human sustainable development, especially optimizing urban management to save fiscal expenditure, boosting economic prosperity, and improving residents' satisfaction for establishing the positive-feedback circulation in the urban space. By transforming from sensing intelligence to cognitive intelligence based on the integration of big data, advanced AI methods and theoretical knowledge, spatial development strategies and design options could be more precisely predicted to witness the beginning of a new model of urban governance.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by the National Natural Science Foundation of China (Grant No. 62394331, 52178044).

ORCID iDs

Tangqi Tu https://orcid.org/0000-0001-9361-3321
Enjia Zhang https://orcid.org/0000-0002-9768-495X
Ying Long https://orcid.org/0000-0002-8657-6988

Data Availability Statement

The raw data for this study, which consists of detailed tables where each row represents an article and includes tags related to data, methods, theory, and other relevant aspects, will be made available upon reasonable request. Please contact the corresponding author for access to the data.

Supplemental Material

Supplemental material for this article is available online.

References

Allam Z and Dhunny ZA (2019) On big data, artificial intelligence, and smart cities. *Cities* 89: 80–91.

Barki JP, Barić A and Bach MP (2022) Urban sustainability and urbanization: a review of research topics. *Croatian Regional Development Journal* 3(2): 1–18.

Batty M (2013) Big data, smart cities and city planning. *Dialogues in Human Geography* 3(3): 274–279.

Benita F and Tunçer B (2019) Exploring the effect of urban features and immediate environment on body responses. *Urban Forestry and Urban Greening* 43: 126365.

Biljecki F, Chow YS and Lee K (2023) Quality of crowdsourced geospatial building information: a global assessment of OpenStreetMap attributes. *Building and Environment* 237: 110295.

Boakye K, Bovbjerg M, Schuna Jr J, et al. (2023) Perceived built environment characteristics associated with walking and cycling across 355 communities in 21 countries. *Cities* 132: 104102.

Chang YW (2022) Capability of non-English-speaking countries for securing a foothold in international journal publishing. *Journal of Informetrics* 16: 101305.

Chen Z, Gong Z, Yang S, et al. (2020) Impact of extreme weather events on urban human flow: a perspective from location-based service data. *Computers, Environment and Urban Systems* 83: 101520.

Choi K and Zhang M (2017) The net effects of the built environment on household vehicle emissions: a case study of Austin, TX. *Transportation Research Part D: Transport and Environment* 50: 254–268.

Cox M and Ellsworth D (1997) Managing big data for scientific visualization. *ACM Siggraph* 97: 1–17.

Ennas G and Guardo MCD (2015) Features of top-rated gold open access journals: an analysis of the scopus database. *Journal of Informetrics* 9: 79–89.

Gan X, Huang L, Wang H, et al. (2021) Optimal block size for improving urban vitality: an exploratory analysis with multiple vitality indicators. *Journal of Urban Planning and Development* 147(3): 04021027.

Godwin C, Chen G and Singh KK (2015) The impact of urban residential development patterns on forest carbon density: an integration of LiDAR, aerial photography and field mensuration. *Landscape and Urban Planning* 136: 97–109.

Hao J, Zhu J and Zhong R (2015) The rise of big data on urban studies and planning practices in China: review and open research issues. *Journal of Urban Management* 4(2): 92–124.

He H, Lin X, Yang Y, et al. (2020) Association of street greenery and physical activity in older adults: a novel study using pedestrian-centered photographs. *Urban Forestry and Urban Greening* 55: 126789.

Huang J, Cui Y, Li L, et al. (2023) Re-examining Jane Jacobs' doctrine using new urban data in Hong Kong. *Environment and Planning B: Urban Analytics and City Science* 50(1): 76–93.

Ito K, Kang Y, Zhang Y, et al. (2024) Understanding urban perception with visual data: a systematic review. *Cities* 152: 105169.

Kandt J and Batty M (2021) Smart cities, big data and urban policy: towards urban analytics for the long run. *Cities* 109: 102992.

Kang CD (2020) Effects of the human and built environment on neighborhood vitality: evidence from Seoul, Korea, using mobile phone data. *Journal of Urban Planning and Development* 146(4): 05020024.

Kim YL (2025) Urban vitality measurement through big data and internet of things technologies. *ISPRS International Journal of Geo-Information* 14(1): 14.

Kim M and Cho GH (2021) Analysis on bike-share ridership for origin-destination pairs: effects of public transit route characteristics and land-use patterns. *Journal of Transport Geography* 93: 103047.

Kim JI, Yu CY and Woo A (2023) The impacts of visual street environments on obesity: the mediating role of walking behaviors. *Journal of Transport Geography* 109: 103593.

Kitchin R (2013) Big data and human geography: opportunities, challenges and risks. *Dialogues in Human Geography* 3(3): 262–267.

Kong L, Liu Z and Wu J (2020) A systematic review of big data-based urban sustainability research: state-of-the-science and future directions. *Journal of Cleaner Production* 273: 123142.

Kousis A1 and Tjortjis C (2023) Investigating the key aspects of a smart city through topic modeling and thematic analysis. *Future Internet* 16(1): 3.

Li D, Deal B, Zhou X, et al. (2018) Moving beyond the neighborhood: daily exposure to nature and adolescents' mood. *Landscape and Urban Planning* 173: 33–43.

Li S, Ma S, Tong D, et al. (2022) Associations between the quality of street space and the attributes of the built environment using large volumes of street view pictures. *Environment and Planning B: Urban Analytics and City Science* 49(4): 1197–1211.

Lim C, Kim KJ and Maglio PP (2018) Smart cities with big data: reference models, challenges, and considerations. *Cities* 82: 86–99.

Liu C, Xu F, Gao C, et al. (2024) Deep learning resilience inference for complex networked systems. *Nature Communications* 15(1): 9203.

Long Y and Huang CC (2019) Does block size matter? The impact of urban design on economic vitality for Chinese cities. *Environment and Planning B: Urban Analytics and City Science* 46(3): 406–422.

Marquet O (2020) Spatial distribution of ride-hailing trip demand and its association with walkability and neighborhood characteristics. *Cities* 106: 102926.

Nedungadi P, Surendran S, Tang KY, et al. (2024) Big data and AI algorithms for sustainable development goals: a topic modeling analysis. *IEEE Access* 12: 188519–188541.

Pan Y, Tian Y, Liu X, et al. (2016) Urban big data and the development of city intelligence. *Engineering* 2(2): 171–178.

Park KC and Lee CH (2019) A study on the research trends for smart city using topic modeling. *Journal of Internet Computing and Services* 20(3): 119–128.

Pei W, Biljecki F and Stouffs R (2024) Techniques and tools for integrating building material stock analysis and life cycle assessment at the urban scale: a systematic literature review. *Building and Environment* 262: 111741.

Tang J, Gao F, Han C, et al. (2021) Uncovering the spatially heterogeneous effects of shared mobility on public transit and taxi. *Journal of Transport Geography* 95: 103134.

UN (United Nations) (2023) The sustainable development goals report 2023: special edition. https://unstats.un.org/sdgs/report/2023/

Wang C and Yin L (2023) Defining urban big data in urban planning: literature review. *Journal of Urban Planning and Development* 149(1): 04022044.

Wang X, Zhang Y, Yu D, et al. (2022) Investigating the spatiotemporal pattern of urban vibrancy and its determinants: spatial big data analyses in Beijing, China. *Land Use Policy* 119: 106162.

Wang Z, Lu F, Liu Z, et al. (2023) Measuring spatial nonstationary effects of POI-based mixed use on urban vibrancy using Bayesian spatially varying coefficients model. *International Journal of Geographical Information Science* 37(2): 339–359.

Wei M (2022) How does the weather affect public transit ridership? A model with weather-passenger variations. *Journal of Transport Geography* 98: 103242.

Wu S, Liang Z and Li S (2019) Relationships between urban development level and urban vegetation states: A global perspective. *Urban Forestry and Urban Greening* 38: 215–222.

Wu J, Ta N, Song Y, et al. (2018a) Urban form breeds neighborhood vibrancy: a case study using a GPS-based activity survey in suburban Beijing. *Cities* 74: 100–108.

Wu J, Wang L and Li W (2018b) Usage patterns and impact factors of public bicycle systems: comparison between city center and suburban district in Shenzhen. *Journal of Urban Planning and Development* 144(3): 04018027.

Xing X, Yu B, Kang C, et al. (2024) The synergy between remote sensing and social sensing in urban studies: review and perspectives. *IEEE Geoscience and Remote Sensing Magazine* 12(1): 108–137.

Yang C, Clarke K, Shekhar S, et al. (2020) Big spatiotemporal data analytics: a research and innovation frontier. *International Journal of Geographical Information Science* 34(6): 1075–1088.

Yang L, Ao Y, Ke J, et al. (2021) To walk or not to walk? Examining non-linear effects of streetscape greenery on walking propensity of older adults. *Journal of Transport Geography* 94: 103099.

Yu H and Peng ZR (2019) Exploring the spatial variation of ridesourcing demand and its relationship to built environment and socioeconomic factors with the geographically weighted poisson regression. *Journal of Transport Geography* 75: 147–163.

Zhang S, Liu X, Tang J, et al. (2019) Urban spatial structure and travel patterns: analysis of workday and holiday travel using inhomogeneous Poisson point process models. *Computers, Environment and Urban Systems* 73: 68–84.

Zhang E, Xie H and Long Y (2023a) Decoding the association between urban streetscape skeletons and urban activities: experiments in Beijing using Dazhong Dianping data. *Transactions in Urban Data, Science, and Technology* 2(1): 3–18.

Zhang Y, Deng Q, Guo M, et al. (2023b) Using street view imagery to examine the association between urban neighborhood disorder and the long-term recurrence risk of patients discharged with acute myocardial infarction in central Beijing, China. *Cities* 138: 104366.

Zheng Y, Lin Y, Zhao L, et al. (2023) Spatial planning of urban communities via deep reinforcement learning. *Nature Computational Science* 3(9): 748–762.

Tangqi Tu, Ph.D., was a postdoctoral researcher in School of Architecture, Tsinghua University. He is currently working in Department of Geography and Spatial Information Techniques, Ningbo University. His research focuses on urban space planning and sustainable urbanization.

Enjia Zhang, Ph.D., graduated from School of Architecture, Tsinghua University. She is currently working in College of Architecture and Urban Planning at Beijing University of Technology. Her research focuses on data-augmented design, urban analytics, and urban science, emphasizing the application of data in urban planning and design.

Ying Long, Ph.D., is now working in School of Architecture, Tsinghua University, as a tenured professor. His research focuses on urban science, including applied urban modeling, urban big data analytics and visualization, data augmented design, and future cities. More information is available at https://www.beijingcitylab.com/longy.