



ARTICLE



<https://doi.org/10.1057/s41599-023-01551-y>

OPEN

Realistic characteristics and driving mechanisms of pseudo-human settlements in Chinese cities

Wenbo Yu¹, Jun Yang^{1,2,3✉}, Feng Wu⁴, Baojie He⁵, Bing Xue⁶, Shaohua Wang⁷, Huisheng Yu¹, Xiangming Xiao⁸ & Jianhong (Cecilia) Xia⁹

Pseudo-human settlements (PHS) reflect the subjective wishes and needs of residents. An in-depth characterization of PHS aids in clarifying the actual development needs and future development directions of the city from the residents' perspective. China is currently undergoing high-speed urbanization and advocates improving the quality of cities. It is particularly important to test the achievements of urbanization through PHS. This study quantitatively analyses spatiotemporal characteristics and the realistic driving mechanism of PHS in 286 Chinese cities through spatial autocorrelation analysis and geographic detectors. The results indicate that (1) the PHS level exhibited notable hierarchical characteristics and spatial differences (high in the eastern coastal regions and low in the western inland regions), with weak positive autocorrelation characteristics; (2) nighttime light index, per capital GRP, and population size are major drivers of heterogeneity of development level and functional structure of PHS; and (3) the functional structure followed an emergence-development-maturity-decline-extinction life cycle. Overall, the development of PHS must play a leading role in high-level areas, and urban decision-makers should clarify and adapt to local preferences of PHS and predict future changes. The findings of this study can help China and other countries promote realistic development by unveiling the characteristics of PHS.

¹School of Humanities and Law, Northeastern University, Shenyang, China. ²Jangho Architecture College, Northeastern University, Shenyang, China.

³Human Settlements Research Center, Liaoning Normal University, Dalian, China. ⁴Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing, China. ⁵School of Architecture and Urban, Planning, Chongqing University, Chongqing, China. ⁶Institute of Applied Ecology, Chinese Academy of Sciences, Beijing, China. ⁷School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China. ⁸Center for Spatial Analysis, University of Oklahoma, Norman, OK, USA. ⁹School of Earth and Planetary Sciences (EPS), Curtin University, Perth, WA, Australia. ✉email: yangjun8@mail.neu.edu.cn

Introduction

The scope, vitality, and attention of a human's daily life are extremely limited (Surya et al., 2020); people can neither contact nor understand the entirety of the outside world. Beyond a person's direct life experiences, one must rely on the information about society produced by the media (Gavin, 2018) to understand the same. Consequently, one's views obtained through information gradually deviate from direct experience, forming via a "pseudo-environment" created by the media with strong parole and notable color (Lippmann, 1997). With the advent of the Web 2.0 era, We Media and social media have developed rapidly, providing infinite opportunities for communications among people (Denissen et al., 2010), overcoming spatial restrictions caused by geographical distances, and transforming the world into an interrelated pseudo-real environment (Tsou, 2015). The traditional "pseudo-environment" has subtly changed and developed into pseudo-human settlements (PHS) with specific functions and services, such as living, socializing, shopping, traveling, and entertainment (Tian et al., 2022a). Compared with a "pseudo-environment," PHS are information-based human settlements constructed by residents by editing and processing with the aid of media. It can be characterized by such features as authenticity, deviation, and informatization, and comprises residents' online human settlement activities. "Authenticity" refers to the PHS reflecting the subjective cognition, willingness, and preference of residents for a living environment. "Deviation" implies that the residents' subjective perception of the human settlement environment may be affected by other residents, and that there is a certain deviation from the real situation. "Informatization" refers to the PHS environment's virtualization and extension of the real human settlement environment. Information that urban residents cannot fully perceive in the real human settlement environment can be obtained through the PHS environment. In-depth explorations of the characteristics of PHS in cities can help discover and meet the subjective needs of residents while promoting the construction and development of more diversified and complicated urban human settlements.

Users generate massive amounts of data on various social media platforms based on their background details and daily activities; these data can facilitate an in-depth exploration of the properties of PHS. Using artificial intelligence, potential trends in the development of PHS and the hidden emotions and opinions of users can be identified (Ghani et al., 2019). Relevant studies have already covered topics on society (Dong et al., 2015; Schonfeld et al., 2021), medical treatments (Liu et al., 2016), sports (Barnaghi et al., 2016), shopping (Fast et al., 2016), and transportation (Chen et al., 2016). Social media data are also an important source of urban knowledge (Shen and Yang, 2014), offering new perspectives for delineating the relationship between PHS and the spatial distributions of social aspects such as urban economy, land use, and human activities (Jendryke et al., 2017). Using geo-tagged pictures, log-ins, tweets, and point-of-interest data provided by social media big data platforms such as Weibo, Flickr, and Facebook, scholars have been able to: (1) confirm the relationship between urban space and residents' emotions and behaviors (Huang et al., 2020; Zhu et al., 2021; Kim et al., 2022); (2) identify the function and structure of real urban space (Hu et al., 2019; Chen et al., 2019; Anjali and Kumar, 2021); and (3) analyze the impact of the urban ecological environment on residents' activities and emotions (Zheng et al., 2019; Yan et al., 2019; Wang et al., 2021b).

Social media use is highly subjective, making it conducive to exploring users' behavior and emotions. Within social media, users are themselves the producers, consumers, and disseminators of content, inevitably causing information pollution, such as rumors and fake news (Zhang and Ghorbani, 2020; Meel and

Vishwakarma, 2020). Social media is more likely to spread popular social events (Guan et al., 2014), reducing its ability to comprehensively and objectively describe the nature of PHS.

In China, Baidu is the most popular online search engine, with a market share of 80.5% and more than one billion users (Jin et al., 2022). The Baidu index, a Chinese keyword analysis tool, contains rich spontaneous search information and provides a new perspective for detecting residents' subjective network concerns (Liu and Liao, 2021). The use of the Baidu index accounts for shortcomings in the social media data used in this study. More importantly, the Baidu index is being increasingly used to analyze the relationship between the attributes of urban reality and PHS across a wide area by providing geographic data tags (Zhao et al., 2020b, 2021; Chen et al., 2021a). At the national scale (i.e., China), Wang and Loo (2019) used the Baidu index and found notable hierarchical differences in the number of Internet news reports obtained by cities. The scope of influence of city names was highly correlated with the socioeconomic development level of those cities (Guo et al., 2022). Zhang and Tang (2016) proposed a housing price attention index based on the Baidu index to explore the spatial pattern of public focus on housing prices, revealing that the spatial interactions with respect to the focus on housing prices between cities were consistent with the results of the gravity model. Li et al. (2021) found a strong positive correlation between the focus on most smog keywords and the PM_{2.5} concentration based on the Baidu index. However, despite their merit, these studies only used some functions of the PHS. Therefore, the overall driving mechanism of PHS based on urban reality attributes requires further discussion.

According to the science of human settlements, PHS are based on real human settlements (Wu, 2001). Tian et al. (2021) and Xue et al. (2021) discuss the correlation characteristics between pseudo and real human settlements at different scales. They found a notable positive correlation between the pseudo-human settlement and real human settlement development levels. Through qualitative analysis, both studies proposed that socioeconomic factors are the leading realistic factors affecting the development level of PHS. The preferences of residents for various functional applications of PHS indicate their real-world requirements. However, in PHS, how real attributes affect functional preferences and how functional preferences change over time remain unclear. Therefore, this study first analyzed the spatiotemporal patterns of PHS in cities at the national scale and examined how various attributes of the real human settlements influence PHS development level and functional structure. The present results could facilitate the identification of the inherent requirements and mechanisms of human settlement construction and coordination of the development direction of cities, based on the subjective development requirements of residents. The study results could also further promote the collaborative construction of regional pseudo-real human settlements and provide novel geographical perspectives for human settlement studies.

Specifically, we performed the following analyses. (1) A statistical aggregation of the Baidu index for commonly used classification keywords was used to establish the pseudo-human settlements index (PHSI) for 286 Chinese cities, while the reality attribute dataset was obtained by integrating multi-source data. (2) Based on spatial analysis and visualization of the PHSI of cities, we obtained the spatiotemporal characteristics of the development level of PHS to clarify the interaction relationship among the regional PHS. (3) Using a combination of *K*-Means cluster analysis and geographic detectors, we explored the mechanism associated with the reality attributes of cities in terms of the development level and functional structure differentiation of PHS.

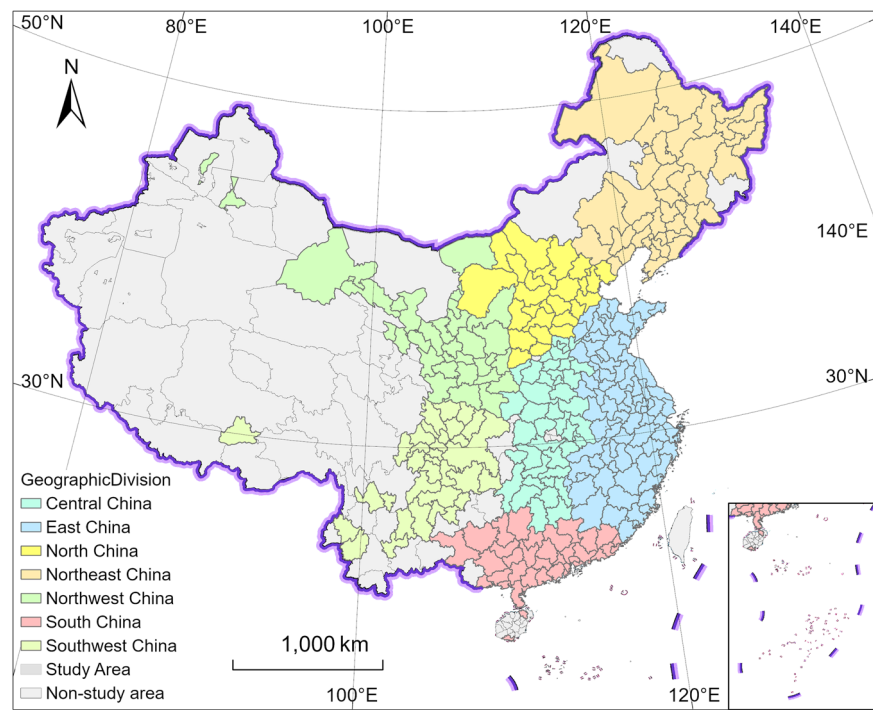


Fig. 1 Spatial distribution and geographic division of cities in the study area. The 286 cities are divided according to China’s seven natural geographical divisions, as depicted in the figure. The divisions usually combine several geographically and culturally close provinces into one area. The map was developed by the authors using ArcGIS Pro Version 3.0.2, and the background map was issued by the Ministry of Natural Resources of the People’s Republic of China (drawing review No. GS (2020) 4619), as shown in the website: <http://bzdt.ch.mnr.gov.cn/>.

Methods

Data

Study area. A total of 286 Chinese cities, including four municipalities, 15 sub-provincial cities, and 267 prefecture-level cities, were selected for the analysis. The cities typically have population of several millions and area of several thousand square kilometers, equivalent in size to metropolitan areas in the United States (Li, 2020). Figure 1 shows the spatial distribution and geographical divisions of the study area.

Data sources. The PHSI was calculated based on the Baidu index. We counted search indices for common classifications of PHS for cities in China from January 1, 2015, to December 31, 2020, yielding a data volume of approximately 20 million entries. Table 1 lists the index system for PHS. The index system is divided into ten functional structure indices involving several main functions, such as home life, socialization, shopping, travel, and entertainment in PHS. Three related keyword indices were counted under each functional structure index, which conformed to the comprehensive requirements of the design of the index system. The index system simultaneously followed the scientific framework of human settlements (Li et al., 2015) based on related studies (Xue et al., 2021; Tian et al., 2022a), satisfying the scientific principle of index system design. The application has a competitive effect with similar applications over a long-term series and has attention instability. Therefore, compared with using specific application names as keywords in related studies, the index system proposed in this study used a more ambiguous index system. Using the function as a keyword partially improved the objectivity of the index system in describing the PHS.

We refer to the following principles in selecting cities’ realistic attributes: being scientific, simplification, and full coverage. First, being scientific, human settlements science believes that pseudo-human settlements are the extension of real human settlements. In human settlements science, social subsystem is the core

Table 1 Index system for the pseudo-human settlements.

Functional index	Keywords	Functional index	Keywords
Audio-visual (F_{AV})	Video Live broadcast Music	Travel traffic (F_{TT})	Map Tourism Public transportation
Chat and socialize (F_{CS})	Chat Communication Community	Financial management (F_{FM})	Bank Stock Payment
Book reading (F_{BR})	Cartoon Novel	News information (F_{NI})	Information Science and technology Headlines
Fashion shopping (F_{FS})	Magazine E-commerce Global shopping Fashion	Home life (F_{HL})	Weather Gourmet food Real estate
Learning and education (F_{LE})	Examination Education Learning	Medical health (F_{MH})	Health preservation Beauty Medical treatment

system influencing real human settlements; population subsystem is the internal driving system of real human settlements; and environment subsystem is the potential influence system of real human settlements. Therefore, this study explores the relationship between the three subsystems of population, society, and environment, and PHS. Second, on simplification, previous studies (Xue et al., 2021; Tian et al., 2022a, 2022b) on real human settlements usually selected dozens of indicators for synthesis to explore the relationship between them and PHS. However, we believe this is not conducive to the interpretation of key issues. Accordingly, we selected several indicators with higher weight in previous studies to explore their relationship with PHS in detail.

Table 2 Data sources and descriptions.				
Data type	Data source	Describe	Time	Indices
Baidu index	Baidu index platform	http://index.baidu.com	2015–2020 (daily)	PHS index
Statistical data	China Urban Statistical Yearbook	https://data.cnki.net/Yearbook/	2016–2021 (annual)	Registered population at the end of the year; Per capita GDP; Industrial structure of regional GDP
Night data	VIRS-NPP	http://www.ngdc.noaa.gov/	2015–2020 (annual)	Statistical value of night light index by city
Air quality	AQI Study	https://www.aqistudy.cn	2015–2020 (daily)	Air quality index
Administrative divisions	National Administrative Division Information Inquiry Platform	http://xzqh.mca.gov.cn/map	2020 (annual)	Geographical division; Longitude and latitude of city center

Third, with regard to full coverage, the human settlement environment is the living environment of all urban residents; consequently, its constituent indicators should also be closely related to urban residents.

To sum up, the selection logic of urban reality attributes in this study primarily follows the existing scientific research on human settlements, and simplifies the complex elements of previous research. Reality attributes of cities can be divided into statistical data, night light data, and air quality index data according to the specific acquisition method. Of these, statistical data encompasses important indicators that reflect the development status of cities in terms of population, resources and environment, economic development, and residential livelihoods. The night light index presents the intensity of human activities at night in a region, reflecting the urban form and expansion from the side (Chen et al., 2017). The air quality index is directly related to the health, daily production, and life of urban residents. Table 2 lists the specific data sources and their descriptions. We tested whether a linear relationship exists between various reality attributes. The results show that the tolerance of all attributes was greater than 0.1, and the variance inflation factor was less than 5, indicating no collinearity between attributes.

Methods

Pseudo-human settlements index (PHSI). The PHSI is composed of ten functional structure indices of PHS. Each functional structure index was described by three specific keyword indices, indicating the attention degree and residents’ preference for the main functions of PHS. There was a considerable difference in the attention among different keyword indices. For instance, the average annual Baidu index of the “weather” keywords was 1014, while the average annual Baidu index for the “information” keywords was only 16. All keywords had the same importance; therefore, the Min–Max standardization method was used to process them. All keywords were considered positive indices. The functional structure indices of PHS were calculated as follows:

$$PHSI_i = \sum_{j=1}^3 \frac{k_{ij} - k_{min}}{k_{max} - k_{min}} \tag{1}$$

where $PHSI_i$ represents the evaluation result of functional structure index i , k_{ij} represents the Baidu index of the j th keyword of functional structure index i , and k_{max} and k_{min} , respectively, represent the maximum and minimum values of the keyword Baidu index for all years and cities.

To maintain the original data information to the greatest extent possible, the PHSI was directly obtained by adding ten functional structure indices, reflecting the development level of PHS, with a value range of [0, 30]. Equation (2) shows the PHSI calculation method. For each index, its proportion, p , in the PHSI reflects the preference degree of the residents for the function of PHS, as

expressed in Eq. (3). The set formed by the proportion, p , of each index was the functional structure of PHS:

$$PHSI = \sum_{i=1}^{10} PHSI_i \tag{2}$$

and

$$p_i = \frac{PHSI_i}{PHSI} \tag{3}$$

Spatial autocorrelation analysis. “Space” is a key topic in geography research. The first law of geography proposed by Tobler (1970) holds that everything is related to other things, but similar things are more closely related. In a virtualized and information-based environment, whether the spatial pattern of PHS conforms to this law can be confirmed by global and local spatial autocorrelation analysis (Bivand and Wong, 2018).

The global spatial autocorrelation of the PHSI can be measured via the Global Moran’s I index. This index reflects the overall correlation of the PHSI in Chinese cities, with a value range of [−1, 1]. The closer this value is to 1, the more notable the positive spatial autocorrelation between regions; the closer it is to −1, the more notable the negative spatial autocorrelation between regions; the closer it is to 0, the more evident are the random spatial distribution characteristics. The Local Indicators of Spatial Autocorrelation (LISA) were used to analyze the local spatial pattern of the average annual PHSI. When the analysis result for a city was statistically significant ($P < 0.05$), the city was categorized into one of four cluster/outlier types according to its spatial pattern, as follows: (1) High–High cluster type: a regional city with a high PHSI with small differences, it is a PHSI hotspot, showing the “diffusion” characteristic; (2) Low–High outlier type: the city has a low PHSI, but the surrounding cities have a high PHSI, showing the “sinking” characteristic; (3) Low–Low cluster type: a regional city with a low PHSI with small differences, it is a PHSI cold spot, showing the “contagious” characteristic; (4) High–Low outlier type: the city itself has a high PHSI, but its surrounding cities have a low PHSI, showing the “polarization” characteristic.

K-means multivariate cluster analysis. The numerical reality attributes and PHS functional structure partially reduced the interpretability of the analysis results. We divided the numerical data into different clusters via the clustering method such that the similarity between individuals in the same cluster was as large as possible, whereas the similarity between different clusters was as small as possible. The K -means unsupervised clustering algorithm has been widely used in related urban fields owing to its simple idea and optimal clustering effect (Delmelle, 2015; Feng et al., 2020). This study used SPSS 26 to concretely implement this method. The K -Means method must set the cluster number, K , in advance; the clustering results under this value of K can be

evaluated based on a silhouette score (Rousseeuw, 1987). The closer the score is to 1, the better the clustering results. In this study, the trend of the silhouette scores with cluster numbers of 2–30 was plotted. The cluster number corresponding to a notable elbow point or local maximum for the silhouette scores curve was selected for a follow-up analysis; therefore, the clustering results were complex and effective.

Geographical detectors. In this study, the geographical detectors proposed by Wang et al. (2010, 2016) were used to analyze the mechanism of the reality attributes for the development level and functional structure differentiation of PHS. Widely used in related urban fields, Geographical Detectors are statistical methods to detect spatial differentiation that reveal the driving factors behind it (Zhao et al., 2020a; Xu et al., 2021). Geographical Detectors include four detectors, namely: (1) The factor detector, revealing to what extent factor X explains the spatial differentiation of attribute Y , expressed as follows:

$$q = 1 - \frac{SSW}{SST} = 1 - \frac{\sum_{h=1}^L N_h \sigma_h^2}{N \sigma^2} \quad (4)$$

where L represents the number of strata of factor X , N_h and N are the number of units in strata h and the entire area, respectively, σ and σ_h are the variances of the Y values in strata h and the entire area, respectively, and SSW and SST are the sum of the variance within the strata and total variance of the entire area, respectively. The range of q is [0, 1].

2. The interaction detector, which assesses whether factors X_1 and X_2 act together to increase or decrease the explanatory power of the dependent variable, Y . The evaluation method first calculates the q value of two factors, X_1 and X_2 , to Y , i.e., $q(X_1)$ and $q(X_2)$, and then calculates the q value when they interact: $q(X_1 \cap X_2)$. By analyzing the correlation among $q(X_1)$, $q(X_2)$, and $q(X_1 \cap X_2)$, the joint effect of X_1 and X_2 was obtained.
3. The risk detector, which is used to judge whether there is a statistically significant difference between the attribute mean values of the two strata.
4. The ecological detector, which is used to compare whether there are significant differences between the two factors and their influences on the spatial distribution of the attributes, as measured by the F statistic:

$$F = \frac{N_{X1}(N_{X2} - 1)SSW_{X1}}{N_{X2}(N_{X1} - 1)SSW_{X2}} \quad (5)$$

$$SSW_{X1} = \sum_{h=1}^{L1} N_h \sigma_h^2, \quad SSW_{X2} = \sum_{h=1}^{L2} N_h \sigma_h^2 \quad (6)$$

where N_{X1} and N_{X2} represent the sample sizes of factors X_1 and X_2 , respectively; SSW_{X1} and SSW_{X2} represent the sum of the variance within the stratum formed by X_1 and X_2 , respectively; and $L1$ and $L2$ represent the number of strata in X_1 and X_2 , respectively. The null hypothesis, H_0 , was $SSW_{X1} = SSW_{X2}$. If H_0 was rejected at the α significance level; this indicated that there was a significant difference in the influence of the two factors, i.e., the effects of X_1 and X_2 on the spatial distribution of attribute Y .

Results

PHSI spatiotemporal pattern characteristics. Figure 2 shows the spatial distribution of the development levels of PHS in cities classified by the K -means method. Table 3 lists the top 20 cities with the average PHSI. The PHSI in Chinese cities had the following characteristics. (1) The development level of PHS had notable hierarchical characteristics. Among the top 20 cities in this index, the majority included municipalities, sub-provincial cities, and provincial capitals. The average values of the PHSI for

the municipalities, sub-provincial cities, and provincial capital cities were 16.50, 11.25, and 10.08, respectively, which were considerably higher than the average value for prefecture-level cities, which was 3.04. (2) East–west and north–south China were delimited by the boundaries of coastal geographic divisions and the Qinling–Huaihe line, respectively, and differences in the spatial distribution of PHSI in the regions were compared. The results showed that the PHSI exhibited notable spatial heterogeneities, high in the east and low in the west, and the north–south variations were not evident. The average PHSI for eastern cities was 4.14, which was considerably higher than the 2.87 for western cities. Among the cities in the east, there were 70 cities with high and medium PHSI levels (39.33% of all eastern cities); however, there were only 19 cities in the west with high and medium PHSI levels (17.59% of all western cities). In contrast, the average PHSIs of the northern and southern cities were 3.43 and 3.88, respectively, which were similar. Among the cities with high and medium PHSI levels, 42 were northern cities (31.11% of all northern cities), and 47 were southern cities (31.13% of all southern cities). Among the eastern coastal cities with good socioeconomic development, the cities with high and medium PHSI values were densely distributed; however, most non-coastal cities had low PHSI level. (3) Fig. 3 shows the time-series of the PHSI. The annual average index in 2015 was the highest, at 3.92, and the average in 2016 was the lowest, at 3.49. The degree of evolution in terms of the average indices for each year was weak over time, showing relatively stable time characteristics and an evolution direction with certain randomness. The standard deviations between the PHSI values of each year were similar to the average value, indicating considerable differences in the development levels of PHS.

PHSI spatiotemporal autocorrelation characteristics. Figure 4 shows the local spatial autocorrelation results and the LISA scatter plot distribution in 2016, 2018, and 2020. Moran's I indices of PHSI fluctuated within the 0.124–0.162 range, thus exhibiting weak positive autocorrelation characteristics. The overall spatial autocorrelation pattern was relatively stable. The Z -score showed that the confidence level of this result was stable, i.e., >99%. The local spatial autocorrelation results showed the following. (1) The cities that exhibited “diffusion” characteristics for the development level of PHS were mainly central cities in the national-level urban agglomerations, i.e., Beijing–Tianjin–Hebei, Yangtze River Delta, and Pearl River Delta. (2) Cities showing “sinking” characteristics were peripheral cities in the national-level urban agglomerations. (3) Cities exhibiting “polarization” characteristics were primarily provincial capitals located in the remote provinces of China, such as Lanzhou and Yinchuan in Northwest China; Shenyang, Changchun, and Harbin in Northeast China; and Chongqing, Chengdu, Guiyang, Kunming, and Nanning in Southwest China. (4) Cities showing “contagious” characteristics were mainly prefecture-level cities in the remote provinces of China, concentrated in old industrial base cities in eastern Heilongjiang Province and ecologically fragile cities in the Shaan–Gan–Ning region.

Specifically, among the regions exhibiting “diffusion” characteristics, the Yangtze River Delta urban agglomeration is the region with the best urbanization foundation in China. Among the seven core cities (Shanghai, Nanjing, Suzhou, Hangzhou, Wuxi, Hefei, and Ningbo) in the urban agglomeration, the average PHSI was 11.88, and the standard deviation was 3.72. The cities formed a good PHS coordinated development model, and all cities had a high level of development. Among the regions exhibiting “sinking” characteristics, Zhangjiakou and Chengde are located on the periphery of the Beijing–Tianjin–Hebei urban

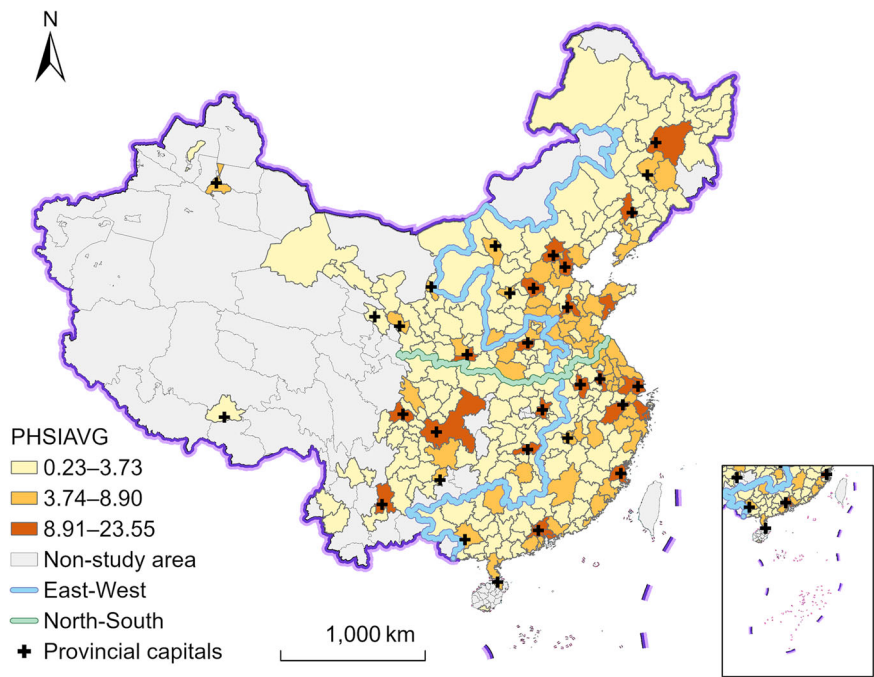


Fig. 2 Spatial distribution of the development levels of PHS. PHSI is divided into three grades according to the K-Means clustering method. The first, second, and third grades range from 11.75 to 24.66, 4.81 to 11.74, and 0.20 to 4.80, representing the high, medium, and low PHS development levels, respectively. The map was developed by the authors using ArcGIS Pro Version 3.0.2, and the background map was issued by the Ministry of Natural Resources of the People’s Republic of China (drawing review No. GS (2020) 4619), as shown in the website: <http://bzdt.ch.mnr.gov.cn/>.

Table 3 Top 20 cities with the average PHSI.					
Rank	City	Average PHSI	In coastal regions?	North or south	City level
1	Beijing	23.55	Yes	North	Municipalities
2	Shanghai	18.59	Yes	South	Municipalities
3	Guangzhou	16.34	Yes	South	Sub-provincial city
4	Shenzhen	15.57	Yes	South	Sub-provincial city
5	Chengdu	15.24	No	South	Sub-provincial city
6	Hangzhou	14.80	Yes	South	Sub-provincial city
7	Zhengzhou	12.95	No	North	Provincial capital city
8	Wuhan	12.85	No	South	Sub-provincial city
9	Chongqing	12.67	No	South	Municipalities
10	Suzhou	12.07	Yes	South	City at prefecture-level
11	Xi'an	11.55	No	North	Sub-provincial city
12	Tianjin	11.19	Yes	North	Municipalities
13	Nanjing	11.03	Yes	South	Sub-provincial city
14	Changsha	10.92	No	South	Provincial capital city
15	Jinan	10.62	Yes	North	Sub-provincial city
16	Qingdao	9.91	Yes	North	Sub-provincial city
17	Hefei	9.87	Yes	South	Provincial capital city
18	Dongguan	9.67	Yes	South	Prefecture-level city
19	Shijiazhuang	9.67	Yes	North	Provincial capital city
20	Fuzhou	9.32	Yes	South	Provincial capital city

agglomeration, and their PHSIs were 3.36 and 2.44, respectively, which were considerably lower than the 23.55 and 11.19 of Beijing and Tianjin, respectively. Compared with the PHS level of cities in the hinterlands of urban agglomeration, the development level of PHS in the cities was suboptimal. Among the regions exhibiting “polarization” characteristics, the three northeastern provinces (Liaoning, Jilin, and Heilongjiang provinces) were once key heavy industry bases in China; however, recently, there have been human settlement challenges, such as population shrinkage, resource depletion, and environmental pollution. Among the three northeastern provinces, the provincial capital cities,

Shenyang, Changchun, and Harbin, had high levels of PHS development. Their average PHSI was 8.70, whereas the average PHSI of other prefecture-level cities was only 1.85, indicating huge differences in PHS development in the three northeastern provinces. Among the regions exhibiting “contagious” characteristics, cities in Gansu Province and Ningxia Autonomous Region, excluding Lanzhou and Yinchuan, had an average PHSI of 0.84 and a standard deviation of 0.35, indicating a lower level of PHS development and less variance. In addition, the cities had relatively low levels of urban, social, and economic development and fragile ecological environments.

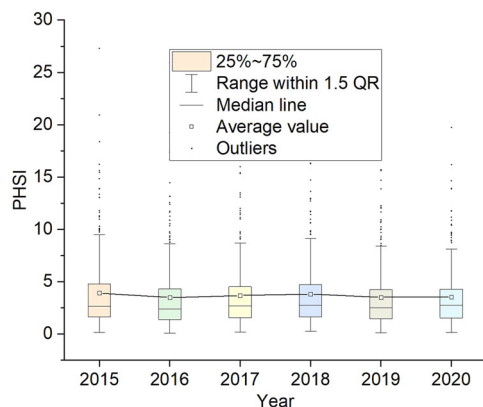


Fig. 3 Time evolution of the pseudo-human settlements index (PHSI).

The graph describes the numerical characteristics of PHSI at each research time node, including the mean, median, outliers, etc., and describes the time evolution characteristics of the annual mean of PHSI. The figure was developed by the authors using Origin Pro Version 2021.

Driving mechanism of the development level of PHS. The geographical location, population size, and air quality index were classified using the classification method defined by national standards. The per capita GRP, night light index, and industrial structure were clustered via the K-Means clustering method, and the cluster number for these reality attributes was set to 5 according to the silhouette scores. The specific classification standards are shown in Supplementary Table S1 and Supplementary Table S2.

The explanatory strength for the single factor of the city reality attributes for the development level differentiation of PHS from high to low was as follows: night light index > population size > per capita GRP > industrial structure > geographical location > air quality index (Fig. 5). Of these, the night light index, population size, and per capita GRP explained the spatial differentiation in the development level of PHS > 0.3, which were the direct factors affecting the development level of PHS. The interaction detector analysis results showed that the interaction between the reality attributes had a stronger explanatory power than a single attribute, that is, the development level of PHS was a

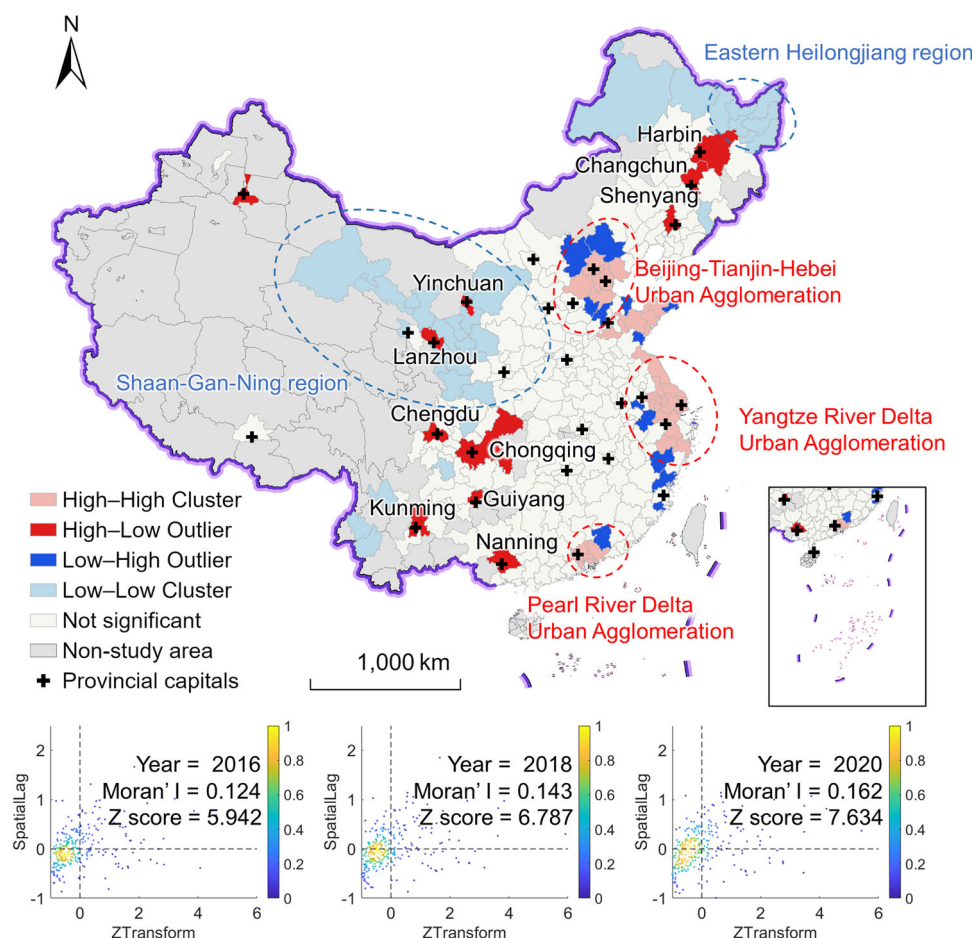


Fig. 4 Spatial autocorrelation analysis of the pseudo-human settlements index (PHSI). The upper and lower parts of the figure, respectively, show the spatial distribution map and scatter diagram of local spatial autocorrelation analysis of PHSI, where the High-High cluster implies that the PHSI of the city and surrounding cities is high (statistically significant) (located in the first quadrant of the scatter map); Low-High outlier means that the PHSI of the city is low while that of its surrounding cities is high (located in the second quadrant of the scatter diagram); Low-Low cluster means that the PHSI of the city and surrounding cities is low (located in the third quadrant of scatter diagram); High-Low outlier represents the PHSI of the city itself is high while that of its surrounding cities is low (located in the fourth quadrant of the scatter diagram). The map was developed by the authors using ArcGIS Pro Version 3.0.2, and the background map was issued by the Ministry of Natural Resources of the People's Republic of China (drawing review No. GS (2020) 4619), as shown in the website: <http://bzdt.ch.mnr.gov.cn/>.

common result of the interaction among various reality attributes. Therefore, the night light index had the highest single-factor explanation strength because it had a more comprehensive ability to express the scope of urban development (Duque et al., 2019; Zheng et al., 2021, 2022) and socioeconomic attributes (Yu et al., 2015; Liu et al., 2021). The interaction results are influenced by the strength of a single factor and a positive correlation between the two factors involved in the interaction. The interaction between the population size and per capita GRP had the highest explanatory power in terms of the differentiation of the development level of PHS, which was 0.682. Combined with the analysis results for the ecological detectors, interactions among the industrial structure, geographical location, and air quality index with other reality attributes could considerably enhance the explanatory strength, indicating that these were potential factors affecting the development of PHS. The strong single explanatory power factor, night light index, had no evident effect on the enhancement of explanatory power, primarily due to its high positive correlation with population size (Townsend and Bruce, 2010; Zeng et al., 2011) and per capita GRP (Chen et al., 2022; Wang and Sun, 2022).

Functional structure types of PHS. Table 4 lists the six functional structure types of PHS determined based on the *K*-means methods. As listed in Table 4, the learning and education function had a proportion exceeding 10% of the average value in each cluster center, indicating that relevant parties have focused more on the learning and education function of PHS in the cities of China.

The polarized degree of each type of functional structure of PHS was measured via the sum of the absolute difference between the

proportion of the function and 10%. Type Is were relatively average, and the sum of the absolute difference was 15.38%, exhibiting the characteristics of a balanced functional structure. The sum of the absolute difference for other types was >20%. Among them, Types IV and V had considerable preferences for learning and educational functions, exhibiting highly polarized functional structure characteristics. In contrast, Types II, III, and IV preferred learning and educational, medical health, fashion shopping, chat, and socialization functions, among others, exhibiting multi-polarization functional structure characteristics.

Based on Supplementary Table S3, similar to the factors affecting the development level of PHS, the functional structure differentiation of PHS was also affected by night light index ($q = 0.255$), population size ($q = 0.240$), per capita GRP ($q = 0.199$), and the proportion of the tertiary industry ($q = 0.169$). From a temporal evolution perspective, the functional structure types of PHS were characterized by a dynamic change process. Among the 286 examined cities, the functional structure types in 243 cities have experienced a transition, resulting in standard deviation ellipses with considerable overlap owing to the stacking of the functional structure types across multiple years. Figure 6(a) shows the annual changes in the number of cities of various types. The trend in the number of cities of various types generally presented an inverted “V” shape that rises to a peak and then declines, indicating that the functional structure type of PHS had a specific life cycle. Among them, Types IV and V were in the emergence and development stages during the study period. The types exhibited an increasing and mature development trends. Types I, and III were characterized by a period of decline and extinction, such that the number of cities within the types exhibited a declining trend and even extinction. Type II underwent a relatively complete life cycle from development to maturity, decline, and even extinction during the study

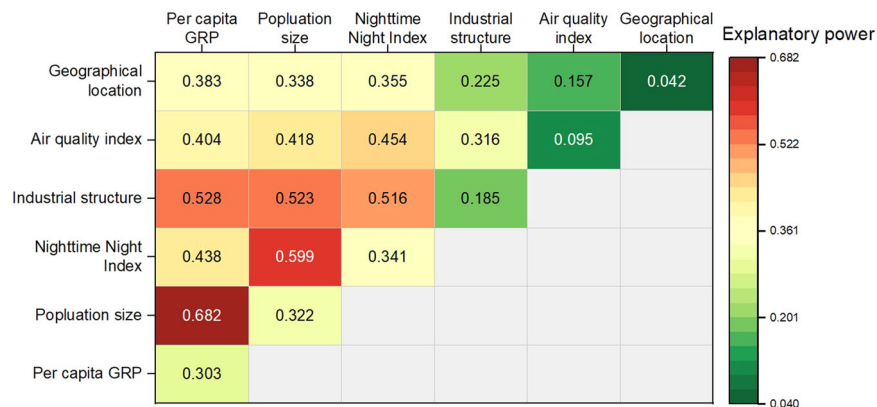


Fig. 5 Explanation strength for the actual attributes on the development level of pseudo-human settlements (PHS). The value of the diagonal cell is the result of the factor detector of the reality attribute, indicating to what extent the attribute explains the spatial differentiation of the pseudo-human settlements index (PHSI). Other cells report the result of the interaction detector, representing the explanatory power of the spatial differentiation of PHSI when the reality attributes of the row and column act together. The figure was developed by the authors using Origin Pro Version 2021.

Table 4 Proportions of functions of each PHS structure type.

Type	Numbers	F_{AV}	F_{CS}	F_{BR}	F_{FS}	F_{LE}	F_{TT}	F_{FM}	F_{NI}	F_{HL}	F_{MH}
I	191	12.37	10.11	8.94	9.15	12.61	6.33	9.57	8.32	11.12	11.48
II	324	7.06	11.54	7.31	11.32	17.90	4.90	9.80	8.73	9.30	12.14
III	389	7.70	11.46	7.91	12.12	13.97	5.31	10.56	7.49	9.21	14.26
IV	297	6.33	15.64	7.48	9.87	20.62	3.86	10.79	9.23	6.97	9.21
V	106	11.00	14.16	8.28	6.47	24.11	4.16	9.78	9.14	7.35	5.56
VI	409	5.88	14.68	8.92	11.20	14.04	4.52	10.23	9.34	8.12	13.08

Note: unit of function *F* is percentage (%).

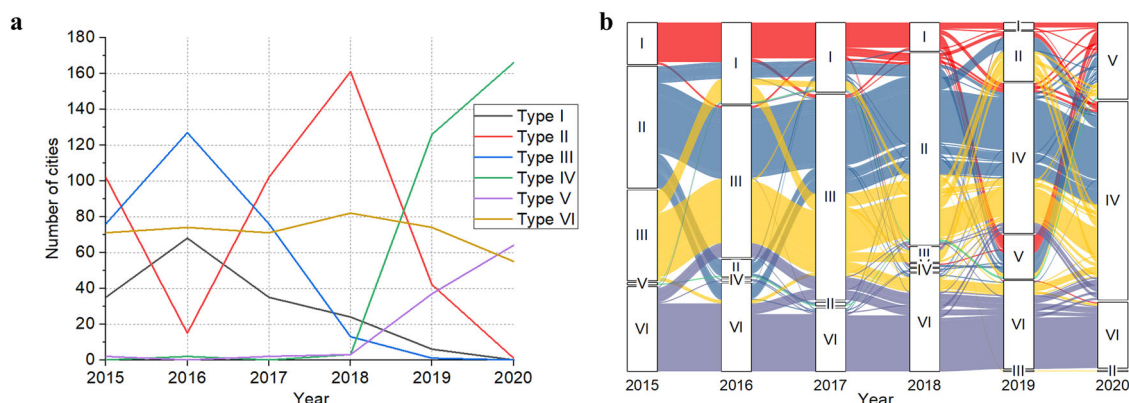


Fig. 6 Time evolution characteristics of the functional structure types of pseudo-human settlements (PHS). Subpanel **a** describes the annual changes in the number of cities represented by each functional structure type of PHS, and subpanel **b** describes the direction and extent of the circulation of each functional structure type of PHS with the year. The figure was developed by the authors using Origin Pro Version 2021.

period. Type VI had a long life cycle, and the number of cities steadily changed during the study period. Figure 6(b) shows the gradual changes in functional structure type. In 2015, the dominant functional structure types were Type I and III, with balanced characteristics. From 2017 to 2018, the dominant functional structure types were Types II and IV, with multi-functional polarization characteristics. From 2019 to 2020, the dominant types were IV and V, with complete polarization characteristics of learning and educational functions.

Discussion

Impact of urban reality policy on PHS development. Based on the spatial and temporal pattern of the PHSI, we found that municipalities, sub-provincial cities, provincial capitals, and some cities in the eastern coastal areas had PHS with a higher development level; these cities, without exception, had a good level of social and economic development. The results of the geographic detector directly confirmed that the social and economic development factors played a decisive role in the development level of PHS. The interaction between the urban population and social and economic factors had an explanatory power of >0.6 for the PHSI, which confirmed the qualitative analysis conclusions of Tian et al. (2021) and Xue et al. (2021). Among them, the population was an external factor in the development of PHS. By the end of 2020, the number of Internet users in China was 989 million, such that the Internet penetration rate reached 70.4% (Yin et al., 2022). The wide popularization of the Internet allowed the population to directly determine the search indices of cities. Social and economic factors are the internal driving factors for the development level of PHS. Cities with high social and economic levels often have more employment opportunities, improved social security, and educational advantages, which are the factors that attract the floating population to remain in the city (Wang and Shen, 2017).

We simultaneously found PHS with an unbalanced and insufficient development status in the examined Chinese cities. Municipalities, sub-provincial cities, provincial capitals, and coastal provinces have been given priority for development due to their geographical advantages, the reform and opening up, and the concept of “Rich former leads the latter, eventually developing together.” The introduction of substantial talent has promoted rapid improvements to the social economy, instating a virtuous circle for the development of PHS. However, the expected role of PHS in driving development only significantly exists in national urban agglomerations. In contrast, some remote Chinese provinces have formed the regional “polarization” and “contagion”

pattern characteristics for the PHS development level. With improvements to China’s comprehensive national strength, China will inevitably promote high-quality development strategies to realize the coordinated development of pseudo and real human settlements in specific regions and across the entire country. Previous research has shown that (Cao et al., 2021; Wang et al., 2021a), due to the implementation and promotion of strategies for developing the western region, revitalizing the northeast, and coordinating the development of Beijing, Tianjin, and Hebei, the coordination degree of development in various parts of China has been characterized by an upward trend. However, overall, considerable opportunities remain for development. To realize the development of real and PHS, we must understand the double-edged sword of spatial aggregation, fully develop the role of high-level areas, narrow the regional gap, and break down institutional obstacles hindering coordinated development through comprehensive reform.

Correlation between PHS functional structure and developmental level. Our results show that the functional structure of the PHS drastically changed over the study period, and it was influenced by multiple socioeconomic properties. As the development level of PHS is a comprehensive index influenced by the interactions of various urban real attributes, it can better reflect differences in urban development status than a single attribute. Therefore, we used the factor detector to detect the explanatory power of functional structure types on PHSI, and the results showed that $q = 0.52$, which was considerably higher than that of a single factor ($q = 0.255$), indicating that the functional structure of PHS is also the result of the interactions among multiple real attributes.

The intensification of polarization characteristics is an important development trend of PHS functional structure in the study period. Therefore, we further analyzed the correlation between the polarization degree of PHS functional structure and PHSI to explore the effect of the integrated properties on the type of functional structure. With PHSI as the independent variable and the degree of polarization as the dependent variable, a scatter plot is presented in Fig. 7.

Pearson correlation analysis revealed that the PHSI was not related to the polarization degree during 2015–2018, which showed that both low PHSI cities and high PHSI cities might have a relatively balanced functional structure of PHS. During 2019–2020, with the rapid increase in polarization degree of PHS functional structure in low PHSI cities, the polarization degree of functional structure and PHSI exhibited significant negative correlation

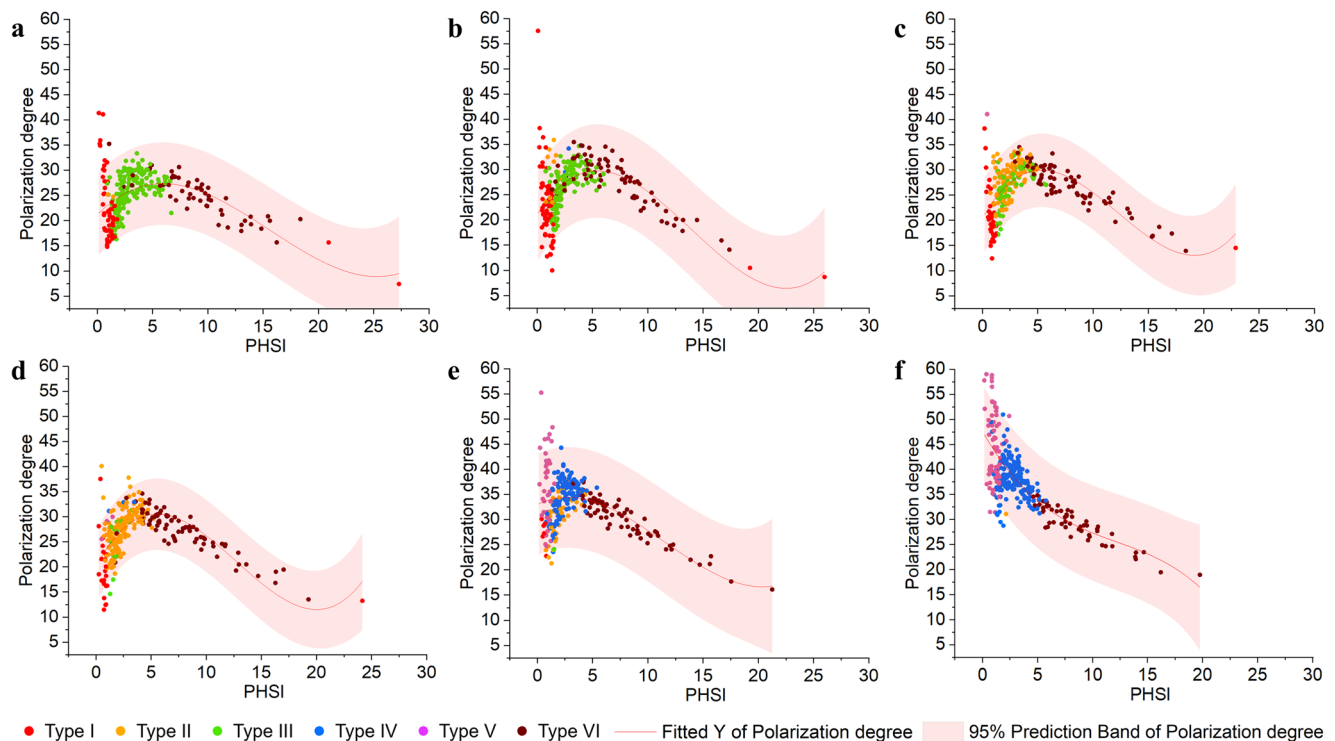


Fig. 7 Correlation between PHSI and polarization degree. Sub-panels a–f depict the cubic polynomial fit plots drawn from 2015–2020 with PHSI as the independent variable and the polarization degree as the dependent variable. The figure was developed by the authors using Origin Pro Version 2021.

characteristics ($P < 0.01$), and the correlation coefficients in 2019 and 2020 were -0.419 and -0.729 , respectively.

The spatial distribution of PHS functional structure types in 2020 is shown in Fig. 8. The functional structure types of cities with high PHS development levels remained stable and were maintained as Type VI. From the perspective of spatial distribution, provincial capital cities, and the core cities in the Yangtze River Delta urban agglomeration, the Pearl River Delta urban agglomeration and the Beijing–Tianjin–Hebei urban agglomeration are mostly Type VI. Conducive social and economic conditions have driven such cities to establish diversified subjective demand structures, such as education, medical care, and social interaction. Therefore, such cities should focus on the coordinated development of diversified urban functions in reality. Cities with a medium PHS development level have a relatively stable process of polarization of learning and education functions; such cities are widely distributed in the middle and east of China. Considering the functional structure of cluster centers, the preference of such cities for PHS learning and education functions increased from 13.97% in 2015 to 17.90% in 2018 to 20.62% in 2020. By contrast, the functional structure of PHS in cities with a low level of PHS development showed a change from equilibrium Type I to fully polarized Type V for learning and education. Considering the functional structure of cluster centers, the preference of such cities for learning and education functions increased suddenly from 12.61 to 24.11%. From the perspective of spatial distribution, such cities are distributed in Northeast China and Gansu Province. For cities with medium and low levels of PHS development, attention should be paid to the subjective requirements of residents for learning and educational functions. In particular, cities with low levels of PHS development should detect changes in the functional structure of PHS in a timely manner and make corresponding development strategy adjustments based on the subjective requirements of residents.

Applications of PHS. PHSI has the following three characteristics. It: (1) is a comprehensive index formed under the influence of various entity attributes. There is a close connection between PHS and the realistic attributes of a city, and it can largely depict the development level of the real human settlement environment of the city; (2) has a high temporal resolution. Statistical yearbooks typically have a time lag of 1–2 years. Therefore, PHSI has high temporal resolution and can achieve real-time evaluation of human settlements compared with the human settlement evaluation studies based on statistical yearbooks (Tian et al., 2016; Zhang and Fan, 2020; Yu et al., 2022); and (3) reflects the subjective preferences and requirements of urban residents for the living environment. Studies evaluating human settlements using statistical yearbooks and satellite remote sensing images (Chen et al., 2021b; Luo et al., 2021) are typically conducted with “land” as the core. PHS establishes a resident-oriented evaluation system based on the Baidu Index, which can help urban planners understand the subjective development requirements and preferences of urban residents.

Therefore, PHS can be applied to regional and urban planning in the following ways: (1) It can provide a monitoring and feedback system for the development of human settlements in cities, nations, and regions. According to the high temporal resolution of PHS, a PHS database with real-time updates and annual data can be established to keep up with time-series changes in urban development levels. Furthermore, using exploratory spatiotemporal analysis methods, regional development status can be detected based on local PHS patterns. The impact of regional development strategies on PHS pattern can be analyzed, and these strategies can be adjusted according to such impacts to achieve more coordinated development. (2) PHS can help urban planners understand current status and shifting trends of subjective preferences of residents for urban functions. The functional structure of PHS reflects the preferences of residents for human settlement environment functions. Therefore, urban

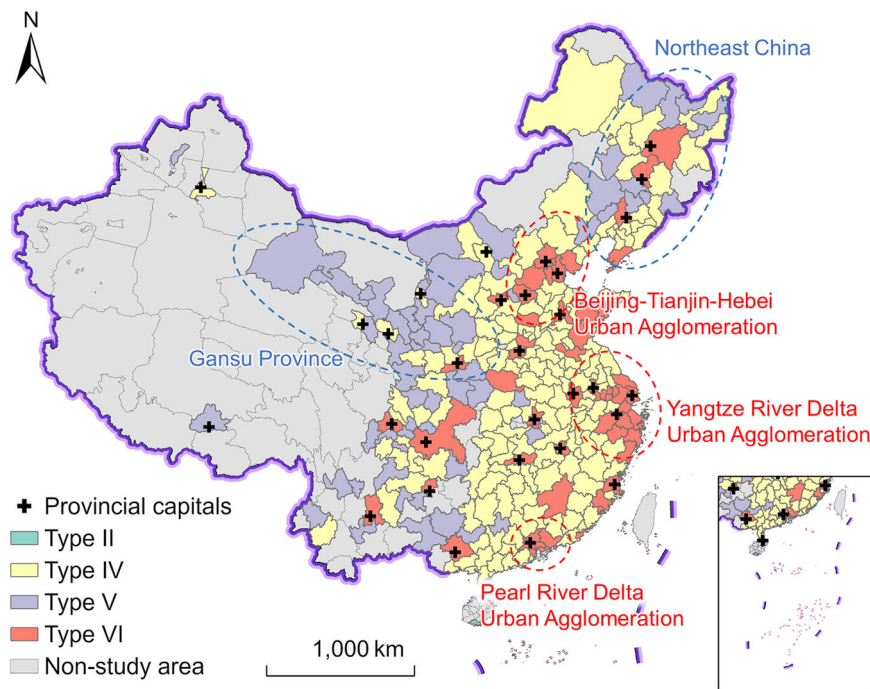


Fig. 8 The spatial distribution of PHS functional structure types. The figure shows the main characteristics of PHS functional structure types in different areas. The map was developed by the authors using ArcGIS Pro Version 3.0.2, and the background map was issued by the Ministry of Natural Resources of the People's Republic of China (drawing review No. GS (2020) 4619), as shown in the website: <http://bzdt.ch.mnr.gov.cn/>.

policy makers can reasonably adjust urban policy frameworks according to the structural characteristics of local PHS to adapt to the subjective preferences of urban residents. Urban decision-makers should pay attention to the life cycle changes of the PHS functional structure, predict changes in the preferences of future urban residents, and adjust development strategies.

Limitations. As the functional structure of Chinese urban PHS is still in a relatively dynamic stage, the classification results of the K-Means clustering method may greatly vary with study period. However, the method can achieve better results in classifying urban reality attributes primarily because the development of urban reality attributes is relatively stable, and the variation range is small. Therefore, the classification of the functional structure of PHS relies on longer time observations to establish more stable results.

Conclusions

The PHS of a city reflect the subjective wishes of residents. China is undergoing high-speed urbanization and advocates improving the quality of its cities. It is particularly important to test the achievements of urbanization in the country through an analysis of PHS. Existing scientific research on human settlements primarily analyzes the relationship between the overall development level of PHS and the real human settlement environment, thereby ignoring the internal interaction and evolution process of self-constituting factors. It especially lacks a quantitative analysis of the driving factors and structural evolution of PHS.

An in-depth characterization of PHS can help clarify the realistic development needs and direction of cities. Through multi-source big data and quantitative analysis methods, this study found that the development level of PHS in Chinese cities has obvious hierarchical characteristics and a “high in east, low in west” spatial pattern. In the central cities of national urban agglomerations and the prefecture-level cities of remote

provinces, hot and cold development spots have been formed, respectively. Nighttime light index, per capital GRP, and population size are undoubtedly major driving factors of PHS development level; furthermore, the interaction between per capita GRP and population size yielded the highest explanatory power. In addition, the structure of PHS was systematically analyzed in this study. The preference of urban residents for PHS function is also primarily affected by the nighttime light index, per capital GRP, and population size of the city, and its functional structure has an evident life cycle.

From the perspective of regional development, this study concludes that national development policy leads to a spatial-temporal imbalance in social and economic factors, affecting the development level of PHS. China's national-level urban agglomerations are in the process of hierarchical diffusion from high-level centers to low-level centers, while the provincial capitals of the remote provinces have certain first polarization characteristics. China is adjusting this unbalanced and inadequate development trend by means of policy support, and has made certain achievements in coordinated development. To further promote coordinated development, it is necessary to break down cooperation barriers through deepening reform. It is also important to guard against the possible paralyzing effect of the “first city” on the development of smaller cities in remote provinces.

From the perspective of urban development, this study concludes that urban planners and decision-makers should understand the development characteristics of the functional structure of PHS, adapt to the current preference characteristics of PHS, predict new growth points in PHS, explore the development blueprint of PHS with local characteristics, and discover the endogenous demand for human settlement development owing to the subjective preferences of residents.

Understanding the development level and PHS functional structural changes is essential for sustainable regional and urban development. The analysis of the realistic characteristics and driving factors of PHS in Chinese cities is not only helpful

to understanding its own endogenous development needs but also has certain reference value for other countries, especially developing countries, that are determining the development path of PHS.

Data availability

All data sources generated or analyzed during this study have been noted in this published article. The datasets analyzed during the current study are available from the corresponding author on reasonable request.

Received: 21 April 2022; Accepted: 31 January 2023;

Published online: 09 February 2023

References

- Anjali A, Kumar S (2021) Research of Urban Area Structure Mixing and Human Activity Space from Citizen Activity Perspectives. In: 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS). pp 1484–1489
- Barnaghi P, Ghaffari P, Breslin JG (2016) Opinion mining and sentiment polarity on twitter and correlation between events and sentiment. In: 2016 IEEE Second International Conference on Big Data Computing Service and Applications (BigDataService). pp 52–57
- Bivand RS, Wong DWS (2018) Comparing implementations of global and local indicators of spatial association. *Test* 27:716–748. <https://doi.org/10.1007/s11749-018-0599-x>
- Cao J, Zhang Y, Wei T, Sun H (2021) Temporal-spatial evolution and influencing factors of coordinated development of the population, resources, economy and environment (PREE) system: evidence from 31 provinces in China. *Int J Env Res Pub He* 18:13049. <https://doi.org/10.3390/ijerph182413049>
- Chen B, Ma W, Pan Y et al. (2021a) PM2.5 exposure and anxiety in China: evidence from the prefectures. *BMC Public Health* 21:429. <https://doi.org/10.1186/s12889-021-10471-y>
- Chen J, Gao M, Cheng S et al. (2022) Global 1 km × 1 km gridded revised real gross domestic product and electricity consumption during 1992–2019 based on calibrated nighttime light data. *Sci Data* 9:1–14. <https://doi.org/10.1038/s41597-022-01322-5>
- Chen S, Huang Y, Huang W (2016) Big Data Analytics on Aviation Social Media: The Case of China Southern Airlines on Sina Weibo. In: 2016 IEEE Second International Conference on Big Data Computing Service and Applications (BigDataService). pp 152–155
- Chen T, Hui ECM, Wu J et al. (2019) Identifying urban spatial structure and urban vibrancy in highly dense cities using georeferenced social media data. *Habitat Int* 89:102005. <https://doi.org/10.1016/j.habitatint.2019.102005>
- Chen W, Zhu K, Wu Q, et al. (2021b) Adaptability evaluation of human settlements in Chengdu based on 3S technology. *Environ Sci Pollut Res* 1–12. <https://doi.org/10.1007/s11356-021-15998-6>
- Chen Z, Yu B, Song W et al. (2017) A new approach for detecting urban centers and their spatial structure with nighttime light remote sensing. *IEEE Trans Geosci Remote Sens* 55:6305–6319. <https://doi.org/10.1109/TGRS.2017.2725917>
- Delmelle EC (2015) Five decades of neighborhood classifications and their transitions: a comparison of four US cities, 1970–2010. *Appl Geogr* 57:1–11. <https://doi.org/10.1016/j.apgeog.2014.12.002>
- Denissen JJA, Neumann L, van Zalk M (2010) How the internet is changing the implementation of traditional research methods, people's daily lives, and the way in which developmental scientists conduct research. *Int J Behav Dev* 34:564–575. <https://doi.org/10.1177/0165025410383746>
- Dong X, Mavroudis D, Calabrese F, Frossard P (2015) Multiscale event detection in social media. *Data Min Knowl Disc* 29:1374–1405. <https://doi.org/10.1007/s10618-015-0421-2>
- Duque JC, Lozano-Gracia N, Patino JE et al. (2019) Spatiotemporal dynamics of urban growth in Latin American cities: an analysis using nighttime light imagery. *Landsc Urban Planning* 191:103640. <https://doi.org/10.1016/j.landurbplan.2019.103640>
- Fast E, Chen B, Bernstein MS (2016) Empath: Understanding Topic Signals in Large-Scale Text. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. ACM, San Jose California USA, pp 4647–4657
- Feng Z, Peng J, Wu J (2020) Using DMSP/OLS nighttime light data and K-means method to identify urban-rural fringe of megacities. *Habitat Int* 103:102227. <https://doi.org/10.1016/j.habitatint.2020.102227>
- Gavin NT (2018) Media definitely do matter: Brexit, immigration, climate change and beyond. *Brit J Polit Int Rel* 20:827–845. <https://doi.org/10.1177/1369148118799260>
- Ghani NA, Hamid S, Hashem IAT, Ahmed E (2019) Social media big data analytics: a survey. *Comput Hum Behav* 101:417–428. <https://doi.org/10.1016/j.chb.2018.08.039>
- Guan W, Gao H, Yang M et al. (2014) Analyzing user behavior of the micro-blogging website Sina Weibo during hot social events. *Physica A* 395:340–351. <https://doi.org/10.1016/j.physa.2013.09.059>
- Guo H, Zhang W, Du H et al. (2022) Understanding China's urban system evolution from web search index data. *EPJ Data Sci* 11:20. <https://doi.org/10.1140/epjds/s13688-022-00332-y>
- Hu Q, Bai G, Wang S, Ai M (2019) Extraction and monitoring approach of dynamic urban commercial area using check-in data from Weibo. *Sustain Cities Soc* 45:508–521. <https://doi.org/10.1016/j.scs.2018.11.039>
- Huang Y, Li J, Wu G, Fei T (2020) Quantifying the bias in place emotion extracted from photos on social networking sites: a case study on a university campus. *Cities* 102:102719. <https://doi.org/10.1016/j.cities.2020.102719>
- Jendryke M, Balz T, McClure SC, Liao M (2017) Putting people in the picture: combining big location-based social media data and remote sensing imagery for enhanced contextual urban information in Shanghai. *Comput Environ Urban* 62:99–112. <https://doi.org/10.1016/j.compenvurbysys.2016.10.004>
- Jin K, Sun S, Li H, Zhang F (2022) A novel multi-modal analysis model with Baidu Search Index for subway passenger flow forecasting. *Eng Appl Artif Intel* 107:104518. <https://doi.org/10.1016/j.engappai.2021.104518>
- Kim S, Lim H-W, Chung S-Y (2022) How South Korean Internet users experienced the impacts of the COVID-19 pandemic: discourse on Instagram. *Humanit Soc Sci Commun* 9:75. <https://doi.org/10.1057/s41599-022-01087-7>
- Li W, Yang G, Li X (2021) Correlation between PM2.5 pollution and its public concern in China: evidence from Baidu Index. *J Clean Prod* 293:126091. <https://doi.org/10.1016/j.jclepro.2021.126091>
- Li X, Li H, Li J, Zhang Y (2015) The geographical study of human settlements: from positivism to humanism. Science Press, Beijing
- Li Y (2020) Towards concentration and decentralization: The evolution of urban spatial structure of Chinese cities, 2001–2016. *Comput Environ Urban Syst* 80:101425. <https://doi.org/10.1016/j.compenvurbysys.2019.101425>
- Lippmann W (1997) Public Opinion. Free Press
- Liu S, Shi K, Wu Y, Chang Z (2021) Remotely sensed nighttime lights reveal China's urbanization process restricted by haze pollution. *Build Environ* 206:108350. <https://doi.org/10.1016/j.buildenv.2021.108350>
- Liu W, Luo X, Gong Z et al. (2016) Discovering the core semantics of event from social media. *Future Gener Comp Sy* 64:175–185. <https://doi.org/10.1016/j.future.2015.11.023>
- Liu Y, Liao W (2021) Spatial characteristics of the tourism flows in China: a study based on the Baidu Index. *ISPRS Int J Geo-Inf* 10:378. <https://doi.org/10.3390/ijgi10060378>
- Luo X, Yang J, Sun W, He B (2021) Suitability of human settlements in mountainous areas from the perspective of ventilation: a case study of the main urban area of Chongqing. *J Clean Prod* 310:127467. <https://doi.org/10.1016/j.jclepro.2021.127467>
- Meel P, Vishwakarma DK (2020) Fake news, rumor, information pollution in social media and web: a contemporary survey of state-of-the-arts, challenges and opportunities. *Expert Syst Appl* 153:112986. <https://doi.org/10.1016/j.eswa.2019.112986>
- Rousseeuw P (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math* 20:53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- Schonfeld J, Qian E, Sinn J et al. (2021) Debates about vaccines and climate change on social media networks: a study in contrasts. *Humanit Soc Sci Commun* 8:1–10. <https://doi.org/10.1057/s41599-021-00977-6>
- Shen J, Yang X (2014) Analyzing urban competitiveness changes in major Chinese cities 1995–2008. *Appl Spat Anal Pol* 7:361–379. <https://doi.org/10.1007/s12061-014-9114-2>
- Surya B, Hadijah H, Suriani S et al. (2020) Spatial transformation of a new city in 2006–2020: perspectives on the spatial dynamics, environmental quality degradation, and socio-economic sustainability of local communities in Makassar City, Indonesia. *Land* 9:324
- Tian S, Li X, Li H et al. (2016) Initial evaluation of provincial-level environmental risks from the perspective of human settlements. *Sustainability* 8:1259. <https://doi.org/10.3390/su8121259>
- Tian S, Li X, Yang J, et al. (2022a) Spatiotemporal evolution of pseudo human settlements: case study of 36 cities in the three provinces of Northeast China from 2011 to 2018. *Environ Dev Sustain*. <https://doi.org/10.1007/s10668-022-02120-0>
- Tian S, Li X, Yang J et al. (2021) Spatio-temporal coupling coordination and driving mechanism of urban pseudo and reality human settlements in the three provinces of Northeast China. *Acta Geogr Sinica* 76:781–798
- Tian S, Yang B, Liu Z et al. (2022b) Coupling coordination of urban pseudo and reality human settlements. *Land* 11:414. <https://doi.org/10.3390/land11030414>

- Tobler WR (1970) A computer movie simulating urban growth in the Detroit region. *Econ Geogr* 46:234. <https://doi.org/10.2307/143141>
- Townsend AC, Bruce DA (2010) The use of night-time lights satellite imagery as a measure of Australia's regional electricity consumption and population distribution. *Int J Remote Sens* 31:4459–4480. <https://doi.org/10.1080/0143160903261005>
- Tsou M-H (2015) Research challenges and opportunities in mapping social media and Big Data. *Cartogr Geogr Inf Sc* 42:70–74. <https://doi.org/10.1080/15230406.2015.1059251>
- Wang B, Loo BPY (2019) The hierarchy of cities in Internet news media and Internet search: some insights from China. *Cities* 84:121–133. <https://doi.org/10.1016/j.cities.2018.07.013>
- Wang J, Han Q, Du Y (2021a) Coordinated development of the economy, society and environment in urban China: a case study of 285 cities. *Environ Dev Sustain*. <https://doi.org/10.1007/s10668-021-01975-z>
- Wang J-F, Li X-H, Christakos G et al. (2010) Geographical detectors-based health risk assessment and its application in the neural tube defects study of the Heshun region, China. *Int J Geogr Inf Sci* 24:107–127. <https://doi.org/10.1080/13658810802443457>
- Wang J-F, Zhang T-L, Fu B-J (2016) A measure of spatial stratified heterogeneity. *Ecol Indic* 67:250–256. <https://doi.org/10.1016/j.ecolind.2016.02.052>
- Wang L, Shen J (2017) Comparative analysis of urban competitiveness in the Yangtze River Delta and Pearl River Delta Regions of China, 2000–2010. *Appl Spat Anal Polic* 10:401–419. <https://doi.org/10.1007/s12061-016-9210-6>
- Wang T, Sun F (2022) Global gridded GDP data set consistent with the shared socioeconomic pathways. *Sci Data* 9:221. <https://doi.org/10.1038/s41597-022-01300-x>
- Wang Z, Zhu Z, Xu M, Qureshi S (2021b) Fine-grained assessment of greenspace satisfaction at regional scale using content analysis of social media and machine learning. *Sci Total Environ* 776:145908. <https://doi.org/10.1016/j.scitotenv.2021.145908>
- Wu L (2001) Introduction to sciences of human settlements. China Architecture and Building Press, Beijing
- Xu L, Du H, Zhang X (2021) Driving forces of carbon dioxide emissions in China's cities: an empirical analysis based on the geodetector method. *J Clean Prod* 287:125169. <https://doi.org/10.1016/j.jclepro.2020.125169>
- Xue Q, Yang X, Wu F (2021) A two-stage system analysis of real and pseudo urban human settlements in China. *J Clean Prod* 293:126272. <https://doi.org/10.1016/j.jclepro.2021.126272>
- Yan L, Duarte F, Wang D et al. (2019) Exploring the effect of air pollution on social activity in China using geotagged social media check-in data. *Cities* 91:116–125. <https://doi.org/10.1016/j.cities.2018.11.011>
- Yin Z, Jiang X, Lin S, Liu J (2022) The impact of online education on carbon emissions in the context of the COVID-19 pandemic—taking Chinese universities as examples. *Appl Energ* 314:118875. <https://doi.org/10.1016/j.apenergy.2022.118875>
- Yu B, Shi K, Hu Y et al. (2015) Poverty evaluation using NPP-VIIRS nighttime light composite data at the county level in China. *IEEE J Sel Top Appl Earth Obs Remote Sens* 8:1217–1229. <https://doi.org/10.1109/JSTARS.2015.2399416>
- Yu W, Yang J, Sun D, et al. (2022) Spatial-temporal patterns of network structure of human settlements competitiveness in resource-based urban agglomerations. *Front Environ Sci* 10. <https://doi.org/10.3389/fenvs.2022.893876>
- Zeng C, Zhou Y, Wang S et al. (2011) Population spatialization in China based on night-time imagery and land use data. *Int J Remote Sens* 32:9599–9620. <https://doi.org/10.1080/01431161.2011.569581>
- Zhang X, Ghorbani AA (2020) An overview of online fake news: Characterization, detection, and discussion. *Inform Process Manag* 57:102025. <https://doi.org/10.1016/j.ipm.2019.03.004>
- Zhang Y, Fan Q (2020) The application of the fuzzy analytic hierarchy process in the assessment and improvement of the human settlement environment. *Sustainability* 12:1563. <https://doi.org/10.3390/su12041563>
- Zhang Z, Tang W (2016) Analysis of spatial patterns of public attention on housing prices in Chinese cities: a web search engine approach. *Appl Geogr* 70:68–81. <https://doi.org/10.1016/j.apgeog.2016.03.004>
- Zhao R, Zhan L, Yao M, Yang L (2020a) A geographically weighted regression model augmented by Geodetector analysis and principal component analysis for the spatial distribution of PM_{2.5}. *Sustain Cities Soc* 56:102106. <https://doi.org/10.1016/j.scs.2020.102106>
- Zhao Y, Li J, Wang J-E (2021) Analysis and prediction of “AI + Education” attention based on Baidu Index—taking Guizhou province as an example. *Future Internet* 13:120. <https://doi.org/10.3390/fi13050120>
- Zhao Y, Li R, Wu M (2020b) Correlation Studies between land cover change and Baidu Index: a case study of Hubei Province. *ISPRS Int J Geo-Inf* 9:232. <https://doi.org/10.3390/ijgi9040232>
- Zheng Q, Weng Q, Wang K (2021) Characterizing urban land changes of 30 global megacities using nighttime light time series stacks. *ISPRS J Photogramm Remote Sens* 173:10–23. <https://doi.org/10.1016/j.isprsjprs.2021.01.002>
- Zheng S, Wang J, Sun C et al. (2019) Air pollution lowers Chinese urbanites' expressed happiness on social media. *Nat Hum Behav* 3:237–243. <https://doi.org/10.1038/s41562-018-0521-2>
- Zheng Y, He Y, Zhou Q, Wang H (2022) Quantitative evaluation of urban expansion using NPP-VIIRS nighttime light and landsat spectral data. *Sustain Cities Soc* 76:103338. <https://doi.org/10.1016/j.scs.2021.103338>
- Zhu X, Gao M, Zhang R, Zhang B (2021) Quantifying emotional differences in urban green spaces extracted from photos on social networking sites: a study of 34 parks in three cities in northern China. *Urban For Urban Green* 62:127133. <https://doi.org/10.1016/j.ufug.2021.127133>

Acknowledgements

This research study was supported by the National Natural Science Foundation of China (grant numbers 41771178, 42030409, and 41671151), the Fundamental Research Funds for the Central Universities (grant number N2111003), Basic Scientific Research Project (Key Project) of the Education Department of Liaoning Province (grant number LJKZ0964), Natural Science Foundation of Guizhou Province (grant number (2019) 1150).

Competing interests

The authors declare no competing interests.

Ethical approval

This article does not contain any studies with human participants performed by any of the authors. Ethical approval was not relevant.

Informed consent

This article does not contain any studies with human participants performed by any of the authors. Informed consent was not relevant.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1057/s41599-023-01551-y>.

Correspondence and requests for materials should be addressed to Jun Yang.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023, corrected publication 2023