

ARTICLE OPEN



Underload city conceptual approach extending ghost city studies

Xiuyuan Zhang¹, Shihong Du¹✉, Hannes Taubenböck^{2,3}, Yi-Chen Wang⁴, Shouhang Du¹, Bo Liu¹ and Yuning Feng¹

Global population growth and land development are highly imbalanced, marked by 43% of population increase but 150% of built-up area expansion from 1990 to 2018. This results in the widely concerned ghost city phenomenon and runs against the sustainable development goals. Existing studies identify ghost cities by population densities, but ignore the spatial heterogeneity of land carrying capacities (LCC). Accordingly, this study proposes a general concept termed underload city to define cities carrying fewer people and lower economic strength than their LCC. The underload city essentially describes imbalanced human-land relationship and is understood in a broader context than the usually applied ghost city. In this study, very high-resolution satellite images are analyzed to obtain land functional structures, and further combined with population and GDP data to derive LCC. We empirically identify eight underload cities among 81 major Chinese cities, differing from previous findings of ghost cities. Accordingly, the proposed underload city considers heterogeneous human-land relationships when assessing city loads and contributes to sustainable city developments.

npj Urban Sustainability (2022)2:15; <https://doi.org/10.1038/s42949-022-00057-x>

INTRODUCTION

The United Nation sustainable development goals (SDGs; UNDP, 2015)¹ closely concern sustainable cities and communities (SDG 11), requiring inclusive and sustainable urbanization with benign human-land relationships². Particularly the SDG 11.3.1 indicator that compares population growth to land consumption is addressed; thus, balanced human-land relationship plays an important role in sustainable city development. However, governments worldwide have difficulties in balancing urban land expansion and population growth rates, as land development and population growth are not always synchronized, especially during the rapid global urbanization^{3,4}; additionally, global urban agglomeration development leads to population concentrations in metropolises but population outflows in other cities⁵. Both can cause imbalanced human-land supply-demand and result in overload or underload of land resources, which deviates from SDG 11. Prior work on urbanization related consequences has mainly focused on land overload issues, such as overlarge population, heat island, traffic congestion, or pollution⁶, but scant attention has been paid to land underload particularly in relation to ghost cities^{7–9}.

Different from the traditional image of ghost towns that are abandoned because of economic or natural resource hardship¹⁰, ghost cities/neighborhoods are generally understood as vacant areas that were newly built but uninhabited^{11–13}. A number of studies have identified ghost cities based on population-related measures, using nightlight satellite images^{14,15}, mobile phone positioning¹⁶, social media vitality⁷, or point-of-interest data¹⁷. The conceptual approaches towards ghost cities in these studies mostly consider population density, but ignore the land's spatial heterogeneity, and thus the specific land carrying capacities (LCC) of diverse cities for people and economy remain undifferentiated^{18–22}. In fact, cities contain distinct functional zones with varying spatial structures, i.e., heterogeneous land functional

structures (LFS), that support variant human socioeconomic activities and have different LCC for people and economy^{23–25}. Employing a unified threshold to extract ghost cities hence abstracts the varying LCC across different cities^{21,26}. Besides these issues, ghost city studies have encountered another conceptual bottleneck. Wade Shepard who proposed the widely recognized definition of ghost cities suggested that there could be no ghost city but only temporary ghost stages in China²⁷. From the perspective, ghost cities would disappear with urbanization development and population influx⁵; thus, ghost city might lose significance and partly lose popularity⁷. Whether or not this will happen, we believe a more general concept rather than the extremes of ghost cities should be developed and relevant in the long term.

To account for the unknown effects raised above, we extend the ghost cities and propose a concept of underload cities. These underload cities are defined to carry fewer people and lower economic strength than their LCC which are featured by LFS. There are three reasons for proposing the concept: 1) underload cities as a synthetical geographic concept consider not only population, but also economy and LCC, which meet the requirement of SDG 11 that highlights human-land relationship² and are different from ghost cities that solely measure population density; 2) underload cities as an imbalanced phenomenon always exist in the process of land construction and human development, which can be more common and last even longer than ghost cities²⁷, and thus have more profound impact on sustainable city development; and 3) underload cities are the opposite to overload cities which have been widely studied and represent another imbalanced phenomenon in urbanization^{5,6}; thus, investigating underload cities can complement existing researches of overload cities. Based on this concept, underload cities essentially represent imbalanced human-land relationships¹⁷, and it can be imperative to identify underload cities for re-integration of land resources,

¹Institute of Remote Sensing and GIS, Peking University, Beijing 100871, China. ²German Remote Sensing Data Center, German Aerospace Center, Wessling 82234, Germany.

³Institute for Geography and Geology, Julius-Maximilians-Universität Würzburg, Würzburg 97074, Germany. ⁴Department of Geography, National University of Singapore, Singapore 117570, Singapore. ✉email: shdu@pku.edu.cn

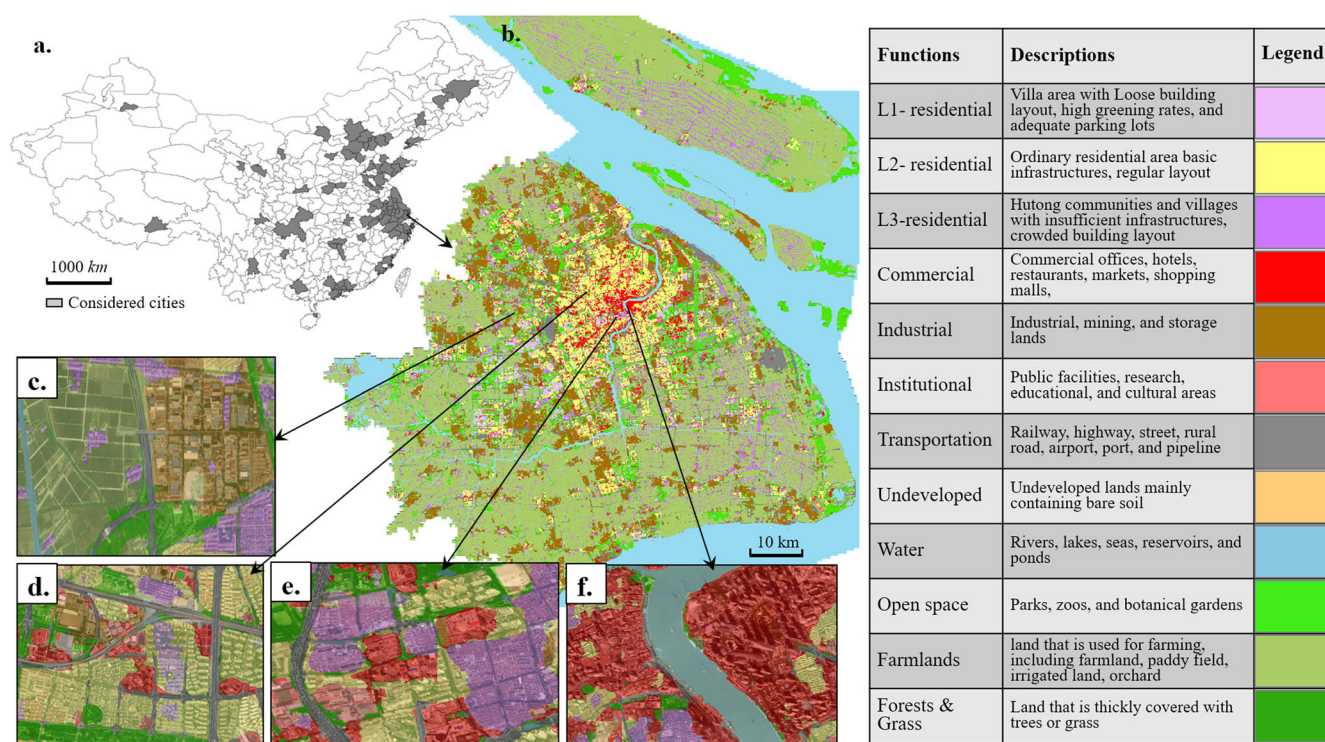


Fig. 1 Land functional zone maps of 81 major cities across China. **a** Distributions of the considered 81 cities whose names and detailed information are provided in the Supplementary Fig. 6. **b** Land functional zone map exemplified for Shanghai. The legend illustrates the 12 functional classes derived. **c–f**, Four selected regions in Shanghai to illustrate the land functional zones in details, where the maps are set at 60% transparency and overlapped with satellite images.

coordination of human-land relationship, and promotion of sustainable city development²⁸.

However, two methodological research gaps need to be bridged for studying underload cities. On the one hand, LFS need to be characterized. LFS represent the land's spatial heterogeneity and refer to land functions and their spatial structures. These are the basic units where people undertake different socio-economic activities, and thus are fundamental for measuring LCC²⁶. Three issues should be considered for extracting LFS: 1) high spatial resolution is needed to enable detailed representation of LFS^{28,29}, but current uses of low and medium-resolution land cover/use data are not sufficient for this purpose; 2) considering land functions' diverse representations and abstract categories, traditional satellite image classification methods cannot recognize land functions accurately²⁶; and 3) considering land structures' spatial complexity and heterogeneity, existing image textures and landscape matrices cannot measure land structures comprehensively and robustly²⁹. To address these challenges, this study derives very high-resolution land functional zone maps, based on which we quantify LFS and cluster different types to express LFS robustly.

On the other hand, LCC needs to be evaluated. LCC refers to the population and economy that a city can carry considering its LFS. It measures the equilibrium state of human-land relationship and thus serves as a benchmark for identifying underload cities. Existing LCC evaluation methods are limited, because they consider either local land functions^{30,31}, or mean population and GDP as the standard¹⁷, but few consider both^{32–34}. To address the limitation, this study presents an evaluation method for LCC, measuring the correlation between LFS and population/GDP in a relational system across cities for a better benchmarking. Based on this technique, we identify China's underload cities by comparing their actual populations and GDP to the estimated LCC.

The study will resolve the two methodological challenges raised above and identify underload cities from 81 major cities in China (Fig. 1a), including all province-level municipalities, provincial capitals, and top 50 cities in GDP, covering a total area of 983,215 km². First, we map land functional zones using very high-resolution (VHR) satellite images. Second, we characterize and cluster their LFS. Third, we evaluate LCC by combining LFS and demographic/economic data. And finally, we identify underload cities, discuss their general patterns, and compare the results with ghost cities and the SDG 11.3.1 indicator, respectively.

RESULTS

Mapping VHR land functional zones

We use VHR satellite images at 2.4 m resolution fused by multiple sensors (see Data in Methods) and multilevel semantic segmentation (see Methods) to map land functional zones. According to the codes of urban land-use classification and planning published by the Ministry of Housing and Urban-rural Development of China, we divided the land of 81 cities into 12 nonoverlapping categories of land functional zones (three residential classes, commercial, industrial, institutional, transportation, as well as undeveloped, water, open space, farmlands, and forest and grass), and generate very-high-resolution land functional zone maps (see Fig. 1 and Supplementary Fig. 7). The overall mapping accuracy is 85.0%, and a Kappa index is measured at 0.82, which are evaluated based on 25,419 test samples (see Supplementary Method). Compared to existing land cover/use data, the generated land functional zone map has two advantages. First, the map has relatively high resolution so that functional zone boundaries and categories can be more accurately represented. Second, the functional categories are closely related to socioeconomic activities, i.e., living, working, etc., which can be more applicable to human-land relationship analysis and LCC evaluation.

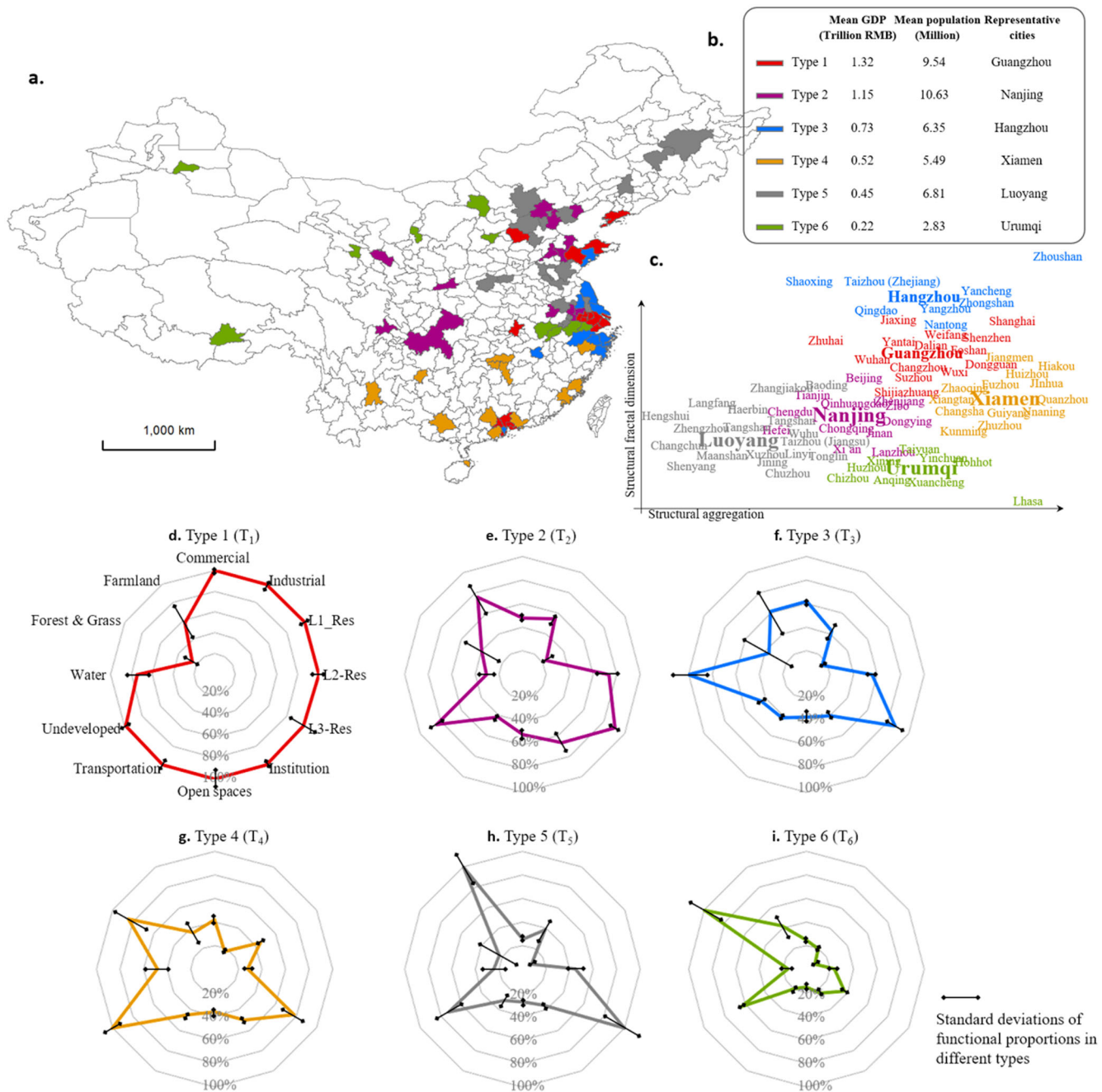


Fig. 2 Six types of cities clustered based on LFS. a Spatial distributions of the six types. **b** Mean population, GDP, and representative cities of the six types. **c** Cities' distributions in the LFS feature space, with structural aggregation as the x-axis and structural fractal dimension as the y-axis to represent the 2D feature space, because of their highest information entropy. A larger font size is used to highlight the six representative cities in the feature space. **d–i** Normalized land functional proportions for the six types. The normalized proportion of a functional category in a specific type refers to its ratio to the largest proportion of the category among all types, and their standard deviations are labeled.

Characterizing and clustering land functional structures

For quantitative representations of heterogeneous land functional structures (LFS), we measure LFS of each city by 64 indices (including functional, structural, and thematic ones) which can be derived from the land functional zone maps (see Methods). These indices are defined to characterize land functional services, spatial structures, and socioeconomic attributes. This is done in reference to existing studies as they find land functional and structural indices can directly influence population distribution and economic strength^{26,30,31}. In this study, we also presume that the

socio-economic attributes, e.g., living environments and industrial types²⁹, influence or even define LCC for population and economy. Then, we cluster LFS types at city level using these indices to robustly express land functions and structures. As the result, we group 81 cities into six types (T₁–T₆ in Fig. 2) using the density peaks clustering algorithm (DPCA), where the type number of 6 is detected by the DPCA considering samples' densities and distances in feature spaces (see Methods and Supplementary Fig. 3).

These six types differ significantly in land functions and structures (see Supplementary Table 2): T₁ Comprehensively-developed cities

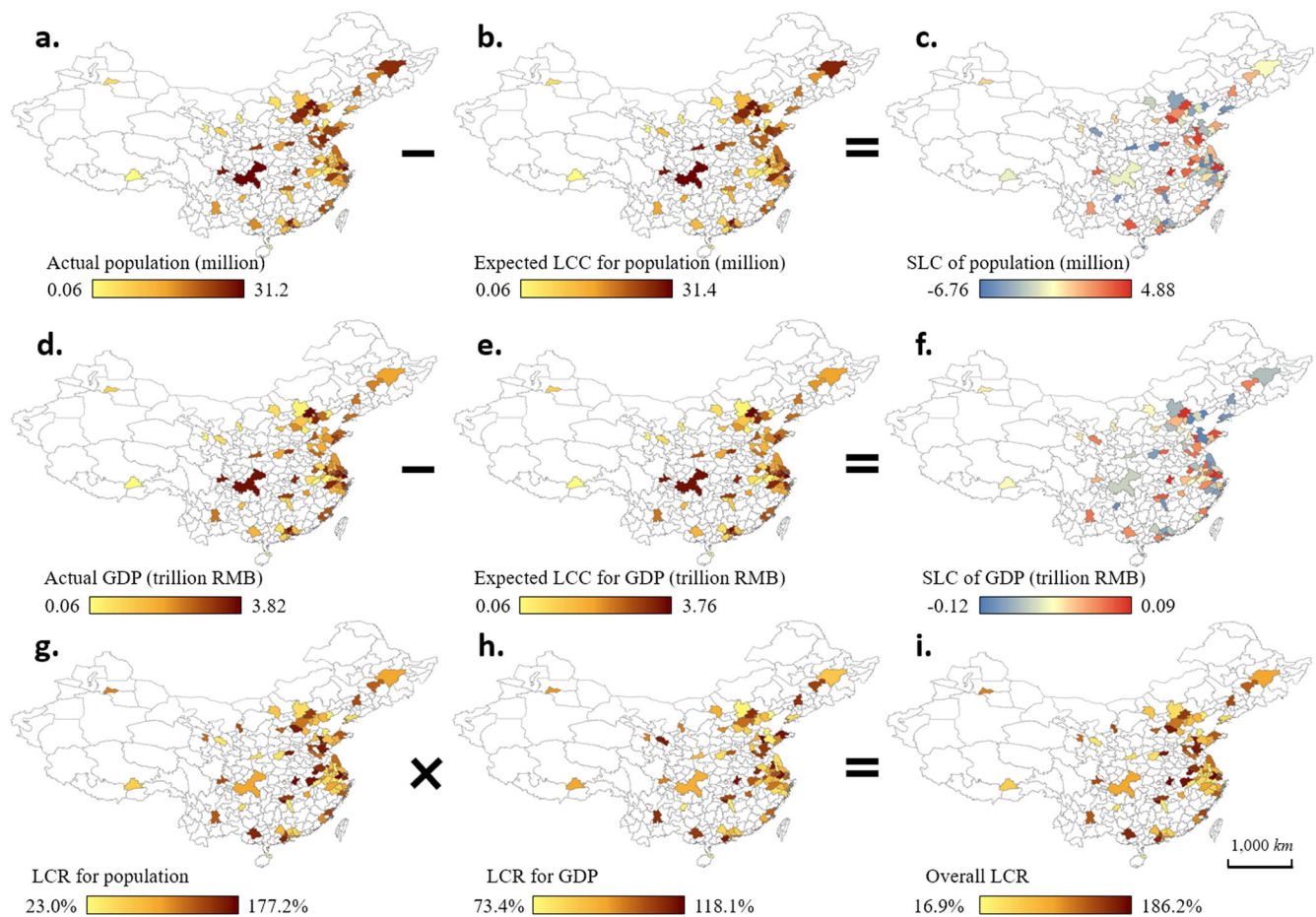


Fig. 3 Land carrying capacity (LCC), and the computations for shortage in land carrying (SLC) and land carrying rate (LCR) for 81 major cities. **a–c** the actual population of the cities, and their expected LCC for and SLC of populations. **d–f** the actual GDP of the cities, and their expected LCC for and SLC of GDP. **g, h** LCR for population and GDP of the cities. **i** overall LCR of the cities where dark brown represents the overload (high carrying rate), and light yellow the underload (low carrying rate).

have the largest proportions for commercial, industrial, L1-residential, L2-residential, open spaces, institutional, and transportation lands. These cities are mainly distributed in developed areas, e.g., the Yangtze river delta, the Pearl river delta, and the Shandong peninsula megalopolis. T_2 Administrative-centered cities feature higher proportions of institutional and residential lands. They mainly are province-level municipalities and provincial capitals. T_3 Coastal cities have the smallest undeveloped lands and are mainly distributed in the eastern coastal areas. T_4 Cities with development potentials have the smallest farmland proportion but the highest proportion of undeveloped lands, indicating large developmental potentials. They are all distributed in southern China. T_5 Rural-agricultural cities have the largest proportions of the categories of L3-residential and farmland, indicating a relatively developed first industry and a backward urbanization level, and they are mainly distributed in northern China. T_6 Ecological cities have the smallest proportions for most kinds of built-up lands, but the largest proportion of relatively natural lands, e.g., forest and grass, and provide important ecological services. Most T_6 cities are located in northwest China. The clustered six types represent LFS of diverse cities robustly, as LFS indices are relatively homogeneous in each type, so that we can describe LFS's correlation with population/GDP type by type and accurately evaluate LCC.

Evaluating LCC based on LFS

We evaluate cities' LCC based on LFS indices, type by type, where six LFS types generated in the prior section are

considered to reduce LFS indices' heterogeneity and improve the robustness of LCC estimation results²⁹. Specifically, we employ stacked autoencoder (see Methods) to integrate and dimensionally reduce LFS indices, and use ordinary least squares to estimate LCC by measuring the correlation between LFS indices and population/GDP (see Methods) in each type of cities (Eq. 1 and 2),

$$LCC_c^{\text{Pop}} = LR_u^{\text{Pop}}(\overrightarrow{\text{FS}}_c)(c \in \text{type}_u), \quad (1)$$

$$LCC_c^{\text{GDP}} = LR_u^{\text{GDP}}(\overrightarrow{\text{FS}}_c)(c \in \text{type}_u), \quad (2)$$

where LCC_c^{Pop} and LCC_c^{GDP} refer to the city c 's LCC for population and GDP respectively, and the city c belongs to u -th LFS type (type_u). $\overrightarrow{\text{FS}}_c$ denotes the LFS indices of the city c which are integrated and dimensionally reduced by stacked autoencoder, LR_u^{Pop} and LR_u^{GDP} are linear regression models which are trained by cities in type_u . As presented in Fig. 3b, e, the LCC results are highly heterogeneous. This proves that different cities have different LCC, and thus ghost city analysis considering a unified population threshold are limited due to aggregation effects. According to LCC, we evaluate the cities' land carrying conditions by two indicators, i.e., shortage in land carrying (SLC) and land carrying rate (LCR) (see Supplementary Table 3).

SLC of population and GDP are defined as differences from the actual population and GDP respectively to their expected

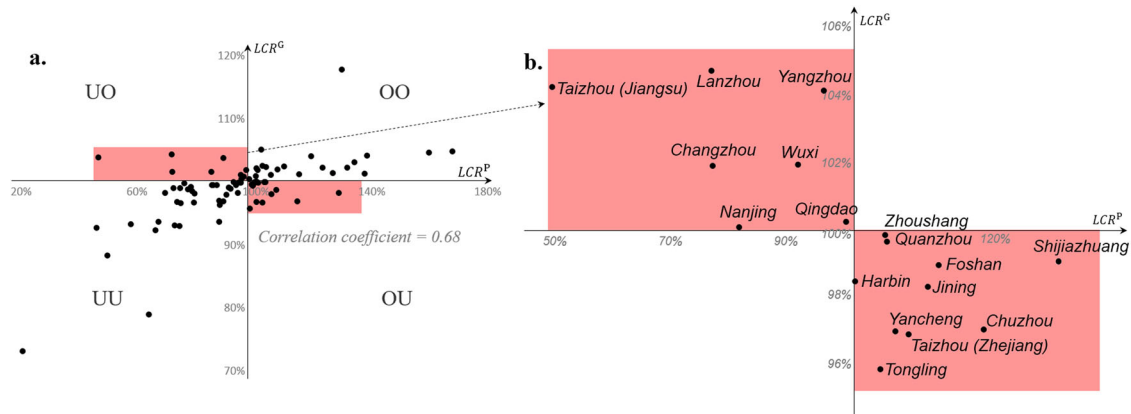


Fig. 4 Distribution of cities in LCR feature space. **a** Points represent the 81 cities in the LCR feature space, x-axis shows LCR for population (LCR^P), and y-axis is LCR for GDP (LCR^G). UO means underload population but overload GDP, while OU means the opposite; OO denotes overload for both, while UU is underload for both. **b** A close-up view of cities distributed in the UO and OU quadrants with city names, LCR^P , and LCR^G labeled.

LCC. We applied the SLC to the 81 major cities (see Supplementary Fig. 6) and their spatial distributions are shown in Fig. 3. In terms of population, we identify Zhuhai as the most underloaded in China. It features an extremely low SLC, i.e., −6.76 million people, as it has a large LCC of 8.78 million people but only 2.02 million residents. Two further eastern cities also have large LCC but carry much smaller populations: Taizhou in the Jiangsu Province (SLC of population = −4.57 million) and Jiaxing in the Zhejiang Province (−4.12 million). Even regional central cities such as Guangzhou (−3.06 million) and Xi-an (−2.63 million) are found to be underloaded for the population. The former is the capital city of the Guangdong Province, which can bear 18.36 million people considering its LCC but only carries 15.31 million residents. The latter is the capital city of the Shaanxi Province, whose LCC is 12.83 million, but it has only 10.20 million residents. The underpopulation of these cities indicates a serious waste of land resources and also imbalanced human-land relations. In terms of economy proxied by GDP, Zhuhai is found again the most underloaded city of China (SLC of GDP = −124.71 billion RMB). Some other cities are also weak in SLC with respect to economy, e.g., Jiaxing (−69.22 billion), Qinhuaungdao (−42.36 billion), Tianjin (−39.53 billion), and Weifang (−36.83 billion). The negative values relate to the lower GDP than their inherent LCC, and thus the results for these cities signify inefficient land utilizations.

With LCR, we aim to measure a city's carrying rate for population (LCR^P) and GDP (LCR^G), respectively. They are calculated as follows: $LCR^P = \frac{Actual^{Pop}}{LCC^{Pop}}$ and $LCR^G = \frac{Actual^{GDP}}{LCC^{GDP}}$, where $Actual^{Pop}$ and $Actual^{GDP}$ refer to the actual population and GDP of the city. Then, we calculate the overall LCR as $LCR = LCR^P \times LCR^G$, where $LCR > 100\%$ means relative overload and $LCR < 100\%$ denotes underload. LCR^P and LCR^G are calculated for 81 cities. To analyze cities' land carrying conditions and to express the relationship between LCR^P and LCR^G , a LCR feature space is drawn (Fig. 4). We find a generally linear correlation of cities' LCR^P and LCR^G (Fig. 4a), with a correlation coefficient of 0.68. This indicates that most cities are over- or underloaded for population and GDP simultaneously, except for 17 cities (Fig. 4b). The cities with underload population but overload GDP are located in the UO quadrant, including Taizhou (Jiangsu), Lanzhou, Changzhou, Nanjing, Wuxi, Yangzhou, and Qingdao. Conversely, the cities with overload population but underload GDP are in the OU quadrant, including Harbin, Tongling, Zhoushan, Quanzhou, Yancheng, Taizhou (Zhejiang), Jining, Foshan, Chuzhou, and Shijiazhuang. The results demonstrate

that different cities have different LCR and even a city can have different carrying conditions for population and economy (LCR^P vs. LCR^G).

Identifying underload cities

We identify underload cities based on LCR. Cities whose $LCR < \overline{LCR} - Std(LCR)$, i.e., $LCR < 71.6\%$ in this study, are defined as underload, where \overline{LCR} denotes the average LCR of 81 cities and $Std(LCR)$ the standard deviation of LCR. Cities whose $\overline{LCR} - Std(LCR) \leq LCR < \overline{LCR} - \frac{1}{2}Std(LCR)$, i.e., $71.6\% \leq LCR < 85.8\%$, are considered as slightly underload; $\overline{LCR} - \frac{1}{2}Std(LCR) \leq LCR_c < \overline{LCR} + \frac{1}{2}Std(LCR)$, i.e., $85.8\% \leq LCR < 114.2\%$, as well-balanced; $\overline{LCR} + \frac{1}{2}Std(LCR) \leq LCR_c < \overline{LCR} + Std(LCR)$, i.e., $114.2\% \leq LCR < 128.4\%$, as slightly overload; $LCR_c \geq \overline{LCR} + Std(LCR)$, i.e., $LCR > 128.4\%$, as overload. This analysis intends to identify underload or overload cities in a relational system across China for a better benchmarking by comparisons. Consequently, the 81 cities are classified into five types (Fig. 5). Eight underload and 15 slightly underload cities are found, and identifying other three types is the additional contribution of the study (see Supplementary Fig. 5).

In our analysis, we identify 15 cities as slightly underload: Huizhou, Shaoxing, Guangzhou, Jinhua, Xuancheng, Nanjing, Lanzhou, Xi-an, Changzhou, Dalian, Huzhou, Luoyang, Zhangjia-kou, Chizhou, and Guiyang. They consist of the first to fifth tiers of cities and are scattered across the country (Fig. 5b), indicating a lack of association between the slightly underload cities and their locations/city tiers. This finding is different from the common perception about cities' loads and development levels. For example, Guangzhou has a large population and developed economy, but it is found to be slightly underloaded for population and GDP when its large LCC is taken into consideration. It is similar for Nanjing and Xi-an, both capital cities for previous dynasties, and now serving as capitals of Jiangsu and Shaanxi Provinces being population hubs with advanced economies. However, they are found to be underloaded when comparing their populations and GDP to LCC.

We further identify eight underload cities: Zhuzhou ($LCR = 67.4\%$), Dongying (65.4%), Haikou (57.9%), Qinhuaungdao (54.1%), Taizhou (Jiangsu) (52.4%), Jiaxing (47.3%), Ma-anshan (46.2%), and Zhuhai (16.91%). Among these cities, Dongying, Ma-anshan, and Jiaxing have been recognized as ghost cities in previous studies^{8,15,35}. The main economic activities in Dongying and Ma-anshan are petroleum and steel industries. Due to the structural crisis of their industries, they have experienced slow economic development and low population growth

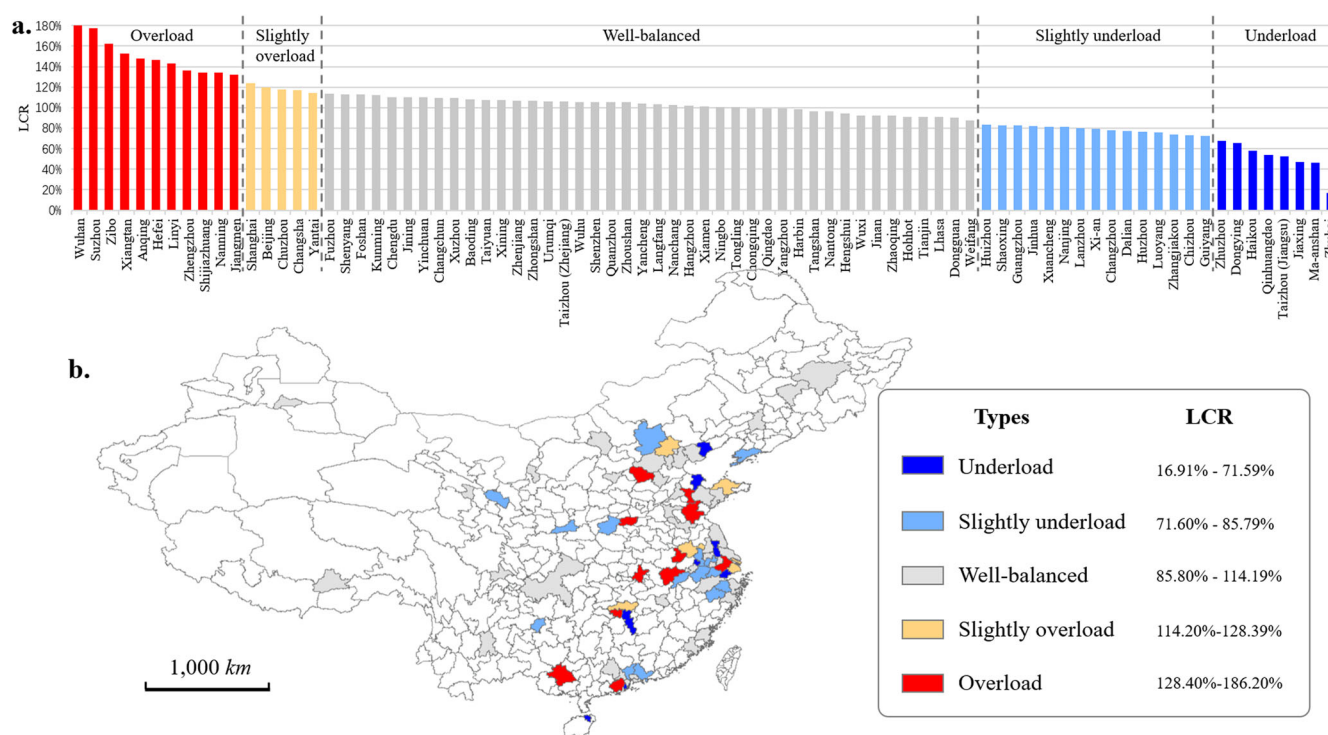


Fig. 5 Cities classified into five types in reference to land carrying rate (LCR), i.e., underload, slightly underload, well-balanced, slightly overload, and overload. **a** Cities' LCR which are ranked in descending order. **b** Spatial distributions of the five types of cities.

recently⁸. Leichtle et al. (2019) revealed the ghost city phenomenon in Dongying with a considerable mismatch between its estimated population capacity and residential number⁸; Jin et al. (2017) profiled Ma-anshan with low urban vitality as a ghost city¹⁵. Additionally, an influential report "Ranking list of ghost city index in China mainland (2014)" ranked Jiaying high as a potential ghost city³⁵. Our findings corroborate this, as Jiaying has a low LCR^p of 53.4% and a low LCR^G of 88.6%, indicating that the city is underloaded for both population and GDP. These comparisons reflect the close relationship between underload and ghost cities, and verify the rationality of the detected underload cities. The remaining five underload cities however have not been noticed in prior work, and they are identified because our approach accounts for heterogeneous LCC.

Analyzing general patterns of the underload cities

We find three general patterns of these underload cities. Firstly, the cities whose original industries are experiencing structural crisis can have slow development or even declines in population and GDP, resulting in LCC surpluses and underload cities, such as Dongying and Ma-anshan. The phenomenon is very common worldwide and occurred in many international cities, e.g., Detroit in the U.S., Hashima in Japan, and Fordlandia in Brazil¹⁵. This underload condition is often irreversible, unless new industries develop.

Secondly, the cities on the edge of urban agglomerations can be influenced by neighboring metropolises, as their population and GDP are affected by the siphon effect from surrounding cities. In this case, we find Qinghuangdao in Beijing-Tianjin-Hebei³⁶, Zhuzhou in Changsha-Zhuzhou-Xiangtan³⁷, Jiaying and Taizhou (Jiangsu) in Yangtze River Delta as underload cities. Similarly, small or medium-sized cities around metropolises, e.g., New York, London, and Tokyo, can also face the same problem³⁸.

Thirdly, the cities supported by major national policies can have fast land development with high urban expansion rates; thus, their LCC skyrocket, but their populations and economy will not rise rapidly in the short term. In this case, two Special Economic Zones of China including Zhuhai and Haikou³⁹ are found as underload cities. This underload condition is essentially a temporary product of the construction phase, as Brasilia in 1960s and Sejong in 2010s, and can be alleviated with population influx and economic growth. Taking Zhuhai as an example, the population increase in Zhuhai has speeded up at a rate of 7% annually since 2017 and may soon catch up with that of built-up expansion⁴⁰, so that the extremely underload phenomenon in Zhuhai may thus be alleviated in the future, according to the current development trend.

Comparing underload cities with ghost cities

We compare the alternative concept of underload cities with the often-used concept of ghost cities^{41,42}, to demonstrate the significance of the broader conceptual approach. First of all, these two concepts have significant differences in definition, as ghost cities mostly consider population density, but underload cities measure population, GDP, and LCC; thus, underload cities cannot be directly related to ghost cities. If we compare both concepts in an empirical manner, we find the following: we calculate population density (PD; the ratio of population to built-up area) to identify the top ten ghost cities (see Supplementary Table 4), and examine the differences between these ten ghost cities and the findings of this study (Fig. 6). Consequently, the ten ghost cities based on low PD are: Lhasa (PD = 0.29 thousand people km^{-2}), Dongying (0.97), Urumqi (1.00), Chizhou (1.11), Zhangjiakou (1.14), Xuancheng (1.33), Nanning (1.60), Chongqing (1.60), Qinghuangdao (1.92), and Jinhua (1.94). Six out of these ten ghost cities are detected in this study as underload or slightly underload cities, including Dongying, Qinghuangdao, Chizhou, Zhangjiakou, Xuancheng, and Jinhua.

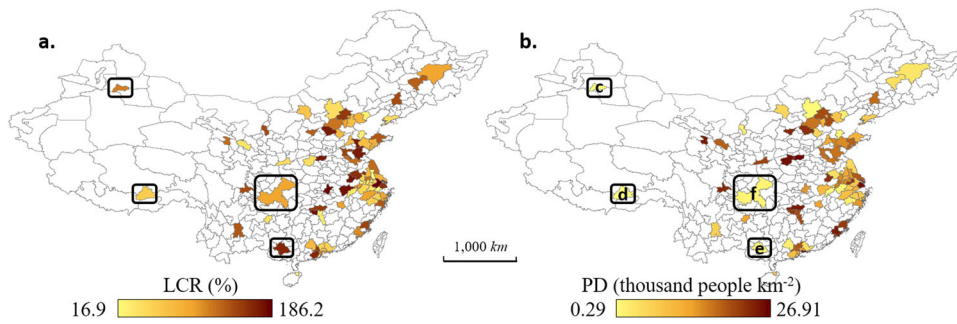


Fig. 6 Comparison between land carrying rate (LCR) and population density (PD), highlighting the differences between underload and ghost cities. **a, b** LCR and PD of 81 cities. **c–f** Urumqi, Lhasa, Nanning, and Chongqing which are identified as ghost cities according to PD but nonunderload by LCR.

Conversely, the remaining four ghost cities, i.e., Lhasa, Urumqi, Nanning, and Chongqing, are not identified as underload. These cities are located in outlying areas in western China. Due to rugged terrains and harsh natural environments, their LCC are limited. Although their low PD characterize them as ghost cities, the populations of these cities may have depleted their limited LCC, resulting in the categorization of them as non-underload. Identifying ghost or underload cities thus has implications and divergences in city planning. Let us take Nanning as an example: based on the underload city concept, Nanning has a high LCR at 133.8% (Fig. 6e), and could be considered overloaded, needing population outflow and incentives for land developments; from the ghost city viewpoint, however, a different suggestion such as more population influx and reduced land constructions could be made due to the low population density of Nanning.

In addition, among the eight underload cities identified in this study, only two are found among the ten ghost cities; while the other six have ordinary PD and are not recognized as ghost cities. These six cities have large LCC which greatly exceed their population and GDP, and thus are classified as underload. Based on the underload city concept, these six cities should increase population and limit land constructions. In contrast, from the ghost city viewpoint, the six cities can continue land constructions, which potentially exacerbates the cities' underload conditions. The comparisons above underscore the conceptual and in consequence differences for underload and ghost cities. Employing the ghost city viewpoint could have adverse impacts on human-land relationship, for its negligence in LCC. Therefore, more conceptual approaches besides ghost cities are needed to assess land carrying conditions for sustainable city planning, and thus underload cities in synopsis allow a complementary perspective and improved analysis.

Comparing LCR to the SDG 11.3.1 indicator

SDG 11.3.1 refers to the ratio of land consumption rate to population growth rate, which measures land use efficiency⁵ and also concerns the human-land relationship^{43,44}. Here, we compare LCR with SDG 11.3.1 to demonstrate the significance of this study to SDG.

Existing studies on SDG 11.3.1 measure land consumptions by impervious surface or artificial land^{45,46}, but they mostly ignore heterogeneous land functions and structures inside impervious surface which however can greatly influence population growth⁴⁷. In consequence, previous measurements of land consumption rates are limited, resulting in aggregated information on SDG 11.3.1. Differently, this study models diverse land functions and their structures (LFS) within cities to measure land consumption from socioeconomic and spatial-structural perspectives, based on which LCR can be estimated to quantify the relationship between land consumption and population with more comprehensive and detailed information. Although SDG 11.3.1 and LCR both produce

aggregated evaluations at the city scale, LCR however considers intracity land-function heterogeneity and generates more meaningful results to explain the human-land relationship. Accordingly, LCR can be regarded as a complement to SDG 11.3.1 and an important indicator for evaluating sustainable city development. Furthermore, using LCR can identify two unsustainable city development modes, i.e., underload and overload cities (see Supplementary Fig. 5); thus, the concept, approaches, and findings of this study provide a different perspective for understanding the imbalanced human-land relationship and unsustainable development modes, potentially contributing to SDG.

DISCUSSION

This study demonstrates the use of VHR satellite data and derived LFS (including land functions and structural types) to evaluate the LCC of cities, and the importance of considering LCC to identify underload cities. This study is distinct from prior works on ghost cities and the SDG 11.3.1 which mainly considers population density but does not relate to the LCC.

This study identifies eight major underload cities across China, among which five cities would not have been detected had the conventional ghost city concept been applied. We also find three general patterns of these underload cities, whose original industries are experiencing structural crisis, whose population and GDP are affected by the siphon effect from surrounding cities, and whose land development is too fast supported by major national policies. Among those, the first pattern of underload cities can be irreversible, but the third is basically a temporary construction phase and can be alleviated with population influx and economic growth.

This study also finds 15 slightly underload cities. Different from the common perceptions that big cities are over developed and congested by large populations, some big cities can be recognized as slightly underload in the study, because they could carry more people and GDP when their LFS and LCC are considered. It is suggested that these underload cities need population and GDP influx and limit further land developments, so as to best use their LCC. The findings of this study provide new insights into human-land relationships, potentially contributing to a more sustainable city development.

However, the empirical verification in this study has two limitations. Firstly, the study focuses on evaluating land loads at the scale of prefecture-level cities but ignores the heterogeneity within them. Prefecture-level cities are essentially administrative units and are usually composed of multiple county-level cities which however may have different land load conditions¹¹. Accordingly, a local evaluation, such as underload counties/neighborhoods, should be developed to reveal heterogeneous human-land relations within cities. Secondly, this study considers a part of Chinese cities, but not include international cities in other

countries, owing to the limitation of acquiring very high-resolution satellite data. In the future, international evaluations on underload cities can be conducted to find the differences in human-land relation among different nations and international cities.

In summary, we hope to work together with other scholars to resolve these limitations, and enrich the conceptual approach by studying underload cities/counties/neighborhoods and supplementing the international verification cases.

METHODS

Data

For demographic and economic data, population and GDP data are retrieved from the “China city construction statistical yearbook in 2019” (<https://www.mohurd.gov.cn/gongkai/fdzdgknr/sjfb/tjxx/jstjnj/index.html>), where GDP represents the gross domestic product, and population refers to the resident population including both permanent (registered) and temporary residents who live there for more than six months in one year.

For land observation, satellite images covering 81 major cities are acquired by splicing and resampling images from different sensors, including SPOT-6, GF-1, and ZY-3. The image product was acquired in 2019 and basically contained three visible bands (i.e., red, green, and blue) with a very high resolution of 2.4 meters. The images have been ortho-rectified and thus provide accurate image features for mapping land functional zones. In addition, built-up area of each city is obtained from “China city construction statistical yearbook in 2019”, where built-up area refers to the developed land that carries intensive human activities.

Multilevel semantic segmentation to map land functional zone

We employ a multilevel semantic segmentation (MSS) proposed by our previous study⁴⁸ to map land functional zones. MSS consists of a semantic segmentation at object level and a neighborhood optimization at block level. For the object-level semantic segmentation, it excels in classifying complex structures with strong internal heterogeneity, confusing boundaries, and multiscale representations^{48,49}; thus, it can recognize diverse land functions. It generally has three processes, i.e., object segmentation, deep feature encoding, and semantic prediction (see Supplementary Method). For object segmentation, a multiresolution segmentation approach⁵⁰ is used to segment satellite images into objects. For deep feature encoding, objects are processed by a densely connected convolutional network (DenseNet)⁵¹ and four dilated convolutions⁵² with different convolution kernels (1×1 , 3×3 , 3×3 , 3×3) and dilated rates (1, 6, 12, 18). For semantic prediction, multiple deep feature maps generated by DenseNet and dilated convolutions are stacked, convoluted, and up-sampled to predict land function categories for objects. In addition, MSS considers a neighborhood optimization at block level, which uses conditional random field to model the neighboring relationship among land functional zones and eliminates the wrong classification results within roadblocks. Technical details of MSS see the reference⁴⁸.

For training the model, we manually delineated and labeled 84,730 samples of land functional zones based on visual interpretations and field investigations, 70% of which are clipped into 94,306 non-overlapped patches of 512×512 pixels and fed into the model, where the pixels hit by manually labeled samples are set as their categories, and others are null values. The hyperparameters set in the training process are illustrated as follows. The TensorFlow system is generally used for implementation; the segmentation scale of multiresolution segmentation is set to 50; a L2-regularization is considered to avoid overfitting; an Adam optimizer with a learning rate of 2×10^{-4} is employed for training⁵³; and cross entropy is used as loss function, i.e., $CELoss = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M -\hat{y}_{ij} \log(y_{ij})$, where N denotes the number of samples, M the number of land function categories, y_{ij} represents the probability that i -th sample belonging to the j -th category, and \hat{y}_{ij} represents the real class with one-hot encoding; the semantic segmentation structure will be trained for 50,000 iterations. The hyperparameter tuning is demonstrated in our previous study⁴⁸.

For evaluating mapping results, we utilize the remaining 30% samples (25,419) and measure the pixel-wise confusion matrix⁵⁴. Overall accuracy, $OA = N_T/N$, is calculated to assess mapping accuracy, where N denotes the pixel number of test samples and N_T the number of accurately recognized pixels. In addition, Kappa index of mapping result is also measured

$Kappa = \frac{OA - P}{1 - P}$, where $P = \frac{\sum_{i=1}^M A_i \times B_i}{N \times N}$, M refers to the number of functional

categories, A_i the pixel number of the i -th category in test samples, B_i the pixel number of the i -th category in mapping results⁵⁴.

Index system for characterizing LFS

For quantitatively expressing LFS per city, we integrate indices related to land use, urban planning, and landscape pattern to propose an index system^{55–57}, which includes functional, structural, and thematic indices (see Supplementary Table 1). Functional indices describe functional types and services, e.g., functional proportion, area, abundance and dominance; structural indices measure spatial structures of diverse land functions, e.g., structural fragmentation, minimum distance, aggregation; and thematic indices characterize socio-economic attributes, e.g., industrial type, living environment, public and ecological service. These indices have been verified effective to characterize land functions and structures^{55,56} and closely related to LCC evaluation^{29–31}. All these indices can be derived from the generated land functional zone maps.

Density peaks clustering algorithm to detect robust LFS type

To reduce the influence of LFS's heterogeneity on modeling human-land relationships, we extract robust LFS types by a state-of-the-art clustering model. We apply the density peaks clustering algorithm (DPCA)⁵⁸, as DPCA excels in overcoming outliers, explicitly selecting clustering centers, and processing aspheric distributions⁵⁹. DPCA recognizes samples with large local densities and large distances to other high-density samples as clustering centers, thus it mainly considers two indicators in clustering, i.e., local density and distance to high-density samples. On the one hand, local density of the i -th sample $\rho_i = \sum_{j=1:N, j \neq i} \tau(d_{ij} - d_c)$, where N denotes the number of samples, d_{ij} the distance from sample i to sample j in the feature space, d_c is a threshold defining the neighborhood range, and $\tau(x) = 1$ when $x < 0$; on the other hand, i -th sample's minimum distance to the samples with higher density is formulated by $\delta_i = \min_{j: \rho_j > \rho_i} (d_{ij})$ or $\delta_i = \max(d_{ij})$ when $\rho_i > \rho_j (j = 1:N, j \neq i)$. Accordingly, the d_{ij} is core to DPCA, and we define d_{ij} as $d_{ij} = \max(d_{ij}^f, d_{ij}^s, d_{ij}^t)$, where d_{ij}^f represents the distance from sample i to j in functional-index space, d_{ij}^s that in structural-index space, d_{ij}^t that in thematic-index space, and distances are measured by the Euclidean distance. This definition of d_{ij} ensures clusters' homogeneity in all feature spaces.

As demonstrated above, we calculate the distances between cities in feature spaces, measure ρ_i and δ_i of each city, and select six cluster centers by $\rho_i \geq 7$ and $\delta_i \geq 0.5$ in this study with considering the cities' distributions in ρ - δ space (see Supplementary Fig. 3). Other cities are assigned to the clusters of their nearest neighbor with higher density.

Combining SAE and OLS to estimate LCC

Estimating LCC needs modeling the human-land relationship which is essentially to measure the correlation between LFS indices and socio-economic attributes, i.e., population and GDP in our case. However, two technical issues arise: 1) how to fuse diverse LFS indices, and 2) how to avoid the “curse of dimensionality”⁶⁰ affecting correlation modeling. We use a stacked auto-encoder (SAE) to resolve these two issues. SAE has been widely used for feature fusion and dimensionality reduction⁶¹. SAE is composed of multiple auto-encoders (AEs)⁶², and each AE consists of three layers: input, hidden and output layers, where the input and hidden layers compose encoder, hidden, and output layers compose decoder (see Supplementary Fig. 4). For encoding, the input features $\vec{x} \in R^N$ are dimensionally reduced into hidden features $\vec{y} \in R^K$, where N refers to the dimension number of input features and K that of hidden features; for decoding, the $\vec{y} \in R^K$ reconstructs $\vec{x}' \in R^N$. Accordingly, AE is formulated as:

$$\vec{y} = f(\vec{w}_x \cdot \vec{x} + b_x), \quad (3)$$

$$\vec{x}' = f(\vec{w}_y \cdot \vec{y} + b_y), \quad (4)$$

where \vec{w}_x , \vec{w}_y , b_x , and b_y are parameters of encoder and decoder respectively, and $f(\cdot)$ represents a sigmoid activation function⁶³. Training AE is essentially to minimize the reconstruction loss between \vec{x} and \vec{x}' , $RecLoss = \|\vec{x} - \vec{x}'\|$, and the hidden features are extracted as the features after dimension reduction. Similarly, SAE is constructed by stacking multiple AEs, with each AE trained layer by layer, and only the encoder part is kept after training. The whole network is finally fine-tuned with minimizing global reconstruction loss. The hyperparameters in the training process are described as follows: three hidden layers are considered

with their dimensional numbers set as 32, 16, and 8; the batch size is set as 16; every AE is trained for 150 iterations; and the whole SAE is fine-tuned for 500 iterations. As a result, the LFS indices can be nonlinearly integrated with dimensions reduced from 64 to 8, resolving the two issues demonstrated above.

We finally use original least squares (OLS) to model the human-land relationship and predict the land carrying capacity (LCC) of each city. For a city c_i ($1 \leq i \leq 81$), it belongs to the u -th LFS type $type_u$ ($1 \leq u \leq 6$), and other cities of $type_u$, $c_j \in type_u$ ($1 \leq j \leq 81, j \neq i$), are considered to train the regression model, i.e., $y_{c_i}^{pop} = \vec{FS}_{c_i} \cdot \vec{w}_u^{pop} + b_u^{pop}$ and $y_{c_i}^{GDP} = \vec{FS}_{c_i} \cdot \vec{w}_u^{GDP} + b_u^{GDP}$, where $y_{c_i}^{pop}$ and $y_{c_i}^{GDP}$ denote the c_i 's population and GDP respectively, \vec{FS}_{c_i} the LFS indices of c_i processed by SAE, and \vec{w}_u^{pop} , b_u^{pop} , \vec{w}_u^{GDP} and b_u^{GDP} are parameters of the two regression models, which can be trained by cities in $type_u$ and estimated by OLS⁶⁴; thus, the LCC of c_i for population and GDP can be predicted by $LCC_{c_i}^{pop} = \vec{FS}_{c_i} \cdot \vec{w}_u^{pop} + b_u^{pop}$ and $LCC_{c_i}^{GDP} = \vec{FS}_{c_i} \cdot \vec{w}_u^{GDP} + b_u^{GDP}$.

DATA AVAILABILITY

All data shown in the figures and tables are publicly available, and the source data are provided with this paper. Furthermore, the produced land functional zone maps of 81 major cities in China, and the estimated land carrying capacities of these cities (see Supplementary Fig. 7) are open to download at <https://geoscape.pku.edu.cn/>.

CODE AVAILABILITY

We acknowledge Rodriguez A. and Laio A. for their density peak clustering algorithm which is publicly available at https://people.sissa.it/~laio/Research/Res_clustering.php. We acknowledge Arcmap (v10.2) for providing secondary development and operation interface. Our code of multilevel semantic segmentation for mapping land functional zones, feature extraction for measuring land functional structure indices, and stacked auto-encoder (SAE) for processing features are maintained centrally by DoLab, Peking University (<https://geoscape.pku.edu.cn/>) and is available on request.

Received: 23 September 2021; Accepted: 24 May 2022;
Published online: 17 June 2022

REFERENCES

- United Nations Development Programme (UNDP). Sustainable development goals (SDG). Retrieved from <https://www.undp.org/content/undp/en/home/sustainable-development-goals.html> (2015).
- Kaur, H. & Garg, P. Urban sustainability assessment tools: a review. *J. Clean. Prod.* **210**, 146–158 (2019).
- Liu, X. et al. High-spatiotemporal-resolution mapping of global urban change from 1985 to 2015. *Nat. Sustain.* **3**, 564–570 (2020).
- Wang, L. et al. China's urban expansion from 1990 to 2010 determined with satellite remote sensing. *Sci. Bull.* **57**, 2802–2812 (2012).
- Fang, C. & Yu, D. Urban agglomeration: an evolving concept of an emerging phenomenon. *Landsc Urban Plan* **162**, 126–136 (2017).
- Hu, J., Wang, Y., Taubenböck, H. & Zhu, X. X. Land consumption in cities: a comparative study across the globe. *Cities* **113**, 103163 (2021).
- Williams, S., Xu, W., Tan, S. B., Foster, M. J. & Chen, C. Ghost cities of China: identifying urban vacancy through social media data. *Cities* **94**, 275–285 (2019).
- Leichtle, T., Lakes, T., Zhu, X. X. & Taubenböck, H. Has Dongying developed to a ghost city? - evidence from multi-temporal population estimation based on VHR remote sensing and census counts. *Comput. Environ. Urban* **78**, 101372 (2019).
- Woodworth, M. D. & Wallace, J. L. Seeing ghosts: parsing China's "ghost city" controversy. *Urban Geogr* **38**, 1270–1281 (2017).
- Lee, J., Park, Y. & Kim, H. W. Evaluation of local comprehensive plans to vacancy issue in a growing and shrinking city. *Sustainability-base* **11**, 4966 (2019).
- Shi, L. et al. Urbanization that hides in the dark - spotting china's "ghost neighborhoods" from space. *Landsc Urban Plan* **200**, 103822 (2020).
- Roger S. A. The Empty City of Ordos, China: A Modern Ghost Town. Retrieved from <https://weburbanist.com/2011/01/10/the-empty-city-of-ordos-china-a-modern-ghost-town/> (2011).
- Shepard, W. Ghost Cities of China, Vol. 1 (Zed Books Asia Studies Archive, 2015). ISBN: 978-1-78360-220-9
- Ge, W., Yang, H., Zhu, X., Ma, M. & Yang, Y. Ghost city extraction and rate estimation in China based on NPP-VIIRS night-time light data. *ISPRS Int. J. GeoInf* **7**, 1–17 (2018).

- Ma, X., Tong, X., Lu, S., Li, C. & Ma, Z. A multisource remotely sensed data oriented method for "ghost city" phenomenon identification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **11**, 2310–2319 (2018).
- Chi, G., Liu, Y. & Wu, H. Ghost cities analysis based on positioning data in China. *Comput. Sci.* **68**, 1150–1156 (2015).
- Jin, X. et al. Evaluating cities' vitality and identifying ghost cities in China with emerging geographical data. *Cities* **63**, 98–109 (2017).
- Xie, S., Gong, J., Huang, X. & Xia, J. A study on coordination of urbanization in the pearl river delta between 1990 and 2015. *Guangdong Land Science* **16**, 18–26 (2017).
- Lane, M. The carrying capacity imperative: assessing regional carrying capacity methodologies for sustainable land-use planning. *Land Use Policy* **27**, 1038–1045 (2010).
- Yan, H., Liu, F., Liu, J., Xiao, X. & Qin, Y. Status of land use intensity in China and its impacts on land carrying capacity. *J. Geogr. Sci.* **27**, 387–402 (2017).
- Agustizar, A., Muryani, C. & Sarwono, S. Agricultural land carrying capacity and shift of land use in upstream of Grompol watershed, central java province. *IOP Conf. Ser. Earth Env.* **145**, 012070 (2018).
- Li, M. Evolution of Chinese ghost cities. *China Perspect.* **1**, 69–78 (2017).
- Zhao, M., Xu, G., Jong, M. D., Li, X. & Zhang, P. Examining the density and diversity of human activity in the built environment: the case of the pearl river delta, China. *Sustainability-base* **12**, 1–21 (2020).
- Yan, S., Wang, X., Chen, J., Zeng, M. & Wang, L. Quantitative research on regional carrying capacity in eastern China's economic zone based on GIS. *Res. Ind.* **15**, 120–130 (2013). In Chinese.
- Tang, H. & Jiang, J. Research on the population carrying capacity of the land resources in the economic area of Zhujiang Delta. *Chin. Geogr. Sci.* **11**, 174–180 (2001).
- Zhang, X., Du, S. & Zhang, J. How do people understand convenience-of-living in cities? A multiscale geographic investigation in Beijing. *ISPRS J. Photogramm. Remote Sens.* **148**, 87–102 (2019).
- Shepard W. The myth of China's ghost cities. *Reuters*, Retrieved from <https://www.reuters.com/article/idUS170445800220150422> (2015).
- Ferris, J., Norman, C. & Sempik, J. People, land and sustainability: community gardens and the social dimension of sustainable development. *Soc. Policy Adm.* **35**, 559–568 (2010).
- Zhang, X., Du, S., Du, S. & Liu, B. How do land-use patterns influence residential environment quality? a multiscale geographic survey in Beijing. *Remote Sens. Environ.* **249**, 112014 (2020).
- Christianawati, A., Prima, F. R., Aryati, S., Salsabila, G. & Thilfatantil, M. H. Analysis of the relationship of land carrying capacity and building area towards the development of Kualanamu International Airport in 2010 and 2017. *J. Degrade. Min. Land Manage.* **7**, 2095–2103 (2020).
- Yao, Q. et al. A comparative analysis on assessment of land carrying capacity with ecological footprint analysis and index system method. *Plos One* **10**, e0130315 (2015).
- Holden, S. T. & Otsuka, K. The roles of land tenure reforms and land markets in the context of population growth and land use intensification in Africa. *Food Policy* **48**, 88–97 (2014).
- Sun, J., Di, L., Sun, Z., Wang, J. & Wu, Y. Estimation of GDP using deep learning with NPP-VIIRS imagery and land cover data at the county-level in conus. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **13**, 1400–1415 (2020).
- Chen, Y., Chen, X., Liu, Z. & Li, X. Understanding the spatial organization of urban functions based on co-location patterns mining: a comparative analysis for 25 Chinese cities. *Cities* **97**, 102563 (2020).
- Investment Times. *Ranking list of ghost city index in China mainland*. Retrieved from <https://km.loupan.com/html/news/201410/1464288.html> (2014).
- Wang, W. Study on the act of government in the urbanization process of Hebei province—taking Qinghuangdao as an example. In *Proceedings of the 2013 International Academic Workshop on Social Science*, 466–469 (2013).
- Zou, Z. et al. Valuing natural capital amidst rapid urbanization: assessing the gross ecosystem product (GEP) of China's 'Chang-Zhu-Tan' megacity. *Environ. Res. Lett.* **15**, 124019 (2020).
- Fujita, K. A world city and flexible specialization: restructuring of the Tokyo metropolis. *Int. J. Urban Reg. Res.* **15**, 269–284 (2010).
- Xu, D. & Liu, J. The analysis of real estate investment value based on prospect theory—with an example in Hainan province. *Rev. de la Fac. de Ing.* **32**, 63–69 (2017).
- Xian, S., Li, L. & Qi, Z. Toward a sustainable urban expansion: a case study of Zhuhai, China. *J. Clean. Prod.* **230**, 276–285 (2019).
- Cooke, P. After the contagion. ghost city centres: closed "smart" or open greener? *Sustainability-base* **13**, 3071 (2021).
- Wang, Q., Li, R. & Cheong, K. C. Shandong's Yintan town and China's "ghost city" phenomenon. *Sustainability-base* **11**, 4584 (2019).
- Melchiorri, M., Pesaresi, M., Florczyk, A. J., Corbane, C. & Kemper, T. Principles and applications of the global human settlement layer as baseline for the land use efficiency indicator—SDG 11.3.1. *ISPRS Int. J. GeoInf* **8**, 96–98 (2019).

44. Schiavina, M. et al. Multi-scale estimation of land use efficiency (SDG 11.3.1) across 25 years using global open and free data. *Sustainability-basel* **11**, 5674 (2019).
45. Jiang, H. et al. An assessment of urbanization sustainability in China between 1990 and 2015 using land use efficiency indicators. *npj Urban Sustainability* **1**, 1–13 (2021).
46. Mohapatra, R. & Wu, C. Modeling Urban Growth at a Micro Level: A Panel Data Analysis. *Int. J. Appl. Geospat. Res.* **6**, 36–52 (2018).
47. Mudau, N. et al. Assessment of SDG Indicator 11.3.1 and Urban Growth Trends of Major and Small Cities in South Africa. *Sustainability-basel* **12**, 7063 (2020).
48. Du, S., Du, S., Liu, B. & Zhang, X. Mapping large-scale and fine-grained urban functional zones from VHR images using a multi-scale semantic segmentation network and object based approach. *Remote Sens. Environ.* **261**, 112480 (2021).
49. Chen, L., Papandreou, G., Kokkinos, I., Murphy, K. & Yuille, A. L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **40**, 834–848 (2018).
50. Baatz, M., & Arno Schape, A. Multiresolution segmentation: An optimization approach for high quality multi-scale image segmentation, Vol. 1 (*Angewandte Geographische Informationsverarbeitung*, 2020).
51. Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4700–4708 (2017). <https://ieeexplore.ieee.org/document/8099726>
52. Yu, F., & Koltun, V. Multi-scale context aggregation by dilated convolutions. *arXiv:1511.07122* (2015). <https://arxiv.org/abs/1511.07122>
53. Kingma, D. P., & Ba, J. Adam: A Method for Stochastic Optimization. *arXiv: Learning*, 1412.6980 (2014). <https://arxiv.org/abs/1412.6980>
54. Silván-Cárdenas, J. & Wang, L. Sub-pixel confusion-uncertainty matrix for assessing soft classifications. *Remote Sens. Environ.* **112**, 1081–1095 (2008).
55. O'Neill, R. V. et al. Indices of landscape pattern. *Landsc Ecol* **3**, 153–162 (1988).
56. Pontius, R. G. Jr, Cornell, J. D. & Hall, C. A. Modeling the spatial pattern of land-use change with GEOMOD2: application and validation for Costa Rica. *Agric Ecosyst Environ* **85**, 191–203 (2001).
57. Gavrilidis, A. A., Ciocănea, C. M., Niță, M. R., Onose, D. A. & Năstase, I. I. Urban landscape quality index—planning tool for evaluating urban landscapes and improving the quality of life. *Procedia Environ. Sci.* **32**, 155–167 (2016).
58. Rodríguez, A. & Laio, A. Clustering by fast search and find of density peaks. *Science* **344**, 1492 (2014).
59. Ding, S., Du, M., Sun, T., Xu, X. & Xue, Y. An entropy-based density peaks clustering algorithm for mixed type data employing fuzzy neighborhood. *Knowl. Based Syst.* **133**, 294–313 (2017).
60. E, W., Ma, C., Wu, L. & Wojtowysch, S. Towards a Mathematical Understanding of Neural Network-Based Machine Learning: What We Know and What We Don't. *CSIAM Trans. Appl. Math.* **4**, 561–615 (2020).
61. Li, W. et al. Stacked autoencoder-based deep learning for remote-sensing image classification: a case study of African land-cover mapping. *Int. J. Remote Sens.* **37**, 5632–5646 (2016).
62. Vincent, P., Larochelle, H., Bengio, Y., & Manzagol, P. A. Extracting and Composing Robust Features with Denoising Autoencoders. Machine Learning, *Proceedings of the Twenty-Fifth International Conference (ICML 2008)*, Helsinki, Finland, 1096–1103 (2008). <https://dl.acm.org/doi/10.1145/1390156.1390294>
63. Yin, X., Jan, G., Egbert, A. L., Jan, V. & Huub, J. S. A flexible sigmoid function of determinate growth. *Ann. Bot.* **91**, 753–753 (2003).
64. Efron, B. Discussion of least square regression, least angle regression. *Ann. Stat.* **32**, 465–469 (2004).

ACKNOWLEDGEMENTS

The work presented in this paper is funded by the National Key Research and Development Program of China (No. 2021YFE0117100), National Natural Science Foundation of China (No. 42001327) and the China Postdoctoral Science Foundation (No. 2019M660003 and No.2020T130005).

AUTHOR CONTRIBUTIONS

Z.X. and D.S. designed the research and conduct the experiments, H.T. provided conceptual refinements, D.S., F.Y., and L.B. collected the land observation data, Z.X. and D.S. led writing of the paper, W.Y. and H.T. revised the paper, and all authors contributed to interpreting the results and writing the paper.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s42949-022-00057-x>.

Correspondence and requests for materials should be addressed to Shihong Du.

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) 2022