International Journal of Remote Sensing

ISSN: 0143-1161 (Print) 1366-5901 (Online) Journal homepage: https://www.tandfonline.com/loi/tres20

An estimation of housing vacancy rate using NPP-VIIRS night-time light data and OpenStreetMap data

Luyao Wang, Hong Fan & Yankun Wang

To cite this article: Luyao Wang, Hong Fan & Yankun Wang (2019) An estimation of housing vacancy rate using NPP-VIIRS night-time light data and OpenStreetMap data, International Journal of Remote Sensing, 40:22, 8566-8588, DOI: 10.1080/01431161.2019.1615655

To link to this article: https://doi.org/10.1080/01431161.2019.1615655

Published online: 19 May 2019.

INTERNATIONAL JOURNAL OF REMOTE SENSING
2019, VOL. 40, NO. 22, 8566–8588
https://doi.org/10.1080/01431161.2019.1615655

An estimation of housing vacancy rate using NPP-VIIRS night-time light data and OpenStreetMap data

Luyao Wang [a,b], Hong Fan [a,b] and Yankun Wang [c]

[a]State Key Lab for Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China; [b]Collaborative Innovation Centre of Geospatial Technology, Wuhan University, Wuhan, China; [c]Research Institute for Smart Cities & Guangdong Key Laboratory of Urban Informatics & Shenzhen Key Laboratory of Spatial Smart Sensing and Services, School of Architecture and Urban Planning, Shenzhen University, Shenzhen, China

ABSTRACT

As an informative proxy measure for a range of socio-economic variables, satellite-derived night-time light (NTL) data have been widely used to investigate the diverse anthropogenic activities and reveal urbanization development. Due to the rapid increase of Chinese urbanization rate, from 25.3% in 1987 to 58.5% in 2017, and 'crazy expansion' of city space, the sick phenomenon – 'Ghost Town' – has been brought out, generally defined as places with high housing vacancy rate (HVR), which will cause the huge waste of the limited land source in China. To investigate the HVR of urban areas in China, this study attempts to establish a hybrid model combining data derived from National Polar-Orbiting Partnership–Visible Infrared Imaging Radiometer Suite (NPP-VIIRS) NTL sensors with OpenStreetMap (OSM) data. By distinguishing non-residential areas and introducing detailed residential building information, we proposed a novel HVR estimation model, thus realizing the estimation of HVR in 31 Chinese provincial cities with different development levels (Tier 1–Tier 3). The results showed the average HVR of Tier 2 cities (0.204) was higher than that of Tier 1 cities (0.189) and Tier 3 cities (0.233). The model was proven more accurate (root mean square error of approximation (RMSE) = 0.022) when compared with previous models. To explore the reasons causing different HVRs in these provincial cities, the relationship between HVR and typical socio-economic factors – gross domestic product (GDP), population, and housing price – was also revealed. Through correlation verification and built of a regression model, HVR was found positively correlated with housing price (0.409), however, negatively correlated with population (−0.829) and GDP (−0.356). The research is an indication of the applicability of using data derived from NPP-VIIRS NTL sensors in reflecting HVR and an exploration to distinguish socio-economic factors influencing HVR in different cities. The model we proposed can potentially provide guidance for urban planners to formulate better land-use plan and rental measures.

ARTICLE HISTORY
Received 17 September 2018
Accepted 21 April 2019

CONTACT Hong Fan hfan3@whu.edu.cn State Key Lab for Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, 129 Luoyu Road, Wuhan 430079, China

1. Introduction

By profoundly involving itself in the world development wave, Chinese economy has experienced significant growth in the past 30 years, resulting in the rapid increase of Chinese urbanization rate from 25.3% in 1987 to 58.5% in 2017. With the considerable growth of urban population, many socio-economic problems have emerged, including traffic jams, air pollution, and high unemployment rate (Luo 2000). Realizing the current facilities of cities cannot support the huge external population, many local governments have enthusiastically supported the massive construction of cities, leading to the 'crazy expansion' of city space. According to the *statistical yearbooks* published by the Chinese National Statistics Bureau (CNSB), the construction land area in the main 151 cities of China has increased more than 20% from 2000 to 2010, which is considerably higher than the increasing rate of United States (4.2%) and England (0.7%; Cook 2015).

The excessive pursuit of city development and construction also caused the phenomenon of 'Ghost Town', which is generally defined as a place with an abundance of buildings but few inhabitants. This phenomenon is a reflection of the high HVR and huge property-value bubble in some cities of China (Nie and Liu 2013). According to some international academic communities, the property-value bubble in China has been one of the biggest potential economic crises globally (Truman 2014). The essence of the high property-value bubble is the imbalance between supply and demand in the real estate market. It causes a huge waste on the limited land resources in China. It also highlights the unreasonable and unsustainable city plans in China, particularly the increasing rate of population in big cities that is far behind the increase in city construction. As the expansion of cities will attract more labour force and bring high financial revenue through land auction, many local governments in China have still put great enthusiasm in rapidly constructing new district and expansions to rural areas, further pushing up the HVR.

According to the economic theory, the HVR in a healthy and sustainable housing market (Wang and Chang 2013) should not be more than 10%. The control of high HVR has been a long-standing policy in developed countries. The United States has attempted to gather information on house vacancy in each state since 1956, through the Census Bureau (Weinberg 2006; Rosen and Smith 1983), and put out a series of strategies to restrict the increase of HVR (Gabriel and Nothaft 1988). Realizing the potential hazard in the real estate market, Chinese government has formulated strategies to control the excessive exploitation in big cities and keep the HVR at a reasonable range (Luo 2008; Yeung and Howes 2006). Many researchers and committees have accomplished considerable exemplary works to describe the HVR in China (Chen 2007; Chen and Wen 2014; Song and Huang 2013; Zhou and Yang 2012). Based on a survey of 609 construction projects in 12 Chinese cities conducted by Credit Lyonnais Securities Asia (CLSA), the average vacancy rate of Chinese cities is thought to be at 15% during 2009 and 2014 (Lu et al. 2018a). As the survey is mainly conducted in large prefecture-level cities with a huge population, the results are deemed lower than the actual HVR in China. According to the research by Chen and Wen (2014), the actual HVR of Chinese cities is 22.4% in 2013, and the HVR of six big cities of Chongqing, Shanghai, Chengdu, Wuhan, Tianjin, and Beijing is 25.6%, 18.5%, 24.7%, 23.5%, 22.5%, and 19.5%, respectively. The HVR in Tier 3 cities is the highest among all districts, which is 23.2%.

The accurate estimation of HVR will aid city planners to come up with more suitable land-use strategies and to adjust the housing price and rental. However, it often depends on the considerable data size through the labour-intensive field survey (Zhang, Jia, and Yang 2016). To optimize the estimation process, the new data source is taken into consideration. With the rapid development of remote sensing technology and more successful applications in solving social problems in urban planning, the remote sensing data has been an important source in the socio-economic research. Satellite-derived night-time light (NTL) data, as one of the typical remote sensing data, is proven to have a high correlation with human activities and has been extensively used in researching human activities (Tan, Zhou, and Bai 2017) and city development over time (Zhang and Seto 2013), gross domestic product (GDP) spatialization in square kilometre (Han et al. 2012), and other fields, including CO_2 diffusion (Ou et al. 2015), disasters distribution analysis (Schultz, Cole, and Molthan 2015), and energy consumption analysis (Stokes, Roman, and Seto 2014). Satellite-derived NTL data are the electromagnetic wave information of overnight light source in the visible near-infrared band through earth surface reflection. The light source was mainly lamplight illume, and thus, NTL data can be used to research human activities. Compared with census data collected by administrative boundaries, the NTL data are timelier with lower cost and flexible spatial scale. The earliest NTL data are the stable NTL data on the Defense Meteorological Satellite Program/Operational Linescan System (DMSP/OLS). The DMSP/OLS NTL data are free for use of the public with coarse spatial resolution (30 arc seconds, 5km $\times$ 5km at nadir), and long period (1992–2013; Liu et al. 2012). Through the DMSP/OLS NTL data, Ju et al. (2017) found that vacancy building clusters at county levels with high proportion of the urbanization acceleration. Lu et al. (2018a) proposed a method to evaluate the 'Ghost City' index of 22 Chinese cities based on the change of Normalized Difference Barren Index (NDBI). However, several limitations of DMSP/OLS data also restrict further application: the range of digital number (DN) values in DMSP/OLS data is small (0–63), which results in the over-saturation in city centres. The DMSP/OLS data are available only in the period of 1992–2003, which is why many studies have used the DMSP/OLS but only focused on the socio-economic problems after 2013 (Lu et al. 2018b); The DMSP/OLS data are obtained by different satellites causing a difference in the equality of data between contiguous years.

In October 2011, Suomi National Polar-Orbiting Partnership (NPP) satellite with the Visible Infrared Imaging Radiometer Suite (VIIRS) is launched by the National Oceanic and Atmospheric Administration (NOAA)/National Geophysical Data Center (NGDC). As the new generation of NTL data, NPP-VIIRS NTL data are superior in spatial and radiometric resolution (15 arc seconds, 0.5 km $\times$ 0.5 km), radiometric detection range, and onboard calibration (Arnone et al. 2013; Shi et al. 2014a, 2014b) compared with DMSP/OLS. Two kinds of products are available on the NPP-VIIRS website (https://ngdc.noaa.gov/eog/viirs/), the 'vcm' and the 'vcmsl'. The 'vcm' is the monthly averaging dataset that excludes those affected by stray light, whereas the 'vcmsl' is the dataset that includes them. The timeliness of NPP-VIIRS data is better than the DMSP/OLS data, and it is still updated each month, which makes it superior in studying issues after 2013, compared with DMSP/OLS data (Lee et al. 2006). Based on the foregoing superiorities mentioned above, the NPP-VIIRS data have been adapted to be a good indicator of human activities (Li et al. 2017), land use, and economics (Ma et al. 2014). For instance,

Shi et al. (2014a) and Xu et al. (2016) extracted land-use information in China through the threshold segmentation of DN values of NPP-VIIRS NTL data.

Compared with other data sources, a distinct advantage of NPP-VIIRS data is their spatially continuous over large areas and relatively high resolution compared with DMSP NTL data in the distinguishing of urban boundaries, which is significant for the HVR estimation, especially in city-scale, due to the huge differences in land policy, population structure, and building levels between urban areas and rural areas (Irwin 2004; Muth 1961; Currit and Easterling 2009). Numerous studies have provided novel insights in distinguishing urban boundaries using NPP-VIIRS data (Zhang et al. 2017; Dou et al. 2017; Yu et al. 2018). Previous methods of urban extraction from NPP-VIIRS data can be generally grouped into three categories, including classification-based method, novel index-based method, and threshold techniques-based method (Li and Chen 2018). Due to the simplicity and considerable accuracy, the threshold method was recognized superior than other methods and has been widely used in related studies (Ying et al. 2016). The optimizing threshold can be determined by the use of some external indices including Normalized Difference Vegetation Index (NDVI), Normalized Difference Built-up Index (NDBI), and Urban Built-up Index (UBI) derived from external data sources like MODIS, Landsat 8 OLI, and OpenStreetMap (OSM; Li and Chen 2018; Jing et al. 2015).

Although NTL data could provide more information, researchers still realized that the lack of residential information restricts the improvement of estimation accuracy. Land observation data, thus, were obtained from the previous studies (Lu et al. 2018a; Zheng et al. 2017; Niu 2018). However, the estimation accuracy was still low, although it has been improved 20% (from 0.359 to 0.603) with the adoption of expensive high-resolution commercial NTL data (Du et al. 2018).

Concluded from previous studies, we realized some serious drawbacks imitated the study of HVR estimation. The first problem is that the light intensity produced by non-residential light sources, like street lamp, automobiles lighting, and objects reflection, has not been eliminated effectively. Although fixed threshold method has been used in previous studies to weaken the influence of unrelated light sources, however, the threshold value was often derived empirically and then subjectivity subtracted by all pixels without considering the regional difference (Han et al. 2019). The second problem is that the identification of full occupancy rate (FOR) regions, which will provide reference information for the HVR estimation, was over empirical in previous studies. In their methods, pixels with highest light intensity were thought to be the HVR regions in some studies, without considering the building coverage information, resulting in pixels with numerous buildings were all subjectively distinguished as FOR regions, even their actual occupancy rate was low (Chen et al. 2017).

Hence, this study attempts to provide a practical solution to these problems and establish a novel fine-scale HVR estimation model with high accuracy and objectivity. Through the combination of NPP-VIIRS NTL data and OSM, this study proposed a novel hybrid model to estimate the housing vacancy rate (HVR) in 31 Chinese cities. First, the urban regions were distinguished through an urban boundary extraction model (UBEM). Then, a lighting threshold extraction model (LTXM) was established to extract the lighting DN values of non-residential areas. The threshold was then used to eliminate the effects of lighting values in non-residential areas on residential units. Afterwards, the revised DN threshold of FOR regions in each city was distinguished through a full occupancy region extraction model

(FOREM) with the consideration of population distribution information and residential building information, including building area and levels. Finally, the housing vacancy of 31 cities was estimated in both grid-scale and city-scale. The results were proven accurate (root mean square error of approximation (RMSE) = 0.022) through the comparison with report data. The relationship between HVR and typical socio-economic factors – GDP, population, and housing price – was also revealed. Through the correlation verification, HVR was found negatively correlated with population (–0.829) and GDP (–0.356), meanwhile, positively correlated with housing price (0.409). The research is an indication of the applicability of using NPP-VIIRS NTL sensors in reflecting HVR. The model we proposed can potentially provide guidance for urban planners to formulate better land-use plan and rental measures.

The context of this study is organized as follows: In Section 2, we introduce the study area and data source used in the research. In Section 3, we present the methods, including the preprocessing of the data source and the establishment of the HVR estimation model. In Section 4, the experiments are conducted in 31 cities in China, and the results are described and compared with report data. The relationship between HVR and socio-economic factors is also revealed. The conclusions and directions for future work are provided in Section 5.

2. Study areas and data source

2.1. Study areas

The study areas are the 31 provincial capital cities in mainland China, including 28 provincial capital cities and 4 direct-controlled municipalities: Tianjin, Beijing, Chongqing, and Shanghai. Considering the huge development heterogeneity among different regions in China, the 31 provincial cities were not all highly developed cities. Actually, Chinese provincial cities have been divided into large (Tier-1), medium (Tier-2), and small (Tier-3) cities based on their development level (CHFS 2018). According to the statistical data (GDP), the 31 cities ranked between No. 1 (Beijing) and No. 295 (Lasa) in 2018 among the all 358 cities of China (ZCFE, 2019).

Due to their superiority in public resource allocation in each province, these provincial cities still attracted a huge scale of population aggregation, thus stimulating the quicker population increasing than other prefecture-level cities. According to *China Statistical Yearbook*, the yearly increasing rate of residential population in the provincial capital cities was 13%, during 2006–2016. Until 2016, the permanent resident population in the 31 provincial capital cities has reached 236.84 million, over 30% of the total residential population in China, resulting in the rapid expansion of their urban areas and immoderate construction. Therefore, the HVR problem in the 31 provincial cities is especially distinct, making them appropriate for our study. The estimation of HVR in these 31 provincial cities will provide significant insights for land policy adjustments and better urban planning in China.

2.2. Data source

2.2.1. NPP-VIIRS NTL dataset and preprocessing

The yearly NPP-VIIRS NTL data were used in this study as composite data in 2015. The data are obtained from the official website of NOAA's National Centres for

Environmental Information (NCEI; https://ngdc.noaa.gov/eog/viirs/index.html). Three kinds of the dataset are provided on the website: raw data records (RDR), sensor data records (SDR), and environment data records (EDR; Li et al. 2013). The SDR is the production containing radiance and was thus used in this study. The SDR data are accessible to the public and are updated each month from May 2011 to April 2018. In the original dataset acquired from the website, the effects of the aurora, fires, boats, and other ephemeral lights have been removed. However, background noises, including negative values and abruptly large values, still exists. Thus, the first step was the removal of the effects of these background noises. In this study, the following preparations were taken before the analysis of the NPP-VIIRS: (1) The minimum value of NPP-VIIRS data should be 0, representing regions without light intensity. However, values of a few pixels were lower than 0, caused by imaging error. In our study, these negative values were reset to 0. (2) Some abruptly large pixel also existed which might be extraordinary noises or pixels associated with the weak light reflected by high reflectance surfaces (e.g. snow-capped mountains). To distinguish these pixels, the maximum radiance value derived from the city centre artificially was first set as the upper threshold and then used to distinguish pixels with larger values. A Max Filter was used in these abnormal pixels to fix their values. In this way, the background noises of NPP-VIIRS NTL data was eliminated effectively.

2.2.2. Residential building information derived from OSM and Resources and Environmental Scientific Data Centre

The residential areas used in this study are derived from the OSM (www.openstreetmap.org), an online map collaboration program established by Steve Coast in July 2004. As an open community formed by millions of cartographers, OSM aims to provide free map data to the public with high accuracy and timeliness. The data are updated and maintained every day by over 1.5 million map editors through aerial images and high-precision Global Positioning System (GPS) data. With its great superiority, some well-known companies, such as Apple Inc, Microsoft, and Foursquare, have been adopting the map service provided by OSM. The data we downloaded from the OSM website are the land coverage vector dataset obtained in January 2016, which has been separated into eight layers, including residential areas. Compared with Landsat 8 data, the OSM data have several notable advantages. First, although the building areas can be extracted through Landsat 8 data, differentiating between the residential and non-residential areas through classification results can be difficult. To improve the estimation accuracy of HVR, the influence of non-residential buildings, such as commercial district, communal facilities, and tourism buildings, should be eliminated. In this way, the OSM data will be a better choice than the Landsat 8 data without the need for the description of the function of these buildings. Second, the resolution of OSM data is higher than Landsat 8 data because the OSM data are the vector data. Moreover, the OSM data are updated every day to monitor continuously the HVR in the target city.

To verify the quality of the building data, we overlaid the building data derived from OSM data with Google Map and found data missing of some buildings distributed far from city centre. To improve the data quality and complete detailed building level information, we adopted the building information data derived from Resources and Environmental Scientific Data Centre (RESDC), Chinese Academy of Sciences (CAS). The

vector RESDC building data have detailed building distribution information and building level information but do not contain building type information or outline information. Through spatial connection with RESDC data in ArcGIS 10.3, the missing data and building level information in OSM were completed. Precision validation was then conducted with the new dataset by overlaying with Google Map and distinguishing 1000 randomly chosen samples. The results showed that the overall accuracy of the data set was approximately 96%.

2.2.3. Auxiliary data

The population data used in this study are the Global Human Settlement Population Grid derived from European Commission (https://ghslsys.jrc.ec.europa.eu). It is a kind of spatial raster dataset that depicts the distribution and intensity of population, expressed as the number of people per cell. The data of four target years 1975, 1990, 2000, and 2015 are provided in the website with two kinds of resolution, 250 m (used in our study) and 1 km, covering main urban areas of large cities in the world. The use of grid-based population data would provide reference information when modelling the relationship between NPP-VIIRS data and HVR. Although there is just four years data, and it could just cover parts of a city, which limits its direct use in the estimation of HVR; however, it will be much more objective and quantitative in distinguishing the FOR regions, thus improving the estimation accuracy with the combination of NPP-VIIRS data.

To reveal the relationship between HVR and socio-economic factors, economic data are derived from the National Bureau of Statistics (http://data.stats. gov.cn/), including the yearly GDP, population, and housing price of the 31 cities in China (excluding Hong Kong, Macau, and Taipei) in 2015.

3. Method

In summary, the proposed approach in this study includes five steps (Figure 1). The first step is the prepossessing of dataset. The population data are reprojected to WGS 1984, the same as the reference of NPP-VIIRS data and OSM data. The grid-based population data and NPP-VIIRS data are resampled into resolution of 500m $\times$ 500m. The OSM data is also integrated with the RESDC data. The second step is the establishment of an UBEM, the purpose of which is to distinguish the urban regions of each city based on a threshold method. The third step is the extraction of non-residential DN values, which will eliminate the influence of unrelated light sources. The fourth step is the distinguishing of FOR regions through grid-based population data and statistic data with the consideration of detailed building information. After that, the HVR of 31 cities is estimated in both grid-scale and city-scale, based on an HVR evaluation index. Finally, the accuracy of our model is assessed in step 4 through the comparison with report data in city-scale. To explore the influencing factors of HVR, regression coefficient test is conducted with the city-scale HVR and three socio-economic factors: GDP, housing price, and population.

3.1. UBEM

The relationship between the NPP-VIIRS data and the HVR of the urban and the rural areas may have a huge difference because of the diverse living densities of the

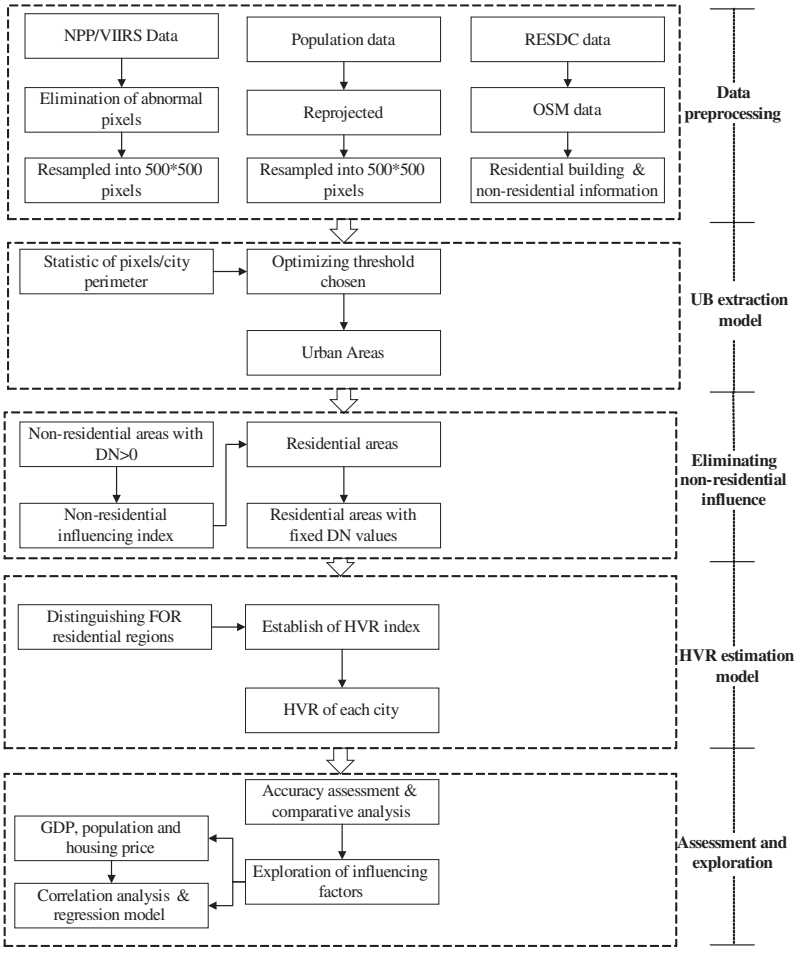

Figure 1. The methodology flowchart of the HVR estimation.

population and building structure between urban and rural areas. Taking into account these differences, it is imperative to differentiate the urban boundary of each city. Although the administrative boundary of each city can be acquired from census data, some rural areas are still situated within the administrative boundary.

The extraction of urban boundary relies mainly on the selection of threshold value from the Day/Night Band of NPP-VIIRS data. Previous studies have proposed many methods to determine the most appropriate threshold value, which can be grouped into four categories (Shi et al. 2014a; Small, Elvidge, and Baugh 2013; Zhang et al. 2017; Dou et al. 2017): (1) experience threshold, the purpose of which is to confirm the threshold value according to specialist experience obtained from previous experiments. The method is simple and practical when the historical experience is reliable; however, the defect is also obvious and too subjective and is useless in new regions without historical knowledge. (2) Statistical analysis method, the purpose of this method is to select a threshold value that makes the areas of residential regions nearer to the statistical yearbook data published by the local government. The accuracy of this

method is considerably high when statistical yearbook data are available. The problem of this method is that areas of residential regions are provided mainly in larger cities, and the timeliness of data in the statistical yearbook is often weak. (3) Spatial comparison method through external data, which is similar to the land-use classification through Landsat products, and the calculation of Normalized Difference Built-up and Vegetation Index (NDBVI) through moderate resolution imaging spectroradiometer (MODIS) products. The high cost and huge complexity limit the realization of this method. (4) Mutation detection method, the principle of this method is based on city structure. The actual urban area is considered to exhibit an aggregation trend that leads to the decrease in the perimeter of the urban area when the light threshold decreased up to N_u. The perimeter increases when the threshold is between N_u and another threshold N_d and continues to decrease when the threshold is larger than N_d. The phenomenon is caused by the threshold decreasing to N_u, which causes the city plaques to split from the inside, leading to an increase of perimeter of urban area. The threshold N_u is considered best threshold to extract urban areas. Through the analysis of the four kinds of methods, the mutation detection method is adopted considering its high convenience and accuracy in this study, the formula is described as follows:

$$\exists N_m,\ \frac{\mathrm{d}f(N_{m-1})}{\mathrm{d}N} \times \frac{\mathrm{d}f(N_{m+1})}{\mathrm{d}N} \geq 0, (N_m = 0, 1, 2, \ldots, N_{\max}) \tag{1}$$

$$N_u = \min(\forall N_m) \tag{2}$$

where N_m represents the DN value being m, ranging from 0 to the maximum value, and satisfies Equation (2). $f(N_{m-1})$ represents the perimeter of urban regions with DN values larger than N_{m-1}. Several N_m may exist because of the complex city structure. To avoid data loss, the minimum N_m is considered as the threshold value N_u.

3.2. Non-residential area extraction model

HVR describes the actual utilization ratio of a residential area, which can be modelled through the relationship between the intensity of a residential area and their light intensity. However, the actual estimation of light intensity of a residential area is often influenced by the lights from non-residential sources (e.g. street lights and background influence). The lights from these non-residential areas may also be mistakenly counted as part of the light intensity value for residential areas. Hence, for a more accurate estimation of HVR, the influence of non-residential areas should be removed.

While it was difficult to distinguish the distribution of residential and non-residential areas just through the rough NPP-VIIRS data. Therefore, OSM data was used to provide residential and non-residential information, including building type information and non-residential objects distribution information. It has to be clarified that the pixels were all located within the urban areas we extracted through urban area boundaries, and pixels contained large lakes or large rivers were also removed. The sample area was a small part of urban area, with size 4.0km $\times$ 3.5km, located in the southeast of city M (here is Wuhan), as shown in Figure 2.

In Figure 2(b), the residential buildings extracted from OSM data in the sample region were overlaid with the NTL data, where several regions with positive (larger than 4, the

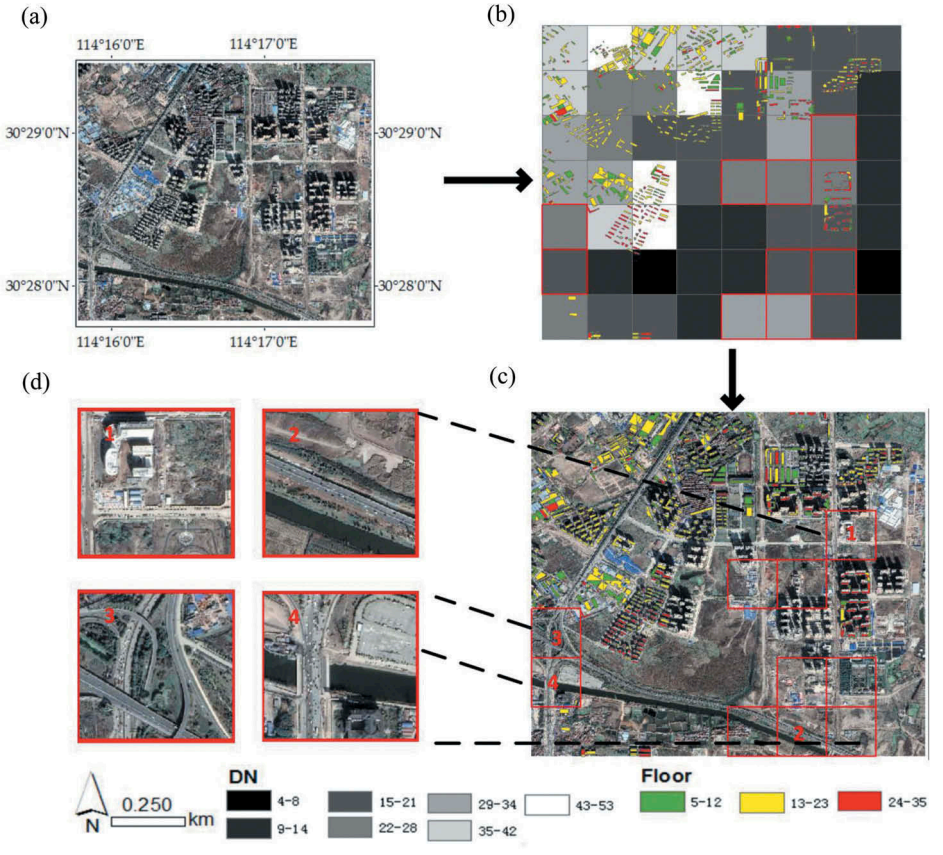

Figure 2. The extraction process of non-residential regions with positive (larger than 4, the minimum value in Wuhan urban area) DN values in a sample area: (*a*) the Google Map image of sample area, (*b*) the OSM data and NPP-VIIRS NTL data of the sample area, (*c*) the extraction of regions with positive DN values and without residential buildings, and (*d*) the details of four non-residential regions.

minimum value in Wuhan urban area) DN values were not covered by residential buildings. These grids were chosen as the target non-residential regions, as shown in Figure 2(c). Four sample grids (Nos. 1, 2, 3, and 4) were chosen and shown in Figure 2(d), where we could find numerous roads and highly reflective non-residential buildings distributed in these grids, which might be the main influencing factors of their high light intensity. The average DN value of these non-residential areas in city *M* was represented as $N_M^{'}$ and estimated below. The influence of non-residential areas was then removed from the residential area:

$$N_M^{'} = \mathrm{Avg}\left(\sum_{i=1}^{i=n} N_{M,i}^{'}\right) \tag{3}$$

$$N_{M,i}^{*} = N_{M,i} - \frac{N_M^{'}}{\mathrm{T}} S_{M,i}^{'} \tag{4}$$

where $N'_{M,i}$ represents the DN value of non-residential area i in city M, $N_{M,i}$ represents the original DN value of residential area i, $N^*_{M,i}$ represents the modified DN value without influence of non-residential area i, $S'_{M,i}$ represents the area of non-residential regions in unit i, and T represents the area of each grid (500m $\times$ 500m).

3.3. Distinguishing of FOR regions

To realize the estimation of HVR, the FOR regions in each city should be distinguished first. In previous studies (Chen et al. 2017), FOR regions were thought to be places with highest light intensity, thus regions with top 20% DN values were empirically set as the FOR threshold value. The method was too subjective to assume a fixed threshold value for each city without considering the residential coverage and population distribution. To make the estimation model more convinced, we used statistical data which described the per-capita living space (PCLS) of urban residents of each city from the China Statistical Yearbook of 2015. In our study, we make the assumption that regions whose PCLS was equal or higher than the average level of the whole city were fully occupied by residents. A FOR region f in city M can be found:

$$\frac{V_{M,f}}{P_{M,f}} \geq Q_M, (i = 1, 2, \ldots, n) \tag{5}$$

where $V_{M,f}$ represents the area of residential buildings in grid f, $P_{M,f}$ represents the population quality in grid f, and Q_M represents the PCLS in city M. In some research (Lu et al. 2018a; Ma et al. 2017), the areas of residential regions were derived from land-use data, which were actually the ground area of a building. However, the residential buildings in urban areas often have more than 30 floors, which means the methods used in previous studies would lose more than 80% of the actual areas of residential buildings, thus could not provide accurate HVR estimation results. In this study, the problem is solved by introducing more detailed residential building information. Considering the level information of the buildings, the residential areas were calculated as below:

$$V_{M,i} = \sum A_{M,i,j} \times L_{M,i,j} \tag{6}$$

where $A_{M,i,j}$ represents the areas of building j in grid i, and its level was represented by $L_{M,i,j}$. In this way, FOR regions were distinguished. However, as the residential areas in these regions may be different, their DN values may also be different. Thus, we could not simply use the average DN values in FOR region to compare with the DN values in other residential regions and estimate their HVR. Hence, we established an index W, to describe the contribution rate of every residential area to DN in each grid:

$$W_{M,i} = \frac{N_{M,i}}{S_{M,i}} \tag{7}$$

$$W_{\text{M\_full}} = 1/N_{M,f} \times \sum \frac{D_{M,f}}{V_{M,f}} \tag{8}$$

where $W_{M,i}$ and $W_{\text{M\_full}}$ represent the evaluation index of grid i and average W of FOR grids, respectively.$N_{M,i}$ represents the DN values of grid i and $N_{M,f}$ represents the DN

values of FOR grids. $S_{M,i}$ represents the area of residential buildings in grid i and $V_{M,f}$ represents the area of residential buildings in FOR grids. Thus, the HVR of grid i can be estimated by:

$$H_{M,i} = \begin{cases} 1 - \frac{W_{M,i}}{W_{\text{M\_full}}} \times 100\%, & W_{M,i} < W_{\text{M\_full}} \\ 0, & W_{M,i} \geq W_{\text{M\_full}} \end{cases} \quad (9)$$

where $H_{M,i}$ represents the HVR of grid i in city M. The city-scale HVR of city M could be estimated:

$$H_M = \frac{\sum H_{M,i} \times V_{M,i}}{\sum V_{M,i}} \quad (10)$$

In this study, we use RMSE (dimensionless) to evaluate the model accuracy through the comparison with historical report data in city-scale:

$$\text{RMSE} = \sqrt{\frac{\sum (y' - y)^2}{n}} \quad (11)$$

where y' represents the estimation value, y represents the reported value, and n represents the observation number. The closer of RMSE to 0, the better our model fitted.

4. Experiment results

4.1. Results of urban areas

The extraction results of a sample city Wuhan, optimal threshold value of which was 4, are shown in Figure 3, where urban areas under different DN values were extracted to describe the urban area distinguishing process more clearly.

Figure 3(a) demonstrates the administrative boundaries of Wuhan City, including urban areas and rural areas. Figure 3(ci) shows the region with DN values more than 2. And its boundary is shown in Figure 3(cii). While Figure 3(di,ei) illustrates the region with DN values more than 3 and 4, their boundaries are depicted in Figure 3(dii,eii). From the captured form of these figures, we can find the area of the region shown in Figure 3(ci) is the maximum with a complete and large boundary, as shown in Figure 3 (cii). It also contains large black regions with low DN values. The area of the region shown in Figure 3(ei) is the minimum compared with Figure 3(ci,di). The boundary of Figure 3(eii) is more cracked than Figure 3(cii,dii) and is considerably closer to the construction of urban areas.

To identify the boundary between urban and rural areas, the UBEM was then conducted in all 31 cities. The threshold value of each city is also different because of the different city structures. The results of other four example cities are shown in Figure 4.

Figure 4(a–d) represents the extraction process of four cities: Hangzhou, Taiyuan, Jinan, and Nanning, with the threshold DN value being 5, 7, 4, and 5, respectively. The urban area ratio of the entire administrative region of the four cities is 63.7%, 58.9%, 31.4%, and 18.6%, respectively. Figure 4 shows the distinct difference in DN values between the urban and the rural areas, with urban area having considerably higher intensity values. However, the coverage of the urban area in some cities is sometimes less than 50%. Therefore, distinguishing between the urban and the rural areas is important.

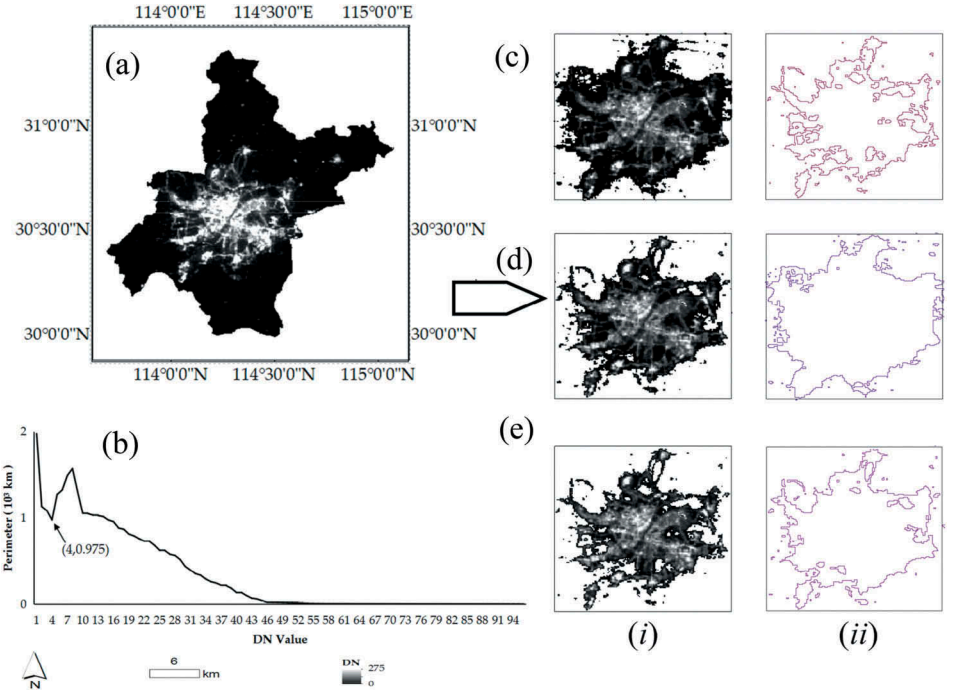

Figure 3. Boundaries of residential areas in Wuhan under different threshold values: (*a*) the original shape of Wuhan, (*b*) the choosing of optimal threshold value, (*c*) shape and boundary when threshold value was 2, (*d*) shape and boundary when threshold value was 3, and (*e*) shape and boundary when threshold value was 4.

4.2. Threshold values of FOR regions

After distinguishing non-residential regions with positive DN values in urban areas, the average reflectance of the non-residential grids in each city was then estimated and used to eliminate its influence on residential grids. Afterwards, the modified DN values combined with population data and building information were used to distinguish the FOR regions in each city. The results of average non-residential reflectance, average FOR threshold values, and ratio of FOR residential areas in each city are shown in Table 1.

Where we can find AN was over 30% of AF in several cities, especially in Tier 1 cities, due to their higher development level and compact city construction, thus indicating the necessity in distinguishing these non-residential regions. The results also showed that there existed huge differences in the FOR threshold values of 31 cities, especially among different tier of cities. AN values of Tier 1 cities were relatively high: between 56.042 (Guangzhou) to 64.855 (Shanghai). AN values of Tier 2 cities were moderate, ranging from 29.085 (Kunming) to 49.485 (Zhengzhou). AN values of Tier 3 cities were lower than Tier 1 and Tier 2 cities. As to the ratio of FOR regions, cities with highest ratio can be distinguished from the results: Beijing (40.852%), Shanghai (38.836%), Guangzhou (41.085%), Chongqing (37.515%), and Hangzhou (39.365%) were Tier 1 or top developed Tier 2 cities in China. Conversely, the ratio of Tier 3 cities was low, ranging from 17.537% (Haikou) to 24.564% (Nanning).

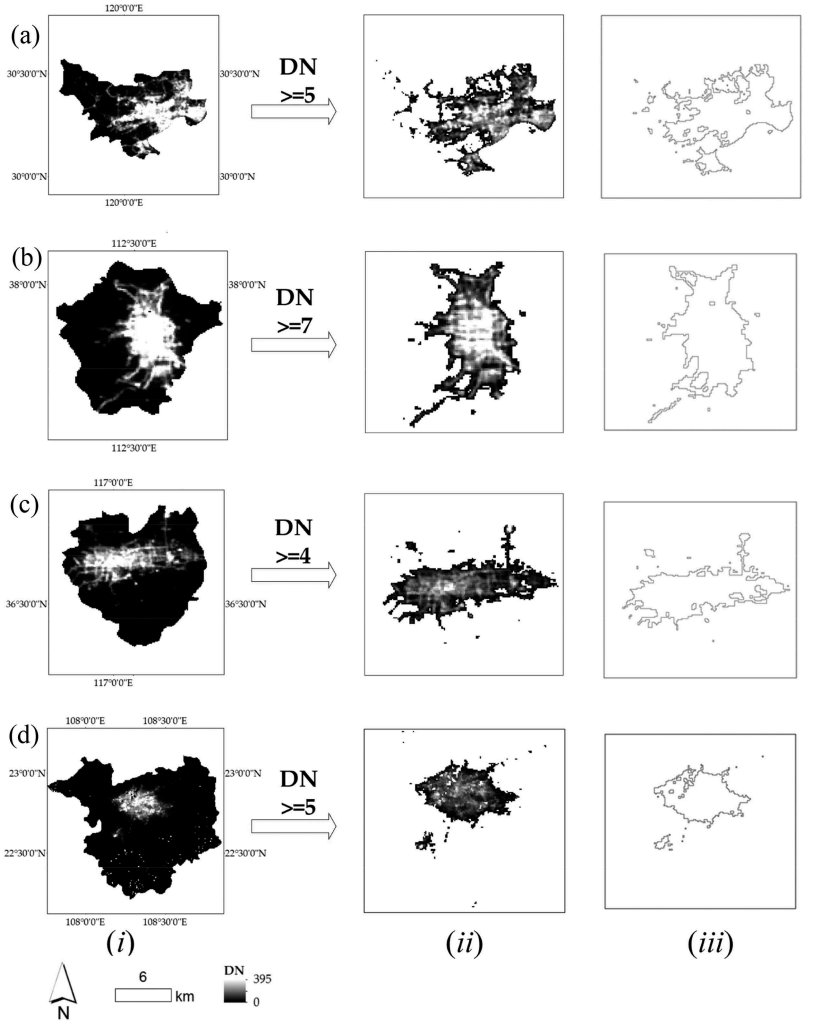

Figure 4. Boundaries of residential areas in four example cities: (*a*) Hangzhou, (*b*) Taiyuan, (*c*) Jinan, and (*d*) Nanning.

4.3. Estimation of HVR and evaluation of model accuracy

The threshold values were then set as a reference for the estimation of HVR. Figure 5 shows the grid-based HVR in eight sample cities. The grids in each city were divided into four categories according to their HVR, low (<0.01), moderate (0.01–0.20), high (0.20–0.50), and extremely high (>0.50). Captured from Figure 5, we can find the high HVR regions were mainly distributed in the core areas of each city, and HVR increased from core urban areas to suburban areas.

Using the model established in Section 3.3, city-scale HVR was then estimated, the results were shown in Figure 6.

In Figure 6, the development levels of cities were described by three colours: red, yellow, and blue, representing T-1 cities, T-2 cities and T-3 cities, respectively.

Table 1. Average non-residential reflectance and FOR in each city.

City Tier	Name	AN	AF	Rate of FOR residential areas (%)
1	Beijing	24.795	57.285	40.852
	Shanghai	22.521	64.855	38.836
	Guangzhou	20.754	56.042	41.085
2	Tianjin	18.083	48.541	31.824
	Shijiazhuang	15.285	41.052	23.164
	Taiyuan	15.716	39.658	25.944
	Huhehaote	12.647	42.874	19.536
	Shenyang	14.842	38.208	30.756
	Changchun	11.882	33.274	23.075
	Ha'erbin	9.411	34.624	20.469
	Nanjing	17.548	40.113	27.300
	Hangzhou	18.523	31.681	39.365
	Hefei	16.274	35.072	22.158
	Fuzhou	10.526	43.942	25.248
	Nanchang	11.286	40.512	21.267
	Jinan	14.268	41.625	29.641
	Zhengzhou	18.524	49.485	34.277
	Wuhan	17.549	48.710	30.671
	Changsha	14.296	42.952	32.255
	Chengdu	13.552	39.632	33.585
	Kunming	10.985	29.085	27.544
	Xi'an	15.644	33.574	31.652
	Lanzhou	12.020	30.856	36.584
	Wulumuqi	10.812	34.633	31.220
	Chongqing	16.285	44.220	37.515
3	Haikou	13.257	35.971	17.537
	Nanning	12.547	27.423	24.564
	Guiyang	11.925	33.845	23.674
	Lhasa	13.651	29.662	21.271
	Xining	14.085	32.526	22.641
	Yinchuan	9.685	31.874	19.542

AN represents the average non-residential reflectance; AF represents the average FOR threshold value.

Deep orange histogram represents the error (described by RMSE, which was dimensionless in the study) between estimated HVR and historical report data from *Analysis of Urban Housing Vacancy of China in 2017* presented by Survey and Research Centre for China Household Finance (CHFS, 2018), where the HVR of major cities in China was described. There were three Tier 1 cities: Beijing, Guangzhou, and Shanghai. The HVR of Beijing is relatively high: 0.206. The HVR of Guangzhou and Shanghai is moderate: 0.180 and 0.175, respectively. The HVR of Tier 2 cities were mainly distributed near the moderate level: 0.175 (Ha'erbin, Fuzhou)–0.234 (Wulumuqi). The HVR of Tier 3 cities is relatively higher. Nanning, with its HVR reaching 0.241, is suffering most serious HVR problem in the 31 provincial cities. The HVR of Guiyang (0.226) was lowest in Tier 3 cities; however, it is still higher than most cities in Tier 1 and Tier 2 cities. The results illustrated that the HVR in the 31 provincial cities ranges between 17% and 25% and the HVR of most cities (26 cities) is higher than 20%, which describes that the housing vacancy problem is very serious in China.

To further explore the HVR in cities with different development levels, the average HVR of Tier 1, Tier 2, and Tier 3 cities was estimated, respectively, and then compared with report data. As shown in Table 2, the results indicated high HVR in Chinese provincial cities, especially in Tier 3 cities with average HVR reaching 23.3%. Table 2

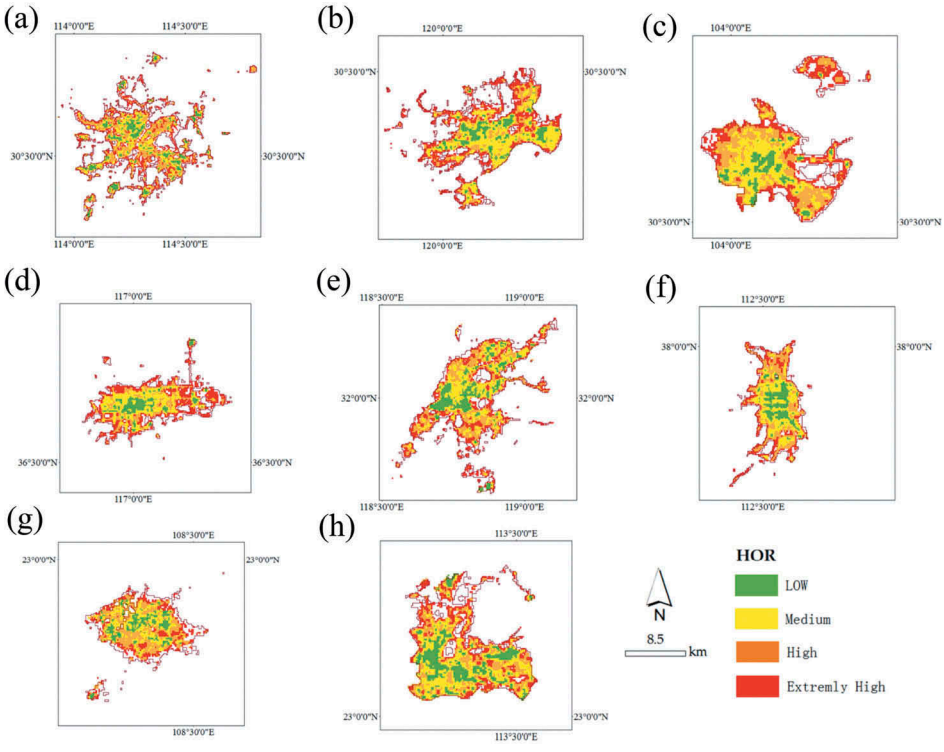

Figure 5. HVR estimation results of some cities in grid level: (*a*) Wuhan, (*b*) Hangzhou, (*c*) Chengdu, (*d*) Jinan, (*e*) Nanjing, (*f*) Taiyuan, (*g*) Nanning, and (*h*) Guangzhou.

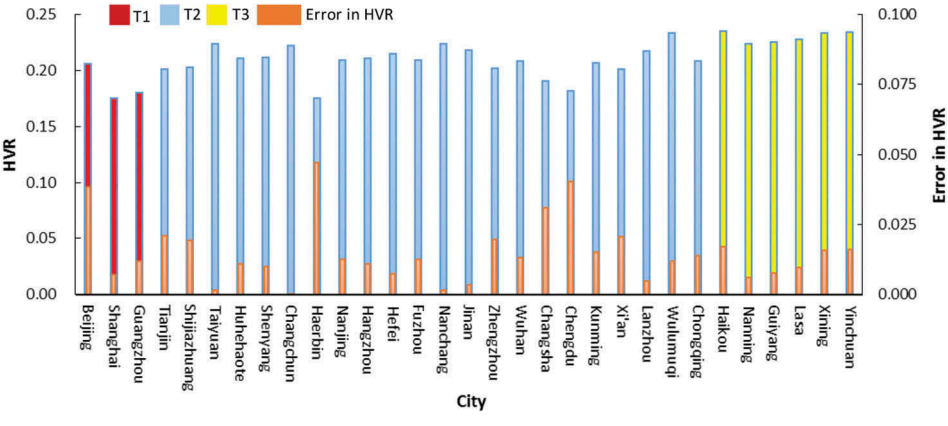

Figure 6. Results of HVR in city-scale and estimation error of HVR compared with report data. HVR and error were dimensionless.

also shows that the accuracy of our model was high through the comparison with Chen's model (2017). And the total accuracy can be improved by eliminating non-residential influence.

Table 2. Average HVR in Tier 1–Tier 3 cities and accuracy assessment.

City tier	Chen's model (a)	With NR (b)	Without NR (c)	Report data	RMSE (a)	RMSE (b)	RMSE (c)
1	0.184	0.144	0.189	0.168	0.135	0.059	0.023
2	0.279	0.199	0.204	0.222	0.157	0.064	0.024
3	0.253	0.201	0.233	0.218	0.161	0.068	0.016
Average	0.265	0.194	0.208	0.216	0.158	0.065	0.022

With NR represents HVR estimation results with influence of non-residential regions; without HR represents HVR estimation results eliminating influence of non-residential regions; RMSE (a), (b), and (c) represent the RMSE of Chen's model, our model containing non-residential influence, and our model eliminating non-residential influence, respectively.

4.4. Exploration of influencing factors of HVR

Based on the estimation results, a huge difference of HVR exists among the cities with different levels of development. However, understanding the pattern through the development level of cities can be difficult. Distinguishing between the profound reasons through some socio-economic factors (e.g. GDP, population, and housing price) is important. In this section, the influencing factors of HVR in different cities are discussed. The details are described in Table 3, where the HVR and normalized socio-economic factors of each city are shown as follows.

To determine the key influence factors of the HVR, a regression model is adopted for each factor, and the results are shown in Figure 7(a–c). The R^2 of each regression result is

Table 3. HVR and normalized socio-economic factors of each city.

City tier	Name	HVR	POP	HP	GDP
1	Beijing	0.206	0.057	0.087	0.206
	Shanghai	0.175	0.066	0.074	0.175
	Guangzhou	0.180	0.041	0.055	0.180
2	Tianjin	0.201	0.038	0.031	0.201
	Shijiazhuang	0.203	0.048	0.03	0.203
	Taiyuan	0.224	0.02	0.028	0.224
	Huhehaote	0.211	0.014	0.019	0.211
	Shenyang	0.212	0.035	0.025	0.212
	Changchun	0.222	0.026	0.025	0.222
	Ha'erbin	0.175	0.045	0.024	0.175
	Nanjing	0.210	0.032	0.044	0.210
	Hangzhou	0.211	0.035	0.057	0.211
	Hefei	0.206	0.057	0.087	0.206
	Fuzhou	0.175	0.066	0.074	0.175
	Nanchang	0.180	0.041	0.055	0.180
	Jinan	0.201	0.038	0.031	0.201
	Zhengzhou	0.202	0.039	0.028	0.032
	Wuhan	0.209	0.040	0.033	0.047
	Changsha	0.191	0.033	0.022	0.037
	Chengdu	0.182	0.057	0.026	0.047
	Kunming	0.207	0.028	0.028	0.017
	Xi'an	0.201	0.039	0.024	0.025
	Lanzhou	0.217	0.018	0.024	0.009
	Wulumuqi	0.234	0.015	0.024	0.011
	Chongqing	0.208	0.027	0.018	0.042
3	Haikou	0.235	0.034	0.048	0.036
	Nanning	0.241	0.036	0.024	0.015
	Guiyang	0.226	0.021	0.019	0.012
	Lhasa	0.228	0.006	0.019	0.002
	Xining	0.234	0.013	0.018	0.005
	Yinchuan	0.234	0.012	0.017	0.006

POP represents the normalized population quantity and HP represents the normalized housing price.

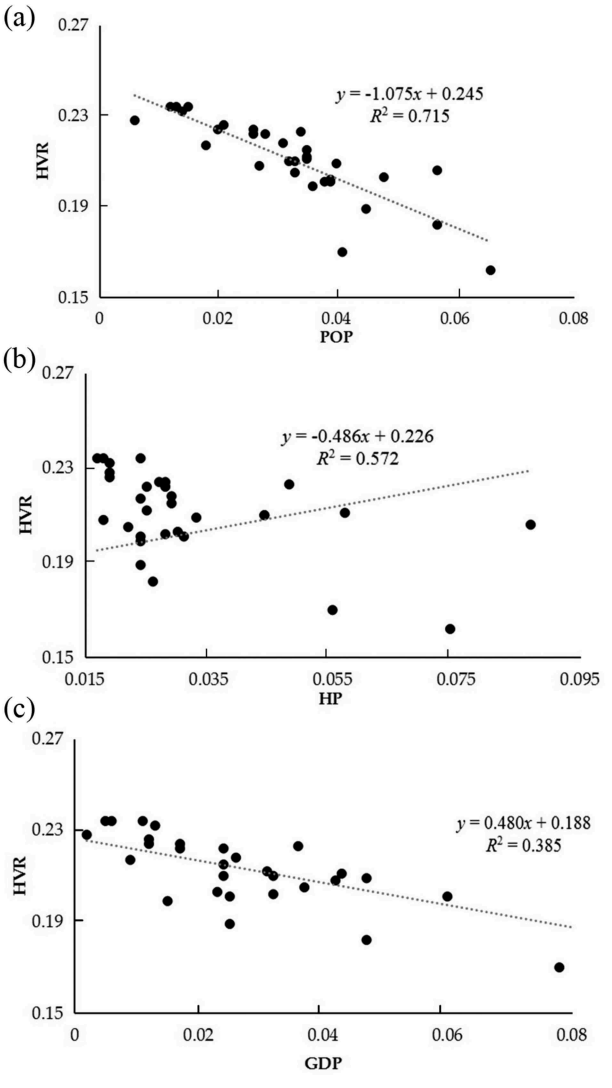

Figure 7. Correlation test between HVR and three socio-economic factors: (*a*) population, (*b*) housing price, and (*c*) GDP.

Table 4. Regression models between HVR and three socio-economic factors.

Variable	Model	R^2
POP	$y = -1.075x+0.245$	0.715
HP	$y = 0.480x +0.188$	0.385
GDP	$y = -0.486x +0.226$	0.572
POP(a) and HP(b) and GDP(c)	$y = -0.829a+0.409b-0.356c+0.236$	0.784

0.715 (population and HVR), 0.385 (housing price and HVR), 0.572 (GDP and HVR). Polynomial regression is used to formulate the relationship of data in Table 4.

In Figure 7, we can find that the HVR shows a strong negative relationship with population (–0.829), which means the population of cities will highly promote the decreasing of HVR due

to higher housing demand. The results illustrate the negative relationship between GDP (−0.356) and HVR, indicating cities with higher development level provides more attraction for residents. The HVR also shows a positive relationship with the housing price (0.409), which means the housing price has seriously influenced the HVR in China.

The results indicate that the HVR in China is highly influenced by the housing price in each city. The huge property bubble in China has deeply influenced the vitality of the substantial economy and aggravated the burden of citizens. Although the increasing population has provided more development potential for cities, the HVR remains higher than 20%. The demographic dividend will also decrease if the high housing price is not effectively limited. However, taking into consideration the city development level and population quantity in urban construction will also help for the controlling of HVR, especially in Tier 3 cities.

5. Conclusion

The rapid development of the Chinese economy has resulted in the significant acceleration of the urbanization rate of Chinese cities. This acceleration has stimulated the implementation of ambitious construction projects and 'crazy expansion' of urban space. Excessive expansion without consideration of the cities' actual development has resulted in millions of empty buildings and streets, which is a huge waste of land resource and financial budget. Previous researchers and committees have attempted to investigate the phenomenon through traditional socio-economic statistical data or web surveys in China, which is labour intensive and involves considerable time. With the rapid development of remote sensing technology, the NPP-VIIRS satellite-derived NTL data have been proven to have a high correlation with human activities and have been widely used in investigating city development (Liu et al. 2012) and energy consumption (Xie and Weng 2016). While OSM data occupied with building level data can provide detailed geographic information system (GIS) information of cities with high accuracy, a combination of NPP-VIIRS NTL data and OSM data has a unique capacity to investigate the HVR in China. The main contributions of this study are concluded:

(1) We proposed a non-residential area extraction model to eliminate the influence of non-residential areas. The model will effectively distinguish non-residential areas with positive DN values and eliminate the negative influence on HVR estimation caused by these regions.
(2) One important process of this study is the distinguishing of FOR regions. Compared with previous studies, more detailed building information, including building area and building level were considered. The establishment of a HVR index – *DPA*, describing the average contribution rate of residential areas to DN values rather than empirical light threshold without the consideration of building information also makes the distinguish process more convinced, thus improving the HVR estimation accuracy.
(3) The relationship between HVR and typical socio-economic factors – GDP, population and housing price – is also revealed.

The research is an indication of the applicability of using NPP-VIIRS sensors in reflecting HVR. The model we proposed can potentially provide guidance for urban planners to formulate better land-use plan and rental measures. This study may potentially provide strategies for the controlling of HVR in cities. In the context, the NPP-VIIRS and OSM data merely covered the year 2017, and the progressive changes in HVR are not considered. Moreover, the HVR of some small-scale county-level cities and small towns is still difficult to estimate because of the low light intensity in these places. In future works, we will focus on these areas and provide more convincing estimation results by introducing new data source and establishing better models.

Author Contributions

Wang Luyao and Fan Hong conceived and designed the main idea and experiments; Wang Luyao and Wang Yankun performed the experiments; Wang Luyao wrote the paper.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was supported by the National Natural Science Foundation of China (Grant Nos. 41471323, 91746206, and 41661086), the National Key Research and Development Program of China (2017YFB0503500) and the Science and Technology Development Project of Guizhou Province Tobacco Corporation of China National Tobacco Corporation (Contract No. 201407).

ORCID

Luyao Wang http://orcid.org/0000-0002-5274-7857

References

Arnone, R., S. Ladner, G. Fargion, P. Martinolich, R. Vandermeulen, J. Bowers, and A. Lawson. 2013. "Monitoring Bio-Optical Processes Using NPP-VIIRS and MODIS-Aqua Ocean Color Products." *Journal of Comparative Neurology* 437 (3): 363–383.

Chen, K., and Y. Wen. 2014. *The Great Housing Boom of China*. Beijing: Social Science Electronic Publishing.

Chen, P. P. 2007. "Analysis of the Vacancy Rate of Commercial Housing in China." *Construction Management Modernization* 7 (4): 34–36.

Chen, Z., B. Yu, Y. Hu, H. Chang, and J. Wu. 2017. "Estimating House Vacancy Rate in Metropolitan Areas Using NPP-VIIRS Nighttime Light Composite Data." *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing* 8 (5): 2188–2197. doi:10.1109/JSTARS.2015.2418201.

CHFS (China Household Finance Survey). 2018. *Analysis of Urban Housing Vacancy of China in 2017*. Chengdu: Southwestern University of Finance and Economics Press.

Cook, I. G. 2015. "China's Urbanization and the World Economy." *Geography* 100 (3): 240–256.

Currit, N., and W. E. Easterling. 2009. "Globalization and Population Drivers of Rural-Urban Land-Use Change in Chihuahua, Mexico." *Land Use Policy* 26 (3): 535–544. doi:10.1016/j.landusepol.2008.08.001.

Dou, Y., Z. Liu, C. He, and H. Yue. 2017. "Urban Land Extraction Using VIIRS Nighttime Light Data: An Evaluation of Three Popular Methods." *Remote Sensing* 9 (175). doi:10.3390/rs9020175.

Du, M., L. Wang, S. Zou, and C. Shi. 2018. "Modeling the Census Tract Level Housing Vacancy Rate with the Jilin1-03 Satellite and Other Geospatial Data." *Remote Sensing* 10 (12). doi:10.3390/rs10121920.

Gabriel, S. A., and F. E. Nothaft. 1988. "Rental Housing Markets and the Natural Vacancy Rate." *Real Estate Economics* 16 (4): 419–429. doi:10.1111/1540-6229.00465.

Han, D., X. Yang, H. Cai, X. Xu, Z. Qiao, C. Cheng, N. Dong, D. Huang, and A. Liu. 2019. "Modelling Spatial Distribution of Fine-Scale Populations Based on Residential Properties." *International Journal of Remote Sensing* 1–14. doi:10.1080/01431161.2019.1579387.

Han, X., Y. Zhou, S. Wang, R. Liu, and Y. Yao. 2012. "GDP Spatialization in China Based on Nighttime Imagery." *Journal of Geo-Information Science* 14 (1): 128–136. doi:10.3724/SP.J.1047.2012.00128.

Irwin, E. G. 2004. "Determinants of Residential Land-Use Conversion and Sprawl at the Rural-Urban Fringe." *American Journal of Agricultural Economics* 86 (4): 889–904. doi:10.1111/j.0002-9092.2004.00641.x.

Jing, W., Y. Yang, X. Yue, and X. Zhao. 2015. "Mapping Urban Areas with Integration of DMSP/OLS Nighttime Light and MODIS Data Using Machine Learning Techniques." *Remote Sensing* 7 (9): 12419–12439. doi:10.3390/rs70912419.

Ju, Y., I. Dronova, Q. Ma, and X. Zhang. 2017. "Analysis of Urbanization Dynamics in Mainland China Using Pixel-Based Night-Time Light Trajectories from 1992 to 2013." *International Journal of Remote Sensing* 38 (21): 6047–6072. doi:10.1080/01431161.2017.1302114.

Lee, T. E., S. D. Miller, F. J. Turk, C. Schueler, R. Julian, S. Deyo, P. Dills, and S. Wang. 2006. "The NPOESS VIIRS Day/Night Visible Sensor." *Bulletin of the American Meteorological Society* 87 (2): 191–200. doi:10.1175/BAMS-87-2-191.

Li, K., and Y. Chen. 2018. "A Genetic Algorithm-Based Urban Cluster Automatic Threshold Method by Combining VIIRS DNB, NDVI, and NDBI to Monitor Urbanization." *Remote Sensing* 10 (2): 277. doi:10.3390/rs10020277.

Li, X., D. Li, H. Xu, and C. Wu. 2017. "Intercalibration between DMSP/OLS and VIIRS Night-Time Light Images to Evaluate City Light Dynamics of Syria's Major Human Settlement during Syrian Civil War." *International Journal of Remote Sensing* 38 (1): 1–18. doi:10.1080/01431161.2017.1331476.

Li, X., H. Xu, X. Chen, and C. Li. 2013. "Potential of NPP-VIIRS Nighttime Light Imagery for Modeling the Regional Economy of China." *Remote Sensing* 5 (6): 3057–3081. doi:10.3390/rs5063057.

Liu, Z., C. He, Q. Zhang, Q. Huang, and Y. Yang. 2012. "Extracting the Dynamics of Urban Expansion in China Using DMSP-OLS Nighttime Light Data from 1992 to 2008." *Landscape & Urban Planning* 106 (1): 62–72. doi:10.1016/j.landurbplan.2012.02.013.

Lu, H., C. Zhang, G. Liu, X. Ye, and C. Miao. 2018a. "Mapping China's Ghost Cities through the Combination of Nighttime Satellite Data and Daytime Satellite Data." *Remote Sensing* 10 (7): 1037. doi:10.3390/rs10071037.

Lu, H., M. Zhang, W. Sun, and W. Li. 2018b. "Expansion Analysis of Yangtze River Delta Urban Agglomeration Using DMSP/OLS Nighttime Light Imagery for 1993 to 2012." *ISPRS International Journal of Geo-Information* 7 (2): 52. doi:10.3390/ijgi7020052.

Luo, K. 2000. "Analysis of Today's Chinese Urbanization Problems." *Journal of Wuhan Urban Construction Institute* 17 (1): 25–31.

Luo, Z. Y. 2008. "The Comparative Analysis on the Difference in Wealth Effect between Financial Assets and Housing Assets in China." *China Soft Science* 44 (4): 40–47.

Ma, T., Y. Zhou, Y. Wang, C. Zhou, S. Haynie, and T. Xu. 2014. "Diverse Relationships between Suomi-NPP VIIRS Night-Time Light and Multi-Scale Socioeconomic Activity." *Remote Sensing Letters* 5 (7): 652–661. doi:10.1080/2150704X.2014.953263.

Ma, X., Z. Ma, X. Tong, and S. Liu. 2017. "Monitoring Ghost Cities at Prefecture Level from Multi-Source Remote Sensing Data." Paper presented at the International Workshop on Remote Sensing with Intelligent Processing, Shanghai.

Muth, R. F. 1961. "Economic Change and Rural-Urban Land Conversions." *Econometrica* 29 (1): 1–23. doi:10.2307/1907683.

Nie, X. Y., and X. J. Liu. 2013. "Types of "Ghost Towns" in the Process of Urbanization and Countermeasures (In Chinese, Chengshihua Zhongde Guicheng Xianxiang Fenxi)." *Journal of Nantong University* 29 (4): 111–117.

Niu, X. 2018. "Estimating Housing Vacancy Rate in Qingdao City with NPP-VIIRS Nighttime Light and Geographical National Conditions Monitoring Data." Paper presented at The ISPRS Technical Commission III Midterm Symposium on "Developments, Technologies and Applications in Remote Sensing", Beijing.

Ou, J., X. Liu, X. Li, M. Li, and W. Li. 2015. "Evaluation of NPP-VIIRS Nighttime Light Data for Mapping Global Fossil Fuel Combustion CO2 Emissions: A Comparison with DMSP-OLS Nighttime Light Data." *PloS One* 10 (9): e0138310. doi:10.1371/journal.pone.0138310.

Rosen, K. T., and L. B. Smith. 1983. "The Price-Adjustment Process for Rental Housing and the Natural Vacancy Rate." *American Economic Review* 73 (4): 779–786.

Schultz, L. A., T. Cole, and A. L. Molthan. 2015. "Supporting Disaster Assessment and Response with the VIIRS Day-Night Band." Paper presented at the 95th American Meteorological Society (AMS) Annual Meeting, Phoenix, AZ.

Shi, K., C. Huang, B. Yu, B. Yin, Y. Huang, and J. Wu. 2014a. "Evaluation of NPP-VIIRS Night-Time Light Composite Data for Extracting Built-Up Urban Areas." *Remote Sensing Letters* 5 (4): 358–366. doi:10.1080/2150704X.2014.905728.

Shi, K., B. Yu, Y. Huang, Y. Hu, B. Yin, Z. Chen, L. Chen, and J. Wu. 2014b. "Evaluating the Ability of NPP-VIIRS Nighttime Light Data to Estimate the Gross Domestic Product and the Electric Power Consumption of China at Multiple Scales&58; A Comparison with DMSP-OLS Data." *Remote Sensing* 6 (2): 1705–1724. doi:10.3390/rs6021705.

Small, C., C. D. Elvidge, and K. Baugh. 2013. "Mapping Urban Structure and Spatial Connectivity with VIIRS and OLS Night Light Imagery." Paper presented at the Urban Remote Sensing Event, Sao Paulo.

Song, Z., and X. Huang. 2013. "The Legislative Issues of Unreasonable Housing Vacancy Rate in China: AnEnviromental Protection Law Perspective." *Journal of Beijing University of Aeronautics & Astronautics* 26 (2): 51–54.

Stokes, E. C., M. O. Roman, and K. C. Seto. 2014. "The Urban Social and Energy Use Data Embedded in Suomi-NPP VIIRS Nighttime Lights: Algorithm Overview and Status." Paper presented at the AGU Fall Meeting, San Francisco.

Tan, L., Y. Zhou, and L. Bai. 2017. *Human Activities along Southwest Border of China: Findings Based on DMSP/OLS Nighttime Light Data*. Social Science Electronic Publishing, Beijing.

Truman, C. 2014. "China's Housing Bubble: The Future of Urbanization and Sustainable Green Development." Paper presented at the International Development Seminar, Phoenix, AZ.

Wang, H., and C. J. Chang. 2013. "Simulation of Housing Market Dynamics: Amenitydistribution and Housing Vacancy." *Simulation Conference,* Washington, DC, 1673–1684.

Weinberg, D. H. 2006. "How the United States Measures Well-Being in Household Surveys." *Journal of Official Statistics* 22 (1): 113–136.

Xie, Y., and Q. Weng. 2016. "World Energy Consumption Pattern as Revealed by DMSP-OLS Nighttime Light Imagery." *Mapping Sciences & Remote Sensing* 53 (2): 265–282. doi:10.1080/15481603.2015.1124488.

Xu, M., C. He, Z. Liu, and Y. Dou. 2016. "How Did Urban Land Expand in China between 1992 and 2015? A Multi-Scale Landscape Analysis." *PloS One* 11 (5): e0154839. doi:10.1371/journal.pone.0154839.

Yeung, C. W., and R. Howes. 2006. "The Role of the Housing Provident Fund in Financing Affordable Housing Development in China." *Habitat International* 30 (2): 343–356. doi:10.1016/j.habitatint.2004.02.007.

Ying, L., T. Delahunty, N. Zhao, and G. Cao. 2016. "These Lit Areas are Undeveloped: Delimiting China's Urban Extents from Thresholded Nighttime Light Imagery." *International Journal of Applied Earth Observation & Geoinformation* 50: 39–50. doi:10.1016/j.jag.2016.02.011.

Yu, B., T. Min, Q. Wu, C. Yang, S. Deng, K. Shi, P. Chen, J. Wu, and Z. Chen. 2018. "Urban Built-Up Area Extraction From Log- Transformed NPP-VIIRS Nighttime Light Composite Data." *IEEE Geoscience & Remote Sensing Letters* 15 (8): 1279–1283. doi:10.1109/LGRS.2018.2830797.

ZCFE (Zhicheng of finance and economics). 2019. "Chinese Cities GDP Ranking in 2018" *Southmoney*, January 28. http://www.southmoney.com/shuju/hysj/201901/2930814.html

Zhang, C., S. Jia, and R. Yang. 2016. "Housing Affordability and Housing Vacancy in China: The Role of Income Inequality." *Journal of Housing Economics* 33: 4–14. doi:10.1016/j.jhe.2016.05.005.

Zhang, Q., and K. C. Seto. 2013. "Can Night-Time Light Data Identify Typologies of Urbanization? A Global Assessment of Successes and Failures." *Remote Sensing* 5 (7): 3476–3494. doi:10.3390/rs5073476.

Zhang, Q., P. Wang, H. Chen, Q. Huang, H. Jiang, Z. Zhang, Y. Zhang, X. Luo, and S. Sun. 2017. "A Novel Method for Urban Area Extraction from VIIRS DNB and MODIS NDVI Data: A Case Study of Chinese Cities." *International Journal of Remote Sensing* 38 (21): 6094–6109.

Zheng, Q., Z. Ying, J. Deng, W. Ke, R. Jiang, and Z. Ye. 2017. ""Ghost Cities" Identification Using Multi-Source Remote Sensing Datasets: A Case Study in Yangtze River Delta." *Applied Geography* 80: 112–121. doi:10.1016/j.apgeog.2017.02.004.

Zhou, J. G., and G. Q. Yang. 2012. "VAR-Model Based Research on the Relationship between Commercial Housing Vacancy Rate and CPI in China." *Journal of Guangdong University of Technology* 12 (2): 45–49.