Cities 103 (2020) 102745

Contents lists available at ScienceDirect

Cities

journal homepage: www.elsevier.com/locate/cities

Spatiotemporal dynamics and the contributing factors of residential vacancy at a fine scale: A perspective from municipal water consumption

Yongting Pan[a,b], Wen Zeng[a], Qingfeng Guan[a,b,*], Yao Yao[a,b,c], Xun Liang[a,b], Hanqiu Yue[a,b], Yaqian Zhai[a,b], Junyi Wang[a,b]

[a] School of Geography and Information Engineering, China University of Geosciences, Wuhan 430078, Hubei province, China
[b] National Engineering Research Center of GIS, China University of Geosciences, Wuhan 430078, Hubei province, China
[c] Department of Data Technology and Products, Alibaba Group, Hangzhou 311121, Zhejiang province, China

ARTICLE INFO

Keywords:
Vacant residences
Municipal water consumption
Spatiotemporal dynamics
Service environment

ABSTRACT

With the unprecedented pace of urbanization in recent decades, urban vacant residences have increased dramatically. The problem of numerous vacant residences has been severely criticized. The lack of comprehensive data obstructs revealing the spatiotemporal dynamics and the contributing factors of the vacant residences at a fine scale. This study proposes a feasible and general-purpose framework to analyze the vacant residences using municipal water consumption data. First, a residential vacancy identification method is proposed to identify the vacant residences, and the spatiotemporal characteristics of the vacant residences are analyzed using the spatial analysis methods. Second, the relationships between the residential vacancies and the service environments of the neighborhoods are examined using the random forest method. The framework was applied in Changshu city, China, and the results revealed a concentric circular spatial structure for residences. The results also indicated that the diversity of the facilities and services (FASs) has the highest impact on residential vacancies. As the degree of enrichment of the FASs approximates 1 or the percentage/diversity of the FASs increase, the residential vacancy length decreases. This study not only draws a detailed picture of vacant residences but also sheds light on policy implications for mitigating residential vacancies.

1. Introduction

Drastic urbanization has taken place in China since the implementation of the reform and opening-up policy in 1978 (Bai, Shi, & Liu, 2014; He, Chen, Mao, & Zhou, 2016; Jiang, Mohabir, Ma, & Zhu, 2017). Real estate development in China has been rising at an unprecedented pace. According to the Urban Construction Statistical Yearbook (2002–2017), the residential area in China has increased by 4.64% year on year, greatly faster than the growth of urban population (1.06% year on year) over the last two decades. Such imbalance between the real estate development and the residential demand has been one of the most important driving forces of the increasing number of vacant residences in China (Jin et al., 2017; Zheng, Wang, & Cao, 2014). The residential vacancy in China is one of the side effects of fast economic growth (Lu, Zhang, Liu, Ye, & Miao, 2018; Shepard, 2015). Shepard (2015) stated that “China is the world's most populated country without a doubt has the world's largest number of empty homes.” The emergence of a large number of vacant residences is problematic, as they (1) decrease the quality of landscapes (Konomi, Sasao, Hosio, & Sezaki, 2017), (2) decrease the vitality of community life (Konomi et al., 2017; Molloy, 2016), (3) increase the risk of fire (Konomi et al., 2017; Kumagai, Matsuda, & Ono, 2016), (4) decrease property values (Morckel, 2014; Nam, Han, & Lee, 2016), and (5) increase the crime rate (Konomi et al., 2017; Molloy, 2016; Morckel, 2014; Nam et al., 2016). Therefore, identifying and analyzing urban vacant residences can contribute to urban planning, urban management and community security.

Up to now, the Chinese government has not published any official statistics or alternative information concerning vacant residences (Chi, Liu, Wu, & Wu, 2015; Zheng et al., 2017). Even in some countries (e.g., USA), the residential vacancy data is periodically collected via field survey, considerable time, labor and capital are required, and the coverage of such survey can hardly be thorough and comprehensive (Chen et al., 2015; Konomi et al., 2017; Kumagai et al., 2016). Hence the available data are employed in giving us new insights into residential vacancy.

* Corresponding author at: School of Geography and Information Engineering, China University of Geosciences, Wuhan 430078, Hubei province, China.
E-mail addresses: panyt@cug.edu.cn (Y. Pan), zengwen@cug.edu.cn (W. Zeng), guanqf@cug.edu.cn (Q. Guan), yaoy@cug.edu.cn (Y. Yao), liangxun@cug.edu.cn (X. Liang), qiuhy_yhqiu@cug.edu.cn (H. Yue), zyq2017@cug.edu.cn (Y. Zhai), bigplato@cug.edu.cn (J. Wang).

https://doi.org/10.1016/j.cities.2020.102745
Received 2 July 2019; Received in revised form 1 March 2020; Accepted 21 April 2020

Remote sensing has been widely applied for monitoring urbanization due to its capability of gathering land surface information with wide and consistent coverage (Deng, Wang, Hong, & Qi, 2009; Xie & Weng, 2016). Nighttime light (NTL) remote sensing images can effectively represent human socioeconomic activity through nighttime luminosity (Fan, Ma, Zhou, Zhou, & Xu, 2014; Zhang & Su, 2016; Zhou, Hubacek, & Roberts, 2015). Thus, based on NTL data, many studies (Chen et al., 2015; Ge, Yang, Zhu, Ma, & Yang, 2018; Ma, Tong, Liu, Li, & Ma, 2018; Zheng et al., 2017; Zheng, Zeng, et al., 2017) extracted the built-up area and proposed a ghost city index to estimate the degree of urban vacancy. Ge et al. (2018) further explored the correlation between the ghost city index and economic indicators at the provincial scale.

The emergence of spatial big data brings opportunities to understand urban space at the microscopic scale (Gao, Janowicz, & Couclelis, 2017; Kang, Liu, Ma, & Wu, 2012; Liu, Wang, Xiao, & Gao, 2012). Widely used location-aware devices (LAD) and location-based services (LBS), such as mobile phones and social media, can generate a tremendous amount of individual trajectory data characterizing human socioeconomic activities. Chi et al. (2015) counted the vacant housing area of each county in China using Baidu positioning data and points of interest (POIs) and found vacant housing for tourist sites in nontourist seasons. Jin et al. (2017) evaluated the vacancy possibilities of various cities in China through urban vitality measured by road junctions, POIs and LBS data.

Although previous studies bring insights into the vacant residences of urban areas based on NTL data or LBS data, there are still limitations on both data: (1) Coarse spatial resolutions. For example, the NTL data have a low spatial resolution (generally 1000 m or 500 m). (2) Low accuracy due to biased representation and indirect inference. In the aspect of biased representation, the NTL data cannot guarantee accurate identification of residential areas owing to urban lighting systems. LBS data are overreliant on how the LAD and positioning applications are used and who uses them. Additionally, indirect relationships between the NTL/LBS data and vacant residences are used to estimate residential vacancies. (3) Limited time span and coarse temporal resolution. The NPP-VIIRS NTL data have been released monthly or annually since 2012. Starting in 1992, DMSP-OLS NTL data were produced every calendar year. The rise of LBS data is completely dependent on the development of location-aware devices and location technologies in recent years. The temporal resolution may vary over time, depending on how the devices and/or services are used.

Due to the above limitations of the data, two issues remain insufficiently studied. First, the long-term spatiotemporal dynamics of vacant residences at fine scales are difficult to describe. Recently, cities in China have experienced both spatial expansion and internal transformation (i.e., urban redevelopment and urban renewal). The small-scale urban redevelopment is increasing dramatically (Long, Mao, Mao, Shen, & Zhang, 2014). However, achieving effective spatiotemporal dynamic analysis at a fine scale is difficult using remote sensing data and commonly used spatial big data. Second, the contributing factors of residential vacancies at fine scales are rarely explained. The factors and their mechanism of influencing residential vacancies are quite complicated with significant spatial heterogeneity. Previous studies only suggest that some economic and social factors affect residential vacancies at urban and even provincial scales. The facilities and services (FASs) that the residents interact with, as the key components of the residential neighborhood service environment (Chaix et al., 2014), must be considered in residential issues (He, Mol, & Lu, 2016; Niu, Dong, Niu, & Deng, 2017; Ströbele & Hunziker, 2017). However, few studies to our knowledge exist that examine the relationships between residential vacancies and the facilities and services.

Compared with the NTL data and LBS data, municipal infrastructure and service data (e.g., consumption of water) have the advantages of fine-grained spatiotemporal resolution, comprehensive coverage of the population and socioeconomic activities, and long temporal span. The municipal infrastructure and service data present new opportunities to objectively measure and model the nature and likely causes for the activity patterns of residents (Lansley, Li, & Longley, 2017; Lansley & Longley, 2016). In this study, we focus on analyzing vacant residences at a fine scale via the municipal water consumption data. First, we explicitly delineate the spatiotemporal dynamics of vacant residences in Changshu, a rapidly developing county-level city in China, thereby revealing the spatial expansion and internal transformation of vacant residences in various periods within the city. Second, we quantitatively examine the relationships between the residential vacancies and the residential neighborhood service factors (NSFs) to explore the influence of the FASs in the residential environment on the residential vacancies at a fine scale. This study attempts not only to fill some gaps in the explicit description of residential vacancies by sketching spatiotemporal dynamics and exploring contributing factors at a fine scale but also to provide the spatiotemporal characteristics of the vacant residences within the city as well as recommendations for mitigating the residential vacancies.

2. Study area and data

2.1. Study area

Changshu, in the hinterland of the Yangtze River Delta, is a county-level city located in the southeast of Jiangsu Province, China (Fig. 1). The spatial coverage of Changshu ranges from 31°31′N to 31°50′N, 120°33′E to 121°03′E. Changshu has been sharply urbanized and industrialized with a booming economy (Li, Long, & Liu, 2010; Zhou, Zhang, Ye, Wang, & Su, 2016). According to the Urban Construction Statistical Yearbook (2004–2013), Changshu had a total administrative area of 1264 km^2 at the end of 2013, and the residential area of Changshu expanded from 15.88 km^2 in 2004 to 34.87 km^2 in 2013. The urban population increased from 630,000 in 2004 to 768,800 in 2012, and the urbanization rate of the population increased from 60.4% in 2004 to 72.0% in 2012, but the population density decreased from 2131 person/km^2 in 2004 to 1197 person/km^2 in 2013. The Changshu Statistical Yearbook (2004–2013) and the China Statistical Yearbook (2004–2013) show that the gross domestic product (GDP) per capita in Changshu increased from 54,314 yuan (RMB) in 2004 to 131,338 yuan in 2013, which is almost 4 times as much as the GDP per capita in China. The proportions of the primary, secondary and tertiary sectors in Changshu are approximately 2%, 57%, and 41%, respectively, of its total GDP each year. The disposable income of urban households (DIUH) per capita and the net income of rural residents (NIRR) per capita in Changshu increased from 15,039 yuan in 2004 to 43,161 yuan in 2013 and from 7517 yuan in 2004 to 21,691 yuan in 2013, respectively. These income levels are almost 1.6 times and 2.4 times the national average. Such intense residential expansion, low population density and prosperous economy make it possible to emerge residential vacancies. Changshu has been recognized in previous studies as a city with a large amount of vacant residences (Chi et al., 2015; Su, 2014).

2.2. Data

This study was conducted using the municipal water consumption data in Changshu from January 2004 to December 2013 provided by the Municipal Water Company of Changshu for research purposes. Each record of the data represents the water consumption history of a particular customer during the study period, including the customer ID (account), customer category, spatial coordinates, and monthly water consumption time series. The data also include the cancelation times of customer accounts but no registration times. Therefore, it is only possible to determine when the water supply service was terminated for a particular customer, but it is impossible to determine when the service started.

It is also worth noting that the water consumption time series were

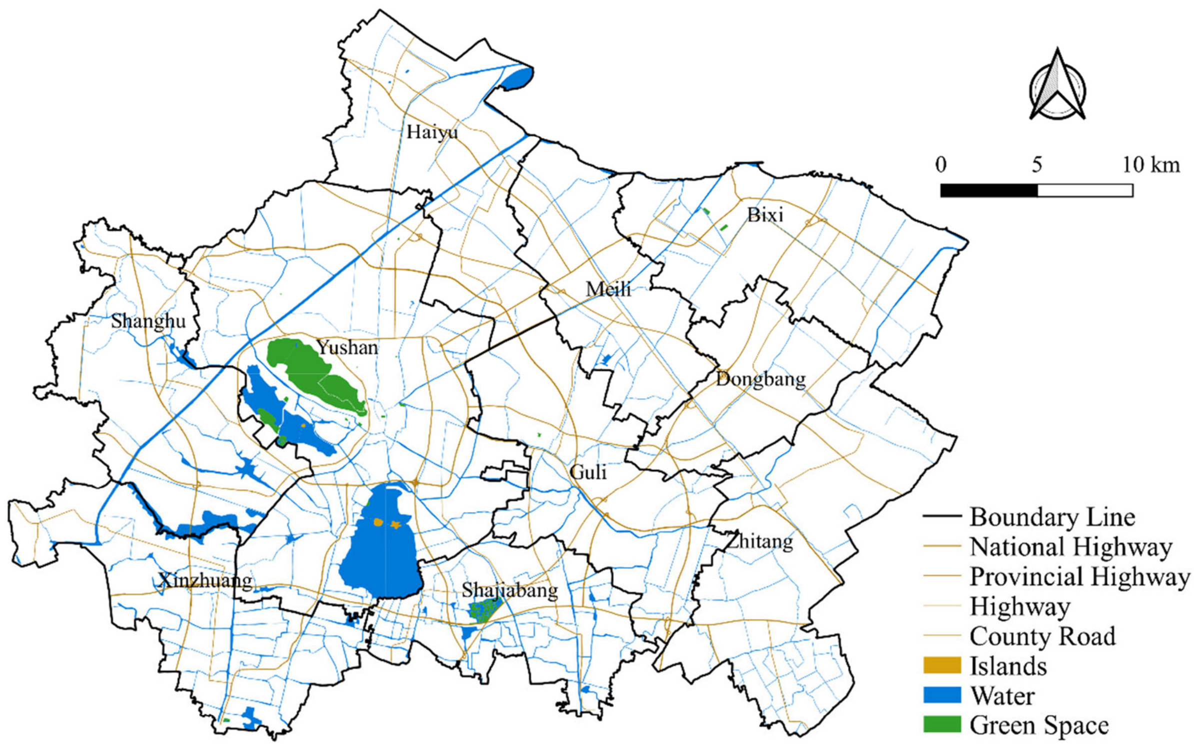


Fig. 1. Case study area: Changshu, Jiangsu, China.

irregular in terms of time span, as the water supply services for customers started and ended at various times. Blanks with various lengths were also found in the time series, some of which were identified as vacancies in this study (the details of the vacancy identification are described in Section 3).

According to the customer category as "resident" or "nonresident", the data were divided into residential water consumption data and nonresidential water consumption data. In total, the residential water consumption data include 286,365 households, growing from 154,248 in 2004 to 276,772 in 2013. The nonresidential water consumption data include a total of 29,096 customers, increasing from 9891 in 2004 to 23,280 in 2013.

For nonresidential water consumption data, the "nonresident" customer category was subdivided into 14 categories of FASs, including government and public service (GOV), industry (IND), business (BUS), financial service (FIN), education facility (EDU), medical service (MED), restaurant (RES), hotel (HOT), shopping (SHP), recreation (REC), life service (LIF), tourist attractions (TOU), automobile service (AUT), and other (OTH). It should be noted though that the nonresidential water consumption data are similar to POIs extracted from various Internet platforms, such as Yahoo! Local, AutoNavi Map and Facebook Places, but without urban-rural biases (Sparks, Thakur, Pasarkar, & Urban, 2019; Zhang, Li, Zhang, Liu, & Du, 2018), as well as spatial and thematic inaccuracy (Bakillah, Liang, Mobasheri, Jokar Arsanjani, & Zipf, 2014; Bordogna, Carrara, Criscuolo, Pepe, & Rampini, 2014).

3. Methodology

3.1. Overview

The residential vacancy analysis framework proposed in this study consists of two main parts: (1) delineating the spatiotemporal dynamics of the vacant residences and (2) exploring the relationships between residential vacancies and neighborhood service factors (NSFs) (Fig. 2). A residential vacancy identification method is proposed using residential water consumption data to recognize the occupied/vacant state of each residence at any time. Spatial analysis methods were employed to understand the spatial distributions of vacant residences at various times. Meanwhile, multisized NSFs representing various characteristics of the FASs within the residential environment were extracted for each residence based on the nonresidential water consumption data. The relationships between residential vacancies and NSFs were examined by an ensemble of classification and regression trees (CART) - random forest.

3.2. Delineating spatiotemporal dynamics of vacant residences

3.2.1. Residential vacancy identification method

To identify whether a residence is vacant or occupied at a certain time, this study developed a residential vacancy identification method (RVIM), which includes two key steps: (1) defining a residential vacancy and (2) establishing a residential state time series.

To our knowledge, currently there is no widely recognized definition of a vacant residence. The definition of vacant residence differs in different countries and studies depending on the standards of usage. As water is a necessity for residential activities, water consumption can be seen as a key indicator for residential vacancy and thus provides a feasible and effective measure of residential vacancy. Based on previous studies (Ami & George, 2014; Haramati & Hananel, 2016; Nam et al., 2016), this study defined a vacant residence as a household with no record of water consumption for several consecutive months or a very small amount of water consumption for each three-month period. This definition involves two variable parameters: the minimum number of months *minMhs* with no water consumption and the minimum water consumption *minWC* for each three-month period. The parameters vary across countries and/or cities. For Changshu, we set *minMhs* as 6 months and *minWC* as 3.6 m^3/month (in nonsummer) or 5.5 m^3/month (in summer) (as the minimum amount of water required by a single-person residence), by referring to the national standards and policies, including the standard of water quantity for the city's residential use (GB/T 50331–2002) and the security law of low-rent

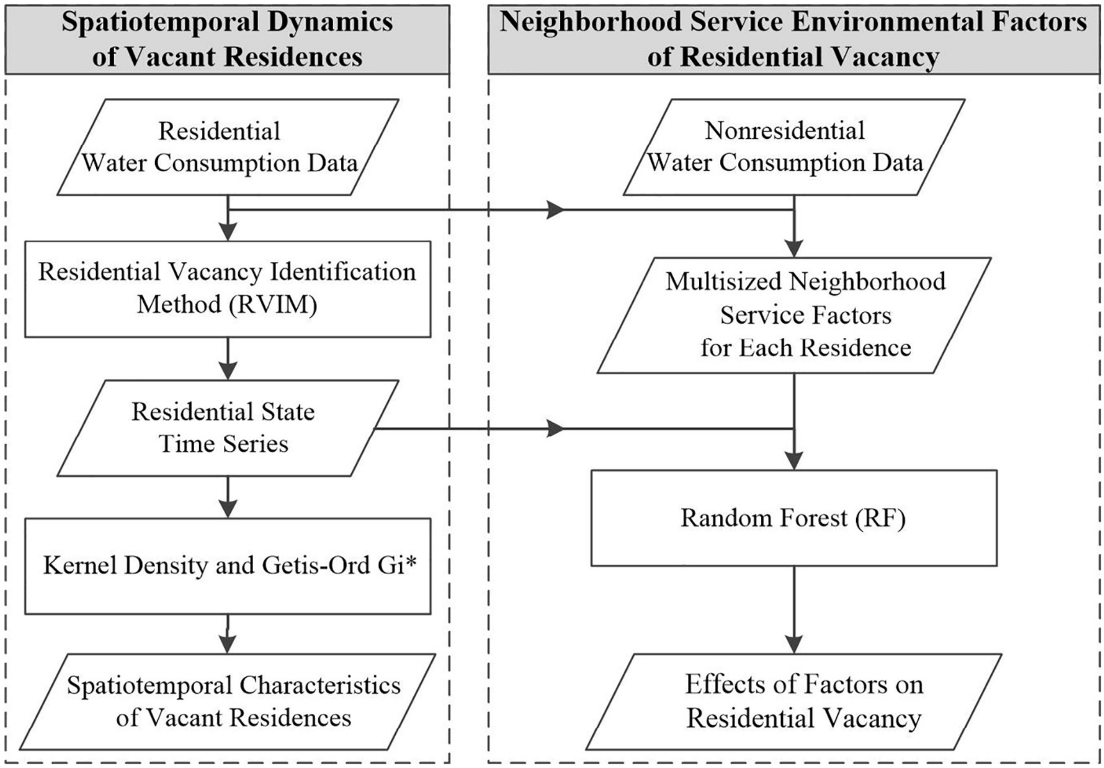


Fig. 2. The methodological framework of residential vacancy analysis.

housing in China. Moreover, a constant parameter in the definition, each three-month period, is noteworthy. If the period of time is too short (e.g., one month), an erroneous judgment is generated owing to abnormal water consumption. If the period of time is too long (e.g., six months), short-term vacancies cannot be detected. The period of time was set to three months to reduce the negative impact of abnormal water consumption and to identify short-term vacancies.

Rule-based systems (RBSs) were employed in this study to establish the residential state time series using the residential water consumption time series. RBSs are a special type of expert system for codifying the problem-solving know-how of human experts (Hayes-Roth, 1985). RBSs are able to directly reproduce deductive reasoning mechanisms as well as to explain how their outcomes have been achieved by a set of condition-action (if-then) rules (Hayes-Roth, 1985; Minutolo, Esposito, & De Pietro, 2017). Hence, RBSs can be used to solve a wide range of possibly complex problems in many application areas, such as healthcare, transportation and security (Hayes-Roth, 1985; Liu, Gegov, & Stahl, 2014). In this study, a collection of conditions was generated from the residential water consumption time series, and a series of actions representing the residential state were achieved.

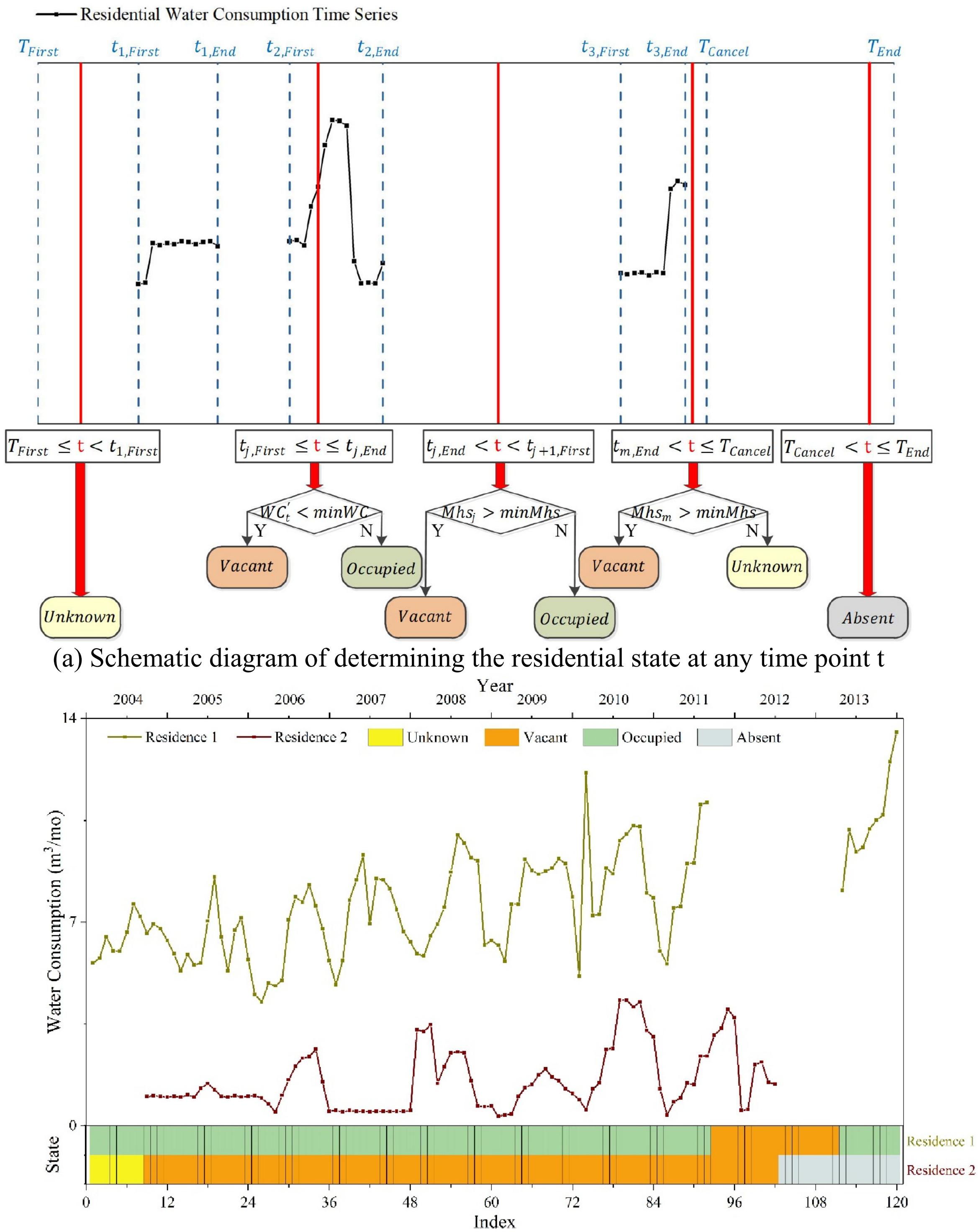
As shown in Fig. 3(a), T_{First} and T_{End} represent the start time and end time during the period of study. For a residential water consumption time series composed of m subsequences (with blanks in between), T_{Cancel} represents the cancellation time of the residential customer account, and $t_{j,\ First}$ and $t_{j,\ End}$ represent the start and end times of the j-th ($1 \le j \le m$) sequence. WC_t' represents the adjusted water consumption at time point t (excluding blanks), which is the mean of the water consumption at three time points $t-1$, t and $t+1$. Mhs_j represents the number of months between the j-th sequence and the (j + 1)-th sequence (i.e., the length of a blank). The following six rules were set to determine RS_t, the residential state at a particular time point t[1] (Fig. 3(a)):

(1) When t is between T_{First} and $t_{1,\ First}$, RS_t is set as a nominal value *Unknown*, indicating that the blank may be caused by any of the following reasons: (a) the residence has not been built yet at this point; (b) the customer account has not been registered yet at this point; or (c) the account has been registered and the residence is vacant;
(2) When t is between $t_{j,\ First}$ and $t_{j,\ End}$, if WC_t' is less than *minWC*, RS_t is set as *Vacant*, indicating that the residence is unoccupied at time point t; otherwise, RS_t is set as *Occupied*;
(3) When t is between $t_{j,\ End}$ and $t_{j+1,\ First}$, if Mhs_j is more than *minMhs*, RS_t is set as *Vacant*; otherwise, RS_t is set as *Occupied*;
(4) When t is between $t_{m,\ End}$ and T_{Cancel}, if Mhs_m is more than *minMhs*, RS_t is set as *Vacant*; otherwise, RS_t is set as *Unknown*, indicating that the blank may be caused by either of the following reasons: (a) the delay of document processing for cancelation at the water supply agency; or (b) the customer account has not been canceled and the residence is vacant.
(5) When t is between T_{Cancel} and T_{End}, RS_t is set as *Absent*, indicating that the customer account has been canceled.
(6) When T_{Cancel} does not exist in the time series and t is between $T_{m,\ End}$ and T_{End}, if Mhs_m is more than *minMhs*, RS_t is set as *Vacant*; otherwise, RS_t is set as *Unknown*, similar to rule (4).

Two examples are shown in Fig. 3(b). The first residential state time series, including *Occupied* and *Vacant*, is obtained by executing rules (2) and (3) on its water consumption time series. The second residential state time series, including *Unknown*, *Vacant* and *Absent*, is gained by executing rules (1), (2) and (5) on its water consumption time series.

Through such RVIM, the numeric residential water consumption time series with blanks and varied lengths were converted into the

[1] A time point can be a particular month, week, day, or hour, depending on the frequency of the time series. For Changshu city in this study, a time point was a month, as the original water consumption time series was collected with a monthly interval.



(b) Examples of the water consumption time series and the residential state time series. The first residential customer account lasted from 2004 to 2013, while the second was canceled in July 2012.

Fig. 3. Schematic diagram (a) and examples (b) of the conversion from the water consumption time series to the residential state time series (i.e., *Absent*, *Vacant*, *Occupied* and *Unknown*).

nominal residential state time series without blanks and with a unified length. Each residential state time series includes several of the four states, i.e., *Absent*, *Vacant*, *Occupied* and *Unknown*.

3.2.2. *Spatiotemporal dynamics of vacant residences*

To better understand the characteristics of the vacant residences in space and time, spatiotemporal distribution analysis methods, including the local indicators of spatial association (or autocorrelation) (LISA) and kernel density estimation (KDE), were applied based on the spatial distribution of customers and their residential state time series.

LISA are statistics that help identify the presence of significant clusters in a spatial dataset by quantifying variations in the local spatial autocorrelation of certain observations relative to the surrounding areas (Anselin, 1995). The Getis-Ord Gi∗ statistic (Getis & Ord, 2010), as a typical LISA, was employed in this study to identify three residential clustering patterns: a relatively long-term (e.g., 11 months) vacant residence surrounded by other relatively long-term vacant residences (i.e., intense concentration of relatively long-term vacant residences, corresponding to the notion of a “hot spot”), an occupied residence surrounded by other occupied residences or a relatively short-term (e.g., 2 months) vacant residence surrounded by other relatively short-term vacant residences (intense concentration of occupied residences or relatively short-term vacant residences, i.e., a “cold spot”), or a lack of a statistically significant contribution to spatial autocorrelation.

Contrary to the Getis-Ord Gi∗, the KDE does not provide a measure of statistical significance. However, when used in combination with the Getis-Ord Gi∗, it offers a medium to visualize the spatial trends by converting discrete information (observation points) into a smoothed and continuous surface of outcome density (Silverman, 2018). In this study, the KDE was used for spatial representation of the quantity of vacant residences. A high kernel density indicates a large number of long-term vacant residences concentrated in space, and vice versa.

The Getis-Ord Gi∗ and KDE were applied for every year of the study period, such that the characteristics of the spatiotemporal dynamics of the vacant residences could be delineated and analyzed.

3.3. Exploring neighborhood service factors of residential vacancy

3.3.1. Multisized neighborhood service factors

To advance our understanding of the impacts of the FASs on residential vacancies, the relationships between the FASs in the neighborhoods of residences and residential vacancies were examined in this study. To do so, multisized neighborhood service factors (NSFs), which represent a variety of characteristics of the FASs within the residential environment, were extracted for each residence, which consists of two key steps: (1) defining the multisized neighborhoods and (2) numerically measuring the NSFs.

In this study, the neighborhood of a residence (referred to as residential neighborhood in the following text) is defined as the area that can be accessed from the residence via walking through the street network within a certain period of time. Such a residential neighborhood represents a behavioral space formed by the resident to carry out various home-centered daily activities, including shopping, entertainment, work/study, medical treatment, and others (Xiao, Chai, & Zhang, 2014). To capture the effects of the FASs at different distances, four residential neighborhoods with different sizes were generated in this study, using 10-min, 20-min, 30-min, and 40-min walking distances over the street network (Fig. 4). The first neighborhood describes the area where the residents hardly notice the cost of time. The second and third represent the areas where strongly concentrated activities occur. The last neighborhood, as the pedestrian limit tolerated by residents, covers almost all daily residence-related activities (Chai, Shen, & Weng, 2008). As suggested by Anderson and Pandy (2001), the average walking speed of a person is 81 m/min, which was used in this study to generate the neighborhoods.

With the four neighborhoods defined by the typical walking distances of each residence, four sets of the NSFs were extracted to represent the characteristics of the FASs within the residential environment that may influence residential vacancy. The NSFs include two major categories: the proportion of each type of the FASs within a neighborhood and the diversity of all the types of FASs within a neighborhood. The descriptions of the NSFs are shown in Table 1.

The proportion the NSFs offer information related to the intensity of a particular type of FASs within a neighborhood (Yan et al., 2013). We used (1) enrichment factor (EF) (Zhang, Lu, Shi, & Pan, 2018), (2) within-district percentage (WDP) (Yan et al., 2013) and (3) areal percentage (AP) (Yan et al., 2013). Suppose a study area includes M types of FASs ($M = 14$ in this study), and N neighborhoods are generated for each residence ($N = 4$ in this study), the proportion of NSFs are calculated as:

$$EF_{ij} = \frac{F_{ij}/FN_j}{FC_i/FC} \tag{1}$$

$$WDP_{ij} = \frac{F_{ij}}{FN_j} \tag{2}$$

$$AP_{ij} = \frac{F_{ij}}{FC_i} \tag{3}$$

where EF_{ij}, WDP_{ij} and AP_{ij} represent the EF, WDP and AP of the i-th ($1 \le i \le M$) type of the FASs within the j-th ($1 \le j \le N$) neighborhood, respectively. F_{ij} is the number of the i-th type of the FASs within the j-th neighborhood. FN_j is the number of all the FASs within the j-th neighborhood. FC_i is the number of the i-th type of the FASs in the city. FC is the number of all the FASs in the city.

Eight diversity measures were used in this study for their extensive adoption in a variety of geospatial applications (McGarigal, 2014): (1) Shannon's diversity index (SHDI) (Gorelick, 2010; Rajaram et al., 2017), (2) Simpson's diversity index (SIDI) (Jaeger, 2000; Mclaughlin et al., 2016), (3) modified Simpson's diversity index (MSIDI) (Gorelick, 2010), (4) Shannon's evenness index (SHEI) (Kvålseth, 2015; Pielou, 1966b), (5) Simpson's evenness index (SIEI) (Kvålseth, 2015; Smith & Wilson, 1996), (6) modified Simpson's evenness index (MSIEI) (Kvålseth, 2015; Smith & Wilson, 1996), (7) splitting index (SPLIT) (Jaeger, 2000), and (8) species richness index (SRI) (Luo, Zhang, Yang, Song, & Cui, 2014; Pielou, 1966a). These factors are denoted with $SHDI_j$, $SIDI_j$, $MSIDI_j$, $SHEI_j$, $SIEI_j$, $MSIEI_j$, $SPLIT_j$ and SRI_j, respectively, reflecting the diversity of the FASs within the j-th neighborhood of a residence, and they are calculated as follows:

$$SHDI_j = -\sum\nolimits_{i=1}^{M} WDP_{ij} \ln WDP_{ij} \tag{4}$$

$$SIDI_j = 1 - \sum\nolimits_{i=1}^{M} WDP_{ij}^2 \tag{5}$$

$$MSIDI_j = -\ln \sum\nolimits_{i=1}^{M} WDP_{ij}^2 \tag{6}$$

$$SHEI_j = -\frac{\sum_{i=1}^{M} WDP_{ij} \ln WDP_{ij}}{\ln M} \tag{7}$$

$$SIEI_j = \frac{1 - \sum_{i=1}^{M} WDP_{ij}^2}{1 - \frac{1}{M}} \tag{8}$$

$$MSIEI_j = -\frac{\ln \sum_{i=1}^{M} WDP_{ij}^2}{\ln M} \tag{9}$$

$$SPLIT_j = -\frac{FN_j^2}{\sum_{i=1}^{M} F_{ij}^2} \tag{10}$$

$$SRI_j = -\frac{M-1}{\ln FN_j} \tag{11}$$

Eventually, four sets of the NSFs for each residence were extracted, corresponding to the 10-min, 20-min, 30-min, and 40-min residential neighborhoods. Each set of factors consists of 42 proportion measures (3 × 14) and 8 diversity measures. In total, 200 NSFs were obtained for each residence.

3.3.2. Relationships between residential vacancy and NSFs

In this study, the random forest (RF)-derived CART (Breiman, 2001) was employed to quantitatively examine the relationships between the residential vacancies and NSFs. A forest is an ensemble of trees, and the

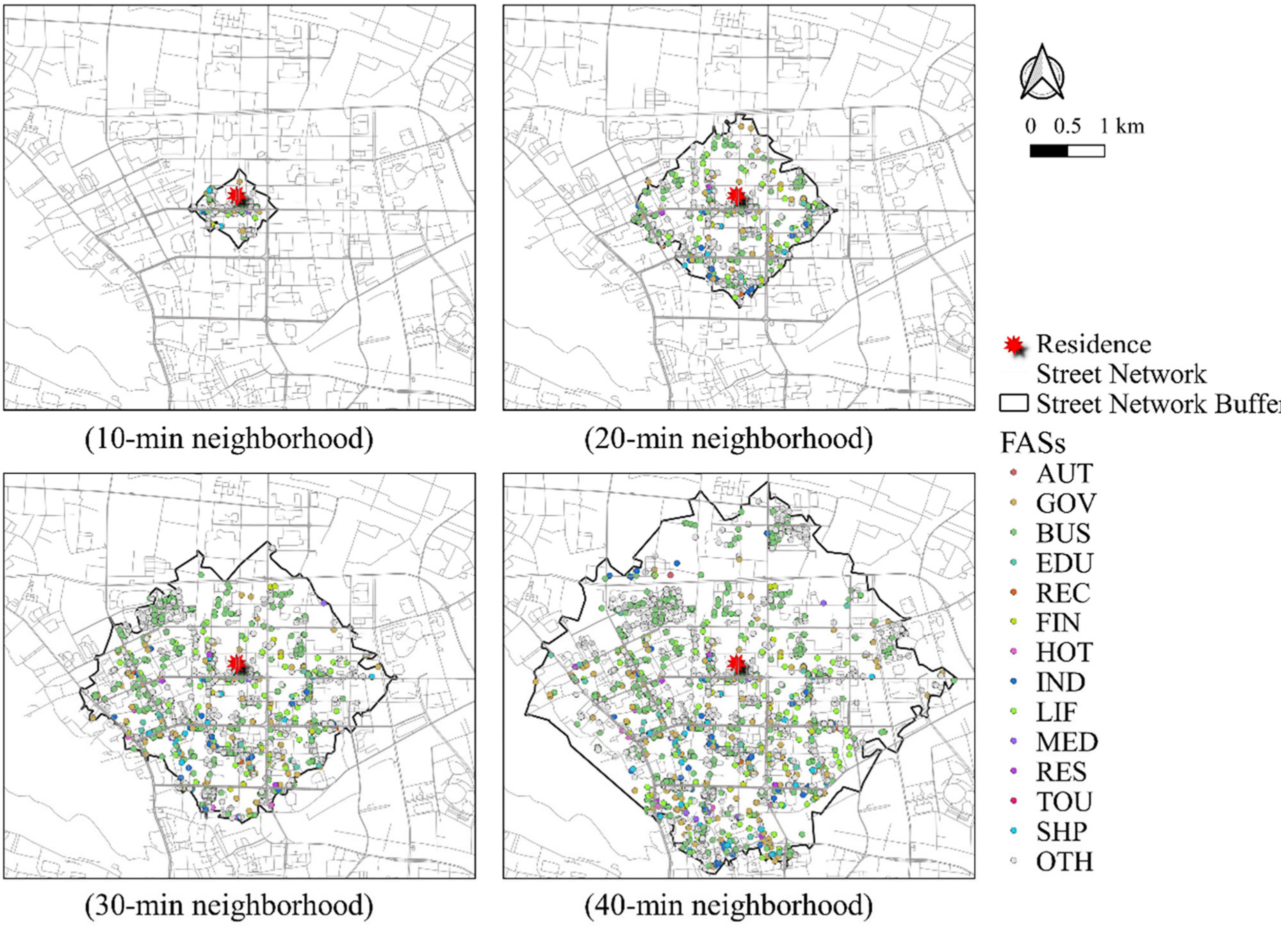


Fig. 4. Examples of multisized residential neighborhoods with the FASs for a residence.

RF has two random processes: (1) each tree is based on a random subset of the observations (e.g. residences in our case), and (2) each split within each tree is determined from a random subset of the candidate variables (e.g. NSFs in our case) (Breiman, 2001). Owing to the stochastic mechanism of the RF, slightly different results might be obtained in each trial (Lu, Im, Rhee, & Hodgson, 2014; Walton, 2015). It has been reported that the RF is not prone to overfitting, with stable performance on multidimensional and highly nonlinearly correlated

Table 1
Description of NSFs.

Factors		Description
Proportion	Enrichment Factor (Zhang, Lu, Shi, & Pan, 2018)	The ratio between the percentage of a certain type of FASs within a certain neighborhood and the percentage of this type of FASs in the entire city
	Within-district Percentage (Yan, Merlin, & Rodriguez, 2013)	The percentage of a certain type of FASs within a certain neighborhood relative to the number of all FASs within the neighborhood (Yan et al., 2013)
	Areal Percentage (Yan et al., 2013)	The percentage of a certain type of FASs within a certain neighborhood relative to the number of this type of FASs in the entire city (Yan et al., 2013)
Diversity	Shannon's Diversity Index (Gorelick, 2010; Rajaram, Castellani, & Wilson, 2017)	The abundance (probability) distributions of each type of FASs within a certain neighborhood (Gorelick, 2010; Rajaram et al., 2017), based on the notion of information (Rajaram et al., 2017)
	Simpson's Diversity Index (Jaeger, 2000; Mclaughlin, Mclaughlin, Mclaughlin, & White, 2016)	The abundance (probability) distributions of each type of FASs within a certain neighborhood, reflecting the dispersion in multiple types of FASs (Mclaughlin et al., 2016)
	Modified Simpson's Diversity Index (Gorelick, 2010)	Analogy with SIDI, and more sensitive to slight changes in the FASs richness (Gorelick, 2010)
	Shannon's Evenness Index (Kvålseth, 2015; Pielou, 1966b)	The typical function of Shannon's diversity measure and the number of FASs categories (Kvålseth, 2015)
	Simpson's Evenness Index (Kvålseth, 2015; Smith & Wilson, 1996)	The typical function of Simpson's diversity measure and the number of the FASs categories (Kvålseth, 2015)
	Modified Simpson's Evenness Index (Kvålseth, 2015; Smith & Wilson, 1996)	The typical function of the modified Simpson's diversity measure and the number of the FASs categories (Kvålseth, 2015), and more sensitive to slight changes in the FASs evenness (Gorelick, 2010)
	Splitting Index (Jaeger, 2000)	The 'effective mesh number' of the FASs within a certain neighborhood based on the cumulative amount distribution of the FASs (Jaeger, 2000; McGarigal, 2014)
	Species Richness Index (Luo et al., 2014; Pielou, 1966a)	The percentage of the number of the FASs categories relative to the amount of all the FASs within a certain neighborhood

data, which is scarce, imbalanced, or noisy (Sarica, Cerasa, & Quattrone, 2017). Such properties make RF well-suited for the 200-dimensional NSFs in this study, and to handle the imbalance problem of residences with different vacancy lengths as well as the multiple correlative problem among multisized NSFs. Additionally, RF has been successfully applied in many geographic studies (Acheson, Volpi, & Purves, 2019; Liao, Dong, Huang, Gartner, & Liu, 2019), and demonstrated highly competitive, if not better, performance compared with other commonly used machine learning techniques (e.g., support vector machines and boosting methods) (Fernandez-Delgado, Cernadas, Barro, & Amorim, 2014) with less costs of computational intensity and parameter tuning (Hof & Wolf, 2014).

The RF also presents an important characteristic - the variable importance (VI). The VI measurement has been extensively exploited in different scenarios, for example, to identify the most relevant factors and to identify the most suitable reason to distinguish particular target classes (Belgiu & Drăguţ, 2016). In this study, after adopting the out-of-bag (OOB) error, mean squared error (MSE) and median absolute error (MedAE) as a global measure and spatial absolute error as a local measure to quantify the accuracy between the predicted residential vacancy length and the actual residential vacancy length (i.e., residential state), we examined the contributions of the NSFs to residential vacancies and assessed their role in influencing disparities in residential vacancies.

4. Results

4.1. Spatiotemporal characteristics of vacant residences

Our analysis was performed in yearly time intervals for a total of 10 years. We identified vacant residences in Changshu by the proposed RVIM and then counted the vacancy length of each residence. Fig. 5 shows that the number of vacant residences grows linearly over the years. Among vacant residences, the residences with vacancy lengths of 12 months represent the largest number, followed by the residences with vacancy lengths of 1 month, 3 months and 2 months. It should also be noted that vacant residences with vacancy lengths greater than 6 months exceed half of the vacant residences.

The results of the hot spot analysis, which implemented the Getis-Ord Gi* statistic, are presented in Fig. 6. The distance band (or threshold distance) of the analysis was set to be 500 m, given that the average edge length of a residential area is between 300 m and 500 m.

As shown in Fig. 6(a), the residences in Changshu had sustained a concentric ring structure from 2004 through 2013: the declining city center, prospering inner city, near suburb, and far suburb. The first observation was the rise of hot spots (please see the details of "hot spots" in Section 3.2.2) in the city center, indicating that the number of relatively long-term vacant residences is increasing as their concentration becomes more intense over the years. In contrast, massive cold spots (please see the details of "cold spots" in Section 3.2.2) could be observed in the inner city outside the declining city center and gradually expanding outward. This implies that numerous occupied residences or relatively short-term vacant residences are concentrated there and progressively expanded to the near suburb. These state that residences in Changshu have been developing with both stagnant recession and prosperous expansion.

We observed that new residential forms had changed mostly by replacement or accretion since 2004 (Fig. 6(b)). The city center gradually declined by replacing statistically insignificant residences with hot spots (Fig. 6(b)A). In particular, the number of relatively long-term vacant residences was not much different from the number of occupied residences or relatively short-term vacant residences in 2004. Then, an increasing number of residents moved away, and increasing residences became vacant. Ultimately, in 2012, large numbers of relatively long-term vacant residences were concentrated in the city center. Based on the pattern, we considered that the declining city center mostly corresponds to an urban village or shanty town. Meanwhile, the Bixi new district as an urban subcenter gradually increased by replacing statistically insignificant residences or hot spots with cold spots (Fig. 6(b)B). Contrary to the city center, an increasing number of residents moved there, and occupied residences increased progressively. In addition, in some areas, hot spots were increasing with increasing residences, such as Shanghu town (Fig. 6(b)C). In this case, we considered that an increasing number of vacant residences might result from the demand or desire of residents to own a home, even if they rarely live there.

We created 10 maps using the kernel density by estimating the density of the vacant residences with different vacancy lengths in the neighborhood of each raster cell (100 m × 100 m raster cell size) within a search radius (or bandwidth) of 500 m. The kernel density scores from 2004 through 2013 are shown in Fig. 7.

The binary structure of vacant residential density, center and periphery can be found from 2004 through 2013 (Fig. 7). This reflects the imbalance in the development of vacant residences with different vacancy lengths in various regions of Changshu. Plainly, almost the entire central area of Yushan is covered with the higher kernel density scores (specifically, the kernel density scores more than 9550). This indicates that most long-term vacant residences in Changshu are gathered in the central area of Yushan. Interestingly, the combination of Fig. 7 and

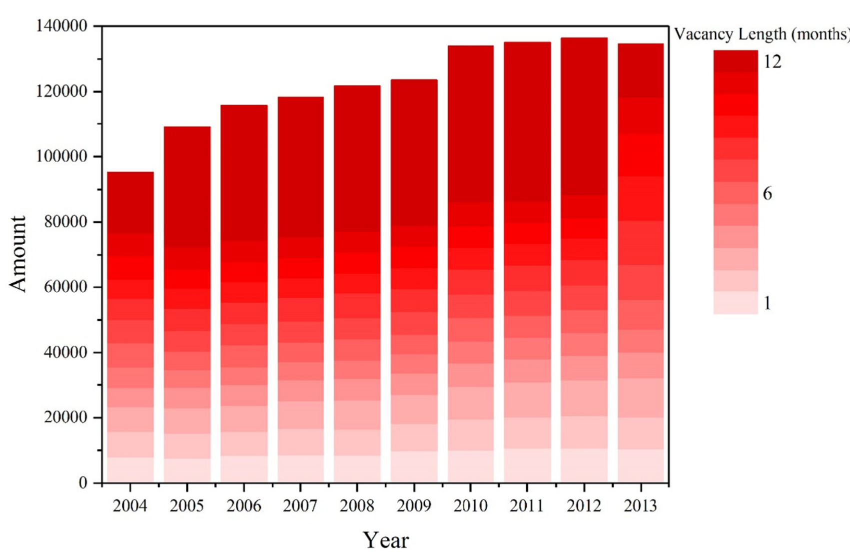
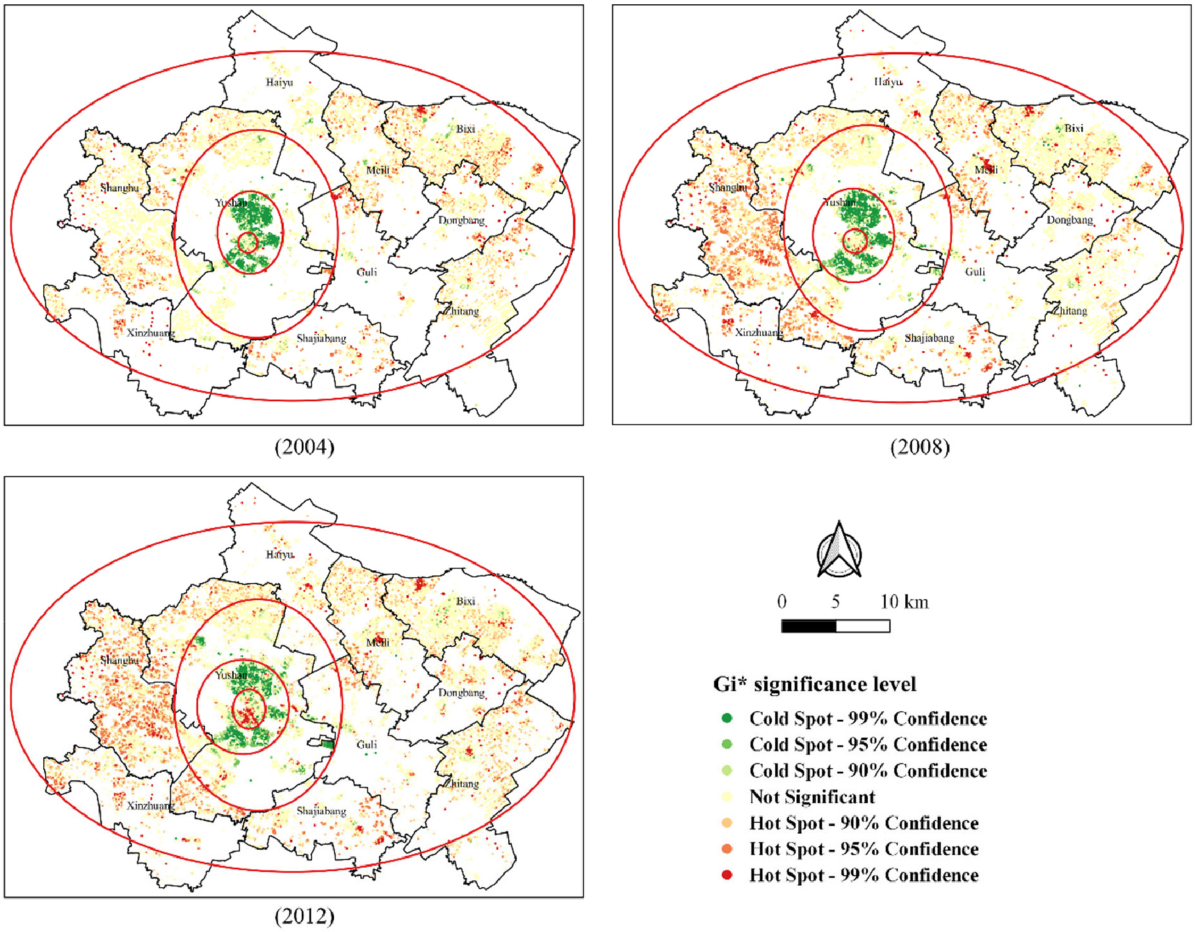
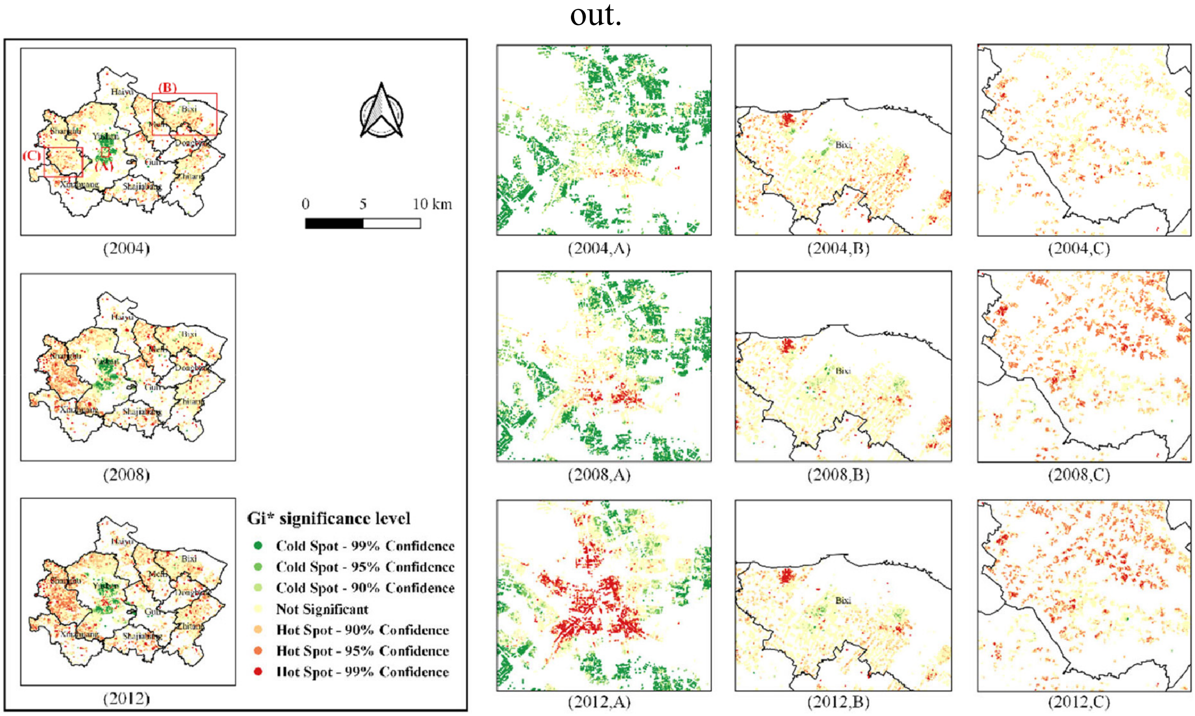


Fig. 5. The number of vacant residences from 2004 through 2013 in Changshu.



(a) The concentric ring structure of residences. The yearly red ellipses represent the declining city center, prospering inner city, near suburb and far suburb from the inside out.



(b) Details of change in three typical residential forms.

Fig. 6. The hot/cold distributions of residences in 2004, 2008, and 2012, in Changshu.

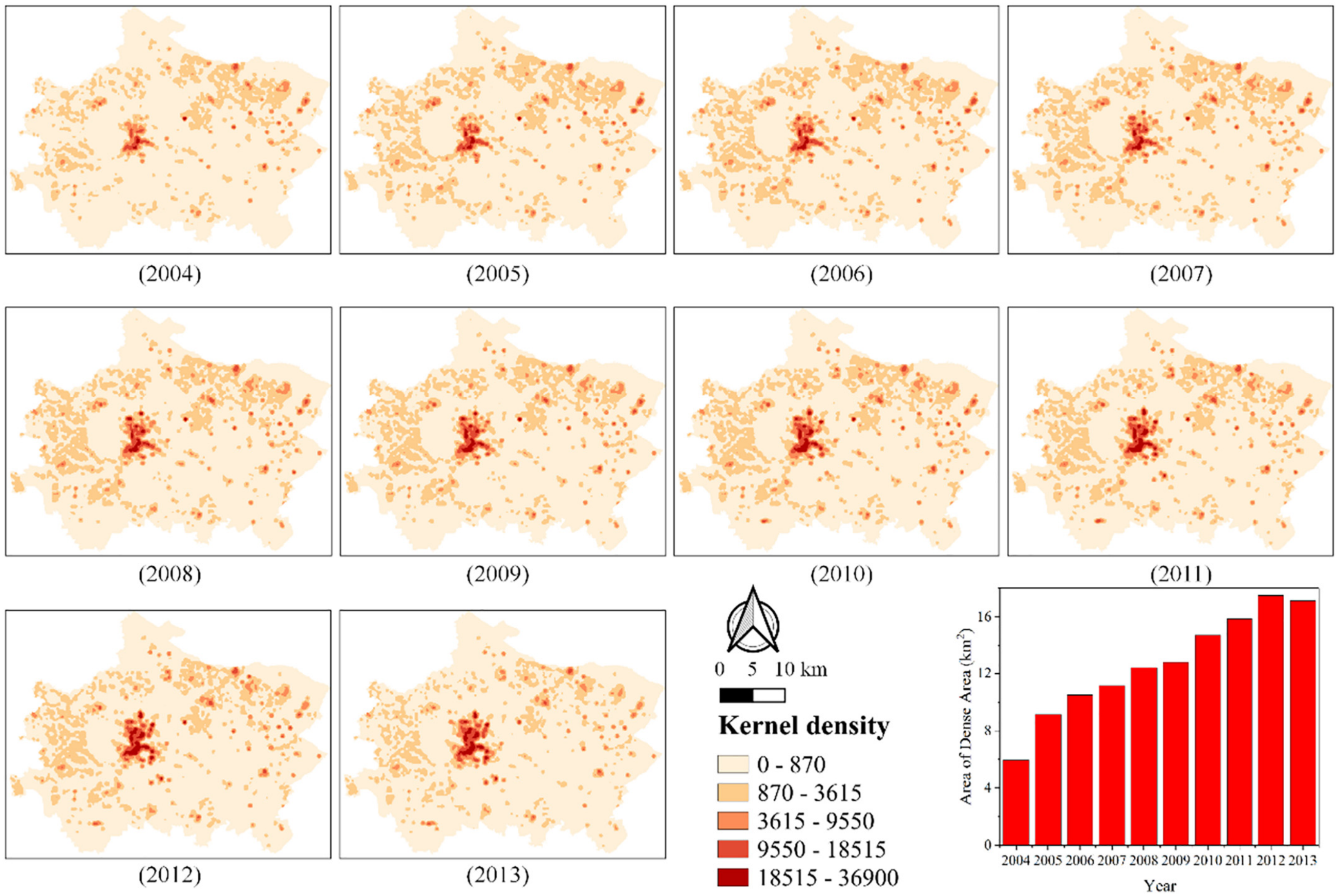


Fig. 7. The kernel density of vacant residences from 2004 to 2013 in Changshu. The kernel density level was graded by natural breaks. The higher the kernel density scores are, the higher the kernel density level.

Table 2
Accuracy assessment results of the relationships analysis.

Year	OOB error	MSE			MedAE		
		Training	Test	Overall	Training	Test	Overall
2004	0.153	0.081	0.127	0.104	0.190	0.256	0.223
2005	0.140	0.087	0.155	0.121	0.189	0.282	0.236
2006	0.122	0.089	0.162	0.126	0.194	0.29	0.242
2007	0.099	0.091	0.165	0.128	0.195	0.292	0.244
2008	0.084	0.093	0.170	0.132	0.201	0.303	0.252
2009	0.071	0.094	0.171	0.133	0.201	0.298	0.250
2010	0.074	0.093	0.170	0.132	0.198	0.294	0.246
2011	0.056	0.095	0.167	0.131	0.201	0.290	0.246
2012	0.033	0.096	0.169	0.133	0.200	0.290	0.245
2013	0.025	0.085	0.132	0.109	0.206	0.270	0.238

Fig. 6(a) shows that the area is just the declining city center and the prospering inner city. The long-term vacant residences in the declining city center are not only large in number but also concentrated. This confirms our inference about the declining city center as an urban village or shanty town. Although there are quite a few long-term vacant residences in the prospering inner city, occupied residences or short-term vacant residences are concentrated there much more than the long-term vacant residences. This supports that residences for investing instead of living might exist in the prospering inner city.

Our study defined the dense area of vacant residences with different vacancy lengths as a region with kernel density scores in the top 2 ranks (i.e., kernel density scores greater than 9550) and area greater than 250,000 m^2. Fig. 7 shows the dynamics of the dense areas of vacant residences in Changshu since 2004. Long-term vacant residences gradually grew in number and then the dense areas steadily expanded. Furthermore, the expansion of the dense areas directly brought about a yearly increase linearly and rapidly in the dense areas.

4.2. Vacancy modeling

This study randomly selected 70% of the residential data as training samples to build the RF model and used the remaining 30% of the residential data to test the model. In particular, considering the imbalance among the residences with different vacancy lengths (see Fig. 5), the proportion of residences with different vacancy lengths in yearly training data and yearly test data is almost identical to that in yearly data. In addition, two parameters need to be specified for the RF: the number of trees in the forest (ntree) and the number of variables randomly sampled at each split (mtry). In this study, ntree and mtry were set to 100 and the square root of the number of input variables, respectively.

Table 2 summarizes the accuracy assessment results of the relationship analysis between the residential vacancies and NSFs. On average, the OOB error is 0.086. The overall MSE and MedAE are 0.125 and 0.242, respectively, with MSE and MedAE of 0.090 and 0.198 on the training data set and 0.159 and 0.287 on the test data set. Fig. 8 exhibits the spatial absolute error between the predicted residential vacancy length and the actual residential vacancy length. Approximately 70% of the residential differences are less than 2 months. That is, most of the differences are slight. In 2012, for instance, we selected three typical areas with predicted residential vacancy lengths of less than 3 months (Fig. 9A), between 5 months and 7 months (Fig. 9B), and

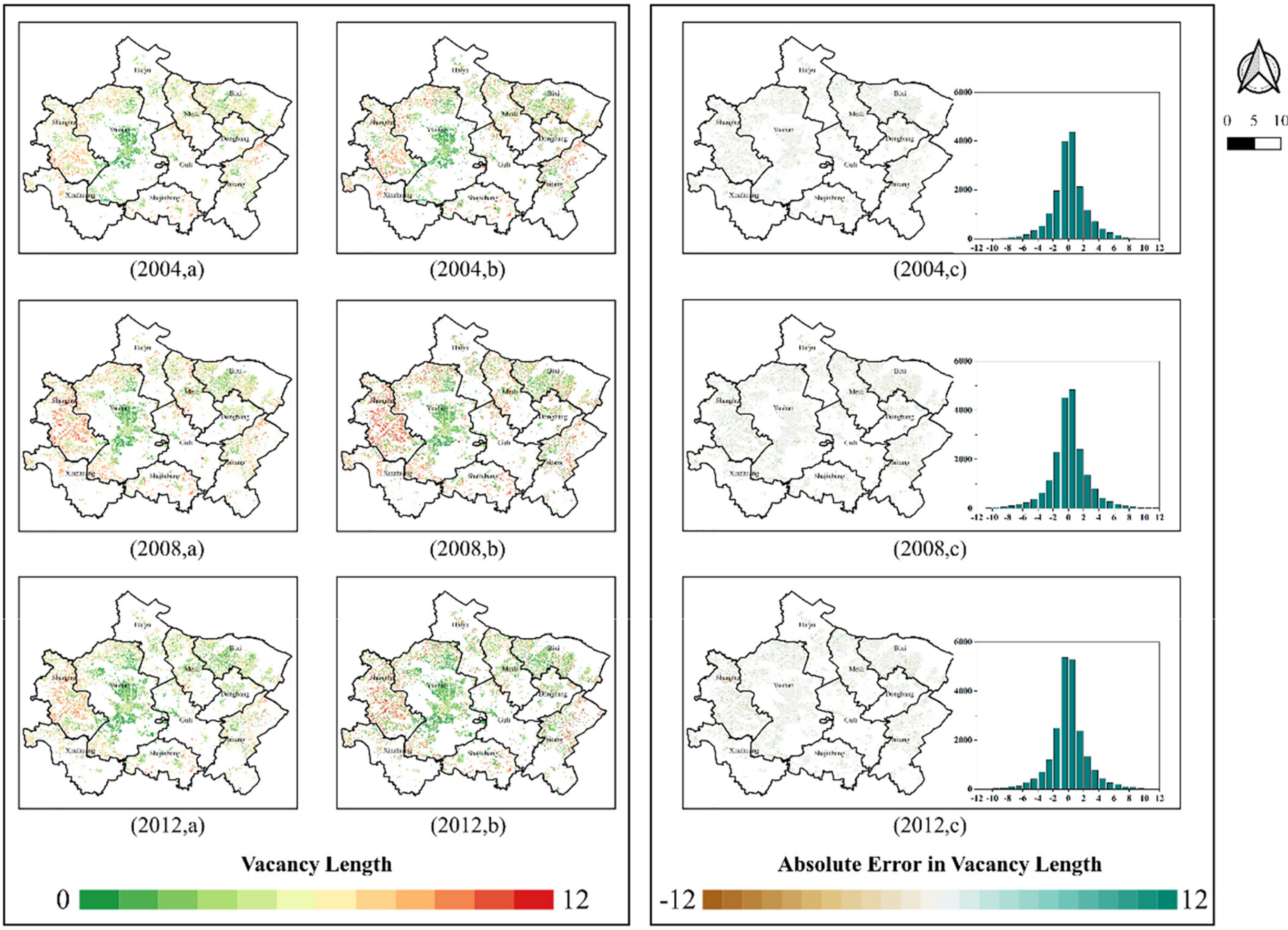


Fig. 8. The spatial distributions of predicted vacancy length (a), actual vacancy length (b) and their differences (c) at 100 m*100 m resolution in 2004, 2008 and 2012 in Changshu.

greater than 10 months (Fig. 9C). In area A, the predicted vacancy length of most residences is relatively short as the number of concentrated occupied residences or short-term vacant residences far exceeded long-term vacant residences. In contrast to area A, the predicted vacancy lengths of most residences are relatively high, with excessively abundant long-term vacant residences concentrated in area C. In area B, the proportion of long-term vacant residences and occupied residences or short-term vacant residences were almost fifty-fifty, and these

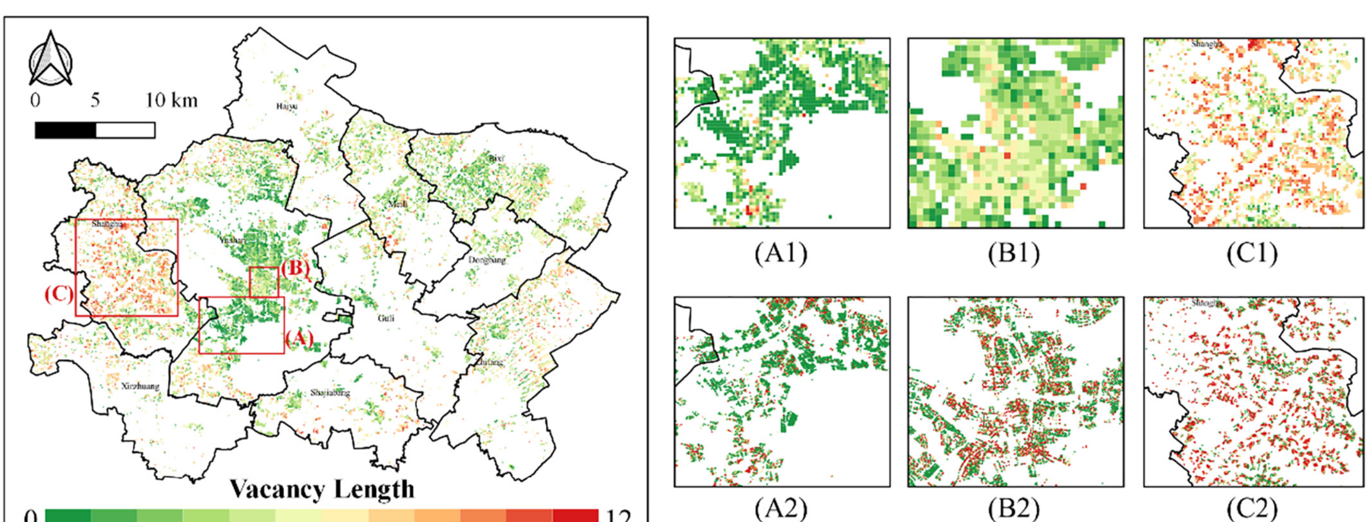


Fig. 9. Details of predicted vacancy length at 100 m*100 m resolution (#1) and actual vacancy length of each residence (#2) in three typical areas in 2012 in Changshu: (A) vacancy length less than 3 months, (B) vacancy length between 5 months and 7 months, and (C) vacancy length greater than 10 months.

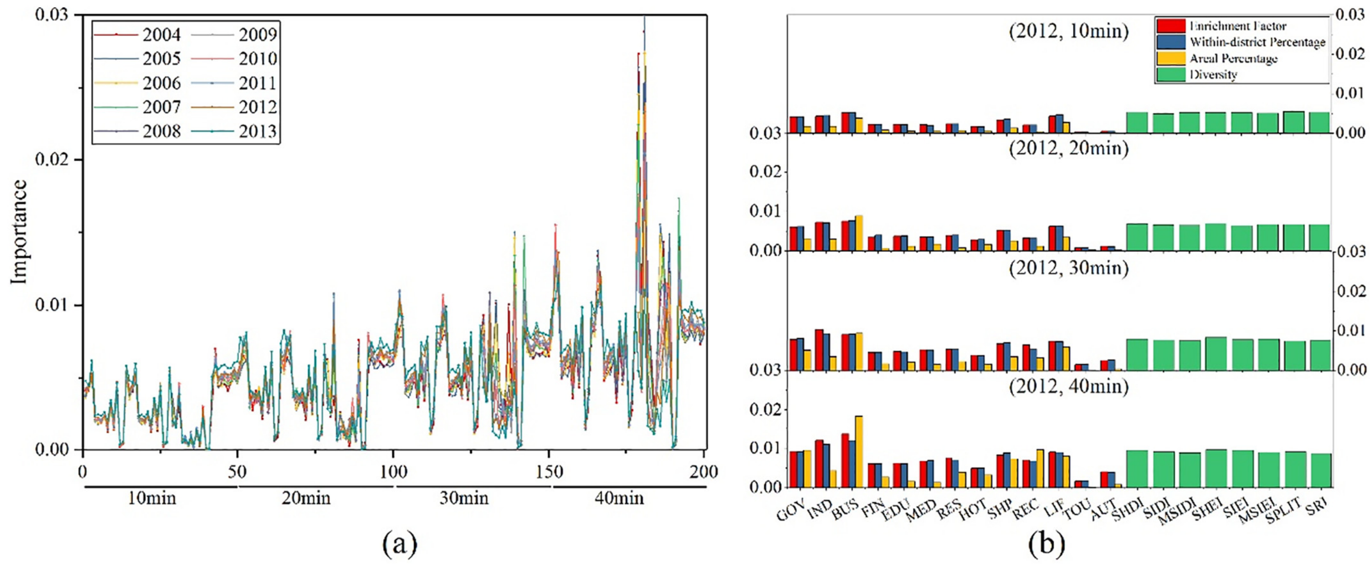


Fig. 10. Importance ranking of factors for residential vacancy within 10-min, 20-min, 30-min and 40-min neighborhoods from 2004 to 2013 (a) and details of 2012 (b) in Changshu. In illustration (a), every importance curve consists of 200 NSFs. In illustration (b), bars with four colors exhibit the importance of the enrichment factor, within-district percentage, areal percentage and diversity, respectively. The first three are the proportion of the 13 FASs (excluding OTH) within these neighborhoods. The last is the 8 kinds of diversities of all the FASs within these neighborhoods.

residences were completely mixed together. The predicted vacancy length of most residences is approximately half a year. These results prove that the predicted results are consistent with the actual situation. In general, the RF model shows satisfying results in the relationship analysis between the residential vacancies and NSFs.

4.3. Effects of factors on residential vacancy

In this study, to understand the contribution of each factor, the VI was measured with the MSE criterion index and can be seen in Fig. 10. A high value of VI suggests that the factor is important to the change of the residential vacancy compared with the other factors. The importance curves are virtually identical from 2004 through 2013 (Fig. 10(a)). This clearly indicates that the contribution of each factor to residential vacancies varies little during this decade.

Taking 2012 as an example, we carried out a detailed analysis of important factors (Fig. 10(b)). As shown, diversity is the most important factor for residential vacancies. That is, the diversity of the surrounding FASs largely determines residential vacancy length. The enrichment factor and within-district percentage play almost the same role. Areal percentage is obviously not very significant. We also noticed the importance of the FASs from the comprehensive perspective of the three kinds of proportions. The three most important FASs are BUS, IND and LIF within the 10-min, 20-min and 30-min neighborhoods and BUS, IND and GOV within the 40-min neighborhood. It is interesting to note that within any neighborhood, BUS and IND are always in the top three. This, we believe, results from the outstanding achievements of Changshu in business and industry (China Communications Institute, 2017; People's Daily, 2017). Also of interest, LIF as the third most important FAS is replaced by GOV within the 40-min neighborhood. Because residents pay attention to short-distance life services, they do not care about long-distance life services. In addition, the two least important FASs are TOU and AUT within any neighborhood. In other words, regardless of distance, tourist attractions and automobile service play negligible roles in residential vacancies.

To further compare the effect of each factor, we extracted three distinct categories of residences according to the residential vacancy length: residences with vacancy lengths less than 3 months (i.e., 0–3 months; one quarter), residences with vacancy lengths between 5 months and 7 months (i.e., almost half a year), and residences with vacancy lengths greater than 10 months (i.e., three quarters), which represent residences with low, medium and high vacancy extents, respectively. And then we compared the factors of the three categories of residences. This study attempts to examine the impacts of the differences in factors and neighborhood changes on residential vacancies. Just as before, taking 2012 for instance, we extracted the three kinds of representative residences. The averaged factor disparities of the three types of residences are illustrated in Fig. 11. Some interesting conclusions can be drawn from this figure.

As the enrichment factor is closer to 1, the residential vacancy length is lower. Enrichment factor has nothing to do with the size of neighborhood. If a neighborhood at any distance is richer or poorer in the FASs than the entire city, there will be more long-term vacant residences within the neighborhood. For example, there are 5000 FASs, including 90 EDU and 45 IND in a city, and there are 500 FASs within a residential neighborhood. According to Eq. (1), we can infer that within the neighborhood, the residential vacancy length is higher and then the vacant residences are more if the number of EDU is much less than 9 or if the IND is much more than 5.

With the increase of the areal percentage and the diversity, the residential vacancy length decreases. As the size of the neighborhood increases, the role that areal percentage plays is more significant, but diversity is more insignificant. If the FAS species diversity is great or most of a certain FAS is concentrated in a neighborhood, there will be few long-term vacant residences within the neighborhood. That seems contradictory, but it is not. First, diversity is a neighborhood summary measure that does not take into account the uniqueness of the individual species, while areal percentage focuses on the homogeneity and even the uniqueness of the neighborhood. A neighborhood may have high diversity yet be comprised largely of common species. Conversely, a neighborhood may have low diversity yet be comprised of especially unique or highly desired species, for instance, a technology park. Second, the areal percentage and the diversity play major roles in different neighborhoods. Diversity works mainly for small-sized neighborhoods, while areal percentage is chiefly aimed at large-sized neighborhoods.

Within a neighborhood at any distance, residences with low vacancy length have virtually the same within-district percentage as residences with high vacancy length. In other words, the proportion of any FAS within any neighborhood has little impact on residential vacancies. For

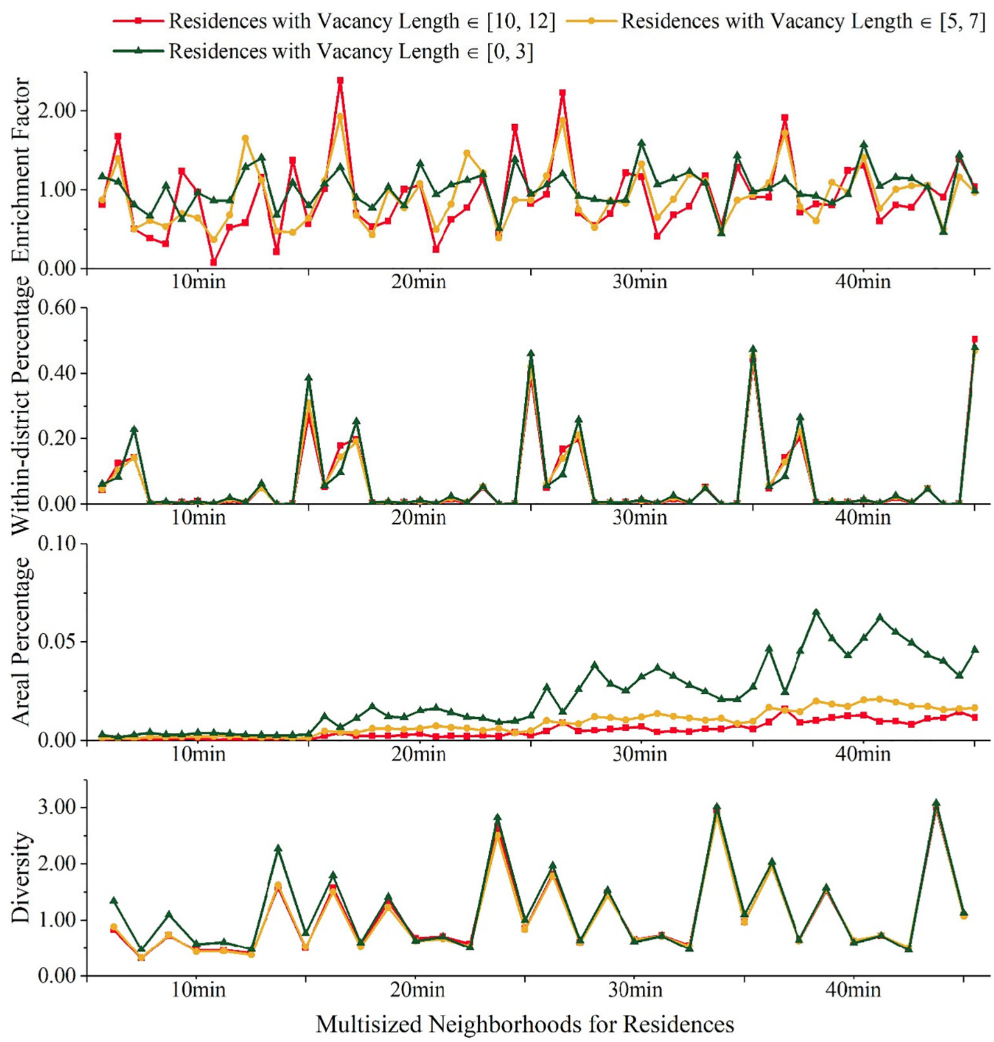


Fig. 11. The dissimilarities of factors and changes in neighborhoods influencing disparities in residential vacancies.

example, there are 100 FASs, including 2 IND and 80 LIF, within a residential neighborhood. Although 0.02 is very small and 0.8 is quite large, they could reflect neither the similarity between the neighborhood and the entire city nor the diversity and uniqueness of the neighborhood. Briefly, within-district percentage could not effectively describe the neighborhood. Hence, the within-district percentage is not significant in residential vacancies at all.

5. Discussion and conclusions

The aim of the present research is to reveal the long-term spatial expansion and internal transformation of vacant residences and to explain the contributing factors of residential vacancies. In addition to delineating the spatiotemporal dynamics of vacant residences at a fine scale via municipal water consumption data, we examined the relationships between residential vacancies and multisized NSFs at a fine scale. First, the RVIM proposed in this study was applied to identify vacant residences. Then, the combination of the LISA and KDE was used to map the spatial distribution of vacant residences and to understand the spatiotemporal characteristics. Second, novel multisized NSFs were extracted for residence, and their relationship analysis with residential vacancies was conducted using the RF.

The hot/cold and kernel density maps show that in 2004 through 2013 in Changshu, residences had a concentric ring structure and dual characteristics, and vacant residences had a binary structure. New residential forms were shaped by replacement or accretion, such as the declining city center and the rising subcenter. In addition, the measures (OOB error of 0.086, overall MSE and MedAE of 0.125 and 0.242, respectively, 70% of residential spatial absolute error less than 2 months) demonstrate the effectiveness of the relationship analysis between residential vacancies and the NSFs. On the basis of that, the importance and impact of the NSFs were clearly revealed. Diversity has the highest importance, followed by enrichment factor and within-district percentage. BUS, IND, and LIF (or GOV) are the three most important FASs, while TOU and AUT are the two least important. Interestingly, the residential vacancy length becomes lower as the enrichment factor is closer to 1 or the areal percentage or the diversity is larger. The contributions of areal percentage and diversity have significant scale (specifically size of the neighborhood) dependence.

As an intensive fine-scale study for detecting and analyzing vacant residences, the contributions of our study mainly lie in three aspects. First, the study focuses on the spatial expansion and internal transformation of vacant residences within a city and pays attention to contributing factors for mitigating residential vacancies rather than estimating the vacancy degree or area of the entire city. Compared with understanding the degree/area of vacancy of the entire city, we are firmly convinced, grasping the spatial expansion, that internal transformation and the contributing factors of the vacant residences within a city are more useful and urgent for decision makers to mitigate residential vacancies.

Second, we propose a feasible and general-purpose vacant residence analysis framework using municipal water consumption data. The framework is not only simple but also effective. Municipal water consumption data can be obtained with ease, especially for decision makers. The methods adopted in this study, including the RVIM, LISA, KDE and RF, are also uncomplicated. Thus, the framework can be implemented effortlessly in terms of both the data and technology. Although less challenging, the framework is highly effective in light of our results. In addition, the framework is general-purpose. Water is consumed by residents throughout China and around the world, so municipal water consumption data could be obtained by local decision makers throughout China and even around the world. Just adjusting *minMhs* and *minWC* based on different national standards and policies, the framework will work effectively across the country and around the world, especially in cities with comprehensive coverage of municipal water service.

Third, this study can contribute to policy making for urban management and planning. Our results provide local decision makers with recommendations for mitigating residential vacancies in the problematic areas by optimizing the arrangement of facilities and services. These hot/cold and kernel density maps send a strong signal to decision makers regarding areas in need of improvement, such as the declining city center, the areas with higher kernel density scores of residential vacancy, and the areas with fast expanding vacant residences. For these problematic areas, considering comprehensively the importance and impact of the NSFs, several recommendations are made. One of the key measures that can be taken is to enhance the FAS diversity within the small-sized neighborhoods of residences (e.g., 10-min street network buffer). It is also important is to ensure that residential neighborhoods are not excessively rich or poor in FASs compared to the entire city. Last but not least, it is recommended to allocate specific functional areas (e.g., a technology park) within large-sized neighborhoods of residences (e.g., 40-min street network buffer). It must be noted that, for other cities in China and the world, the recommendations for mitigating residential vacancies derived from the analysis of our study are likely to be different from the ones for Changshu. Nevertheless, that does not erase the vital contribution of our study, which by adopting this study's framework decision makers are able to identify problematic areas and make appropriate recommendations based on local conditions.

The present study gains insight into vacant residences and their NSFs as well as contributes to academic research and practical application. However, we also admit that there are some potential biases. Municipal water consumption data, for instance, may be amiss due to the damage or insensitivity of water meters. Likewise, we recognize several limitations that still exist in this study, which will be highlighted in our future research. On the one hand, the value of some of the impacts of residential vacancies has yet to be quantified, such as physical environmental characteristics and socioenvironmental context. Further research into how they contribute to residential vacancies is critical for mitigating residential vacancies. On the other hand, some residents, especially in rural areas, may use more mineral water or well water rather than tap water. Further research of integrating multisource municipal infrastructure and service data to analyze residential vacancies is needed, especially for rural areas or the suburbs.

Funding

This research was funded by the National Natural Science Foundation of China (Grant No. 41671408, 41801306, 41901332), the Natural Science Foundation of Hubei Province (Grant No. 2017CFA041) and the Application Foundation Frontier Project of Wuhan City (Grant No. 20180104011293).

CRediT authorship contribution statement

Yongting Pan:Conceptualization, Data curation, Formal analysis, Methodology, Software, Writing - original draft.**Wen Zeng**:Investigation, Resources, Funding acquisition.**Qingfeng Guan**:Conceptualization, Methodology, Project administration, Supervision, Writing - review & editing, Funding acquisition.**Yao Yao**:Writing - review & editing, Funding acquisition.**Xun Liang**:Writing - review & editing, Funding acquisition.**Hanqiu Yue**:Writing - review & editing.**Yaqian Zhai**:Writing - review & editing.**Junyi Wang**:Visualization.

References

Acheson, E., Volpi, M., & Purves, R. S. (2019). Machine learning for cross-gazetteer matching of natural features. *International Journal of Geographical Information Science*, 1–27.

Ami, S., & George, A. (2014). *Empty homes in England: Get the data by local authority*.

Anderson, F. C., & Pandy, M. G. (2001). Dynamic optimization of human walking. *Journal of Biomechanical Engineering, 123*(5), 381–390.

Anselin, L. (1995). Local indicators of spatial association—LISA. *Geographical Analysis, 27*(2), 93–115.

Bai, X., Shi, P., & Liu, Y. (2014). Society: Realizing China's urban dream. *Nature News, 509*(7499), 158.

Bakillah, M., Liang, S., Mobasheri, A., Jokar Arsanjani, J., & Zipf, A. (2014). Fine-resolution population mapping using OpenStreetMap points-of-interest. *International Journal of Geographical Information Science, 28*(9), 1940–1963.

Belgiu, M., & Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. *Isprs Journal of Photogrammetry & Remote Sensing, 114*, 24–31.

Bordogna, G., Carrara, P., Criscuolo, L., Pepe, M., & Rampini, A. (2014). A linguistic decision making approach to assess the quality of volunteer geographic information for citizen science. *Information Sciences, 258*, 312–327.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.

Chai, Y., Shen, J., & Weng, G. (2008). A study on activity space of shopping of Shanghai residents: Temporal and spatial characteristics and relative influencing factors. *Economic Geography, 28*(2), 221–227.

Chaix, B., Simon, C., Charreire, H., Thomas, F., Kestens, Y., Karusisi, N., & Pannier, B. (2014). The environmental correlates of overall and neighborhood based recreational walking (a cross-sectional analysis of the RECORD study). *International Journal of Behavioral Nutrition and Physical Activity, 11*(1), 20.

Chen, Z., Yu, B., Hu, Y., Huang, C., Shi, K., & Wu, J. (2015). Estimating house vacancy rate in metropolitan areas using NPP-VIIRS nighttime light composite data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 8*(5), 2188–2197.

Chi, G., Liu, Y., Wu, Z., & Wu, H. (2015). *Ghost Cities Analysis Based on Positioning Data in China. ArXiv Preprint ArXiv:1510.08505*.

China Communications Institute (2017). *China industrial development reports of top 100 counties and top 100 areas*.

Deng, J. S., Wang, K., Hong, Y., & Qi, J. G. (2009). Spatio-temporal dynamics and evolution of land use change and landscape pattern in response to rapid urbanization. *Landscape and Urban Planning, 92*(3–4), 187–198.

Fan, J., Ma, T., Zhou, C., Zhou, Y., & Xu, T. (2014). Comparative estimation of urban development in China's cities using socioeconomic and DMSP/OLS night light data. *Remote Sensing, 6*(8), 7840–7856.

Fernandez-Delgado, M., Cernadas, E., Barro, S., & Amorim, D. (2014). Do we need hundreds of classifiers to solve real world classification problems. *Journal of Machine Learning Research, 15*(1), 3133–3181.

Gao, S., Janowicz, K., & Couclelis, H. (2017). Extracting urban functional regions from points of interest and human activities on location-based social networks. *Transactions in GIS, 21*(3), 446–467.

Ge, W., Yang, H., Zhu, X., Ma, M., & Yang, Y. (2018). Ghost City extraction and rate estimation in China based on NPP-VIIRS night-time light data. *ISPRS International Journal of Geo-Information, 7*(6), 219.

Getis, A., & Ord, J. K. (2010). The analysis of spatial association by use of distance statistics. *Perspectives on spatial data analysis* (pp. 127–145). Springer.

Gorelick, R. (2010). Combining richness and abundance into a single diversity index using matrix analogues of Shannon's and Simpson's indices. *Ecography, 29*(4), 525–530.

Haramati, T., & Hananel, R. (2016). Is anybody home? The influence of ghost apartments on urban diversity in Tel-Aviv and Jerusalem. *Cities, 56*, 109–118.

Hayes-Roth, F. (1985). Rule-based systems. *Communications of the ACM, 28*(9), 921–932.

He, C., Chen, T., Mao, X., & Zhou, Y. (2016). Economic transition, urbanization and population redistribution in China. *Habitat International, 51*, 39–47.

He, G., Mol, A. P. J., & Lu, Y. (2016). Wasted cities in urbanizing China. *Environmental Development, 18*, 2–13.

Hof, A., & Wolf, N. (2014). Estimating potential outdoor water consumption in private urban landscapes by coupling high-resolution image analysis, irrigation water needs and evaporation estimation in Spain. *Landscape & Urban Planning, 123*, 61–72.

Jaeger, J. A. G. (2000). Landscape division, splitting index, and effective mesh size: New measures of landscape fragmentation. *Landscape Ecology, 15*(2), 115–130.

Jiang, Y., Mohabir, N., Ma, R., & Zhu, P. (2017). Sorting through neoliberal variations of ghost cities in China. *Land Use Policy, 69*, 445–453.

Jin, X., Long, Y., Sun, W., Lu, Y., Yang, X., & Tang, J. (2017). Evaluating cities' vitality and identifying ghost cities in China with emerging geographical data. *Cities, 63*, 98–109.

Kang, C., Liu, Y., Ma, X., & Wu, L. (2012). Towards estimating urban population distributions from mobile call data. *Journal of Urban Technology, 19*(4), 3–21.

Konomi, S., Sasao, T., Hosio, S., & Sezaki, K. (2017). Exploring the use of ambient WiFi

signals to find vacant houses. *European conference on ambient intelligence* (pp. 130–135). Springer.
Kumagai, K., Matsuda, Y., & Ono, Y. (2016). Estimation of housing vacancy distributions: Basic Bayesian approach using utility data. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 41*, 709.
Kvålseth, T. O. (2015). Evenness indices once again: Critical analysis of properties. *Springerplus, 4*(1), 1–12.
Lansley, G., Li, W., & Longley, P. (2017). Representing population dynamics from administrative and consumer registers. *Geographical information science research UK (GISRUK)*.
Lansley, G., & Longley, P. (2016). Deriving age and gender from forenames for consumer analytics. *Journal of Retailing and Consumer Services, 30*, 271–278.
Li, Y., Long, H., & Liu, Y. (2010). Industrial development and land use/cover change and their effects on local environment: A case study of Changshu in eastern coastal China. *Frontiers of Environmental Science & Engineering in China, 4*(4), 438–448.
Liao, H., Dong, W., Huang, H., Gartner, G., & Liu, H. (2019). Inferring user tasks in pedestrian navigation from eye movement data in real-world environments. *International Journal of Geographical Information Science, 33*(4), 739–763.
Liu, H., Gegov, A., & Stahl, F. (2014). Categorization and construction of rule based systems. *International conference on engineering applications of neural networks* (pp. 183–194). Springer.
Liu, Y., Wang, F., Xiao, Y., & Gao, S. (2012). Urban land uses and traffic 'source-sink areas': Evidence from GPS-enabled taxi data in Shanghai. *Landscape and Urban Planning, 106*(1), 73–87.
Long, Y., Mao, M., Mao, Q., Shen, Z., & Zhang, Y. (2014). Fine-scale urban modeling and its opportunities in the "big data" era: Methods, data and empirical studies. *Human Geography, 29*(3), 7–13.
Lu, H., Zhang, C., Liu, G., Ye, X., & Miao, C. (2018). Mapping China's ghost cities through the combination of nighttime satellite data and daytime satellite data. *Remote Sensing, 10*(7), 1037.
Lu, Z., Im, J., Rhee, J., & Hodgson, M. (2014). Building type classification using spatial and landscape attributes derived from LiDAR remote sensing data. *Landscape & Urban Planning, 130*(1), 134–148.
Luo, X., Zhang, J., Yang, J., Song, W., & Cui, W. (2014). Assessing the benthic habitat quality in the Huanghe (Yellow River) estuary and its adjacent areas using AMBI and M-AMBI. *Acta Oceanologica Sinica, 33*(9), 117–124.
Ma, X., Tong, X., Liu, S., Li, C., & Ma, Z. (2018). A multisource remotely sensed data oriented method for "ghost City" phenomenon identification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 11*(7), 2310–2319.
McGarigal, K. (2014). FRAGSTATS help. *Documentation for FRAGSTATS. Vol. 4*.
Mclaughlin, J. E., Mclaughlin, G. W., Mclaughlin, J. S., & White, C. Y. (2016). Using Simpson's diversity index to examine multidimensional models of diversity in health professions education. *International Journal of Medical Education, 7*, 1.
Minutolo, A., Esposito, M., & De Pietro, G. (2017). Optimization of rule-based systems in mHealth applications. *Engineering Applications of Artificial Intelligence, 59*, 103–121.
Molloy, R. (2016). Long-term vacant housing in the United States. *Regional Science & Urban Economics, 59*, 118–129.
Morckel, V. C. (2014). Spatial characteristics of housing abandonment. *Applied Geography, 48*, 8–16.
Nam, J., Han, J., & Lee, C. (2016). Factors contributing to residential vacancy and some approaches to management in Gyeonggi Province, Korea. *Sustainability, 8*(4), 367.
Niu, Y., Dong, L. C., Niu, Y., & Deng, F. (2017). Resident-defined measurement scale for a city's products. *Landscape and Urban Planning, 167*, 177–188.
People''s Daily (2017). 2017 National Comprehensive Strength top 100 counties. *Top 100 areas, and top 1000 towns*.
Pielou, E. C. (1966a). Species-diversity and pattern-diversity in the study of ecological succession. *Journal of Theoretical Biology, 10*(2), 370–383.
Pielou, E. C. (1966b). The measurement of diversity in different types of biological collections. *Journal of Theoretical Biology, 13*(1), 131–144.
Rajaram, R., Castellani, B., & Wilson, A. N. (2017). Advancing shannon entropy for measuring diversity in systems. *Complexity, 2017*.
Sarica, A., Cerasa, A., & Quattrone, A. (2017). Random Forest algorithm for the classification of neuroimaging data in Alzheimer's disease: A systematic review. *Frontiers in Aging Neuroscience, 9*, 329.
Shepard, W. (2015). *Ghost cities of China: The story of cities without people in the world's most populated country.* Zed Books Ltd.
Silverman, B. W. (2018). *Density estimation for statistics and data analysis.* Routledge.
Smith, B., & Wilson, J. B. (1996). A consumer's guide to evenness indices. *Oikos,* 70–82.
Sparks, K., Thakur, G., Pasarkar, A., & Urban, M. (2019). A global analysis of cities' geosocial temporal signatures for points of interest hours of operation. *International Journal of Geographical Information Science,* 1–18.
Ströbele, M., & Hunziker, M. (2017). Are suburbs perceived as rural villages? Landscape-related residential preferences in Switzerland. *Landscape and Urban Planning, 163*, 67–79.
Su, X. (2014). *2014 ghost City index rankings: There will be 50 ghost cities*.
Walton, J. T. (2015). Subpixel urban land cover estimation. *Photogrammetric Engineering & Remote Sensing, 74*(10), 1213–1222.
Xiao, Z., Chai, Y., & Zhang, Y. (2014). Overseas life circle planning and practice. *Planners, 18*(10).
Xie, Y., & Weng, Q. (2016). Updating urban extents with nighttime light imagery by using an object-based thresholding method. *Remote Sensing of Environment, 187*, 1–13.
Yan, S., Merlin, L., & Rodriguez, D. (2013). Comparing measures of urban land use mix. *Computers Environment & Urban Systems, 42*(7), 1–13.
Zhang, M., Lu, X., Shi, D., & Pan, H. (2018). Toxic metal enrichment characteristics and sources of arid urban surface soil in Yinchuan City, China. *Journal of Arid Land, 10*(4), 653–662.
Zhang, Q., & Su, S. (2016). Determinants of urban expansion and their relative importance: A comparative analysis of 30 major metropolitans in China. *Habitat International, 58*, 89–107.
Zhang, X., Li, W., Zhang, F., Liu, R., & Du, Z. (2018). Identifying urban functional zones using public bicycle rental records and point-of-interest data. *ISPRS International Journal of Geo-Information, 7*(12), 459.
Zheng, H., Wang, X., & Cao, S. (2014). The land finance model jeopardizes China's sustainable development. *Habitat International, 44*, 130–136.
Zheng, Q., Deng, J., Jiang, R., Wang, K., Xue, X., Lin, Y., & Shahtahmassebi, A. R. (2017). Monitoring and assessing "ghost cities" in Northeast China from the view of nighttime light remote sensing data. *Habitat International, 70*, 34–42.
Zheng, Q., Zeng, Y., Deng, J., Wang, K., Jiang, R., & Ye, Z. (2017). "Ghost cities" identification using multi-source remote sensing datasets: A case study in Yangtze River Delta. *Applied Geography, 80*, 112–121.
Zhou, N., Hubacek, K., & Roberts, M. (2015). Analysis of spatial patterns of urban growth across South Asia using DMSP-OLS nighttime lights data. *Applied Geography, 63*, 292–303.
Zhou, R., Zhang, H., Ye, X., Wang, X., & Su, H. (2016). The delimitation of urban growth boundaries using the CLUE-S land-use change model: Study on Xinzhuang Town, Changshu City, China. *Sustainability, 8*(11), 1182.