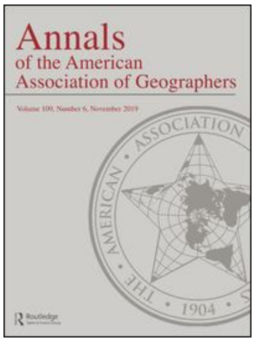


Annals of the American Association of Geographers

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/raag21

Sensing Mixed Urban Land-Use Patterns Using Municipal Water Consumption Time Series

Qingfeng Guan , Sijing Cheng , Yongting Pan , Yao Yao & Wen Zeng

Sensing Mixed Urban Land-Use Patterns Using Municipal Water Consumption Time Series

Qingfeng Guan,[*] Sijing Cheng,[*,†] Yongting Pan,[*] Yao Yao,[*] and Wen Zeng[*]

[*]*School of Geography and Information Engineering, China University of Geosciences*
[†]*PowerChina Huadong Engineering*

The biased population coverage and short temporal lengths of newly emerged data sets (e.g., data sets of social media, mobile phones, and smart cards) obstruct the effective analysis of long-term dynamics of landuse patterns, particularly in small and developing cities. This study proposed a framework to delineate and analyze mixed land-use patterns and their evolution using municipal water consumption data. A two-step classification strategy was designed based on the rotation forest scheme to differentiate the socioeconomic types of customers (e.g., residence, commerce, public facility, manufacturing, and recreation) using multiple features extracted from the various forms of water consumption time series. The spatial distributions of the socioeconomic functions were then derived, and the mixed land use was measured using a diversity index based on information entropy. Such an approach was applied to Changshu, a typical developing county-level city in China, for the period 2004 to 2013. The results showed that the urbanization of Changshu experienced both spatial expansion and intensification, with a slightly declining rate of growth in recent years. Apart from the city center, two subcenters have emerged for industrial development. The degree of land-use mixture has increased with urban growth, indicating a maturing of urbanization. This study explored the approach of identifying individual socioeconomic functions by the consumption patterns of municipal services and demonstrated that municipal service data sets can reveal land-use patterns and dynamics at a fine spatial resolution to evaluate urban planning and management, with the advantages of large population coverage and long-term temporal lengths. *Key Words: land-use patterns, mixed land use, municipal water consumption, rotation forest, social sensing.*

Urban land-use patterns are often complex and significantly heterogeneous across cities (Bennett, Tang, and Wang 2011; Meentemeyer et al. 2013; Zhu et al. 2015). Urban land-use patterns provide key information for urban planning and management, because they reflect past and current socioeconomic status and human–environment interactions (Seto, Sánchez-Rodríguez, and Fragkias 2010) and affect location-allocation decisions (Bourne 1976) and citizen behavior (Kitamura, Mokhtarian, and Daidet 1997; Boarnet 2011). Proper land-use combinations and arrangements are essential to the sustainability and health of urban communities (Tian, Liang, and Zhang 2017).

A large number of methods have been developed to delineate and analyze urban land-use patterns. With the advantages of wide spatial and temporal coverages, remote sensing techniques have been widely used in the past for land-use mapping (Huang, Lu, and Zhang 2014; Wen et al. 2016; C. Wu, Zhang, and Zhang 2016). Besides the classical pixel-oriented and object-oriented classification methods for remote sensing images (Blaschke 2010; Zhang, Du, and Wang 2015), scene classification methods, which use scenes rather than pixels or objects as the basic classification units (L. Wang, Sousa, and Gong 2004), have been applied to identify urban functional zones (i.e., areas primarily for one or more socioeconomic functions, such as residence, industry, commerce, and recreation) and their changes (Zhang and Du 2015). In addition, textural classification methods, using both remote sensing images and parcel attributes from geographic information systems, have been used in urban land-use mapping (S. S. Wu et al. 2009). Because the land-use data extracted from remotely sensed images contain only physical properties (Bratasanu, Nedelcu, and Datcu 2011), lacking socioeconomic properties that have strong correlations with human activity (X. Liu et al. 2017), such methods have limitations in detecting and understanding socioeconomic environments (Y. Liu et al. 2015).

Recently, the rapid growth of big data, along with the booming development of artificial intelligence,

Annals of the American Association of Geographers, 0(0) 2020, pp. 1–19
Initial submission, January 2019; revised submissions, August and December 2019; final acceptance, March 2020
Published by Taylor & Francis, LLC.

has stimulated the studies of urban land-use patterns from the perspectives of social sensing and urban computing through the use of various new data sets (Elwood, Goodchild, and Sui 2012; Zheng et al. 2014; Y. Liu et al. 2015; Singleton, Spielman, and Folch 2018). Traffic data, mobile phone data, and social media data are the typical newly emerged data sets that have been applied in urban land-use analysis (Y. Jiang, Li, and Cutter 2019). For instance, mobile phone data (Toole et al. 2012; Y. Jiang, Li, and Cutter 2019) and taxi trajectory data (L. Yu et al. 2012) have been used to analyze urban functional zones and land-use variations. Temporal patterns of tweets (Soliman et al. 2017) and temporal frequency trends of geotagged social media messages (Y. Wang et al. 2016) have also been used to identify land-use types. Such new data sets enable the monitoring and analysis of socioeconomic activities in urban areas at fine spatiotemporal resolutions (Y. Liu et al. 2015).

Mixed land use often exists in modern urban areas. Studies have pointed out that the functional structure of the urban environment is composed of mixed land-use patterns (Yuan, Zheng, and Xie 2012; Chen et al. 2017; X. Liu et al. 2017; Niu et al. 2017). Furthermore, mixed land use has close relationships with energy consumption, transportation patterns, resident behavior, public health, and environmental problems, such as noise and air pollution (Song and Knaap 2004; Ribeiro et al. 2015; H. Yang, Song, and Choi 2016; Tian, Liang, and Zhang 2017). Several indexes have been proposed to measure the degree of mixture of residential and nonresidential functions in land units (van den Hoek 2008; Tian, Liang, and Zhang 2017). Dovey and Pafka (2017) adopted the live/work/visit triangle to analyze mixed land use and focused on the interconnections among various functions and the unique characteristics of New York, Barcelona, and Bogotá. X. Liu et al. (2017) integrated multisource geospatial big data (e.g., social media data, taxi trajectories, points of interest, and remote sensing images) to characterize mixed-use buildings in urban areas.

Two issues, however, are associated with the use of the newly emerged data sets in the studies of mixed land-use patterns in urban areas:

1. Limited population coverage and biased representations: The population coverage of the commonly used types of big geodata, including social media data, mobile phone data, and smart card data, largely depends on the people who use certain types of smart devices or Internet services (Kwan 2016). Therefore, such data sets can only reflect the socioeconomic behaviors of a limited proportion of the urban population. Particularly, in small and developing cities, where smart devices and Internet services might not be as widely used as in large cities, such data sets could be greatly insufficient, leading to biased and inaccurate representations of urban socioeconomic functions.
2. Limited temporal lengths: Because social media data, mobile phone data, and smart card data have only been systematically collected in recent years, their time spans are relatively short, which limits the analysis of long-term (e.g., over ten years) spatiotemporal dynamics.

As one of the most fundamental municipal services in cities, municipal water is used by the majority of citizens and covers a wide range of socioeconomic activities in many regions, where municipal services are well established and maintained. For example, the Chinese government has made great efforts and progress in improving the coverage rate of municipal water supply in both cities and rural areas. The rural water supply rate in Jiangsu Province, in which the study area of this investigation is located, has reached 96.9 percent (Luo 2008). Municipal water supply has covered the majority proportion of cities in China, including small and developing ones. Therefore, compared with the newly emerged data sets (e.g., social media, mobile phones, and smart cards), municipal water consumption data have two main advantages (for cities with comprehensive municipal water services):

1. Larger population coverage and less bias: Municipal water is usually provided or supported by government agencies, because it is a necessity for all citizens and nearly all human activity (Horner, Zhao, and Chapin 2011). Therefore, municipal water consumption data have much greater population coverage and are less biased in characterizing the population and human activity.
2. Longer temporal spans: Municipal water consumption data have been collected for decades in many cities and therefore can enable the analysis of long-term dynamics.

Efforts have been made to explore the spatial distribution of water consumption. For example, Villar-Navascués and Pérez-Morales (2018) identified the determinants of water consumption, focusing on the variables related to urban land use and socioeconomic factors. Chang et al. (2017) also studied the influence of the spatial dependence of water-use patterns using a set of spatially explicit data of water consumption, socioeconomic activities, and biophysical environments.

Previous studies have demonstrated that water consumption is closely related to socioeconomic, cultural, and water supply conditions (L. Shi et al. 2018), and different types of customers exhibit different water consumption patterns (J. Yang et al. 2015; Gui, Li, and Gao 2016). Few studies, however, have used municipal water consumption data to identify the socioeconomic functions of urban lands and analyze mixed patterns of land use.

Because the socioeconomic functions of urban lands can be identified through the classification of the water consumption patterns of individual customers, the land-use patterns and spatiotemporal dynamics can be delineated and analyzed. To extract and analyze mixed land-use patterns and their evolution in cities, including small and developing ones, this study aimed to develop a framework to identify the socioeconomic types of individual municipal water customers and analyze the long-term dynamics of mixed land-use patterns. Such an approach was applied to Changshu, a typical developing county-level city in China, in which its mixed land-use pattern and its evolution from 2004 to 2013 were analyzed.

Study Area and Data

Located at 31°31′–31°50′ N, 120°33′–121°03′ E in Jiangsu Province, Changshu is a county-level city in the eastern coastal area of China, where the economy has soared over the last three decades (Figure 1A). It covers an area of 1,264 km^2, with a population of 1,068,700 as of 2016. Industry accounts for the largest proportion of gross domestic product, and the textile industry has developed well and is the leading industry in Changshu. In recent years, with the establishment of economic development zones, the high-tech industry has gradually developed. Changshu can be seen as an example of small cities where municipal services have been well established and maintained. The municipal water service is provided and maintained by the Changshu Municipal Water Company, which supplied the water consumption data for this study.

The municipal water consumption data used in this study included the consumption records of 405,768 customers, covering a time span from January 2004 to October 2013. Each record included the customer ID, customer name, address, coordinates, reading date, and meter reading. All of the meter reading records for a particular customer ID were arranged as the water consumption time series for this customer (examples shown in Figure 1B).

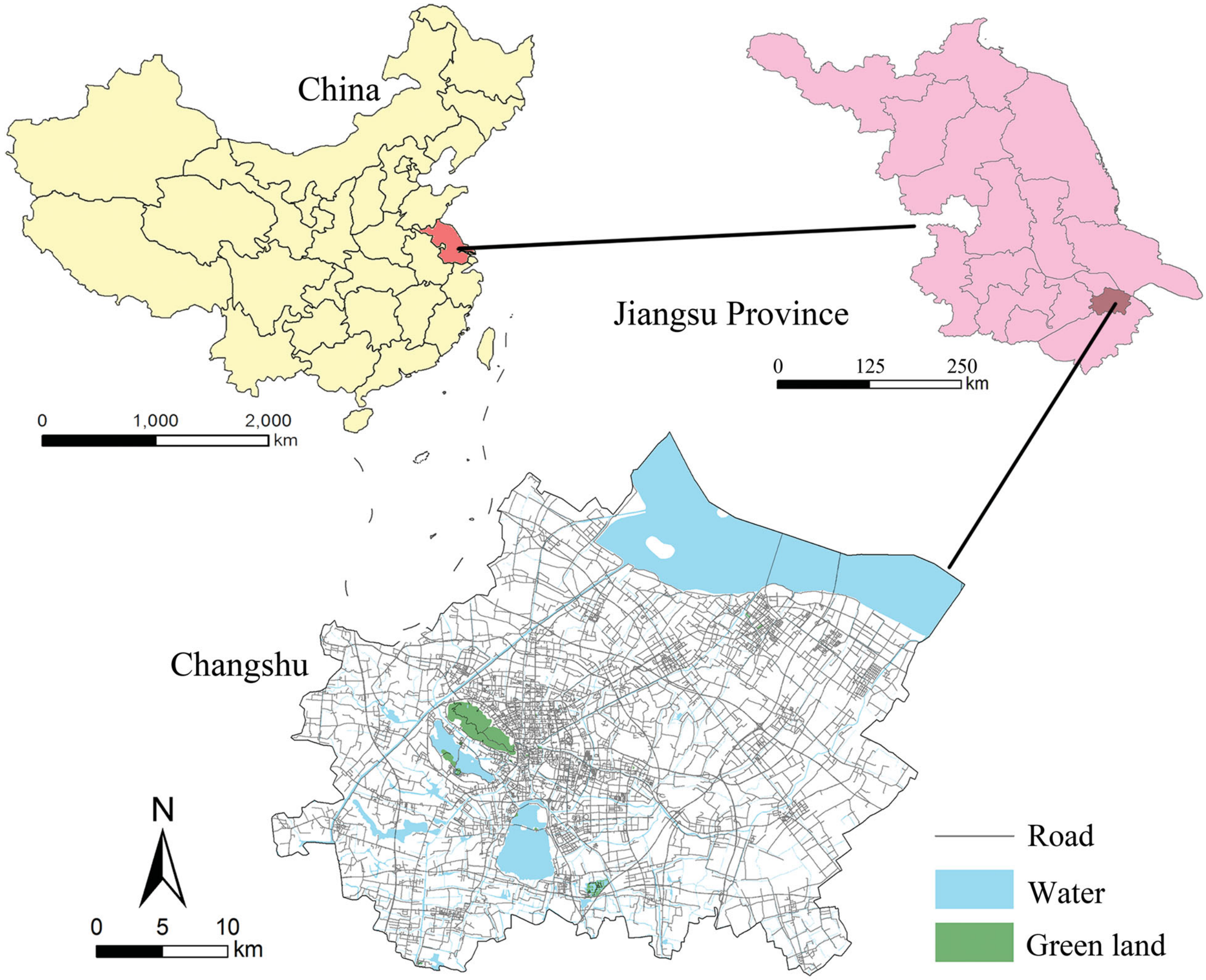
Note that a customer ID corresponds to a particular user of a particular meter. As requested by the municipal service management protocol, if the user of a meter changes, the old customer ID must be canceled and a new ID must be assigned. Therefore, for a particular ID, the customer type (or socioeconomic function of the customer) is consistent throughout the entire water consumption time series associated with that ID. The original data, however, did not include an explicit coding of the customer type. To delineate the land-use conditions, a method was required to determine the socioeconomic types of individual customers, which was one of the main focuses of this study.

Methods

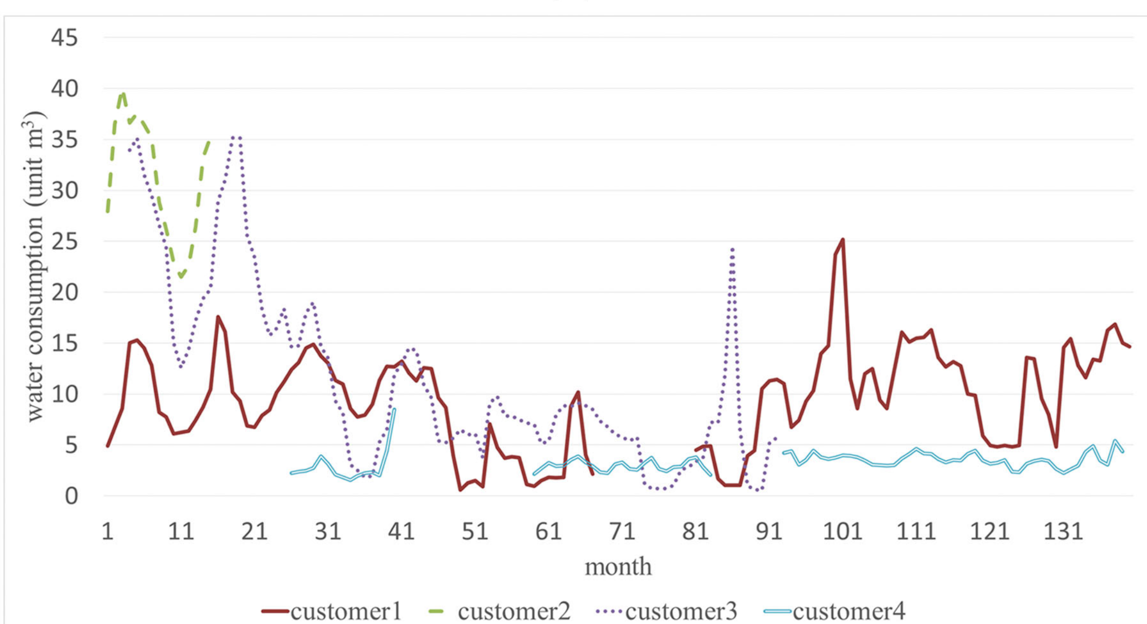
This study proposed a framework to delineate and analyze mixed land use in urban areas based on the classification of water consumption patterns (Figure 2). First, anomalous values in the original water consumption time series were detected and processed and the original time series was converted into three forms of time series with various temporal granularities and spans. A collection of features was then extracted to represent various aspects of consumption patterns. Second, customers, characterized by the water consumption time series features, were classified via a two-step rotation forest classifier. In the first step, the customers were classified into residents and nonresidents. In the second step, nonresidential customers were further classified into four types of socioeconomic functions: commerce, public facility, manufacturing, and recreation. Finally, customers were projected onto their geospatial locations and the urban land-use patterns were delineated and analyzed based on grid partitioning. Specifically, the mixed patterns of land use were measured and the spatiotemporal dynamics over an extensive period of time were analyzed. The detailed procedures are described in the following sections.

Multigranularity Time Series and Feature Extraction

The interval of water consumption records may vary among cities, ranging from monthly and quarterly



(A)



(B)

Figure 1. Study area and examples of water consumption time series. (A) Study area: Changshu, Jiangsu Province, China. (B) Examples of water consumption time series with various spans and vacancies.

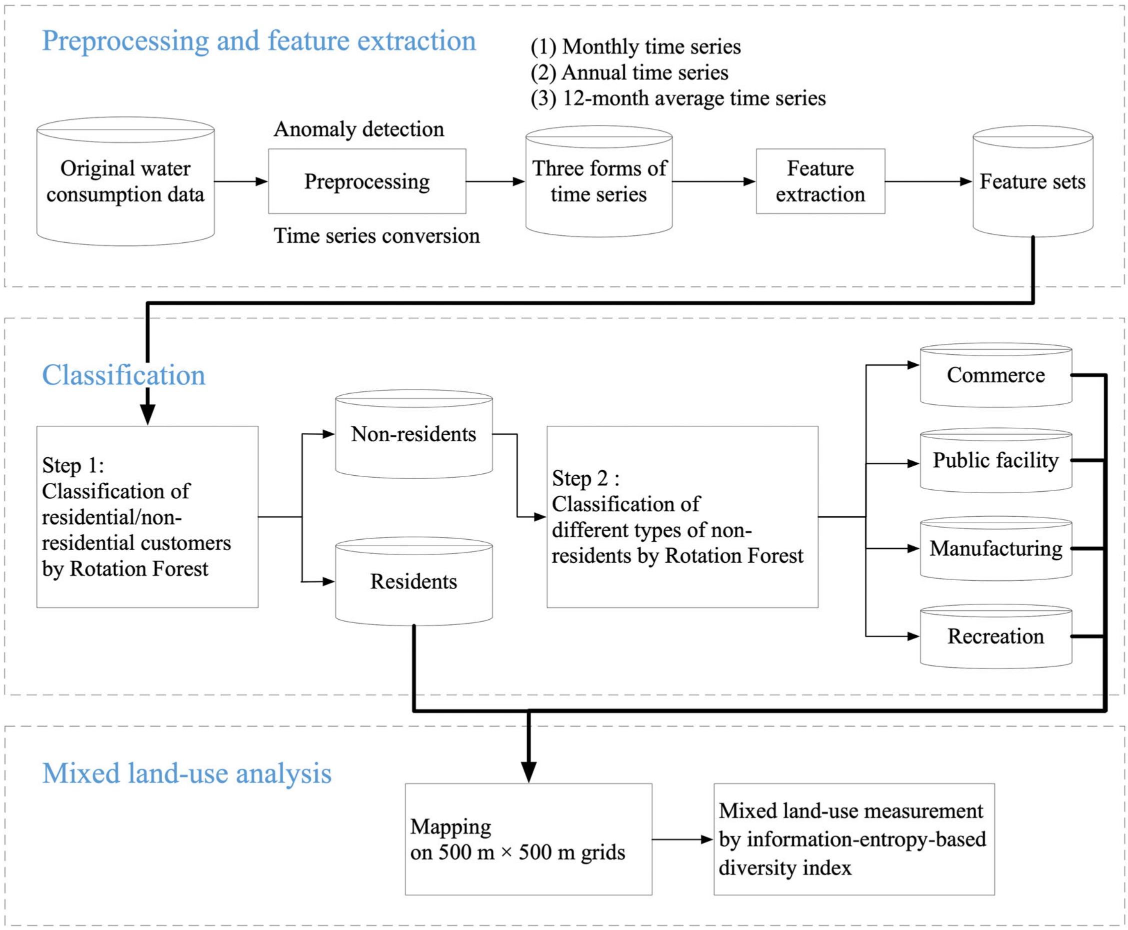


Figure 2. Methodological framework.

(where manual reading is used) to hourly and even minutely (where automatic reading is used). Monthly recording was used in this study, because it was the most commonly used interval in the past decades in China. The methodology proposed in this study can also be applied to water consumption data with other intervals.

Before identifying the socioeconomic types of individual customers, the original water consumption time series must be preprocessed and features must be extracted to represent the consumption patterns, which were used for the following classification of customers.

Anomaly Detection and Processing. Owing to reading errors and unusual events, anomalous values existed in the original time series and required removal, because they would otherwise sabotage the analysis of the long-term general patterns of water consumption and degrade the quality of classification. The selection of an anomaly detection method depends on the statistical characteristics of the data, especially the frequency distribution (M. Liu and Zhang 2011). It was found that most anomalous time series of water consumption contained a very small number of extreme values at random times, exhibiting long-tailed distributions.

Therefore, head–tail breaks, a clustering algorithm for data with a heavy-tailed distribution (Bingham and Spradlin 2011), was used to detect anomalous values. The head–tail breaks method recursively divides the data into a head section (i.e., large values) and a tail section (i.e., small values) around the mean value, until the head no longer exhibits a heavy-tailed distribution. Through such recursive breaks, the extreme values (treated as anomalous values in this study) in the water consumption time series were identified. Each anomalous value was then replaced by the average value of its neighbors along the time series. It is worth noting again that the extreme values in a consumption time series might contain important information regarding an individual customer (e.g., abnormal events). For a group of customers of the same socioeconomic type, however, such extreme values would reduce the accuracy of classification; therefore, removal was required for the purpose of this study.

Multigranularity Time Series Conversion. The water consumption time series of different customers are usually irregular and inconsistent, with variations in time span and vacancies or blanks (as shown in Figure 1B). Therefore, time series classification and clustering methods, such as dynamic time warping (Chen et al. 2017), are not applicable to the identification of the socioeconomic types of individual customers.

To cope with the irregularity and inconsistency of time series, this study adapted a classification approach based on the features extracted from the time series. To obtain more comprehensive information, three forms of time series with various temporal granularities and spans were generated from the original time series: (1) a monthly consumption time series covering the entire time span of the study, in which each element represented the water consumption of a particular month; (2) an annual consumption time series covering the entire time span, in which each element represented the total water consumption of a particular year; and (3) a twelve-month average consumption time series, in which each element represented the average consumption of a particular month in all years. For example, the average water consumption for January of all years in the study period was used to represent the typical water consumption in January. These three time series forms with various temporal resolutions and degrees of aggregation contained a wide range of information, including monthly variation, seasonal variation, and annual variation, that could be extracted as numerical features for the following classification. It is worth noting that time series with lengths less than one year were eliminated because they were too short to reflect seasonal variations and trends.

Feature Extraction. As shown in Table 1, a total of fifty-one features were extracted from the three time series forms, including statistical and shape features (features 1–15 in the table), time domain features (features 16–18 in the table), frequency domain features (feature 19 in the table), and features based on statistical models and other models (features 20–28 in the table). All of these numerical features were normalized into the range of [0, 1].

Some important features were calculated as follows. The seasonal indexes reflect the seasonal variations in the time series (J. L. Shi 2001). Because water consumption can be influenced by seasons and other temporal factors, these indexes play an important part in the analysis. First, $\overline{x_k}$, the average of each period, is calculated as

$$\overline{x_k} = \frac{\sum_{i=1}^{n} x_{ik}}{n}, k = 1, 2, \ldots, m, \tag{1}$$

where m represents the number of periods (with vacancies in between), n represents the number of months in each period, and x_{ik} represents the value of the ith month in the kth period. The overall average $\overline{x}$ is

$$\overline{x} = \frac{\sum_{i=1}^{n} \sum_{k=1}^{m} x_{ik}}{nm}. \tag{2}$$

The seasonal index S_k is

$$S_k = \frac{\overline{x_k}}{\overline{x}}. \tag{3}$$

Mann–Kendall is a nonparametric method used to test the trend of a time series. This method contains three possible hypotheses (negative trend, nonzero trend, and positive trend), which are determined by calculating the Mann–Kendall statistics (for details, please refer to Moran and Kendall 1973).

Two-Step Rotation Forest for Socioeconomic Type Classification

Once the features were extracted from the water consumption time series for each customer, customer socioeconomic type was then identified through a two-step classification procedure using rotation forest.

Rotation Forest. Stemming from the widely used random forest, rotation forest is an ensemble learning strategy. Ensemble learning consists of a set of hypotheses in which each hypothesis "votes" for the prediction (Dietterichl 2002). Knowledge learned from base classifiers can be diverse, and mislabeled samples can be modified by additional classifiers and samples. Therefore, multiple-classifier ensembles can achieve better performance.

Rotation forest was first proposed by Rodríguez, Kuncheva, and Alonso (2006). It is a method for generating classifier ensembles based on feature extraction. It has been demonstrated that a greater

Table 1. Features of water consumption time series

No.	Group	Feature	Monthly time series	Annual time series	Twelve-month average time series
1	Statistical and shape features	Length of series	√	√	
2		M	√		√
3		M (excluding outliers)	√		√
4		*SD*		√	√
5		Range	√		√
6		Median	√		√
7		Kurtosis			√
8		Skewness			√
9		Burstiness statistic		√	√
10		High–low mu statistic	√		√
11		Negative log-likelihood of data derived from a Gaussian distribution	√		√
12		Proportion of data points within *p* standard deviations of the mean	√		√
13		Moment of the distribution for the input time series	√		√
14		Simple mean-stationarity metric	√		
15		Coefficient of variation	√		√
16	Time domain features	Seasonal indexes (four seasons)			√
17		Trend (Mann–Kendall method)		√	√
18		Autocorrelation	√		
19	Frequency domain features	Fitting parameters for the Fourier transform	√		
20	Features based on statistical models and other models	Fit (normally distributed or not)			√
21		Fitting parameters of the normal distribution			√
22		Fitting parameters of the geometric distribution			√
23		Goodness of a polynomial fit to a time series (three different parameters)			√
24		Sliding window measure of stationarity (two different parameters)		√	√
25		Entropy	√		
26		Approximate entropy	√		
27		Custom Pearson skewness measures	√		
28		Custom Bowley skewness measures	√		

diversity of classifiers results in better performance (Margineantu and Dietterich 1997). Rotation forest takes advantage of the features of random forest and improves them (Akar 2018). As shown in Figure 3, before the sampling of each subsample, the sample feature set is randomly grouped and principal component analysis (PCA) is applied to transform the features to make the subsamples different, further diversifying the base classifiers and, thus, improving the accuracy (Xia et al. 2014).

The steps of rotation forest are as follows. Given a data point described by n features, $X = [x_1, x_2, \ldots, x_n]^T$, where X represents a data set with N training objects in the form of an $N \times n$ matrix. $Y = [y_1, y_2, \ldots, y_n]^T$ is a vector of class labels for the data, where y_i assumes a value from the set of class labels $\{w_1, w_2, \ldots, w_c\}$. $D_1, D_2, \ldots, D_L$ represent the classifiers in the ensemble, where L is picked in advance and F represents the feature set.

The training set for classifier D_i is constructed through the following steps.

1. Split F randomly into K subsets (K is a parameter of the algorithm). To maximize the chance for high diversity, the subsets are made to be disjointed. In addition, each feature subset contains $M = n/K$ features.
2. $F_{i,j}$ represents the jth subset of features for the training set of classifier D_i. For each such subset, a nonempty subset of classes is randomly selected and bootstrap sampling of the objects is performed for approximately

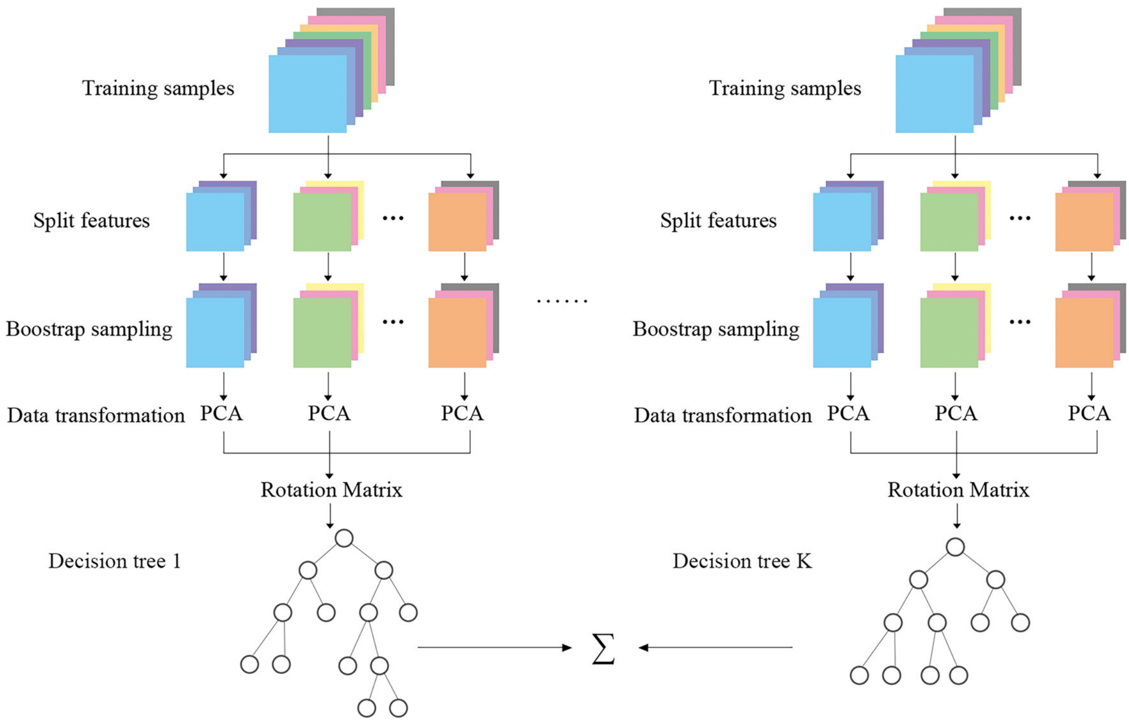


Figure 3. Process of rotation forest. *Note:* PCA = principal component analysis.

60 to 80 percent of the total samples (the commonly used proportions for machine learning methods). PCA is applied using only the M features in $F_{i,j}$ and the selected subset of X. The coefficients of the principal components are stored, $a(1)_{i,j}, \ldots, a(m_j)_{i,j}$, where each has a dimension of M $\times$ 1. Note that it is possible for some eigenvalues to be zero; therefore, $M_j \leq M$. PCA is applied to a subset of classes, instead of to the entire set, to avoid identical coefficients if the same feature subset is chosen for different classifiers.

3. The obtained vectors are organized, with coefficients in a sparse rotation matrix R_i, as

$$R_i = \begin{bmatrix} a_{i,1}^{(1)}a_{i,1}^{(2)}\ldots a_{i,1}^{(M1)}, & [0] & \cdots & [0] \\ [0] & a_{i,2}^{(1)}a_{i,2}^{(2)}\ldots a_{i,2}^{(M2)}, & \cdots & [0] \\ \vdots & \vdots & \ddots & \vdots \\ [0] & [0] & \cdots & a_{i,K}^{(1)}a_{i,K}^{(2)}\ldots a_{i,K}^{(MK)}, \end{bmatrix}. \quad (4)$$

The rotation matrix has a dimension of $n \times \sum_j M_j$. The columns of R_i (i.e., the features) are rearranged such that they correspond to the original features, which results in a rearranged rotation matrix, $R^a{}_i$ (size $N \times n$). Then, the training set for classifier D_i is represented by $XR^a{}_i$.

In the classification phase for a given x, let $d_{i,j}(xR^a{}_i)$ represent the probability assigned by classifier D_i to the hypothesis that x is derived from class w_j. The confidence for each class w_j is calculated by the average combination method

$$\mu_j(x) = \frac{1}{L}\sum_{i=1}^{L} d_{i,j}(xR_i^a), \;\; j = 1, \ldots, c. \quad (5)$$

Then, x is assigned to the class with the largest confidence.

Customer Socioeconomic Type Classification. In this study, the classification of customer socioeconomic types consisted of two steps. With such a two-step classification approach, the idea of hierarchical classification was adopted, which can effectively reduce the complexity of classification to improve the accuracy. Two-step classification was implemented using Weka 3.8, a freely available software package for data mining and machine learning (see https://www.cs.waikato.ac.nz/ml/weka/).

The first step aimed to separate residents from the rest of the customers, because all residential water consumption had similar features that were sufficiently unique to be identified. For the training of the first classification, a total of 30,000 customers with known resident or nonresident labels (determined by customer name) were selected as samples and divided into the training set (60 percent), validation set (20 percent), and test set (20 percent). Trained using the training sample set and validated using the validation set, the rotation forest achieved an accuracy of 83 percent for the test sample set, which was higher than the accuracy of 81 percent achieved by random forest. By applying the trained model to the entire data set, 405,768 customers were classified, resulting in approximately 91.5 percent of customers as residents and 8.5 percent as nonresidents.

The second step then considered the variations among types of nonresidential water consumption. The nonresidential customers were further classified into commerce, government and public facilities, manufacturing, and recreation. The number of sample customers with known socioeconomic functions (determined by customer name) varied greatly across the classes. To balance the sample size by increasing the sample sizes for the classes with fewer customers, random sampling with replacement was used such that some customers might be sampled more than once. Finally, approximately 20,000 samples were obtained, and the proportions of the training, validation, and test sets were also 6:2:2. After training and validation, the classification achieved an accuracy of 76 percent over the test set. The classification results of all nonresidential customers showed that the proportions of commerce, government and public facilities, manufacturing, and recreation were 29.6 percent, 8.1 percent, 54.8 percent, and 7.5 percent, respectively.

Delineating Mixed Land-Use Patterns

After all customers' socioeconomic types were identified, the land-use patterns of all years were delineated by projecting the socioeconomic types at their corresponding locations. Because the locations of customers are points, the spatial area of the land use of each customer is unknown. Therefore, instead of analyzing the areal structure of land use as performed in many other studies (Eck and Koomen 2008; Tian, Liang, and Zhang 2017), this study focused on the intensity structure of land use in each land unit. The concept of land-use intensity is often used in agricultural and environmental studies, referring to land-based production in a broad sense (including agriculture, grazing, and forestry), which can be measured by various metrics of inputs (capital, labor, and technologies) and outputs or effects (agricultural products and biodiversity) per land unit (Kuemmerle et al. 2013). This study expanded the concept and defined the intensity of a certain type of land use as the intensity of the corresponding socioeconomic activities, which can be reflected by the amount of water consumption of the corresponding type of customers within a land unit. In other words, the land-use patterns were derived from the structural composition of water consumption within a group of land units.

Two sets of maps were produced to represent the mixed land-use patterns using the results of the two classification steps: one for residential and nonresidential types from the first classification and the other for all five types of socioeconomic types from the second classification. The mapping procedure was as follows: (1) the entire study area was divided into regularly shaped $500\,\mathrm{m} \times 500\,\mathrm{m}$ grids (a total of 5,485 grids) and (2) the amounts of water consumed through the various socioeconomic activities and their percentages within each grid were calculated to indicate the intensities of the socioeconomic activities in a land unit. In addition, the growth of total water consumption over the years was calculated for each grid to represent the expansion and intensification processes of socioeconomic activities.

After the first step of classification, the customers were divided into residents and nonresidents. Based on the ratio of residential to nonresidential water consumption in each grid, the land-use pattern of residential–nonresidential functions was obtained. To understand the distributions of urban residential and nonresidential land use, kernel density analysis (Portnov and Zusman 2014) was deployed to identify the city centers.

After the second step of classification, the nonresidential customers were further divided into four types: commerce, government and public facility, manufacturing, and recreation. All five types of customers (i.e., residents and four types of nonresidential customers) were used to analyze mixed land use. To quantify the mixed degree of landuse, the landuse diversity index based on information entropy (L. I. Jiang and Guo 2002) was used. Instead of the areas of land-use types, the water consumption amounts of various types of customers were used to measure the mixed degree in every 500 m × 500 m grid. The diversity index was calculated as follows.

The total water consumption within a grid is denoted as S, number of land-use types is denoted as m, and amount of water consumption for the ith type within the grid is denoted as S_i.

$$S = \sum_{i=1}^{m} S_i, \quad i = 1, \; 2, \ldots, m. \tag{6}$$

Therefore, the proportion of the ith type within the grid can be calculated as

$$P_i = \frac{S_i}{S} \quad \sum_i P_i = 1. \tag{7}$$

According to the theory of information entropy, the land-use diversity index H is defined as

$$H = -\sum_{i=1}^{m} (P_i)\ln(P_t). \tag{8}$$

This index describes the complexity of land-use types, in which a larger H indicates a higher degree of land-use mixing in the area. $H = 0$ indicates unmixed land use. When $S_1 = S_2 = \ldots = S_m$, $H_{\max} = \ln(m)$, and the mixed land use is well proportioned. In general, more types and better proportions indicate that the mixed land-use diversity is higher.

Results

The results showed that the number of municipal water customers (especially residential customers) gradually increased during the period of 2004 to 2013 in Changshu (Figure 4A). Among the four categories of nonresidential customers, manufacturing and commercial customers saw large increases during the time period (Figure 4B).

Overall Water Consumption Patterns

The spatial distributions of water consumption for all years were represented by a series of maps, depicting the spatiotemporal dynamics of water consumption during the period of 2004 to 2013. Over the ten years, the proportion of grids with nonzero water consumption increased from 46.73 percent to 55.31 percent of the 5,485 grids, and the average water consumption per grid increased from 1897.36 m^3 to 2674.3 m^3, indicating not only the spatial expansion but also the intensification of socioeconomic

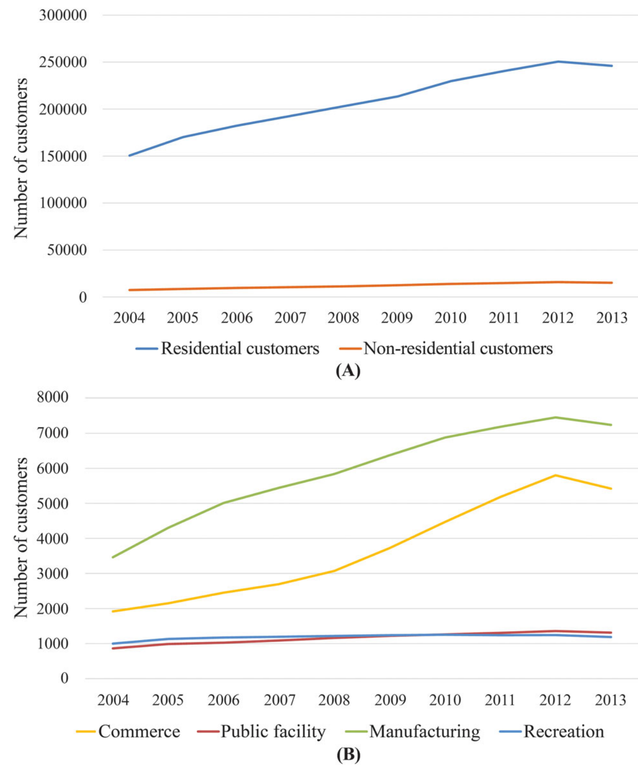


Figure 4. Profile of the number of customers for the various types of socioeconomic functions from 2004 to 2013. (A) Changes in the numbers of residential and nonresidential customers. (B) Changes in the numbers of four types of nonresidential customers.

activities in Changshu (Figure 5). Furthermore, the areas with higher water consumption (hot spots) are clearly shown in Figure 5A, indicating the spatial concentration of socioeconomic activities. Figure 5B shows that the areas with rapid water consumption growth were concentrated around the city center in the early years (i.e., 2004–2007) and then gradually spread out over the years. In particular, the subcenters in the northeast and south experienced high water consumption growth in later years. Overall, the growth of water consumption was faster in the early years, indicated by the rapid expansion of nonzero consumption grids (i.e., newly appeared grids in Figure 5B) and the large number of grids with high growth rates (i.e., red grids in Figure 5B).

Residential Land Use versus Nonresidential Land Use

The ratio of residential/nonresidential water consumption can be used as an indicator to identify typical residential areas in the city (Zelinsky and Sly 1984). The higher the ratio, the more likely the area

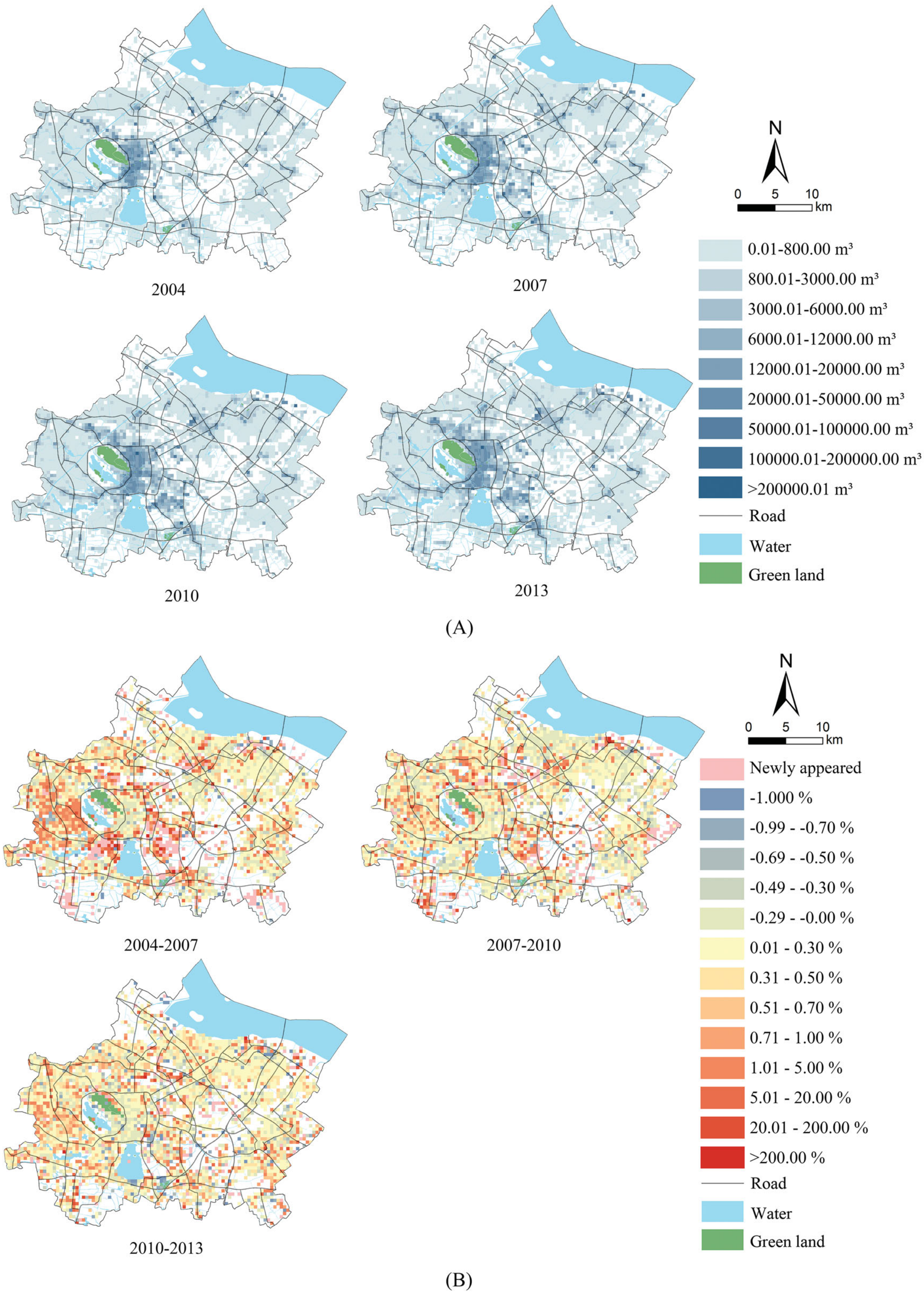


Figure 5. Spatial distributions and growth rates of water consumption from 2004 to 2013. (A) Spatial distributions of water consumption in various years, where darker blue indicates higher water consumption. (B) Growth rates of water consumption from 2004 to 2013, where cooler color (i.e., negative growth rate) indicates larger decrease of water consumption and warmer color (i.e., positive growth rate) indicates larger increase of water consumption.

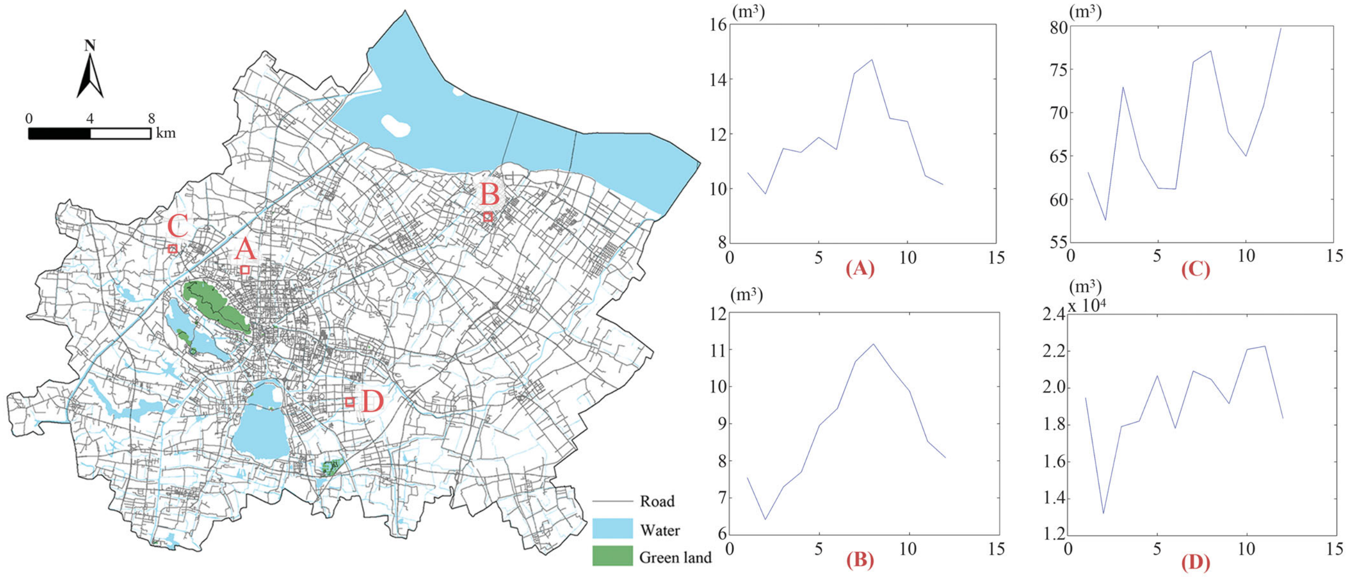


Figure 6. Water consumption time series of typical residential and nonresidential grids. Grids A and B represent typical residential areas, Grid C is an industrial area, and Grid D is a commercial area.

is a residential community. As shown in Figure 6, Grid A (located in the Longjiang community) and Grid B (located in the Hejiabang community) are typical residential areas, because their ratios of residential/nonresidential consumption are 0.996 and 0.956, respectively. On the contrary, Grid C (located in an industrial park in the town of Yushan) and Grid D (located in the Changshu high-tech industrial development zone) are typical nonresidential areas, because their ratios are 0.088 and 0, respectively. The twelve-month average consumption time series of these grids are also shown in Figure 6. The water consumption curves of the residential grids (e.g., Grids A and B) are similar, in terms of both amount and temporal variation. The curve usually has a spike during the period from June to October (Figures 6A, 6B), which is a typical seasonal characteristic of residents. However, the water consumption curves of nonresidential grids (e.g., Grids C and D) are relatively various, because different nonresidential functions result in different water consumption patterns (Gui, Li, and Gao 2016). Some exhibit apparent seasonality, whereas other fluctuate. For example, Grid C is an industrial area where some material factories are located, whereas Grid D is a commercial area where some electronic corporations are located, resulting in quite different water consumption curves (Figures 6C, 6D). In general, water consumption patterns are largely related to the socioeconomic activities; hence, they are related to the land-use types.

The results of the kernel density analysis showed that there was one large residential center in the central area of the city, in which the residential land use had been gradually intensified over the years, indicated by the increase in residential water consumption in this area (Figure 7A). As for nonresidential land use, a main center was located in the central area of the city in the early years, later split into several smaller subcenters around the central area (Figure 7B). Meanwhile, some subcenters appeared in the northeast and south and expanded gradually over the years. In addition, the distribution and evolution of the overall water consumption (Figure 7C) were more similar to those of nonresidential socioeconomic functions, indicating that nonresidential users had a greater influence than residential users in terms of water consumption.

The central area of Changshu was clearly the core of the city, where a large number of residents, manufacturing, and retail activities were located. The subcenters in the northeast and south (Regions 1 and 2 in Figure 7C) emerged in later years where nonresidential activities were clustered, resulting in a higher proportion of nonresidential land use. The gradual expansion and intensification of socioeconomic activities occurred at these subcenters during the time span of the study. These phenomena were

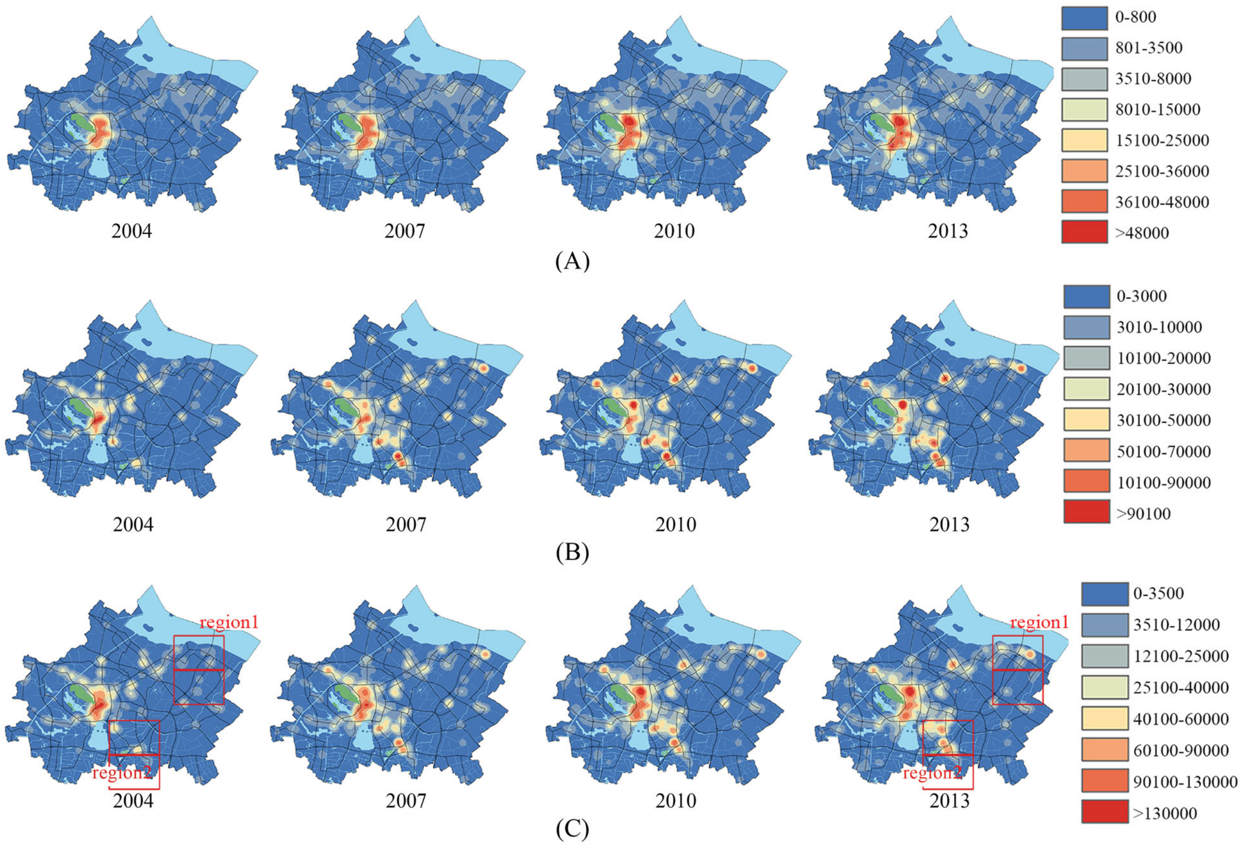


Figure 7. Concentrated areas of (A) residential water consumption, (B) nonresidential water consumption, and (C) overall water consumption during 2004 to 2013. Because nonresidential water consumption is much larger than residential water consumption, different intervals are used in (A), (B), and (C).

inseparable from the development of Changshu. Explanations in detail are given later.

Mixed Urban Land-Use Patterns

The land-use diversity analysis showed that highly mixed grids were concentrated in the centers of the city and towns, indicating a higher degree of urbanization with relatively well-balanced work–residence mixture and more convenient access to various facilities for residents (Figure 8). Furthermore, some highly mixed grids were distributed along the major roads, because convenient accessibility is one of the key factors in the concentration and intensification of nearly all types of socioeconomic activities.

Temporally, the degree of land-use mixing increased throughout the study period. The proportion of mixed grids increased from 33 percent in 2004 to 42 percent in 2013. Meanwhile, the average land-use diversity index for all nonzero water consumption grids increased from 0.26 to 0.34. In addition, Figure 8 shows that mixed land use expanded from the central area to its surrounding areas. The increase in land-use mixing mainly occurred in the central area of the city in the early years and then scattered into other areas (especially the northeastern and southern subcenters) in later years. Such a trend in land-use mixing was consistent with the spatiotemporal dynamics of the overall water consumption, indicating the maturity of urbanization following the spatial expansion of urbanization.

We also identified the dominant socioeconomic function of each grid according to the proportions of socioeconomic types. As shown in Figure 9, the following observations were found.

1. Resident-dominated grids occupied the largest proportion of the urban area.
2. Grids dominated by recreational functions (including the catering industry, shopping malls, and other leisure venues) only comprised a very small part of the

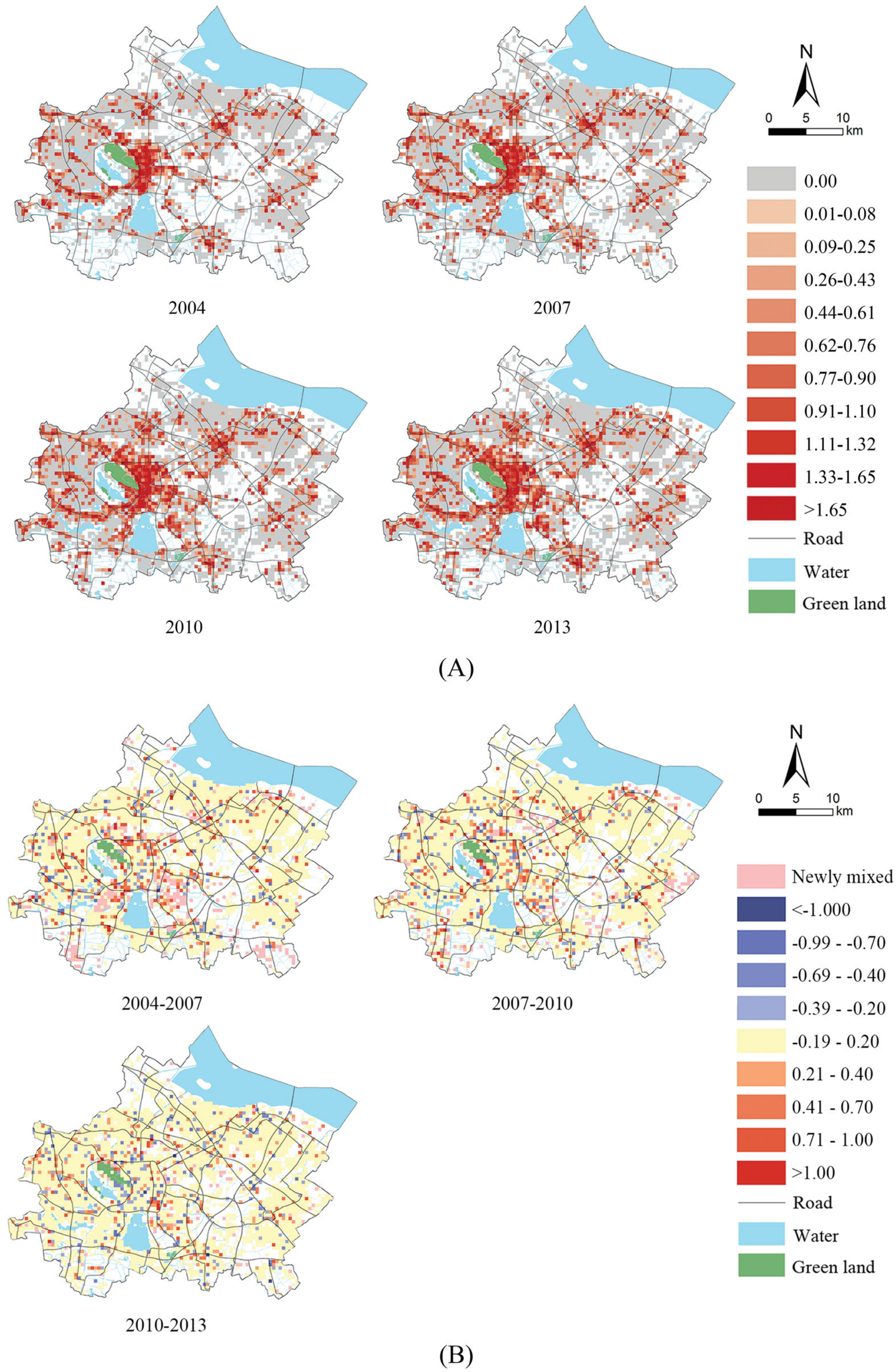


Figure 8. Mixed land-use patterns during 2004 to 2013. (A) Mixed land-use patterns in various years measured by the land-use diversity index, where darker red indicates a higher degree of land-use mixing. (B) Changes in land-use diversity indexes from 2004 to 2013, where cooler color (i.e., negative value) indicates a larger decrease in land-use mixing, and warmer color (i.e., positive values) indicates a larger increase in land-use mixing.

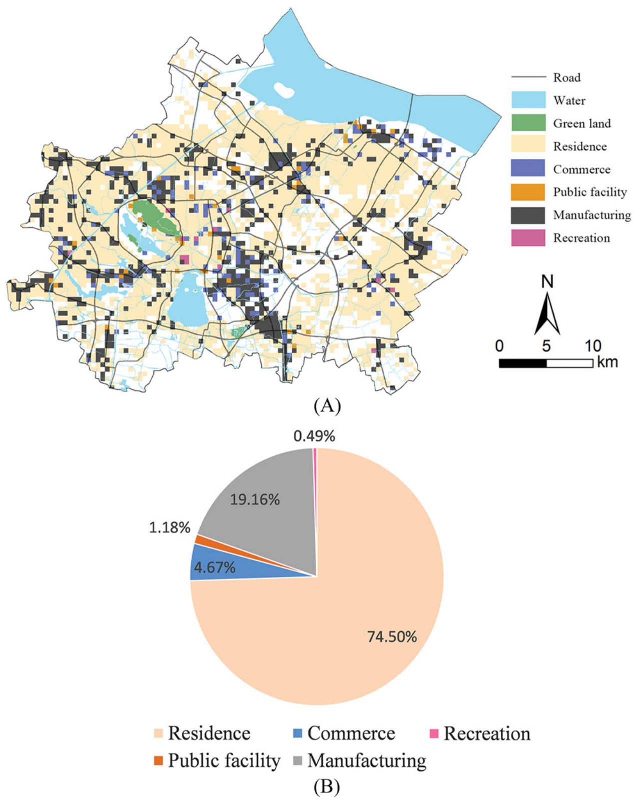


Figure 9. Spatial distribution and areal proportions of dominant landuse types in 2013. (A) Spatial distribution of dominant land-use types in 2013. (B) Areal proportions of dominant landuse types in 2013.

urban area and were mainly distributed in the city center and town centers.

3. Grids dominated by public facilities (e.g., government agencies and infrastructures) were also rare, because public facilities were often mixed in residential or various industrial areas but did not occupy a dominant proportion. This type of grid was mainly located in the city center and town centers.
4. Manufacturing-dominated grids accounted for the second largest proportion of the urban area and were distributed in several areas; for example, in the northeastern and southern subcenters and along the main roads.
5. Commerce-dominated grids were the second largest nonresidential land use, with a distribution similar to that of the manufacturing grids.

Discussion and Conclusions

Mixed urban land-use patterns are formed during the process of urban development and are mainly affected by socioeconomic activities. This study proposed a framework to delineate and analyze mixed land-use patterns and dynamics by identifying the socioeconomic types of individual municipal water customers and measuring land-use mixing using information entropy. Through the application of the proposed framework to a typical county-level city in China, Changshu over the ten-year period of 2004 to 2013, the following observations were made.

Rapid Urbanization with a Declining Rate. The urbanization of Changshu experienced both spatial expansion and intensification during the period from 2004 to 2013. The development rate was high in the early years and gradually declined in the later period. The expansion of the central area and the developments along the river were consistent with the findings from other studies regarding the urbanization of fast developing areas in China (Chen et al. 2013). Many studies have demonstrated that Changshu is a city with rapid urbanization (Cao and Zhang 2010). Rapid development often leads to problems, however, such as insufficient facilities and large differences between urban and rural areas, eventually resulting in a decrease in the pace of urbanization (T. Yu, Zhang, and Luo 2010). Our study revealed a large difference in urbanization between the central area and suburban or rural areas and a decrease in the urbanization rate in recent years in Changshu, in accordance with previous studies.

Rise of Subcenters for Industrial Development. Besides the main center of Changshu, two subcenters gradually emerged in the northeast and south over the study period. The central area, Yushan town, was the most developed part of Changshu, with the largest population and highest gross domestic product. The textile industry has been well established in Yunshan, which was known as the textile town famous for winter clothing manufacturing (Zhao 2013). The formation of subcenters in the northeast and south was due to the development of national-level economic and technological development zones. In recent years, high-tech industries, such as automobile and machinery manufacturing, have been rapidly growing in the development zones, resulting in a higher proportion of nonresidential land use.

Increasing Degree of Land-Use Mixture. The land use of the central area of Changshu was highly mixed, and the degree of land-use mixture has increased along with urban growth, indicating the maturing process of urbanization. The textile industry in Changshu, especially in Yushan town, started as family-owned manufacturing businesses in the early years. With the growth of the textile industry, the center of Changshu gradually developed into an area in which a large number of residents were integrated into the functions of the textile industry, including production and sales. This increasing degree of mixture was partly caused by the spatial concentration of industry, especially in Yushan town (central area of Changshu) and Bixi District (subcenter in the northeast; Ding et al. 2014).

Methodologically, this article presents a new approach to quantitatively delineating and analyzing mixed urban land use. Compared with existing data-driven urban land-use studies, the main contributions of this article can be described in two aspects.

First, this article proposes a framework to identify the socioeconomic types of individual customers and analyze mixed land-use patterns using municipal water consumption data. Through the transformation from the original time series to multigranular time series and the extraction of features, the proposed method can handle the irregular and inconsistent time series of municipal services. Furthermore, with the adoption of the idea of hierarchical classification, the two-step rotation forest used in this framework can effectively improve the accuracy of classifying complex consumption patterns.

In the application to Changshu, the municipal water customers were first classified into residents and nonresidents with an accuracy of 0.83, and the nonresidents were further classified into four socioeconomic types with an accuracy of 0.76. Such classification accuracies are comparable with those of other studies on land-use classification using social sensing data. The accuracies obtained using a single data source are often below 0.8. For example, in the Toole et al. (2012) study, the land-use classification based on mobile phone activities using random forest achieved an accuracy of 0.54, whereas the accuracy for nonresidential land use was 0.4. Based on points of interest, Yao et al. (2017) achieved an accuracy of 0.55 using frequency-inverse document frequency, an accuracy of 0.74 using probabilistic latent semantic analysis, and an accuracy of 0.67 using latent Dirichlet allocation. Therefore, our study demonstrated that through the classification of individual customer socioeconomic functions based on municipal water consumption patterns, urban land-use patterns can be delineated at fine scales.

Second, this article presents the analysis of long-term urban land-use dynamics, whereby commonly used big data sources, such as social media, can only reflect short-term land-use patterns and remote sensing data can hardly reflect socioeconomic information. In addition, previous data-driven urban land-use studies mostly focused on large or developed cities, because of the easy availability and large population coverage of social media and public transit data for large cities. This study takes advantage of the fact that municipal water consumption data cover a major proportion of the population and a wide range of socioeconomic activities, which provides an opportunity for the study of small or developing cities and towns where smart devices and Internet services are less widely used. In addition, this article sheds light on the overall distributions and long-term evolution of mixed land use in cities, as well as the relevant formation mechanisms. These elements help to not only understand the process of urbanization but also discover the characteristics of a specific city, which is key to the sustainable development and management of cities.

This study also has limitations. There are still some developing cities in China where municipal water is not as widely available as in the study area presented in this article, and the water consumption data might not cover a major proportion of the population. In future work, we will discuss the bias of water consumption data and consider the use of multisource data fusion to improve the population coverage. Moreover, we have not obtained municipal water consumption data from other cities. If this method is applied to other cities, we need to fully consider the spatial and temporal heterogeneities among cities. In addition, the classification granularity of the socioeconomic types in this study is relatively coarse. In our future work, more detailed classification will be introduced to differentiate the different types of industrial and commercial functions.

Acknowledgments

We sincerely thank the editors and anonymous reviewers for their constructive comments and suggestions that significantly strengthened this article. Dr. Wen Zeng (zengwen@cug.edu.cn) serves as the corresponding author for this article.

Funding

This work was supported by the National Natural Science Foundation of China (Grant Nos. 41671408, 41801306), Natural Science Foundation of Hubei Province (Grant No. 2017CFA041), and Special Fund for Foundation and Frontier of Applications of Wuhan (Grant No. 2018010401011293).

ORCID

Qingfeng Guan http://orcid.org/0000-0002-7392-3709
Yongting Pan http://orcid.org/0000-0002-7726-9302
Yao Yao http://orcid.org/0000-0002-2830-0377
Wen Zeng http://orcid.org/0000-0002-1550-7926

References

Akar, Ö. 2018. The rotation forest algorithm and object-based classification method for land use mapping through UAV images. *Geocarto International* 33 (5):538–53. doi: 10.1080/10106049.2016.1277273.

Bennett, D. A., W. Tang, and S. Wang. 2011. Toward an understanding of provenance in complex land use dynamics. *Journal of Land Use Science* 6 (2–3):211–30. doi: 10.1080/1747423X.2011.558598

Bingham, A., and D. Spradlin. 2011. *The long tail of expertise*. Santa Monica, CA: Pearson Education.

Blaschke, T. 2010. Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry & Remote Sensing* 65 (1):2–16. doi: 10.1016/j.isprsjprs.2009.06.004.

Boarnet, M. G. 2011. A broader context for land use and travel behavior, and a research agenda. *The Journal of the American Planning Association* 77 (3):197–213. doi: 10.1080/01944363.2011.593483.

Bourne, L. S. 1976. Urban structure and land use decisions. *Annals of the American Association of Geographers* 66 (4):531–35. doi: 10.1111/j.1467-8306.1976.tb01108.x.

Bratasanu, D., I. Nedelcu, and M. Datcu. 2011. Bridging the semantic gap for satellite image annotation and automatic mapping applications. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 4 (1):193–204. doi: 10.1109/JSTARS.2010.2081349.

Cao, G., and P. Zhang. 2010. The partial urbanization phenomenon in developed regions: Changshu example. *Planners* 4:16.

Chang, H., M. R. Bonnette, P. Stoker, B. Crow-Miller, and E. Wentz. 2017. Determinants of single family residential water use across scales in four western U.S. cities. *The Science of the Total Environment* 596–597:451–64. doi: 10.1016/j.scitotenv.2017.03.164.

Chen, Y., X. Liu, X. Li, X. Liu, Y. Yao, G. Hu, X. Xu, and F. Pei. 2017. Delineating urban functional areas with building-level social media data: A dynamic time warping (DTW) distance based *k*-medoids method. *Landscape & Urban Planning* 160:48–60. doi: 10.1016/j.landurbplan.2016.12.001.

Chen, Y., L. Xia, S. Wang, X. Liu, and B. Ai. 2013. Simulating urban form and energy consumption in the Pearl River Delta under different development strategies. *Annals of the Association of American Geographers* 103 (6):1567–85. doi: 10.1080/00045608.2012.740360.

Dietterichl, T. G. 2002. Ensemble learning. In *Handbook of brain theory & neural networks*, ed. L. Yann and B. Yoshua, 125–42. Cambridge, MA: MIT Press.

Ding, Z. S., Y. Wang, Z. Y. Shang, L. I. Ya-Ru, X. Y. Song, and X. J. Chang. 2014. The spatial characteristics of producer service agglomeration in town: The case of Changshu in Jiangsu Province. *Scientia Geographica Sinica* 253 (9):122–37.

Dovey, K., and E. Pafka. 2017. What is functional mix? An assemblage approach. *Planning Theory & Practice* 18 (2):249–67. doi: 10.1080/14649357.2017.1281996.

Eck, J. R. V., and E. Koomen. 2008. Characterising urban concentration and land-use diversity in simulations of future land use. *The Annals of Regional Science* 42 (1):123–40. doi: 10.1007/s00168-007-0141-7.

Elwood, S., M. F. Goodchild, and D. Z. Sui. 2012. Researching volunteered geographic information: Spatial data, geographic research, and new social practice. *Annals of the Association of American Geographers* 102 (3):571–90. doi: 10.1080/00045608.2011.595657.

Gui, C., C. Li, and K. Gao. 2016. A case study of water consumption change and water-use pattern for city industries. *Journal of Shenzhen University Science & Engineering* 33 (1):49–54. doi: 10.3724/SP.J.1249.2016.01049.

Horner, M. W., T. Zhao, and T. S. Chapin. 2011. Toward an integrated GIScience and energy research agenda. *Annals of the Association of American Geographers* 101 (4):764–74. doi: 10.1080/00045608.2011.567938.

Huang, X., Q. Lu, and L. Zhang. 2014. A multi-index learning approach for classification of high-resolution remotely sensed images over urban areas. *ISPRS Journal of Photogrammetry & Remote Sensing* 90 (4):36–48. doi: 10.1016/j.isprsjprs.2014.01.008.

Jiang, L. I., and Q. S. Guo. 2002. Analysis of dynamic evolvement in urban land-use composition based on Shannon entropy. *Resources and Environment in the Yangtze Basin* 11 (5):393–97.

Jiang, Y., Z. Li, and S. L. Cutter. 2019. Social network, activity space, sentiment, and evacuation: What can social media tell us? *Annals of the American Association of Geographers* 109 (6):1795–810. doi: 10.1080/24694452.2019.1592660.

Kitamura, R., P. L. Mokhtarian, and L. Daidet. 1997. A micro-analysis of land use and travel in five neighborhoods in the San Francisco Bay Area. *Transportation* 24 (2):125–58. doi: 10.1023/A:1017959825565.

Kuemmerle, T., K. Erb, P. Meyfroidt, D. Müller, P. H. Verburg, S. Estel, H. Haberl, P. Hostert, M. R. Jepsen, T. Kastner, et al. 2013. Challenges and opportunities in mapping land use intensity globally.

Current Opinion in Environmental Sustainability 5 (5):484–93. doi: 10.1016/j.cosust.2013.06.002.

Kwan, M. P. 2016. Algorithmic geographies: Big data, algorithmic uncertainty, and the production of geographic knowledge. *Annals of the Association of American Geographers* 106 (2):274–82.

Liu, M., and J. Zhang. 2011. Time series abnormality diagnosis method. *Chinese Journal of Health Statistics* 28 (4):478–80.

Liu, X., J. He, Y. Yao, J. Zhang, H. Liang, H. Wang, and Y. Hong. 2017. Classifying urban land use by integrating remote sensing and social media data. *International Journal of Geographical Information Science* 31 (8):1675–96. doi: 10.1080/13658816.2017.1324976.

Liu, Y., X. Liu, S. Gao, L. Gong, C. Kang, Y. Zhi, G. Chi, and L. Shi. 2015. Social sensing: A new approach to understanding our socioeconomic environments. *Annals of the Association of American Geographers* 105 (3):512–30. doi: 10.1080/00045608.2015.1018773.

Liu, Y., F. Wang, Y. Xiao, and S. Gao. 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. doi: 10.1016/j.landurbplan.2012.02.012.

Luo, Z. 2008. 2008 Jiangsu Provincial Government Work Report. Accessed June 4, 2020. http://www.gov.cn/test/2008-02/18/content_892000.htm

Margineantu, D. D., and T. G. Dietterich. 1997. Pruning adaptive boosting. Paper presented at International Conference on Machine Learning, Nashville, TN, July 8.

Meentemeyer, R. K., W. Tang, M. A. Dorning, J. B. Vogler, N. J. Cunniffe, and D. A. Shoemaker. 2013. Futures: Multilevel simulations of emerging urban–rural landscape structure using a stochastic patch-growing algorithm. *Annals of the Association of American Geographers* 103 (4):785–807. doi: 10.1080/00045608.2012.707591.

Moran, P. A. P., and M. G. Kendall. 1973. Rank correlation methods. *Journal of the Royal Statistical Society* 41 (3):399–400. doi: 10.2307/1402637.

Niu, N., X. Liu, H. Jin, X. Ye, Y. Liu, X. Li, Y. Chen, and S. Li. 2017. Characterizing mixed-use buildings based on multi-source big data. *International Journal of Geographical Information Science* 32 (4):738–56. doi: 10.1080/13658816.2017.1325489.

Portnov, B. A., and M. Zusman. 2014. Spatial data analysis using kernel density tools. In *Encyclopedia of business analytics and optimization*, ed. J. Wang, 2252–64. Hershey, PA: IGI Global.

Ribeiro, A. I., A. Pires, M. S. Carvalho, and M. F. Pina. 2015. Distance to parks and non-residential destinations influences physical activity of older people, but crime doesn't: A cross-sectional study in a southern European city. *BMC Public Health* 15 (1):1–12. doi: 10.1186/s12889-015-1879-y.

Rodríguez, J. J., L. I. Kuncheva, and C. J. Alonso. 2006. Rotation forest: A new classifier ensemble method. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28 (10):1619–30. doi: 10.1109/TPAMI.2006.211.

Seto, K. C., R. Sánchez-Rodríguez, and M. Fragkias. 2010. The new geography of contemporary urbanization and the environment. *Annual Review of Environment and Resources* 35 (1):167–94. doi: 10.1146/annurev-environ-100809-125336.

Shi, J. L. 2001. On the unity of the methods of calculating seasonal indexes. *Journal of East China Shipbuilding Institute (Social Sciences)* 1 (2):45–47.

Shi, L., P. Xu, C. Wang, T. Guan, Y. Zhang, and H. Xu. 2018. A review of applying spatial modelling and GIS in residential water use. *IOP Conference Series Materials Science and Engineering* 392 (6):062106.

Singleton, A. D., S. E. Spielman, and D. Folch. 2018. *Urban analytics*. Thousand Oaks, CA: Sage.

Soliman, A., K. Soltani, J. Yin, A. Padmanabhan, and S. Wang. 2017. Social sensing of urban land use based on analysis of Twitter users' mobility patterns. *PLoS ONE* 12 (7):e0181657. doi: 10.1371/journal.pone.0181657.

Song, Y., and G.-J. Knaap. 2004. Measuring the effects of mixed land uses on housing values. *Regional Science and Urban Economics* 34 (6):663–80. doi: 10.1016/j.regsciurbeco.2004.02.003.

Tian, L., Y. Liang, and B. Zhang. 2017. Measuring residential and industrial land use mix in the peri-urban areas of China. *Land Use Policy* 69:427–38. doi: 10.1016/j.landusepol.2017.09.036.

Toole, J. L., M. Ulm, D. Bauer, and M. C. Gonzalez. 2012. Inferring land use from mobile phone activity. Paper presented at KDD '12—The 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Beijing, August 12.

van den Hoek, J. W. 2008. The MXI (Mixed-use Index) as tool for urban planning and analysis. Paper presented at Envisioning Corporate Real Estate in the Urban Future Conference, Brussels, May 26.

Villar-Navascués, R. A., and A. Pérez-Morales. 2018. Factors affecting domestic water consumption on the Spanish Mediterranean coastline. *The Professional Geographer* 64:1–13.

Wang, L., W. P. Sousa, and P. Gong. 2004. Integration of object-based and pixel-based classification for mapping mangroves with IKONOS imagery. *International Journal of Remote Sensing* 25 (24):5655–68. doi: 10.1080/014311602331291215.

Wang, Y., W. Teng, T. Ming-Hsiang, L. Hao, J. Wei, and F. Guo. 2016. Mapping dynamic urban land use patterns with crowdsourced geo-tagged social media (Sina-Weibo) and commercial points of interest collections in Beijing, China. *Sustainability* 8 (11):1202. doi: 10.3390/su8111202.

Wen, D., X. Huang, L. Zhang, and J. A. Benediktsson. 2016. A novel automatic change detection method for urban high-resolution remotely sensed imagery based on multiindex scene representation. *IEEE Transactions on Geoscience and Remote Sensing* 54 (1):609–25. doi: 10.1109/TGRS.2015.2463075.

Wu, C., L. Zhang, and L. Zhang. 2016. A scene change detection framework for multi-temporal very high resolution remote sensing images. *Signal Processing* 124:184–97. doi: 10.1016/j.sigpro.2015.09.020.

Wu, S. S., X. Qiu, E. L. Usery, and L. Wang. 2009. Using geometrical, textural, and contextual information of land parcels for classification of detailed urban land use. *Annals of the Association of American Geographers* 99 (1):76–98. doi: 10.1080/00045600802459028.

Xia, J., P. Du, X. He, and J. Chanussot. 2014. Hyperspectral remote sensing image classification based on rotation forest. *IEEE Geoscience and Remote Sensing Letters* 11 (1):239–43. doi: 10.1109/LGRS.2013.2254108.

Yang, H., J. Song, and M. Choi. 2016. Measuring the externality effects of commercial land use on residential land value: A case study of Seoul. *Sustainability* 8 (5):432. doi: 10.3390/su8050432.

Yang, J., Y. Li, N. F. Zhang, J. F. Yang, K. Kuang, Y. H. Hu, and W. G. Qi. 2015. Analysis of urban residential water consumption based on smart meters and fuzzy clustering. Paper presented at the IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing, Liverpool, UK, October 26.

Yao, Y., X. Li, X. Liu, P. Liu, Z. Liang, J. Zhang, and K. Mai. 2017. Sensing spatial distribution of urban land use by integrating points-of-interest and Google Word2Vec model. *International Journal of Geographical Information Science* 31 (4):825–48. doi: 10.1080/13658816.2016.1244608.

Yu, T., J. Zhang, and X. Luo. 2010. The research on urbanization quality of county-level cities in eastern developed area of China—A case study of Changshu City. *Urban Studies* 17 (11):7–12.

Yuan, J., Y. Zheng, and X. Xie. 2012. Discovering regions of different functions in a city using human mobility and POIs. KDD '12—The 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Beijing, August 12. doi: 10.1145/2339530.2339561.

Zelinsky, W., and D. F. Sly. 1984. Personal gasoline consumption, population patterns, and metropolitan structure: The United States, 1960–1970. *Annals of the Association of American Geographers* 74 (2):257–78. doi: 10.1111/j.1467-8306.1984.tb01452.x.

Zhang, X., and S. Du. 2015. A linear Dirichlet mixture model for decomposing scenes: Application to analyzing urban functional zonings. *Remote Sensing of Environment* 169:37–49. doi: 10.1016/j.rse.2015.07.017.

Zhang, X., S. Du, and Y.-C. Wang. 2015. Semantic classification of heterogeneous urban scenes using intrascene feature similarity and interscene semantic dependency. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 8 (5):2005–14. doi: 10.1109/JSTARS.2015.2414178.

Zhao, G. 2013. Changshu Yushan creates a textile and clothing special service industry cluster. *Textile and Clothing Weekly* 46:98.

Zheng, Y., L. Capra, O. Wolfson, and H. Yang. 2014. Urban computing: Concepts, methodologies, and applications. *ACM Transactions on Intelligent Systems and Technology* 5 (3):1–55. doi: 10.1145/2629592.

Zhu, J., X. Wang, L. Zhang, H. Cheng, and Z. Yang. 2015. System dynamics modeling of the influence of the TN/TP concentrations in socioeconomic water on NDVI in shallow lakes. *Ecological Engineering* 76 (5210):27–35. doi: 10.1016/j.ecoleng.2014.06.030.

QINGFENG GUAN is a Professor in the School of Geography and Information Engineering at China University of Geosciences, Wuhan, China. E-mail: guanqf@cug.edu.cn. His research interests include big spatio-temporal data analytics and mining, spatio-temporal modeling, spatial computational intelligence, and high-performance spatial computing.

SIJING CHENG received her Master degree from the School of Geography and Information Engineering at China University of Geosciences, Wuhan, China, and she is currently a Software Engineer in the PowerChina Huadong Engineering Co., Ltd., Hangzhou, China. E-mail: chengsijing@cug.edu.cn. Her research interests include big spatio-temporal data analytics and mining, and urban computing.

YONGTING PAN is a PhD candidate in the School of Geography and Information Engineering at China University of Geosciences, Wuhan, China. E-mail: panyt@cug.edu.cn. Her research interests include big spatio-temporal data analytics and mining, and urban computing.

YAO YAO is an Associate Professor in the School of Geography and Information Engineering at China University of Geosciences, Wuhan, China. E-mail: yaoy@cug.edu.cn. His research interests include big spatio-temporal data analytics and mining, urban computing, and social sensing.

WEN ZENG (corresponding author) is a Professor in the School of Geography and Information Engineering at China University of Geosciences, Wuhan, China. E-mail: zengwen@cug.edu.cn. His research interests include analytical and modeling algorithms for urban networks, big spatio-temporal data analytics and mining.