Computers, Environment and Urban Systems xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

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

Integrating landscape metrics and socioeconomic features for urban functional region classification

Hanfa Xing[a,b], Yuan Meng[b,*]

[a] School of Geography, South China Normal University, Guangzhou, Guangdong, China
[b] College of Geography and Environment, Shandong Normal University, Jinan, Shandong, China

ARTICLE INFO

Keywords:
Urban functional region
Classification
Landscape metrics
Crowdsourced data
Socioeconomic features

ABSTRACT

Urban functional region classification plays an important role in urban planning, resource management and sustainable development. While previous studies have solely focused on delineating regions by urban morphology or deriving functional types from human activities, the sufficient depiction and integration of these characteristics are seldom discussed. This paper inherently considers both urban morphology and human activities to classify urban functional regions. Building-based and region-based landscape metrics derived from building-level blocks are delineated to measure urban morphology, while socioeconomic features are extracted from crowdsourced data related to human activities using a topic model and semantic scaling method. These characteristics are then fused by applying random forest to measure different functions. This approach is tested in Futian district, Shenzhen, Guangdong. The result shows an overall classification accuracy of 0.818, approximately 0.15 higher than that utilising only landscape metrics or only socioeconomic features. This result indicates the effectiveness of the delineated characteristics to depict urban landscapes and socioeconomic information and the reliability of integrating these features for urban functional region classification.

1. Introduction

Urban functional regions refer to the areas in which human activities occur, including residential, commercial, and industrial land. They serve as an essential indicator in designing urban environments, managing natural resources, and detecting rapid urbanisation (Anderson, 1976; Treitz, Howarth, & Gong, 1992; Wentz, Stefanov, Gries, & Hope, 2006). Classifying urban functional regions is crucial for uncovering the characteristics of urban cities, which comprise not only urban regions delineated from physical properties but also socioeconomic functions driven by human activities (Pei et al., 2014; Xiuyuan, Zhang, Du, & Wang, 2017).

Significant progress has been made in sensing physical properties to depict urban forms. Conventionally, this relies heavily on remote sensing images, where spectral and textural characteristics are extracted for presenting urban land use information (Lu & Weng, 2006; Weng, 2012). Current analyses of functional region detection focus on three levels of perspectives, namely, pixel, object and scene-based classification (Gong & HOWARTH, 1990; Ma et al., 2017; Xiuyuan, Zhang & Du, 2015). Although the pixel and object-based classification approaches have gained popularity in deriving land cover features, they fail to capture detailed functional information because of the complex urban form (Hu, Yang, Li, & Gong, 2016). On the other hand, scene-based classification shows advantages in identifying urban functional patterns (Xiuyuan, Zhang & Du, 2015). However, it can only decompose and measure the whole picture of the mixed scenes, whereas detailed urban morphology, which is often strongly correlated with the functional type, cannot be captured (Vanderhaegen & Canters, 2017).

In fact, remotely sensed images can to some extent reveal the morphology of urban regions, which is defined as the presence, size, shape and spatial distribution among urban landscapes, including different constructed and open spaces. Quantifying landscapes with effective metrics may help improve the delineation of urban morphology. The relationship between urban landscapes and functions have been discussed in many studies (Baus, Kováč, Pauditšová, Kohutková, & Komorník, 2014; Li, Peng, Yanxu, & Yi'na, 2017; Van de Voorde et al., 2016). Typically, building-level blocks are one of the vital aspects in functional regions (Steiniger, Lange, Burghardt, & Weibel, 2008). Many studies have investigated the potential of buildings with different metrics in urban patterns analysis. Wu et al. (2018) proposed an Extended Minimum Spanning Tree method to describe both spatial proximity and building characteristics to discover local urban patterns. In addition, Chen, Xu, and Devereux (2014) and Liu, Hu, and Li (2017) have designed 2- and 3-dimensional scales of landscape metrics to effectively

* Corresponding author.
E-mail addresses: xinghanfa@sdnu.edu.cn (H. Xing), mengyuan@stu.sdnu.edu.cn (Y. Meng).

https://doi.org/10.1016/j.compenvurbsys.2018.06.005
Received 5 February 2018; Received in revised form 12 April 2018; Accepted 15 June 2018
0198-9715/ © 2018 Elsevier Ltd. All rights reserved.

reflect building structures. The landscape metrics proposed for buildings enable further investigations into functional regions. For example, urban block perimeters vary among different functional types: industrial and public regions usually compromise a low density of large individual buildings, while residential regions often contain continuous blocks with small areas, which leads to a much larger value of the perimeter. In fact, an attempt has been made to distinguish different functional regions with building-based metrics (Vanderhaegen & Canters, 2017), proving the ability to describe urban landscapes and to detect functional regions. Despite the achievement of depicting the ground components in urban regions, the functional types that are highly influenced by the interaction between environmental landscape and human activities have not been further explored (Tu et al., 2017).

The massive amounts of crowdsourced data, such as points of interest (POIs), social media data, trajectories and mobile phone signals potentially provide information about human activities and socioeconomic information, which have been investigated to indicate urban functions in many studies (Yu et al., 2015; Zheng, Capra, Wolfson, & Yang, 2014). Yuan, Zheng, and Xie (2012) employed POIs and human mobility to infer functional regions by identifying the distributions of mobility patterns. Gao, Janowicz, and Couclelis (2017) analysed the cooccurrence patterns of POI types to discover functional regions at a semantic level. Semantic information in crowdsourced data can reflect socioeconomic features such as commercial, industrial and residential information, which are considered significant indicators of urban patterns. Yao, Li et al. (2017) proposed a POI-based framework to extract semantic information using a Word2Vec model for sensing functional types. Lansley and Longley (2016) introduced the latent Dirichlet allocation (LDA) topic model to mine functional distributions using Twitter messages. However, due to the spatial heterogeneity of crowdsourced data (Fonte, Bastin, See, Foody, & Lupia, 2015), regions containing less human activities can hardly be differentiated without identifying urban morphology (e.g., people seldom post activity information or generate points of interest in undeveloped open regions); thus, capturing only socioeconomic features cannot meet the requirement of effectively detecting urban functions.

Therefore, combining both urban morphology and human activities has the potential to effectively reveal urban functions. Previous studies have attempted to fuse remote sensing images and crowdsourced data to monitor land cover/land use applications. Liu et al. (2017) proposed a land use word dictionary by extracting natural features from high spatial resolution images and semantic features from social media data. Hu et al. (2016) generated urban functional regions based on the spatial distribution of POIs and on spectral information from satellite images. Moreover, a recent attempt has been made to measure both landscape patterns and human activities from remote sensing images and cell phone data (Tu et al., 2018). Since they delineated functional zones based on regular grids with 2000-m resolution using only four landscape metrics, the urban morphology, which is highly related to the community or building scale, cannot be captured.

Although these studies have integrated both physical properties and socioeconomic information, which result in significant improvements of the functional region classification, challenges still remain in the following aspects. First, they simply extracted and fused a variety of natural features from remote sensing images and failed to quantitively measure the landscape to describe the detailed urban morphology. Second, as crowdsourced data are usually generated irregularly with spatial heterogeneity, effectively quantifying indicators is required to relate the unorganised information with certain functional types. Third, the integration of urban landscapes and socioeconomic information to better delineate the functional regions remains undiscussed.

Faced with the above issues, this paper proposes a functional region classification approach by integrating landscape metrics and socioeconomic features, which includes three major steps. In the first step, landscape metrics are adopted to capture urban physical properties. Particularly, metrics with both building level and functional region level are considered in order to obtain comprehensive descriptions of urban landscapes at a fine scale. In the second step, a variety of semantic information from crowdsourced data is automatically calculated for characterising socioeconomic patterns. Faced with the spatial heterogeneity of crowdsourced data, semantic scaling is proposed to normalise semantic information and extract socioeconomic features. Finally, by utilising both landscape metrics and socioeconomic features, random forest is applied to gain better insight into the diversity of functional types.

The major advantage of the methodology is that it inherently integrates landscape and socioeconomic characteristics. In particular, by calculating landscape metrics from building-level blocks and extracting function-related topics via crowdsourced data, the presented approach not only delineates the underlying urban morphological and social properties of urban forms but also provides an initial insight into identifying different functional types. To validate the presented approach, an analysis of variance is proposed to evaluate the contribution of each landscape metric and socioeconomic feature to functional region classification. Furthermore, comparative experiments are proposed utilising solely landscape metrics and socioeconomic features to prove our hypothesis.

The contributions of our work involve three main aspects. First, urban regions are delineated by extracting physical properties from a landscape perspective. We quantify the spatial morphology of urban regions by constructing building-based and region-based metrics. Second, functional types are determined by socioeconomic features derived from crowdsourced data. By calculating different topics using the topic model and a semantic scaling method, socioeconomic information can be viewed from the perspective of human activities and shows potential to be applied as one of the effective data sources for urban function arrangement. Third, we inherently consider landscape metrics and socioeconomic features to automatically identify different urban functions. As most studies mainly focused solely on urban landscape or social information, the combination of both can give a better understanding of functional region classification.

The remainder of this paper is organised as follows. In Section 2, the methodology for functional region classification is presented considering both landscape metrics and socioeconomic features. Section 3 introduces the study area and experimental data. The classification results are presented in Section 4. The utilisation of experimental data, the selection of both landscape metrics and socioeconomic features, and the identification of mixed function regions are discussed in Section 5. Section 6 concludes the presented work and discusses future studies.

2. Methodology

2.1. Overview

The framework of urban functional region classification is displayed in Fig. 1, including three major steps. In the first step, we utilise building-level blocks to delineate the urban morphology (Section 2.2). In particular, the basic measurement of building-level blocks consists of four indices including the building area, edge, height and structure (Section 2.2.1). These indices are combined for calculating building-based metrics and region-based metrics (Section 2.2.2). The constructed landscape metrics are used to delineate physical properties from the landscape perspective, capture urban morphology information, and further provide essential natural characteristics for functional region classification.

In the second step, crowdsourced data are used to extract socioeconomic information (we choose POIs as crowdsourced data in the following experiment) (Section 2.3). First, using an LDA (latent Dirichlet allocation) topic model, we mathematically extract topics containing words with probabilities from crowdsourced data. Specifically, the number of topics is determined by a perplexity algorithm, with a lower perplexity value representing a more appropriate topic

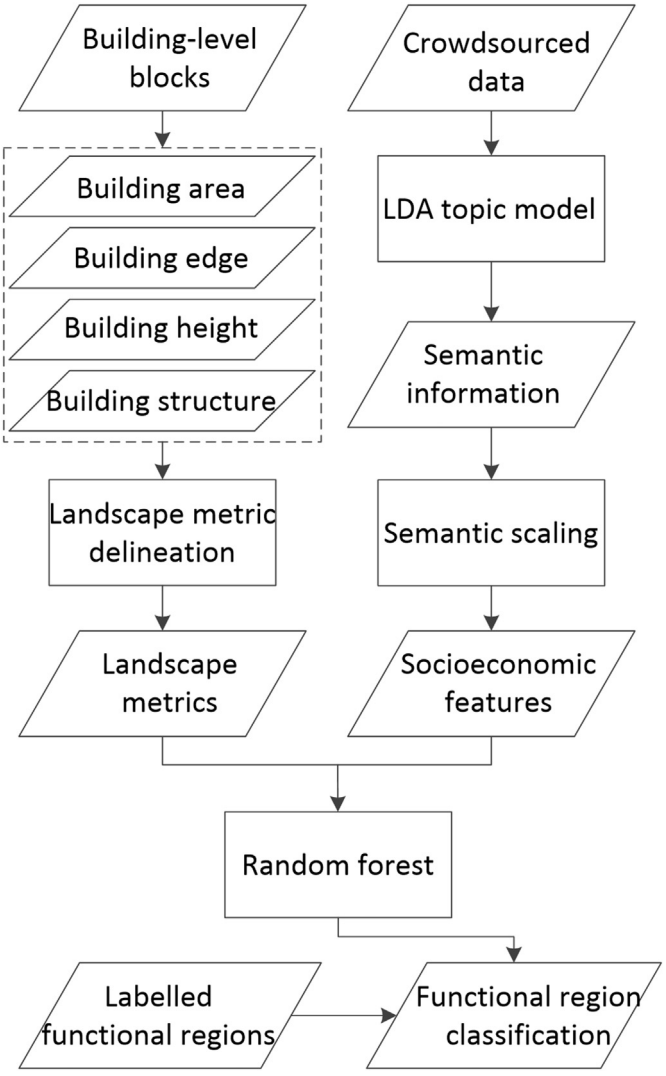


Fig. 1. Framework of the urban functional region classification process.

number (Section 2.3.1). Then, function-related topics are selected by inferring words with high probabilities. Finally, we propose a semantic scaling method to group topics belonging to the same function, in which the probabilities are scaled to $[0, 1]$ (Section 2.3.2). The grouped topics in each functional type are considered as socioeconomic features and will be further used for identifying urban functions from the perspective of human activities.

In the third step, we automatically classify functional regions based on the characteristics of urban morphology and human activities (Section 2.4). More specifically, a random forest model is utilised to inherently integrate both landscape metrics and socioeconomic features. In the model, two parameters are required to be pre-defined, including the number of features used at each node to generate a tree and the number of trees to be grown. As a result, models with different combinations of parameters are evaluated to improve the classification accuracy.

2.2. Delineation of the landscape metrics

To delineate urban landscape characteristics from building-level blocks, effective measurements are proposed in Section 2.2.1, along with utilising these measurements to calculate landscape metrics in Section 2.2.2.

2.2.1. Utilisation of building-level blocks

Generally, basic measurements are needed for capturing landscape morphology. Therefore, utilising building-level blocks, four indices are utilised for measurement in each functional region, including building area, building edge, building height and building structure. Fig. 2 illustrates some typical samples of the four indices. It should be noted that while building height and structure are available in the experimental data in Section 3, building area and edge are spatially calculated from this data source. Moreover, there is a variety of available data containing urban landscape information, which will be discussed in Section 5.1.1.

2.2.1.1. Building area. Building area is used to explain the extent to which functional regions are being occupied with building-level blocks. Fig. 2(a) shows the measurement of building area, in which the regions in red define the target area. Generally, many of the residential and commercial lands contain larger building areas, while building areas on public lands are relatively lower. By including buildings in the same functional region, landscape metrics such as total building area, mean building area, standard deviation and maximum building area can be considered and delineated.

2.2.1.2. Building edge. Building edge is defined as the length of outlines in building-level blocks, which is introduced to demonstrate their shape complexities. As shown in Fig. 2(b), the red lines highlight the length to be measured. Given the same building area, a longer building edge length usually has a more complex shape. General metrics including

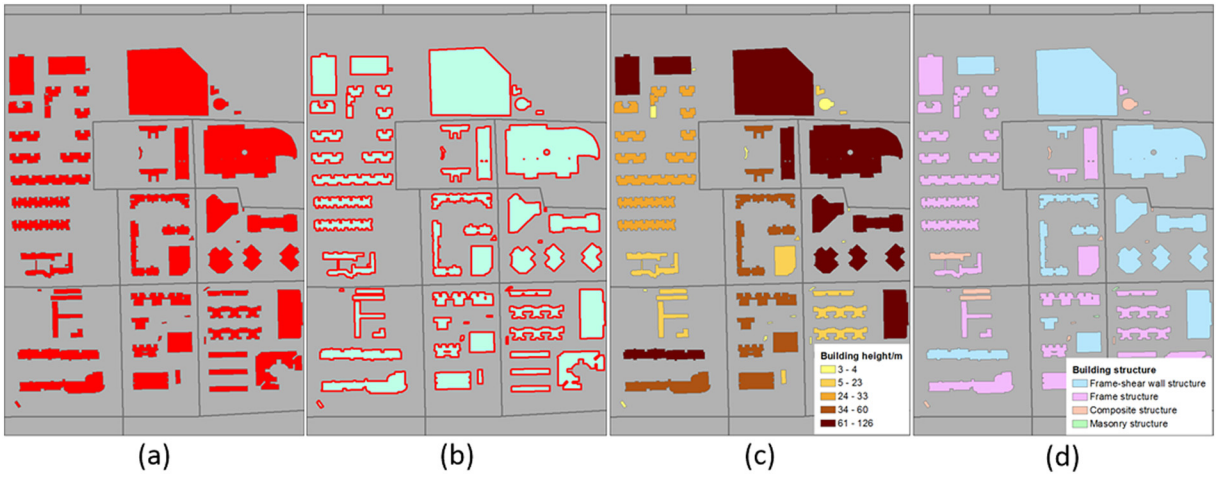


Fig. 2. Typical samples for building-level block measurement. (a) Building area. (b) Building edge. (c) Building height. (d) Building structure.

total building edge, mean building edge and standard deviation can be proposed. Furthermore, the landscape shape index (LSI) is also considered based on the building length for functional region classification.

2.2.1.3. Building height. Although building compositions are well-measured in the horizontal direction based on building area and building edge, landscape metrics require further investigation in the vertical direction. In general, buildings are usually designed with variable height levels to meet the needs of different functional regions. For example, buildings used for residence in a central city usually contain more floors and are thus much taller, while industry-based buildings are designed to be shorter. Accordingly, building height is considered another measurement for landscape metric delineation. A typical sample of building distribution with variable height is shown in Fig. 2(c). Mean building height and standard deviation are considered effective landscape metrics, as well as building structure index proposed by Chen et al. (2014).

2.2.1.4. Building structure. Due to the existence of variable building structures in different functional regions, six kinds of building structures are introduced, including frame-shear wall structure, frame structure, composite structure, masonry structure, tube structure, and steel structure. Fig. 2(d) displays the distribution of buildings with a variety of structural materials, where one can see the dissimilarity of the dominant building structure in each of the functional regions.

2.2.2. Delineating landscape metrics from building-level blocks

Based on the measurement of building-level blocks, 18 metrics are considered for describing the urban form in a landscape perspective. Table 1 provides an overview of the proposed metrics, which are divided into building-based metrics and region-based metrics.

2.2.2.1. Building-based metrics. By measuring the building area, five landscape metrics including total building area, percentage of building area, mean building area, maximum building area, and building area standard deviation can be calculated in each of the functional regions (Table 1). Total building area is defined as the sum of the building area in the same function region. It is the same case for mean building area, maximum building area and building area standard deviation, where the mean value, maximum value and standard deviation are calculated. For the percentage of building area, the ratio of the total building area to the functional region area is calculated to analyse building distribution.

In addition, the number of buildings in each functional region is counted to delineate the two landscape metrics of total building number and building density. In particular, building density is defined as the ratio of total building number to functional region area.

Regarding building edge, three landscape metrics including total building edge, mean building edge, and building edge standard deviation are employed for characterising the urban landscape. Total building edge is calculated by summing the length of the buildings in the same functional region, while the mean value and standard deviation are obtained for the last two metrics.

In terms of measuring building height, four landscape metrics including mean building height, building height standard deviation, mean building structure index and building structure index standard deviation are proposed. The first two metrics are obtained by calculating the mean value and standard deviation of the total building height in each of the functional regions. It is the same with the last two metrics, in which building structure index is calculated as follows (Chen et al., 2014):

$$\text{Building structure index} = \frac{\text{Building area}}{\text{Building height}} \tag{1}$$

As for measuring building structure, six kinds of structural materials discussed in Section 2.2.1 are considered as another landscape metric. In particular, the structural material that is most used in one functional region is considered the dominant type to indicate this metric.

2.2.2.2. Region-based metrics. Three landscape metrics, mean patch fractal dimension, degree of landscape division and landscape shape index, are considered to explore the spatial distribution and relationship between functional regions and the building locations. Mean patch fractal dimension provides information on the irregularity of building distribution in functional regions, in which fractal dimension is calculated as follows (McGarigal & Marks, 1995):

$$\text{Fractal dimension} = \frac{2^* \ln P}{\ln A} \tag{2}$$

since fractal dimension is utilised to describe building-level landscape, variables P and A are defined as the edge and area of a typical building, respectively. In addition, an average value is computed for all buildings in each of the functional regions as the mean patch fractal dimension.

Degree of landscape division, proposed by Jaeger (2000), is used to describe the probability that two randomly chosen buildings in the same functional region are located in the same dissected region. It is defined as:

$$D = 1 - \sum_{i=1}^{n} \left(\frac{A_i}{A_f} \right)^2 \tag{3}$$

where D refers to the degree of landscape division. n indicates the total number of buildings in each functional region. A_i and A_f refer to the area of the ith building and the corresponding functional region, respectively.

Landscape shape index (Patton, 1975; Rutledge, 2003) is employed to characterise the spatial heterogeneity of buildings in functional regions, which can be derived as:

$$LSI = \frac{\sum_{i=1}^{n} P_i}{2^* \sqrt{\pi \sum_{i=1}^{n} A_i}} \tag{4}$$

where LSI presents the value of the landscape shape index. In general, more complex building distributions will have a larger LSI value. P_i and A_i refer to the edge and area of the ith building in one functional region, separately.

Table 1
Landscape metrics delineated from building-level blocks.

Building-based metrics	Total building area
	Percentage of building area
	Mean building area
	Maximum building area
	Building area standard deviation
	Total building number
	Building density
	Total building edge
	Mean building edge
	Building edge standard deviation
	Mean building height
	Building height standard deviation
	Mean building structure index
	Building structure index standard deviation
	Structural material
Region-based metrics	Mean patch fractal dimension
	Degree of landscape division
	Landscape shape index

2.3. Extraction of the socioeconomic features

Although functional regions have been characterised with landscape metrics, the socioeconomic information still needs to be explored. Crowdsourced data usually contain massive amounts of text messages for capturing socioeconomic features. However, effective methods are

necessary for text modelling. Furthermore, the proportions of these messages corresponding to each kind of functional region are far from balanced, as commercial-related information usually occupies a large proportion of the data (Lloyd & Cheshire, 2017). Faced with these issues, semantic information is calculated from crowdsourced data using the LDA topic model in Section 2.3.1. To eliminate the influence of the uneven proportion of semantic information, a semantic scaling method is proposed in Section 2.3.2.

2.3.1. Calculating semantic information using the LDA topic model

Crowdsourced data with text messages are challenging to model because no standard principles exist to restrict unorganised texts. Some studies tried to utilise a list of words or to calculate word frequency for text modelling (Michelson & Macskassy, 2010; Yao et al., 2017) and ignored mining the potential semantic information. The LDA topic model is designed to generate the main topics and investigate latent semantic information inferred from the text (Blei, Ng, & Jordan, 2003). It has been extensively applied in the field of natural language processing and has gained satisfactory results in mining urban function information (Lansley & Longley, 2016; Ríos & Muñoz, 2017).

Given the smoothing parameters α and β, LDA assumes that k topics are associated with N words in the document corpus $\boldsymbol{D}$. Each document d exhibits the patterns of topics with different proportions, and each topic contains a set of words with a variety of distributions.

$$p(\theta, \boldsymbol{z}, \boldsymbol{w} \mid \alpha, \beta) = p(\theta \mid \alpha) \prod_{n=1}^{N} p(z_n \mid \theta) p(w_n \mid z_n, \beta) \tag{5}$$

where $\boldsymbol{w}$ and $\boldsymbol{z}$ represent a set of words and the inferred topics, respectively, in each document d. $\boldsymbol{\theta}$ refers to the proportions of topics in document d.

In this context, each functional region is regarded as a document. Every word in the crowdsourced data is given a probability of falling within the generated topics, and each topic contains a certain proportion of the functional regions. Based on this topic-generated model, it is necessary to determine the appropriate topic number k. We follow the perplexity algorithm in (Blei et al., 2003) for the topic number evaluation, with a lower perplexity value representing better performance. Nevertheless, the selected topic number is not always strongly related to human judgement (Chang, Gerrish, Wang, Boyd-Graber, & Blei, 2009; Contreras-Piña & Ríos, 2016), and thus, effective measures are required to modify topics corresponding to each of the functional types.

2.3.2. Extracting socioeconomic features with semantic scaling

To extract socioeconomic features based on the semantic information in Section 2.3.1, two challenges should be settled first. One is eliminating the negative effect of perplexity when the number of topics does not fully correspond to the functional types. The other is generalising the uneven proportions of topics in each functional type. Consequently, a semantic scaling method is proposed to solve these problems and thus obtain socioeconomic features for functional region classification.

Faced with the first challenge, each topic is evaluated to ensure its representativeness. The semantic information in the topics can be constructed from those words with highest probabilities. Thus, only topics related to each of the functional types are selected, while others are removed from this study.

Regarding the second problem, the probabilities of the remaining topics in each document are extracted for semantic scaling. Topics corresponding to the same type in one functional region are grouped, with their probabilities being summed and averaged as follows:

$$\overline{P}_i(f) = \frac{\sum_{j=1}^{n} P_i^j(f)}{n} \tag{6}$$

where $\overline{P}_i(f)$ presents the averaged probability of topics related to functional type f in the ith functional region. $P_i^j(f)$ and n indicate the probability of the jth topic and the total number of topics corresponding to functional type f in the ith functional region, respectively. As a result, each functional type is connected to a specific probability, which is derived from the calculated topics. Then, these averaged probabilities in the same functional region are scaled to [0,1], in which the scaled probabilities are considered as socioeconomic features for functional region classification.

$$\overline{\boldsymbol{P}}_i' = \frac{\overline{P}_i(f) - min(\overline{\boldsymbol{P}}_i)}{\max(\overline{\boldsymbol{P}}_i) - min(\overline{\boldsymbol{P}}_i)} \tag{7}$$

where $\overline{\boldsymbol{P}}_i'$ indicates the vector of scaled probabilities of all functional types in the ith functional region. The number of values in vector $\overline{\boldsymbol{P}}_i'$ equals the quantity of functional types. $max(\overline{\boldsymbol{P}}_i)$ and $min(\overline{\boldsymbol{P}}_i)$ refer to the maximum and minimum value of the vector of averaged probabilities in the ith functional region, respectively.

2.4. Functional region classification with landscape metrics and socioeconomic features

To mine the complexity of functional characteristics, machine learning-based classification approaches are considered for grouping functional regions with different landscape metrics and socioeconomic features. Here, we employ a random forest algorithm (Ho, 1995), which has been proven to be a sufficient approach for classification in land cover/land use applications (Liaw & Wiener, 2002; Pal, 2005; Rodriguez-Galiano, Ghimire, Rogan, Chica-Olmo, & Rigol-Sanchez, 2012).

Random forest is an ensembled learning approach for classification, which generates multiple non-associated decision trees using bootstrap sampling on training data and predicts testing data based on a majority vote of all decision trees. This classifier is able to handle high dimensional data based on the large number of constructed trees, which meets the requirement of distinguishing functional types with multiple features. As a result, given the calculated landscape metrics and socioeconomic features as variables, parts of the regions are labelled with specific functional types for classification.

Specifically, two user-defined parameters are required, including the number of features used at each node to generate a tree and the number of trees to be grown. For the number of features, only the selected features are examined for the best split at each node to generate new trees and thus reduce both the complexity of the algorithm computation and the correlation between the two trees. The pre-trained random forest classifier consists of N trees, which is the other user-defined parameter, for predicting new samples. To improve the performance of the functional region classification, selection strategies of these parameters should be proposed. The accuracy assessment of different parameter combinations is displayed in Section 4.3.

3. Study area and data

The study area for this research is located in the Futian district (Fig. 3(a)) of Shenzhen, Guangdong province. With a total area of 78.66 km^2 and a permanent population of 1,501,700, it has been considered as the centre of Shenzhen, with advanced political, economic and cultural functions. To analyse the types and distribution of urban functional regions in this area, 411 functional regions (Fig. 3(b)) are divided based not only on the provided high-resolution image in Fig. 3(a) but also on the main roads provided by OpenStreetMap. Based on these functional regions, four functional types including commercial land, industrial land, residential land and service/public land are involved in the classification.

To delineate the landscape metrics, building-level blocks are utilised, which are provided by the Urban Planning, Land & Resources Commission of Shenzhen Municipality (Municipality Oceanic Administration of Shenzhen) in http://www.szgeoinfo.com:5001/msmap/flex/bcode.html. We extracted 16,799 building-level blocks

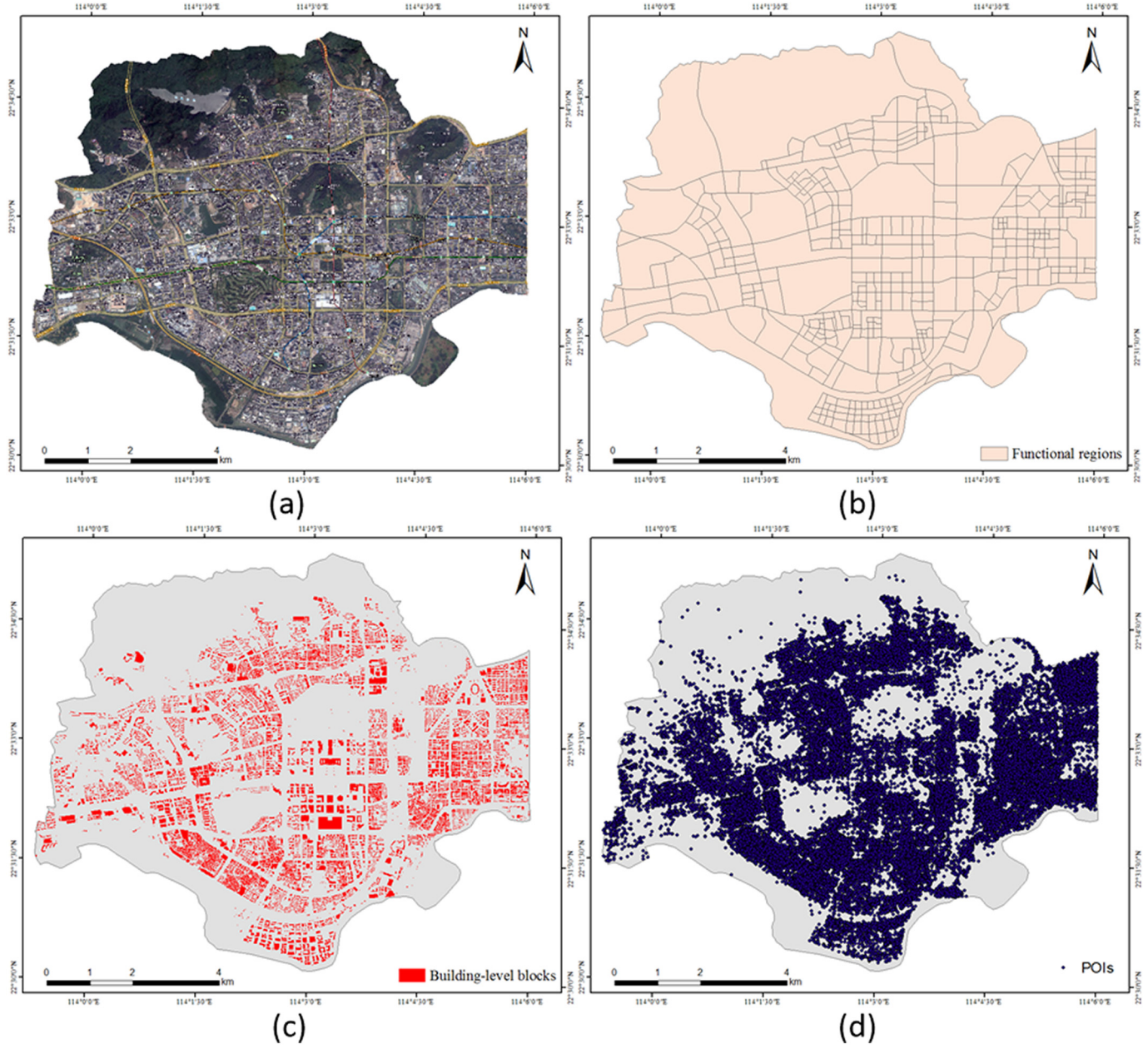


Fig. 3. Study area and data. (a) High-resolution image in the study area. (b) Functional regions derived from road networks. (c) Building-level blocks obtained from the Urban Planning, Land Resources Commission of Shenzhen Municipality (Municipality Oceanic Administration of Shenzhen). (d) Spatial distribution of POIs provided by AutoNavi.

with spatial distribution (Fig. 3(c)), height and structural material information for the following study.

POIs from AutoNavi are employed to extract socioeconomic features from crowdsourced data. We obtained a total of 139,921 POIs via APIs from http://lbs.amap.com/api/, in which three category levels are included. The third category level in each of the POIs is utilised, as it contains more detailed textual information for indicating socioeconomic features. The spatial distribution of POIs is shown in Fig. 3(d). One should note that as these POIs are collected officially by AutoNavi, and the quality can be ensured to some extent.

4. Results

4.1. Landscape metrics and socioeconomic features for classification

Based on the method proposed in Sections 2.2 and 2.3, 18 landscape metrics and 4 socioeconomic features are proposed for functional region classification. Section 4.1.1 shows the delineated landscape metrics and Section 4.1.2 displays the selection of socioeconomic features in the functional regions.

4.1.1. Delineating the landscape metrics

The delineated landscape metrics are visualised in Fig. 4. One can see that percentage of building area shows relatively symmetrical median values, indicating a medium level of building occupation in most of the functional regions. For the remaining metrics, the values for the degree of landscape division present a long head, while the others, except for structural material, show a long tail of values. The long head for the degree of landscape division reveals that only a few functional regions contain larger buildings that can occupy most of the region area. Similarly, the long tails in the other metrics also reveal the irregularity of building occupation and distribution at different

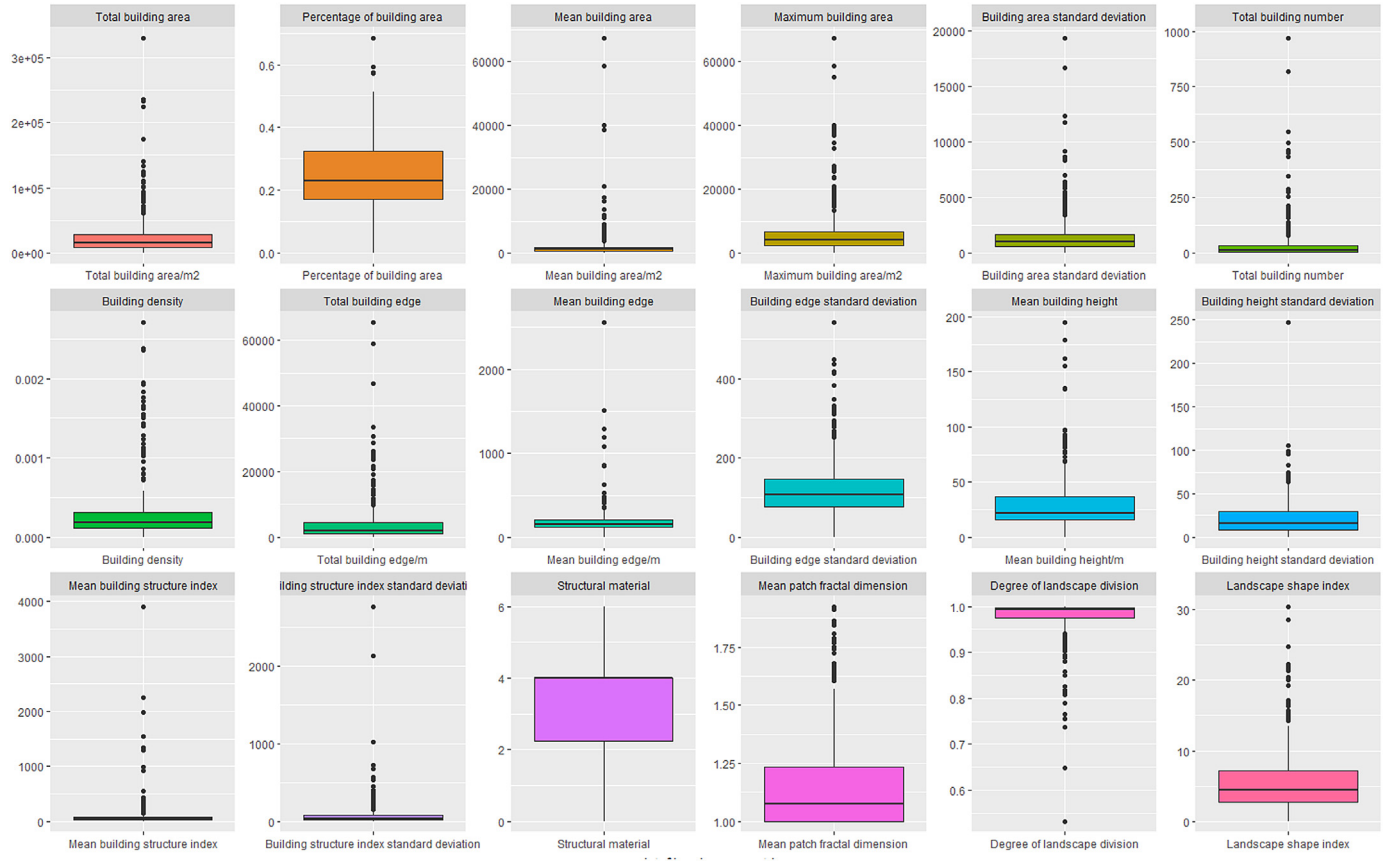


Fig. 4. Boxplot visualisation of the landscape metrics.

perspectives. In terms of structural material, six kinds of structures are coded into [1,6], where most of the buildings are built with the second, third and fourth material, corresponding to composite structure, frame-shear wall structure and frame structure, respectively.

The above 18 landscape metrics show the variation in building distribution, the diversity of building forms, and the arrangement of the function regions and the containing buildings. Therefore, these delineated metrics can be regarded as effective information to depict the urban morphology and to further classify different functions.

4.1.2. Extracting the socioeconomic features

The third-level categories in the POIs are utilised, with categories belonging to the same functional region being considered as one document in the LDA. To determine the number of topics, the topic numbers from 1 to 300 are assessed by calculating the perplexity values. As shown in Fig. 5, the perplexity value decreases dramatically when the topic number is less than 44. It is followed by a continuous increase when the topic numbers become larger. Accordingly, 44 topics are generated from the categories in the POIs.

Based on the POI categories with high probabilities in each topic, 19 topics related to functional types are selected, with 6 topics corresponding to commercial land, 5 and 4 topics belonging to industrial land and residential land, and the remaining 4 topics indicating service/public land. The probabilities of the topics in the functional regions are extracted and scaled for each functional type and are considered as socioeconomic features for the classification, including commercial indicator, industrial indicator, residential indicator and service/public indicator. In this regard, the presented methodology can quantify socioeconomic information related to human activities and to some extent reveals the potential to fuse both landscape metrics and socioeconomic features in functional region classification.

4.2. Evaluation of landscape metrics and socioeconomic features

The evaluation of landscape metrics and socioeconomic features is required to assess the importance and contribution of each feature for functional region classification. Through an analysis of variance (ANOVA) in 411 regions with labelled functional types, the relationship between landscape metrics, socioeconomic features and different functional types is analysed, which is shown in Table 2.

In building-based metrics, four metrics including total building area, mean building edge, building edge standard deviation and building structure index standard deviation, are well evaluated with corresponding *Sig.* values of less than 0.01. The remaining region-based metrics except for maximum building area, which all have *Sig.* values equal to 0.000 and relatively high values of *F*, also qualify as effective variables for functional region classification. Only maximum building area, with an *F* value of 2.505 and a *Sig.* value of 0.059, is less sufficient compared with the other metrics.

As for the region-based metrics, all features in the region-based metrics highly contribute to functional region classification, with *Sig.* = 0.000 and *F* equal to 11.669, 10.624 and 30.685 for mean patch fractal dimension, degree of landscape division and landscape shape index, respectively.

Regarding the landscape metrics, four socioeconomic features are significantly correlated to the functional types (*Sig.* = 0.000), with *F* values equal to 36.560, 109.823, 155.515 and 48.057.

The above analyses evaluate the efficiency of our presented landscape metrics and socioeconomic features, determine the contribution of each feature to delineate the functional types, and further reveal the possibility and usability of these features in inferring functional regions.

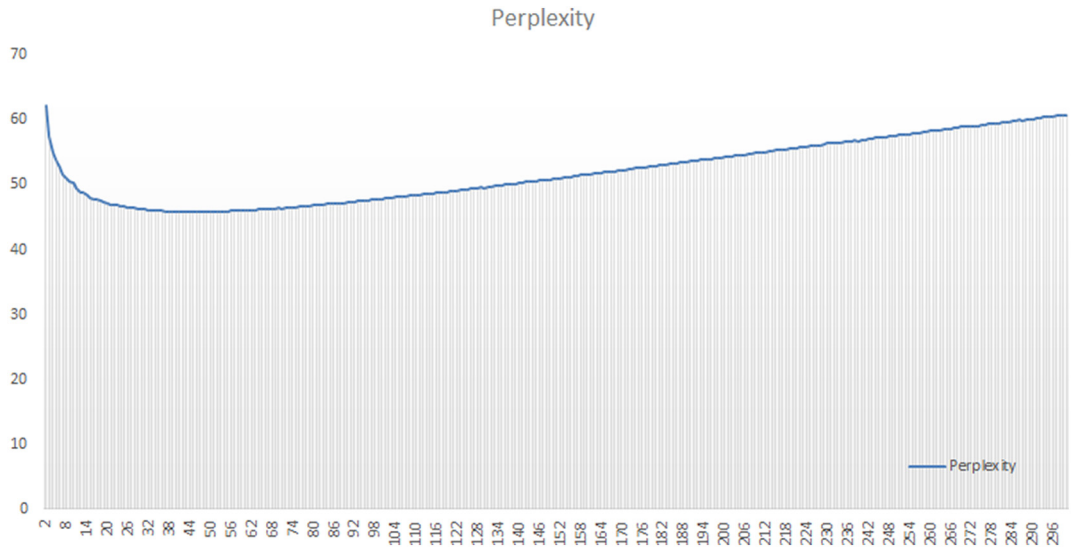


Fig. 5. Perplexity values for the different numbers of topics.

Table 2
Result of ANOVA with landscape metrics and socioeconomic features.

Features			*F*	Sig.
Landscape metrics	Building-based metrics	Total building area	3.976	0.008
		Percentage of building area	28.634	0.000
		Mean building area	12.886	0.000
		Maximum building area	2.505	0.059
		Building area standard deviation	6.750	0.000
		Total building number	17.296	0.000
		Building density	22.323	0.000
		Total building edge	13.542	0.000
		Mean building edge	4.931	0.002
		Building edge standard deviation	5.365	0.001
		Mean building height	22.505	0.000
		Building height standard deviation	12.756	0.000
		Mean building structure index	23.357	0.000
		Building structure index standard deviation	4.257	0.006
		Structural material	14.378	0.000
	Region-based metrics	Mean patch fractal dimension	11.669	0.000
		Degree of landscape division	10.624	0.000
		Landscape shape index	30.685	0.000
Socioeconomic features		Commercial indicator	36.560	0.000
		Industrial indicator	109.823	0.000
		Residential indicator	155.515	0.000
		Service/public indicator	48.057	0.000

4.3. Functional region classification and validation

After both landscape metrics and socioeconomic features are well evaluated, random forest is utilised with those features for functional region classification. As the number of trees and features are required for building the random forest model, different parameter combinations are proposed with the number of trees as (500, 1000, 2000) and the number of features as [2, 5]. The actual functional types in each of the regions are derived and reclassified from http://www.szgeoinfo.com:5001/msmap/flex/landuse.html provided by the Urban Planning, Land & Resources Commission of Shenzhen Municipality (Municipality Oceanic Administration of Shenzhen) (Fig. 6(a)). With the adoption of functional types, the corresponding accuracies are shown in Table 3, where the accuracies vary from 0.782 to 0.818. We take the accuracy of 0.818 as the best performance and choose the corresponding number of trees and features as 500 and 4, separately.

Utilising 411 functional regions in Futian, 300 and 111 of them are divided into training data and testing data for classification, respectively. An accuracy assessment is then applied based on the comparison of the classified functional types and the corresponding ground truth. Table 4 shows the overall accuracies and kappa coefficients with different combinations of landscape metrics and socioeconomic features. When applying only landscape metrics or only socioeconomic features, an overall accuracy of 0.664 is obtained, with kappa coefficients of 0.458 and 0.506, respectively. However, by combining these features, the overall accuracy increases dramatically to 0.818, and the kappa coefficient increases to 0.734. This result indicates the advantages of integrating landscape metrics and socioeconomic features for functional region classification.

The classification results based on landscape metrics, socioeconomic features and all the features are displayed in Fig. 6(b), (c) and (d). Differences can be seen in three highlighted typical areas. In area 1, service/public land is misclassified as residential land using socioeconomic features, while it can be correctly identified with landscape metrics and all features. Area 2, which is occupied with commercial land (Fig. 6(a)), is classified correctly with both socioeconomic features and all features. However, landscape metrics fail to capture the commercial characteristics in several regions and classify them as industrial land. As area 3 has failed to be classified with landscape metrics or socioeconomic features, combining all features correctly classifies the functional types. The above classification results indicate that it is reasonable to combine these features for improving functional region classification accuracy.

Table 5 shows the classification accuracy in each functional type. Regarding the commercial land and industrial land classification, both the producer's and user's accuracies are generally higher with socioeconomic features than those with landscape metrics. On the other hand, landscape metrics show significant advantages in classifying residential land and service/public land in terms of producer's and user's accuracies. The comparison is consistent with the classification result in Fig. 6. As a result, functional regions are classified with much higher accuracies when considering both landscape metrics and socioeconomic features. However, one issue is that service/public land shows a lower

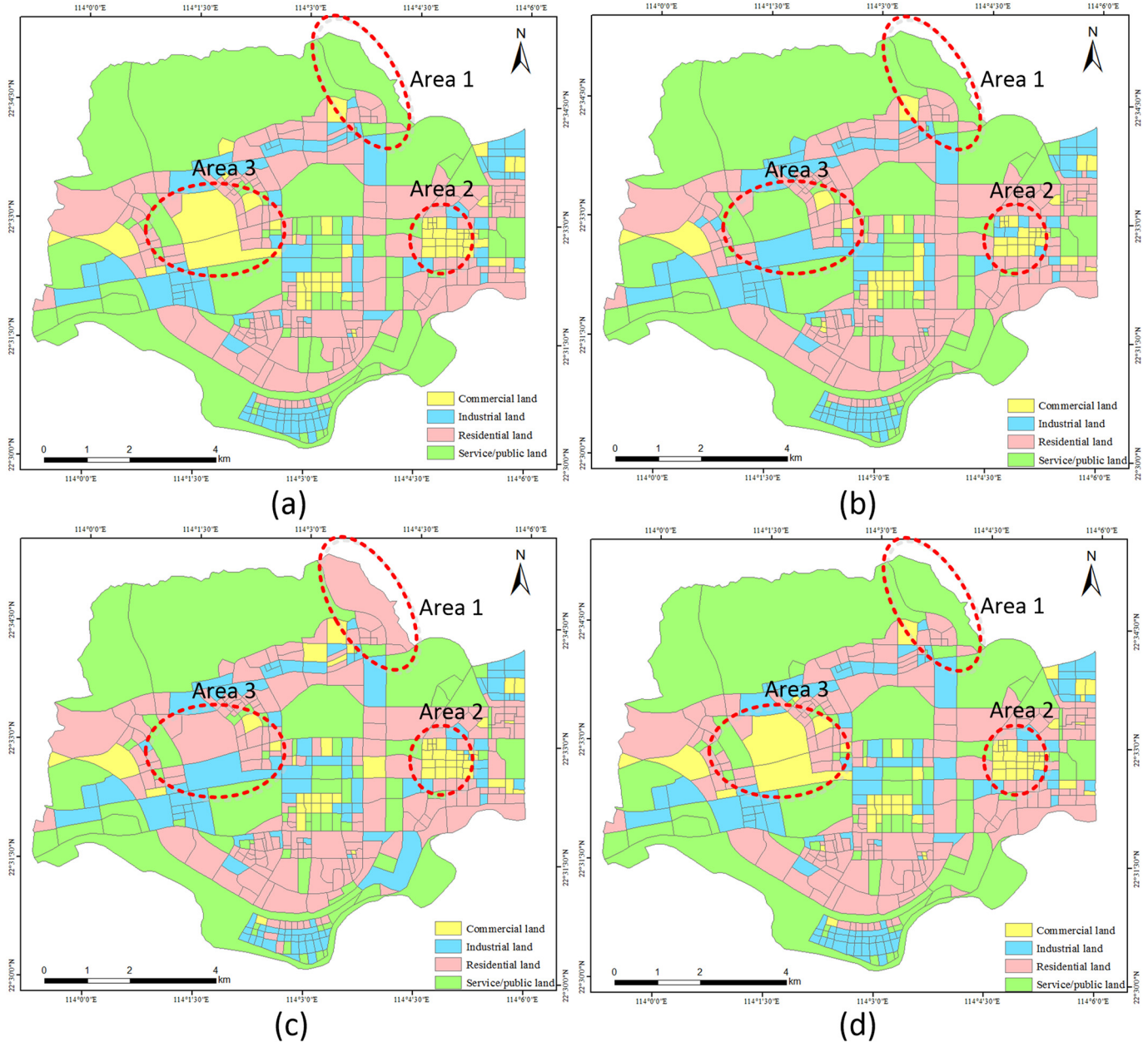


Fig. 6. Functional region classification. (a) Classification results as ground truth. (b) Classification results with landscape metrics. (c) Classification results with socioeconomic features. (d) Classification results with landscape metrics and socioeconomic features.

Table 3
Classification accuracies of random forest with different parameter combination.

Number of trees	Number of features	Accuracy
500	2	0.782
1000	2	0.782
2000	2	0.8
500	3	0.8
1000	3	0.8
2000	3	0.809
500	4	0.818
1000	4	0.791
2000	4	0.809
500	5	0.791
1000	5	0.791
2000	5	0.791

Table 4
Overall accuracy and kappa coefficient of the functional region classification.

Landscape metrics	Socioeconomic features	Overall accuracy	Kappa
√		0.664	0.458
	√	0.664	0.506
√	√	0.818	0.734

user's accuracy with all features compared with the accuracy with landscape metrics. This is caused by the insufficient information provided by socioeconomic features. We will discuss this problem in the next section.

Table 5
Classification accuracy of each functional region type (PA for producer's accuracy; UA for user's accuracy).

		Landscape metrics	Socioeconomic features	Landscape metrics and Socioeconomic features
Commercial land	PA	0.435	0.522	0.696
	UA	0.514	0.632	0.8
Industrial land	PA	0.45	0.75	0.85
	UA	0.391	0.577	0.708
Residential land	PA	0.885	0.788	0.904
	UA	0.73	0.788	0.922
Service/public land	PA	0.533	0.333	0.667
	UA	0.8	0.385	0.667

5. Discussion

5.1. Utilisation of landscape data and crowdsourced data for functional regions classification

The approach is proposed based on the utilisation of urban landscape data (i.e., building-level blocks) and crowdsourced data (i.e., POIs) to capture both physical and socioeconomic properties. Despite the satisfactory result in the functional region classification, the challenges of utilising these data remain undiscussed and should be considered.

5.1.1. Landscape data

Among the existing urban landscape data, building-level blocks are selected to capture and analyse the natural urban form. However, challenges remain in the proposed data. The first is the availability of building-level blocks. As data used in this study are officially published by the Futian government with spatial and structural information, it is convenient to obtain the sources and measure the required indices. However, the problem is that building-level blocks are not available in many of the cities. Second, this kind of data is restricted to delineating building-related landscapes and fails to capture other types of landscapes. Basically, the physical information in functional regions consists of buildings, roads, green land and other built-up areas. Thus, using solely building information may lead to biases and errors in presenting some of the functional types.

Faced with the first issue, available sources are necessary to measure the required indices. In fact, spatial distributions of buildings are provided as open data by Internet maps such as AutoNavi and Baidu maps and thus have the potential to delineate landscape information including building area and building edge. Building height is also displayed in these open data. Moreover, high-resolution images serve as another available source to recognise urban buildings (Du, Zhang, & Zhang, 2015; Xueliang, Zhang, Xiao, Feng, & Yuan, 2017), with textural and spectral characteristics providing the spatial and structural building information (Blaschke, 2010). In addition, building height information can be calculated from airborne Lidar data.

In terms of the second problem, future studies will investigate the usability of remote sensing techniques to depict urban forms (Alshehhi & Marpu, 2017; Huang, Wen, Li, & Qin, 2017). Existing studies have proposed effective approaches to extract urban land use, road networks and green land (Alshehhi, Marpu, Woon, & Dalla Mura, 2017; Li, Stein, Bijker, & Zhan, 2016; Yuqin et al., 2016), which may be considered to help understand urban landscapes comprehensively.

5.1.2. Crowdsourced data

Socioeconomic features generated from crowdsourced data are quantified to depict the social economy characteristics of functional regions. In fact, crowdsourced data provide an effective data source to delineate urban functions, as they are generated by citizens from different urban facilities and highly related to human activities. Here, two issues will be discussed in utilising crowdsourced data, including data accessibility and socioeconomic features extraction.

The availability of crowdsourced data has been discussed in many studies (Liu et al., 2015), which concluded the large quantity and open accessibility. Generally, these data are freely provided by many Internet companies and can be automatically collected in large quantities via APIs. As a result, this process of data collection reduces both time and cost in functional region classification. Among the crowdsourced data, POIs are selected in this paper to capture detailed socioeconomic features. The result reveals that POIs show advantages in identifying commercial and residential land, whereas no significant improvements are displayed in service/public land classification (Table 5). This may be caused by the heterogeneity of POI distribution (Xing, Meng, Hou, Cao, & Xu, 2017), which leads to the lack of efficient data in service/public land. The limitation of POIs further inspires the employment of multiple crowdsourced data sources, such as social media check-ins and trajectory data. For example, an attempt has been made to classify urban land use with Twitter messages (Frias-Martinez & Frias-Martinez, 2014), which has the potential to integrate this data source for extracting socioeconomic features.

Socioeconomic features extraction fills the gap between crowdsourced data and functional regions. With the contribution of semantic information, functional region types can be characterised from a human activity perspective. Although the presented topic model method can automatically calculate socioeconomic features, it is restricted to the relatively complex strategies including text processing and topic number selection. Faced with this issue, developing a prototype system with the integration of the presented approaches is an effective way to reduce time and cost in terms of feature selection. We suggest that the proposed socioeconomic features, which are quantified from crowdsourced data, are able to delineate functional regions, as well as guide urban construction and other socioeconomic applications.

5.2. Improvements of the landscape metric and socioeconomic feature selection

Although both landscape metrics and socioeconomic features are well evaluated in Section 4.2, the analysis of variance (shown in Table 2) indicates that several features are not representative enough to distinguish functional types. For example, maximum building area in building-based metrics contains a *Sig.* value of no more than 0.1, indicating a low significance level in classifying functional regions. Another example can be found in Table 5, where user's accuracies of 0.8 and 0.385 are obtained in service/public land when using landscape metrics and socioeconomic features, respectively. However, as these features are utilised together, a lower user's accuracy of 0.667 is calculated. These problems call for more effective methods to select and utilise landscape metrics and socioeconomic features for functional region classification.

Faced with the uncertainty of these features, methods such as principal component analysis (PCA), Laplacian eigenmaps and analytic hierarchy process (AHP) with specific feature selection strategies can be considered for classification improvement. These approaches enable feature filtering, especially in landscape metrics selection. Meanwhile, complementing the feature selection, a more effective automatic classification method is necessary, where neural networks have the potential to address the complexity of functional characteristics.

5.3. Identification of mixed functional regions

The presented approach is proposed based on identifying single functional types including commercial, industrial, residential and service/public land. One important issue of the current classification results is that it fails to capture the mixture of urban functions. Generally, mixed functions refer to the combination of different functional types in one region. Measuring mixed functions is essential to promote urban

vibrancy and design for future urban development (Liu et al., 2018; Yue et al., 2017).

In the experiment, mixed functions exist in several regions of the study area. As the current methodology focuses on identifying single functional types, the mixed functions are reclassified according to the dominant functional type. To some extent, this leads to the inaccuracy and uncertainty of the final result. This problem might be solved by modifying the presented functional region classification model. In particular, mixed functions can be labelled by functional region in the model training process so that the classification model can be retrained to distinguish mixed functions from single function types.

The modification of the functional region classification model might meet the requirement to infer mixed functions but measuring the urban function mixture is a much more complex process. To further investigate mixed function patterns, several issues should be considered. The first is the measures of landscape metrics and socioeconomic features. Currently, the variation in these features is evaluated to solely correlate to certain functional types. As mixed functions contain different combinations of functional types, it remains a question of how to quantify this uncertainty and further differentiate the corresponding mixture by using the extracted features. Second, effective approaches are required to measure the degree of function mixture. Various measures have been proposed in existing studies based on either distance or quantity (Song, Merlin, & Rodriguez, 2013), but landscape and socioeconomic information related to urban functions have never been discussed.

6. Conclusions and future work

Urban functional region classification serves as a significant component in uncovering urban characteristics. Many studies have considered identifying functional regions by extracting physical properties from remote sensing images or socioeconomic information from crowdsourced data. The integration of both features provides an alternative perspective to better detect functional regions and improve classification accuracy. However, few studies have focused on sufficiently depicting landscape and socioeconomic information and fusing them in functional regions. In this study, the use of landscape metrics and socioeconomic features has been investigated for functional region classification. The delineation of landscape metrics relies on building-level blocks, which enables four building measurements including building area, building edge, building height and building structure to describe urban forms. To extract socioeconomic features, crowdsourced data are utilised to calculate semantic information from the LDA topic model, providing explicit descriptions of human activities. In the experiment, building-level blocks and POIs are selected for the Futian district in Shenzhen. As a result, a more detailed depiction of urban functions is obtained by calculating 18 landscape metrics and 4 socioeconomic features. Through the analysis of variance, the potential relationship between the obtained characteristics and functional types are recognised. Moreover, the presented approach automatically classifies urban functions by combining both landscape metrics and socioeconomic features. In particular, an overall accuracy of 0.818 is obtained, which is approximately 0.15 higher than that with only landscape metrics or only socioeconomic features. These finding will help better understand urban functions and provide useful information for urban planning and regional decision making.

The proposed approach offers interesting perspectives for capturing physical properties and human activities to investigate urban forms and planning. In future studies, we will improve the current performance of functional region classification in the following aspects. First, although satisfactory results have been obtained based on the existing study area, the relatively small study area with limited functional regions may not be representative enough to evaluate the limitations of the presented approach. Thus, we expect to include a much larger study area that contains more functional regions in future studies. Second, functional types, including commercial, industrial, residential and service/public land, are considered for classification, which is not uncommon in many of the previous studies (Frias-Martinez & Frias-Martinez, 2014; Pei et al., 2014; Ríos & Muñoz, 2017). However, these types ignore the more complicated functions and thus fail to depict detailed urban functional patterns. For example, commercial land contains a variety of functions such as restaurants, shopping malls and theatres. This will be improved in future studies. Third, regarding the scale of functional regions, parcels split by road networks are adopted as basic units, the effectiveness of which has been supported in previous studies (Liu et al., 2017; Xingjian, Liu & Long, 2016; Yuan et al., 2012). Nevertheless, mixed land use may exist at this level of units, which leads to the misclassification of functional types. In fact, as finer spatial units such as buildings and streets have been utilised to better detect urban functions in many studies (Chen et al., 2017; Zhu, Wang, Wu, & Liu, 2017), these have the potential to be utilised to substitute road-network-based regions in future studies.

Apart from single functional types, quantifying mixture regions will be considered to cope with the complicated and heterogeneous functional patterns. Dynamic changes in functional regions might also be investigated by utilising effective spatial-temporal data, which might help organise and manage urban forms. Moreover, the integration of landscape metrics and socioeconomic features reveals their potential to be applied in land cover applications. Instead of solely capturing the natural environment, land cover is also greatly influenced by human activities (Meng, Hou, & Xing, 2017). The presented approach provides opportunities to understand and classify different land cover/land use types comprehensively and effectively.

Conflicts of interest

The authors declare no conflict of interest.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (Grant No. 41501420), and the Science and Technology Program of Guangzhou, China (No: 201803030034 and 201802030008).

References

Alshehhi, R., & Marpu, P. R. (2017). Hierarchical graph-based segmentation for extracting road networks from high-resolution satellite images. *ISPRS Journal of Photogrammetry and Remote Sensing, 126*, 245–260.

Alshehhi, R., Marpu, P. R., Woon, W. L., & Dalla Mura, M. (2017). Simultaneous extraction of roads and buildings in remote sensing imagery with convolutional neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing, 130*, 139–149.

Anderson, J. R. (1976). *A land use and land cover classification system for use with remote sensor data. Vol. 964*. US Government Printing Office.

Baus, P., Kováč, U., Pauditšová, E., Kohutková, I., & Komorník, J. (2014). Identification of interconnections between landscape pattern and urban dynamics—Case study Bratislava, Slovakia. *Ecological Indicators, 42*, 104–111.

Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing, 65*, 2–16.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research, 3*, 993–1022.

Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J. L., & Blei, D. M. (2009). Reading tea leaves: How humans interpret topic models. *Advances in Neural Information Processing Systems,* 288–296.

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

Chen, Z., Xu, B., & Devereux, B. (2014). Urban landscape pattern analysis based on 3D landscape models. *Applied Geography, 55*, 82–91.

Contreras-Piña, C., & Ríos, S. A. (2016). An empirical comparison of latent sematic models for applications in industry. *Neurocomputing, 179*, 176–185.

Du, S., Zhang, F., & Zhang, X. (2015). Semantic classification of urban buildings combining VHR image and GIS data: An improved random forest approach. *ISPRS Journal of Photogrammetry and Remote Sensing, 105*, 107–119.

Fonte, C. C., Bastin, L., See, L., Foody, G., & Lupia, F. (2015). Usability of VGI for validation of land cover maps. *International Journal of Geographical Information Science, 29*, 1269–1291.

Frias-Martinez, V., & Frias-Martinez, E. (2014). Spectral clustering for sensing urban land use using twitter activity. *Engineering Applications of Artificial Intelligence, 35*, 237–245.

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*, 446–467.

Gong, P., & Howarth, P. (1990). The use of structural information for improving land-cover classification accuracies at the rural-urban fringe. *Photogrammetric Engineering and Remote Sensing, 56*, 67–73.

Ho, T. K. (1995). Random decision forests. *Proceedings of the Third International Conference on Document Analysis and Recognition* (pp. 278–282). IEEE.

Hu, T., Yang, J., Li, X., & Gong, P. (2016). Mapping urban land use by using landsat images and open social data. *Remote Sensing, 8*, 151.

Huang, X., Wen, D., Li, J., & Qin, R. (2017). Multi-level monitoring of subtle urban changes for the megacities of China using high-resolution multi-view satellite imagery. *Remote Sensing of Environment, 196*, 56–75.

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

Lansley, G., & Longley, P. A. (2016). The geography of twitter topics in London. *Computers, Environment and Urban Systems, 58*, 85–96.

Li, H., Peng, J., Yanxu, L., & Yi'Na, H. (2017). Urbanization impact on landscape patterns in Beijing City, China: A spatial heterogeneity perspective. *Ecological Indicators, 82*, 50–60.

Li, M., Stein, A., Bijker, W., & Zhan, Q. (2016). Urban land use extraction from very high resolution remote sensing imagery using a Bayesian network. *ISPRS Journal of Photogrammetry and Remote Sensing, 122*, 192–205.

Liaw, A., & Wiener, M. (2002). Classification and regression by random forest. *R News, 2*, 18–22.

Liu, M., Hu, Y.-M., & Li, C.-L. (2017). Landscape metrics for three-dimensional urban building pattern recognition. *Applied Geography, 87*, 66–72.

Liu, X., He, J., Yao, Y., Zhang, J., Liang, H., Wang, H., & Hong, Y. (2017). Classifying urban land use by integrating remote sensing and social media data. *International Journal of Geographical Information Science,* 1–22.

Liu, X., & Long, Y. (2016). Automated identification and characterization of parcels with OpenStreetMap and points of interest. *Environment and Planning B: Planning and Design, 43*, 341–360.

Liu, X., Niu, N., Liu, X., Jin, H., Ou, J., Jiao, L., & Liu, Y. (2018). Characterizing mixed-use buildings based on multi-source big data. *International Journal of Geographical Information Science, 32*, 738–756.

Liu, Y., Liu, X., Gao, S., Gong, L., Kang, C., Zhi, Y., ... Shi, L. (2015). Social sensing: A new approach to understanding our socioeconomic environments. *Annals of the Association of American Geographers, 105*, 512–530.

Liu, Y., Meng, Q., Zhang, J., Zhang, L., Jancso, T., & Vatseva, R. (2016). An effective building neighborhood green index model for measuring urban green space. *International Journal of Digital Earth, 9*, 387–409.

Lloyd, A., & Cheshire, J. (2017). Deriving retail Centre locations and catchments from geo-tagged twitter data. *Computers, Environment and Urban Systems, 61*, 108–118.

Lu, D., & Weng, Q. (2006). Use of impervious surface in urban land-use classification. *Remote Sensing of Environment, 102*, 146–160.

Ma, L., Li, M., Ma, X., Cheng, L., Du, P., & Liu, Y. (2017). A review of supervised object-based land-cover image classification. *ISPRS Journal of Photogrammetry and Remote Sensing, 130*, 277–293.

Mcgarigal, K., & Marks, B. J. (1995). *Spatial pattern analysis program for quantifying landscape structure.* Gen. Tech. Rep. PNW-GTR-351US Department of Agriculture, Forest Service, Pacific Northwest Research Station.

Meng, Y., Hou, D., & Xing, H. (2017). Rapid detection of land cover changes using crowdsourced geographic information: A case study of Beijing, China. *Sustainability, 9*, 1547.

Michelson, M., & Macskassy, S. A. (2010). Discovering users' topics of interest on twitter: A first look. *Proceedings of the fourth workshop on Analytics for noisy unstructured text data* (pp. 73–80). ACM.

Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing, 26*, 217–222.

Patton, D. R. (1975). A diversity index for quantifying habitat" edge". *Wildlife Society Bulletin, 3*, 171–173.

Pei, T., Sobolevsky, S., Ratti, C., Shaw, S.-L., Li, T., & Zhou, C. (2014). A new insight into land use classification based on aggregated mobile phone data. *International Journal of Geographical Information Science, 28*, 1988–2007.

Ríos, S. A., & Muñoz, R. (2017). Land use detection with cell phone data using topic models: Case Santiago, Chile. *Computers, Environment and Urban Systems, 61*, 39–48.

Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing, 67*, 93–104.

Rutledge, D. T. (2003). *Landscape indices as measures of the effects of fragmentation: Can pattern reflect process?* .

Song, Y., Merlin, L., & Rodriguez, D. (2013). Comparing measures of urban land use mix. *Computers, Environment and Urban Systems, 42*, 1–13.

Steiniger, S., Lange, T., Burghardt, D., & Weibel, R. (2008). An approach for the classification of urban building structures based on discriminant analysis techniques. *Transactions in GIS, 12*, 31–59.

Treitz, P. M., Howarth, P. J., & Gong, P. (1992). Application of satellite and GIS technologies for land-cover and land-use mapping at the rural-urban fringe: A case study. *Photogrammetric Engineering and Remote Sensing, 58*, 439–448.

Tu, W., Cao, J., Yue, Y., Shaw, S.-L., Zhou, M., Wang, Z., ... Li, Q. (2017). Coupling mobile phone and social media data: A new approach to understanding urban functions and diurnal patterns. *International Journal of Geographical Information Science, 31*, 2331–2358.

Tu, W., Hu, Z., Li, L., Cao, J., Jiang, J., Li, Q., & Li, Q. (2018). Portraying urban functional zones by coupling remote sensing imagery and human sensing data. *Remote Sensing, 10*, 141.

Van de Voorde, T., van der Kwast, J., Poelmans, L., Canters, F., Binard, M., Cornet, Y., ... Williams, B. (2016). Projecting alternative urban growth patterns: The development and application of a remote sensing assisted calibration framework for the greater Dublin area. *Ecological Indicators, 60*, 1056–1069.

Vanderhaegen, S., & Canters, F. (2017). Mapping urban form and function at city block level using spatial metrics. *Landscape and Urban Planning, 167*, 399–409.

Weng, Q. (2012). Remote sensing of impervious surfaces in the urban areas: Requirements, methods, and trends. *Remote Sensing of Environment, 117*, 34–49.

Wentz, E. A., Stefanov, W. L., Gries, C., & Hope, D. (2006). Land use and land cover mapping from diverse data sources for an arid urban environments. *Computers, Environment and Urban Systems, 30*, 320–346.

Wu, B., Yu, B., Wu, Q., Chen, Z., Yao, S., Huang, Y., & Wu, J. (2018). An extended minimum spanning tree method for characterizing local urban patterns. *International Journal of Geographical Information Science, 32*, 450–475.

Xing, H., Meng, Y., Hou, D., Cao, F., & Xu, H. (2017). Exploring point-of-interest data from social media for artificial surface validation with decision trees. *International Journal of Remote Sensing, 38*, 6945–6969.

Yao, Y., Li, X., Liu, X., Liu, P., Liang, Z., Zhang, J., & Mai, K. (2017). Sensing spatial distribution of urban land use by integrating points-of-interest and Google Word2Vec model. *International Journal of Geographical Information Science, 31*, 825–848.

Yao, Y., Liu, X., Li, X., Zhang, J., Liang, Z., Mai, K., & Zhang, Y. (2017). Mapping fine-scale population distributions at the building level by integrating multisource geospatial big data. *International Journal of Geographical Information Science, 31*, 1220–1244.

Yuan, J., Zheng, Y., & Xie, X. (2012). Discovering regions of different functions in a city using human mobility and POIs. *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 186–194). ACM.

Yue, Y., Zhuang, Y., Yeh, A. G., Xie, J.-Y., Ma, C.-L., & Li, Q.-Q. (2017). Measurements of POI-based mixed use and their relationships with neighbourhood vibrancy. *International Journal of Geographical Information Science, 31*, 658–675.

Zhang, X., & Du, S. (2015). A linear dirichlet mixture model for decomposing scenes: Application to analyzing urban functional zonings. *Remote Sensing of Environment, 169*, 37–49.

Zhang, X., Du, S., & Wang, Q. (2017). Hierarchical semantic cognition for urban functional zones with VHR satellite images and POI data. *ISPRS Journal of Photogrammetry and Remote Sensing, 132*, 170–184.

Zhang, X., Xiao, P., Feng, X., & Yuan, M. (2017). Separate segmentation of multi-temporal high-resolution remote sensing images for object-based change detection in urban area. *Remote Sensing of Environment, 201*, 243–255.

Zheng, Y., Capra, L., Wolfson, O., & Yang, H. (2014). Urban computing: Concepts, methodologies, and applications. *ACM Transactions on Intelligent Systems and Technology (TIST), 5*, 38.

Zhu, D., Wang, N., Wu, L., & Liu, Y. (2017). Street as a big geo-data assembly and analysis unit in urban studies: A case study using Beijing taxi data. *Applied Geography, 86*, 152–164.