

RECOGNITION OF URBAN FUNCTIONAL REGIONS AT STREET SCALE BASED ON LJ1-01 NIGHT-TIME LIGHT REMOTE SENSING AND MOBILE BIG DATA

Qiming Zhou^{1,3}, Yang Zhang^{1,2}, Duo Gao⁴, Bo Sun^{1,*}

¹ Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, P.R.China - (qm.zhou, yang.zhang2, sunbo)@siat.ac.cn

² University of Chinese Academy of Sciences, P.R.China - yang.zhang2@siat.ac.cn

³ Department of Geography, Hong Kong Baptist University, Hong Kong S.A.R. - qiming@hkbu.edu.hk

⁴ TalkingData Co., Ltd. - ted.gao@tendcloud.com

KEY WORDS: Night-time Light Remote Sensing, Mobile Big Data, Urban Functional Regions, Residential Area, LJ1-01

ABSTRACT:

Night-time light (NTL) remote sensing data has been widely used in the analysis of human activities at night for a large extent, such as light pollution monitoring and recognition of urban functional regions. In most previous studies, the spatial resolutions of NTL remote sensing data are rather low (e.g., 500 m or coarser). Besides, remote sensing classification of land use rather than land cover types is often a hard task. Due to the reasons, it is difficult to meet the demand of urban refined management. In this study, mobile big data and LuoJia1-01(LJ1-01) NTL remote sensing satellite data with higher spatial resolution are adopted for extracting urban functional regions at the street-level scale. Taking Shenzhen city as a case, mobile big data (i.e., the volume of mobile devices) with the accuracy of approximate 150 m (i.e., 7-bit GeoHash encoding format) is integrated with NTL remote sensing data. After a hot spot analysis, the correlation between the two types of data are analysed. The typical supervised classification algorithms including KNN, SVM and random forest are used to distinguish urban functional regions, which are represented by five types, namely residential, business and commercial, industrial, transportation and other areas. The results show the effectiveness in extracting land use types in cities by combination of LuoJia1-01 night-time light remote sensing and mobile big data. This study provides a basis for monitoring night light pollution of residential area, urban planning and the construction of smart cities.

1. INTRODUCTION

With an increasing usage of mobile phone in modern cities, mobile big data has been widely utilized for urban analysis, such as estimation of travel demand (Toole et al., 2015), analysis of the distribution of economy activities (Chang et al., 2014), home detection (Vanhoof et al., 2018), classification of land use and land cover types (Mao et al., 2016) and understanding of individual mobility patterns. Apart from mobile phone data, night-time light (NTL) remote sensing which reflects human activities is important for studying social issues such as poverty, environment, and ecology (Hu et al., 2017). This kind of data is utilized as an important supplementary data for the census (Calabrese et al., 2013). Relevant studies are focused on urban information extraction (Li and Chen, 2018), urban built-up areas analysing (Ouyang et al., 2016) night-time light pollution (Kuechly et al., 2012), identification of land use (Aubrecht and León Torres, 2016) and analysis of economic activity (Doll et al., 2006).

Due to the coarse resolution of NTL remote sensing data, traditional NTL data are inappropriate for classifying urban land use in details. In previous studies, recognition of urban functional regions based on night-time light remote sensing was conducted at city-level scale due to the restriction of the coarser spatial resolutions. Therefore, residential areas and commercial areas are difficult to distinguish from each other on NTL remote sensing images.

The identification of land use in cities (e.g., distinguishing residential areas from the others) is rather difficult only based on remote sensing with coarse resolutions (Aubrecht and León

Torres, 2016). Pei et al. (2014) and Tu et al. (2017) used mobile data to classify urban land use, and got better results. But the data process is complex and the precision is a little bit low. It needs more various and high-resolution data to get more accurate classification results.

For getting a better understanding of urban functional regions, the new generation NTL remote sensing data with higher resolution (LuoJia1-01) and mobile phone data are introduced in this study. In a certain extent, mobile phone data and NTL remote sensing data can both reflect human activities.

2. STUDY AREA AND DATA

2.1 Study Area

Shenzhen is chosen as the study area. It is a coastal city in Guangdong Province, South China (113.43 E ~ 114.38 E; 22.24 N ~ 22.52 N). Demarcated from Hong Kong by the Shenzhen River in the south, it is bordered by Dongguan and Huizhou in the north, Daya Bay and Dapeng Bay in the east and the Lingding Sea in the west. The total area of Shenzhen is near 2,000 square kilometres. The number of permanent residents has reached to 12,528,300 by the end of 2017. Shenzhen has jurisdiction over 10 districts, including Luohu, Futian, Nanshan, Yantian, Bao'an, Longgang, Longhua, Pingshan, Guangming and Dapeng new district. It is designated as a national pilot city for comprehensive reform. Being an economic centre in southern China, it is also the centre for science and technology innovation, a regional finance centre and a trade and logistics centre (China Pictorial, 2018).

* Corresponding author, Email: sunbo@siat.ac.cn

For more detailed study, Futian district is selected since a rich variety of urban functional areas. Located in the central part of Shenzhen, Futian district is the seat of municipal government office and the chief seat of commerce. It covers an area of 78.66 square kilometers with 10 sub-district offices, including Yuanling, Nanyuan, Futian, Shatou, Meilin, Huaifu, Xiangmi Lake, Lianhua, Huaqiangbei and Fubao. It has a population of 1.33 million. The population density is one of the highest in Shenzhen.

2.2 LJ1-01

Luojia1-01 (or LJ1-01), a new generation of NTL remote sensing satellite launched on June 2, 2018, offers global coverage within 15 days. It is the first remote sensing satellite focusing on night-time light and the first low-orbit satellite with earth observation and satellite navigation enhancement functions. It has a spatial resolution of 130 m and a swath of 250 km. The data is free to download at Hubei Data and Application Centre for High-Resolution Earth Observation System. In order to ensure data quality, a cloud-free NTL remote sensing image on September 3, 2018 is adopted. Figure 1 shows the original NTL image in Shenzhen.

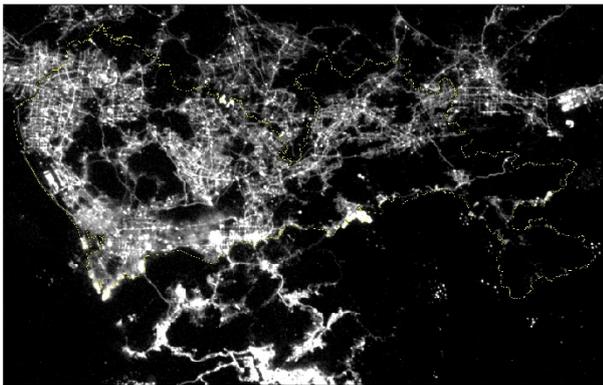


Figure 1. LJ1-01 night-time light remote sensing original image (Sept.3, 2018)

2.3 Mobile Location Big Data

The volume of mobile devices in Shenzhen at night from 22:00 pm to 6:00 am in the third quarter (Q3) of 2017 (i.e., July-September, 2017) is utilized in this study. This data describes the number of mobile devices in a certain area, with higher positioning accuracy (around 10 m) provided by built-in GPS unit. The location information is encoded by GeoHash. In this study, a 7-bit GeoHash coding data is utilized with a specific precision of approximate 150-meter spatial resolution on the ground. For example, a piece of data "ws07k9f" in GeoHash format represents the area with latitudes ranging from 23.0809 N to 23.0822 N, and longitude ranging from 113.0520 E to 113.0534 E. All data are decoded to x-y coordinate data by computer programming and imported into Geographical Information System software (ArcGIS) for further analysis. Figure 2 shows the distribution of the volume of mobile devices in Shenzhen.

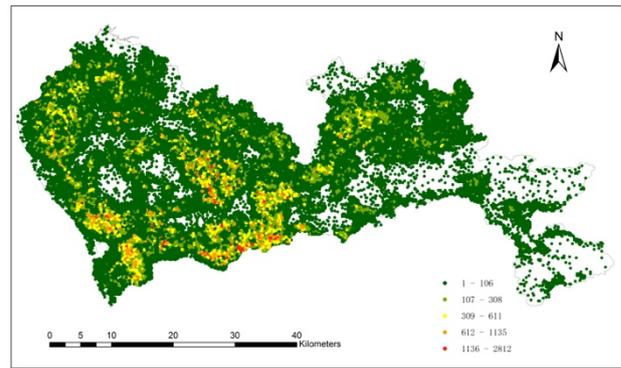


Figure 2. The distribution of the volume of mobile devices in Shenzhen (2017, Q3)

3. METHOD

3.1 Data Pre-processing

Since original mobile big data is encoded by GeoHash, to obtain accurate location information, each encoded string is converted to the corresponding latitude and longitude coordinates. After that, data cleaning is conducted. Each row in the converted file contains three columns: longitude, latitude, and number of mobile devices. The data is clipped in ArcGIS with the boundary of Shenzhen municipal administrative region.

Point density analysis are carried out for mobile data to increase awareness of hotspot areas where mobile devices are concentrated. Accordingly, the densely populated areas of Shenzhen can be recognized, so as to get a general impression of the distribution of human activities.

As for NTL remote sensing data, geometric correction of the imagery is carried out by point-to-point registration to matching existing road vector data. Digital Number (DN) value of NTL remote sensing image represents the intensity of night-time light in Shenzhen.

Given the difference of spatial resolutions between NTL data and mobile data, 130m-resolution NTL remote sensing data are resampled based on nearest neighbor interpolation to match the mobile data. It means all analyses are carried out at the unified spatial resolution with grid size of 150 m * 140 m.

3.2 Sampling and Labelling

The supervised classification method is employed for extracting urban functional regions. 260 samples evenly distributed in Shenzhen are randomly selected from Google earth as ground truth. Those samples are labelled by field survey and manual interpretation based on high-resolution imagery. Figure 3 shows the distribution of the ground truth samples. Among the samples, 70% is used for model training and 30% is used for result validation.

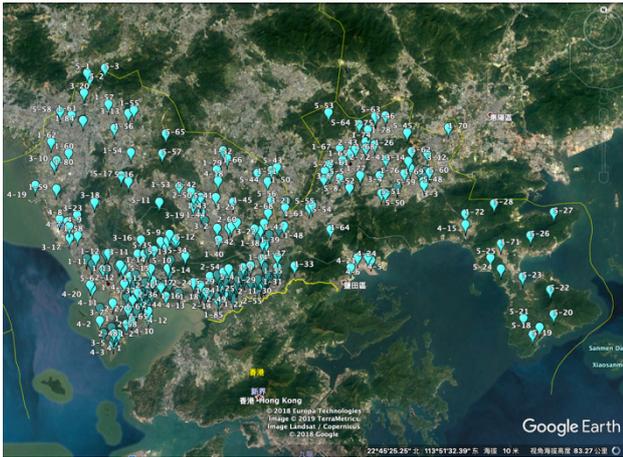


Figure 3. The placemarks manually interpreted based on Google earth

3.3 Recognition of Urban Functional Regions and validation

From the previous studies, an assumption hereby is that intense lighting in that context is associated with a high likelihood of commercial and/or industrial presence, commonly clustered in certain parts of a city (such as central business districts and/or peripheral commercial zones), while areas of low light intensity, in turn, can be considered more likely residential (Aubrecht and León Torres, 2016).

To represent urban functional regions, five land use types are classified, namely residential area (R), business and commercial area (B), industrial area (I), transportation area (T) and others (O). The commonly used supervised classification algorithms are compared, including k-nearest neighbors (KNN), random forest (RF) and support vector machine (SVM). Mobile big data, including volumes of mobile devices during night time and NTL remote sensing data collected at the same season are utilized for training and prediction. In order to compare the classification models, 10-fold cross validation method is adopted for accuracy assessment. The most efficient algorithm with the best accuracy is chosen to predict urban functional regions for the whole study area. The process of recognizing urban functional regions and the steps are shown in Figure 4.

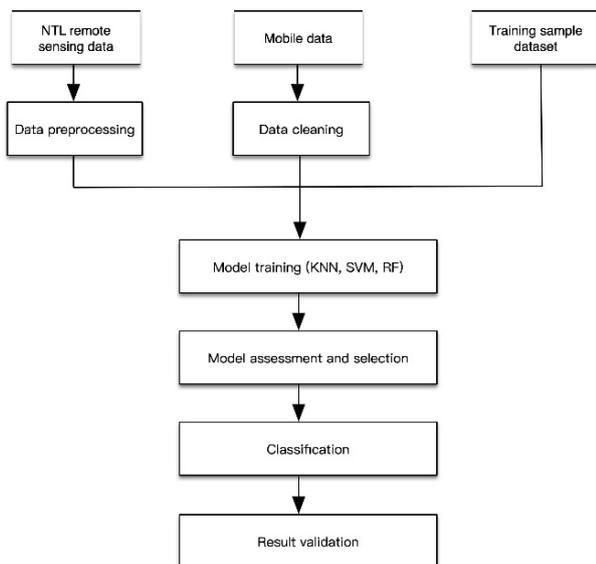


Figure 4. Process of recognition of urban functional regions

4. RESULTS AND DISCUSSION

4.1 Mobile Device Data Processing Results

After data cleaning and hotspot analysis, Figure 5 shows the distribution of population in Shenzhen during the night represented by mobile big data. In hotspot analysis, a neighborhood is defined around the center of each raster pixel, the number of points in the neighborhood is added, and then the area of the neighborhood is divided to obtain the density of the points. From Figure 5, population are concentrated in several districts, including Futian, Nanshan, Baoan, Luohu, Longgang and Longhua new districts. Among them, concentration in Futian, Luohu and Nanshan is more obvious, which account for a higher proportion of dense population associated with higher degree of human activities at night.

4.2 LJI-01 Night-time Light Remote Sensing Data Processing Results

Figure 6 shows the light intensity at night in Shenzhen. Compared with the distribution of population, night-time light intensity shows a general consistent distribution pattern. It should be pointed out that the light intensity at some places is very high with smaller population. These places are mainly located in ports, highways and industrial areas.

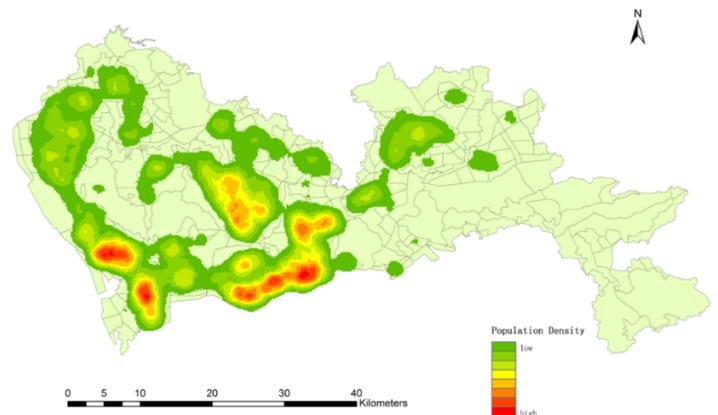


Figure 5. Distribution pattern of human activities at night based on mobile big data

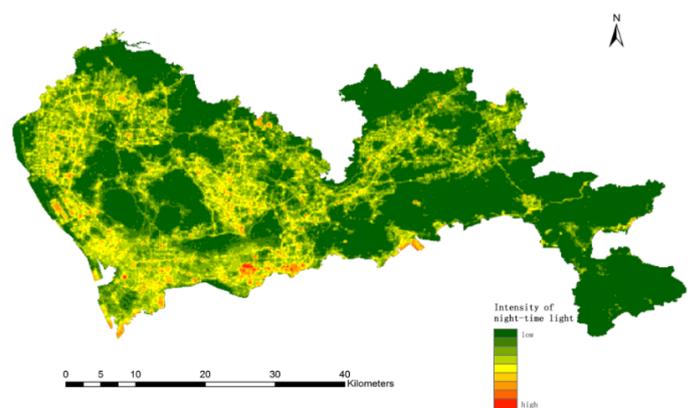


Figure 6. Intensity of night-time light retrieved from remote sensing in Shenzhen

4.3 The Result of Combining Two Types of Data to Classify

From the comparative analysis of mobile data and night-time light remote sensing data in Shenzhen, the two types of data show a consistency in general. The sample data are trained by three typical and commonly used classification algorithms. The results based on 10-folder cross validation are shown in Table 1. Among the three classifiers, random forest algorithm is the better one. Therefore, it is applied to the classification of the entire study area. The classification results are shown in Figure 7. Figure 8 shows the accuracy assessment result by confusion matrix. From the results, some misclassification exists, especially for business/commercial and industrial types. While, majority of uninhabited areas and residential areas on the map

can match with actual situation on the ground. Given transportation type owns the characteristics of high luminous intensity and smaller population, iconic wharf, airport as well as some brighter streets are classified into this category. For a better illustration, the classification result of Futian district is presented as shown in Figure 9.

Classifier	Overall Accuracy
KNN	0.68
RF	0.72
SVM	0.71

Table 1. 10-folder cross validation results by different classifiers

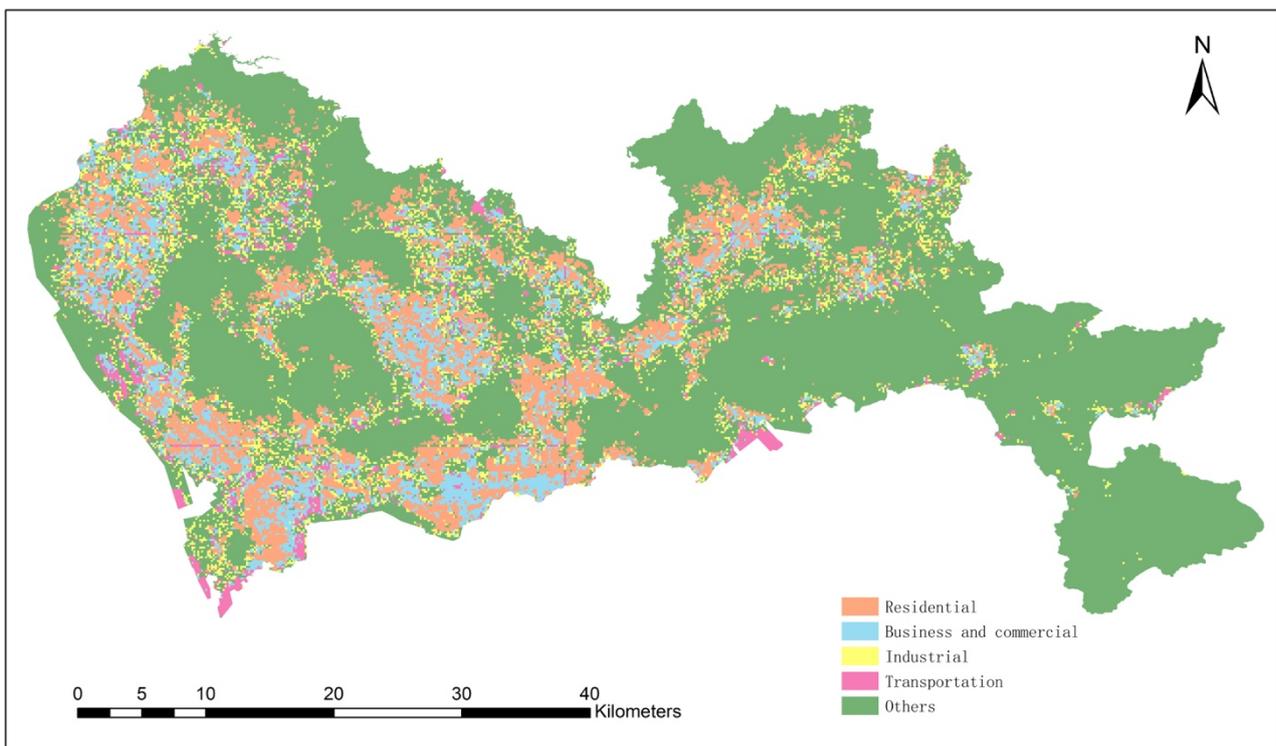


Figure 7. Classification results of functional regions in Shenzhen

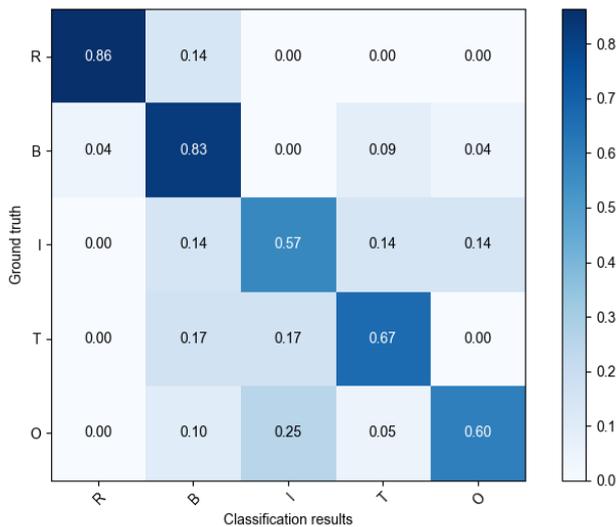


Figure 8. Normalized confusion matrix of supervised classification (RF in this case)

Taking Futian district as a case, more details can be observed. The boundaries of uninhabited areas (e.g., forest, reservoirs, parks) are well matched with the actual situation on the ground. Residential area patches generally reflect the distribution in Futian district, although some patches are misclassified into other categories. Since Futian district is a highly developed region with complex land use attributes, land use patches may have various attributes, which means the functional region types would be a certain overlap.

It should be pointed out that the two datasets used in the study (i.e., night-time light remote sensing and mobile data) are not acquired in the same year. Some uncertainties may be introduced. Besides, the number of training samples is relatively small, which may influence the effectiveness of the classification model.

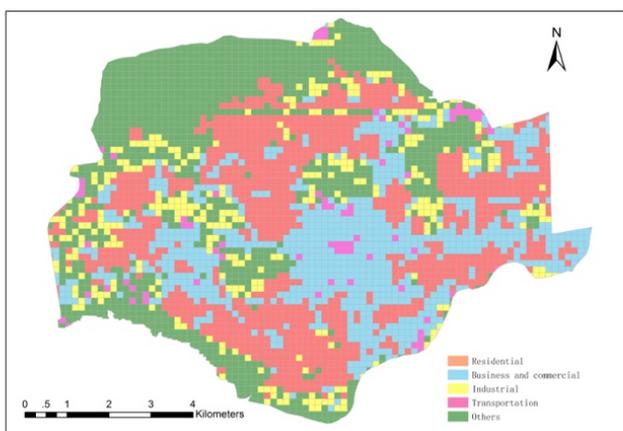


Figure 9. Case study of functional region recognition in Futian district.

5. CONCLUSIONS

A pilot study on combining high-resolution NTL remote sensing and mobile big data has been carried out to distinguish urban functional regions in Shenzhen city. After pre-processing of mobile big data and night-time light remote sensing, hot spot distribution of human activities is analysed. With ground truth data, the supervised classification method is used to classify

land use into five categories. Business and commercial areas as well as industrial areas can be well distinguished from residential areas. In particular, the identification of residential areas is of great significance to urban planning and management. The proposed method can be used to identify the concentrated residential areas of urban residents, reasonably allocate municipal facilities and so on.

In the future, mobile data will be further enriched, for example, detailed time range data and day-time data, so as to get more accurate results. More external data will also be introduced to get a refined classification result of urban functional regions. In addition, training data will be expanded and the accuracy of the training data could be further verified to better supervise the classification algorithm and obtain more accurate result.

ACKNOWLEDGEMENTS

The authors would like to thank Wuhan University for providing LuoJia1-01 night-time light remote sensing data. This research was supported by Hong Kong Baptist University Faculty Research Grant (Grant No.: FRG/17-18/030).

REFERENCES

- Aubrecht, C., León Torres, J.A., 2016. Evaluating multi-sensor nighttime earth observation data for identification of mixed vs. residential use in urban areas. *Remote Sensing*, 8(2), 114.
- Calabrese, F., Diao, M., Di Lorenzo, G., Ferreira, J., Ratti, C., 2013. Understanding individual mobility patterns from urban sensing data: a mobile phone trace example. *Transportation Research Part C: Emerging Technologies*, 26, 301-313.
- Chang, X., Lu, G., Yue, Y., Li, Q., 2014. Estimating the distribution of economy activity: a case study in Jiangsu Province (China) using large scale social network data. *2014 IEEE International Conference on Data Mining Workshop (ICDMW)*. IEEE Computer Society.
- China Pictorial. 2018. Shenzhen: 40 Years of Reform and Opening up. <http://china-pictorial.com.cn/shenzhen-40-years-of-reform-and-opening-up>. accessed 03/04/2019
- Doll, C.N.H., Muller, J.P., Morley, J.G., 2006. Mapping regional economic activity from night-time light satellite imagery. *Ecological Economics*, 57(1), 75-92.
- Hu, K., Qi, K., Guan, Q., Wu, C., Yu, J., Qing, Y., Zheng, J., Wu, H., Li, X., 2017. A scientometric visualization analysis for night-time light remote sensing research from 1991 to 2016. *Remote Sensing*, 9(8), 802.
- Kuechly, H.U., Kyba, C.C.M., Ruhtz, T., Lindemann, C., Wolter, C., Fischer, J., Hölker, F., 2012. Aerial survey and spatial analysis of sources of light pollution in Berlin, Germany. *Remote Sensing of Environment*, 126(11), 39-50.
- Li, K., Chen, Y., 2018. A Genetic algorithm-based urban cluster automatic threshold method by combining VIIRS DNB, NDVI, and NDBI to monitor urbanization. *Remote Sensing*, 10(2), 277.
- Mao, H., Thakur, G.S., Bhaduri, B.L., 2016. Exploiting mobile phone data for multi-category land use classification in Africa. *Acm Sigspatial Workshop on Smart Cities & Urban Analytics*. ACM.

- Ouyang, Z., Fan, P., Chen, J., 2016. Urban built-up areas in transitional economies of southeast Asia: spatial extent and dynamics. *Remote Sensing*, 8(10), 819.
- 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(9), 1988-2007.
- Toole, J. L., Colak, S., Sturt, B., Alexander, L.P., Evsukoff, A., González, M.C., 2015. The path most traveled: travel demand estimation using big data resources. *Transportation Research Part C: Emerging Technologies*, S0968090X15001631.
- Tu, W., Cao, J., Yue, Y., Shaw, S., Zhou, M., Wang, Z., Chang, X., Xu, Y., 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*, 1-28.
- Vanhoof, M., Reis, F., Ploetz, T., Smoreda, Z., 2018. Assessing the quality of home detection from mobile phone data for official statistics. *Journal of Official Statistics*, 34, 935–960.