

PAPER • OPEN ACCESS

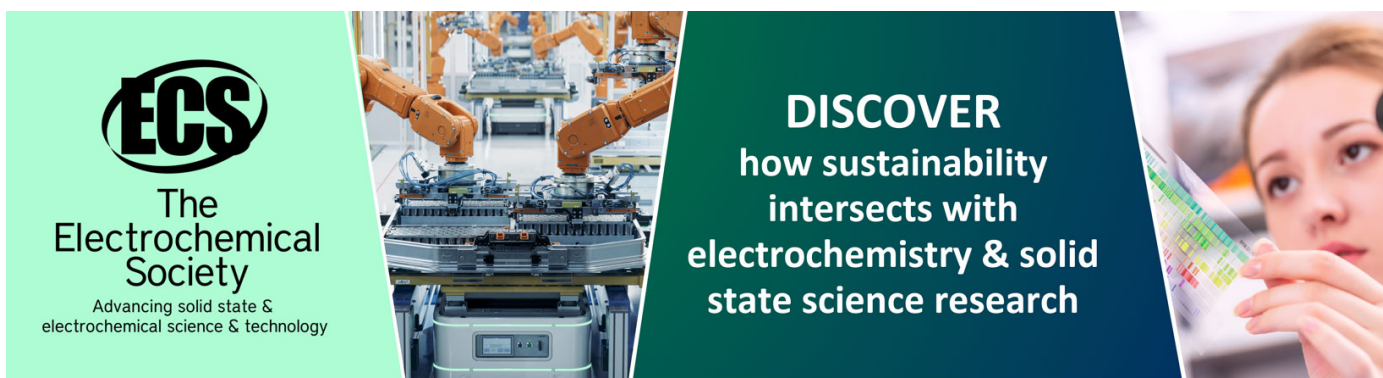
Correlation Analysis between NPP-VIIRS Nighttime Light Data and POIs Data —a Comparison Study in Different Districts and Counties of Nanchang

To cite this article: Yuqian Wang *et al* 2021 *IOP Conf. Ser.: Earth Environ. Sci.* **693** 012103

View the [article online](#) for updates and enhancements.

You may also like

- [Fishing boat detection using Sentinel-1 validated with VIIRS Data](#)
Marza Ihsan Marzuki, Rinny Rahmania, Penny Dyah Kusumaningrum *et al.*
- [Estimating global oilfield-specific flaring with uncertainty using a detailed geographic database of oil and gas fields](#)
Zhan Zhang, Evan D Sherwin and Adam R Brandt
- [Environmental and political implications of underestimated cropland burning in Ukraine](#)
Joanne V Hall, Sergiy V Zibtsev, Louis Giglio *et al.*



ECS
The
Electrochemical
Society
Advancing solid state &
electrochemical science & technology

DISCOVER
how sustainability
intersects with
electrochemistry & solid
state science research

Correlation Analysis between NPP-VIIRS Nighttime Light Data and POIs Data —a Comparison Study in Different Districts and Counties of Nanchang

Yuqian Wang^{1,2*}, Yue Li¹, Xuepeng Song¹, Xin Zou¹ and Jiajun Xiao¹

¹Faculty of Geomatics, East China University of Technology, Nanchang 330013, China

²Key Laboratory for Digital Land and Resources of Jiangxi Province, East China University of Technology Nanchang 330013, China

Email: neo@ecut.edu.cn

Abstract. The study of urban spatial structure is very important for understanding the relationship between people and urban infrastructure as well as urban planning. This paper selected the 2016-2019 NPP-VIIRS NTL data and POI data of Nanchang City to conduct a pixel-level correlation analysis of Nanchang urban districts and counties. Firstly, the NPP-VIIRS and POI datasets should be respectively preprocessed to synthesize NPP-VIIRS annual average dataset and the POI kernel density dataset. Secondly, these datasets were respectively subjected to logarithmic transformation. Finally, they were divided into administrative districts for linear regression analysis. Experimental results showed that the correlation coefficients between NPP-VIIRS NTL data and POIs data gradually decreased from urban to rural districts.

1. Introduction

Urban spatial structure is a very hot research field in urban geography. A good urban spatial structure can promote the harmonious development between people and urban land. However, rapid development of cities will cause many problems. The unplanned expansion of urban will affect a series of problems against the optimization of urban functions and industrial layout. Nighttime light (NTL) remote sensing can capture the weak light signals at night. It indirectly reflects the intensity of human activities on the surface, and provides a special perspective for the study of urban structures. NTL data has been widely used in regional economic development imbalance[1], rural power consumption[2], natural disaster assessment[3], and identification of urban centers[4], and so on. Points-of-interest (POIs) contain rich attribute information and high precision. Besides, they are easily obtained. Kai et al. used multi-source datasets including POIs to evaluate the poverty[5]; Tu et al. used POIs, community media check-in and mobile phone records to develop urban vitality indicators[6]; Chen et al. used POIs data to map the urban spatial structure[7]. At present, a large number of studies have used NTL data and POIs data to analyze the urban spatial structure. Ge et al. used NTL data and POIs data to explore the city center[8]; Liu et al. used NTL data and POIs data to map the urban night leisure space[9]; Sun et al. used multi-source data such as NTL data and POIs data to classify built-up areas[10]. This paper takes Nanchang as the study area, and divides Nanchang into districts and counties under administrative divisions. Then the correlations between NTL data and POIs data of each districts and counties are quantitatively estimated by Pearson correlation coefficients under linear regression analysis. Then the reasons for the inconsistency between NTL data and POIs data are discussed, which is helpful to quantitatively analyze the urban internal spatial structure.



2. Study Area and Datasets

2.1. Study Area

Nanchang is the capital of Jiangxi Province. It is located in East China, north of Jiangxi Province. According to the website of Nanchang Municipal People's Government, Nanchang City currently governs three counties (Nanchang, Jinxian, Anyi) and six districts (Donghu, Xihu, Qingyunpu, Qingshanhu, Xinjian and Honggutan) as shown in figure 1.

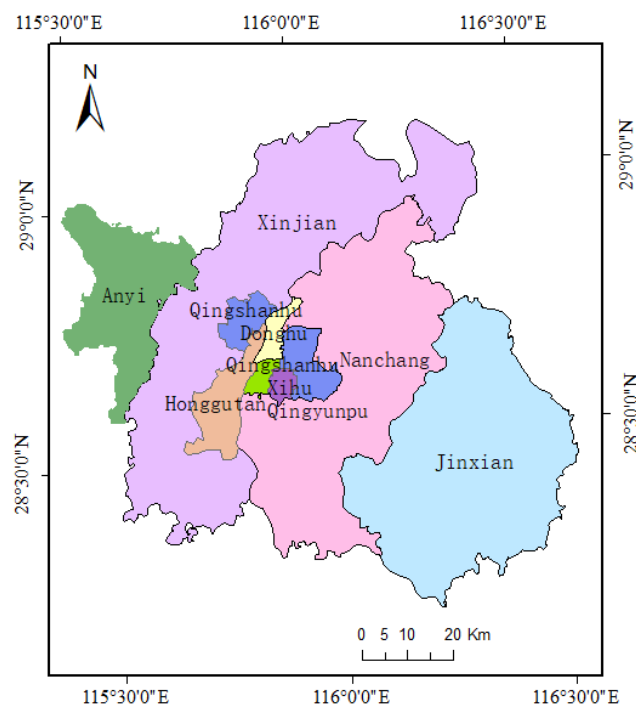


Figure 1. Nanchang city map

2.2. Datasets

The datasets used in this study included NPP-VIIRS NTL data, POIs data and administrative boundary data. NPP-VIIRS NTL datasets (2016-2019) were provided by NOAA/NGDC (<https://eogdata.mines.edu/>). NPP-VIIRS NTL data of 2016 was the annual average data, and NPP-VIIRS NTL datasets of 2017-2019 were monthly composite data. There were 100,478 POIs data of 14 types collected by Python on Gaode Map (<https://lbs.amap.com/>) in 2019 of Nanchang City. The administrative division boundary data of Nanchang City was downloaded in the National Catalogue Service for Geographic Information (<https://www.webmap.cn/>)

3. Methods

3.1. Data Preprocessing

3.1.1. Synthesis of npp-viirs annual average data

The 2016 annual average data has removed unstable light sources and background noise, while the 2017-2019 monthly data has not been standardized and still has noise. Hence the original data should be processed. Zhou et al. developed a method of synthesizing annual night light data based on NPP-VIIRS monthly NTL data[11]. Firstly, images of June 2017 and February 2019 were removed out of the data set as the study area was seriously covered by clouds. Then all remained images were cropped respectively by the vector boundary of the administrative divisions of Nanchang City. Radiance values less than 0 were set to 0, which meant the non-light area. All NTL images were reprojected to an Albers Conic Equal Area projection with a spatial resolution of 500m. The following

steps were taken to eliminate unstable regions with the assumption that if a place had been bright (NTL larger than 0) for two consecutive years, it would be a stable light brightness area. The radiance values of NTL data greater than 0 were set to 1 with others set to 0. Then binary images of 2016-2019 were got. The stable light brightness area of 2017 was got by multiply the binary image of 2016 with 2017. The NTL image of 2017 was multiplied by the stable light brightness area of 2017 to eliminate unstable regions. NTL images of 2018-2019 were similarly treated to eliminate unstable regions. Extremely high radiance values were considered as false. The most economically developed area in China is Shanghai, so the radiance value in Shanghai should be the highest in China. The stable highest radiance value of 192.684 in Shanghai in 2016 was chosen as the threshold to detect the false extremely high value in Nanchang City[11]. Then mean filtering was used to calculate the average value of the pixels around the extremely high values. Then extremely high values were replaced by their average value. Finally, all NTL values were rounded-off. The annual average NTL data of 2019 was shown as in Figure 2.

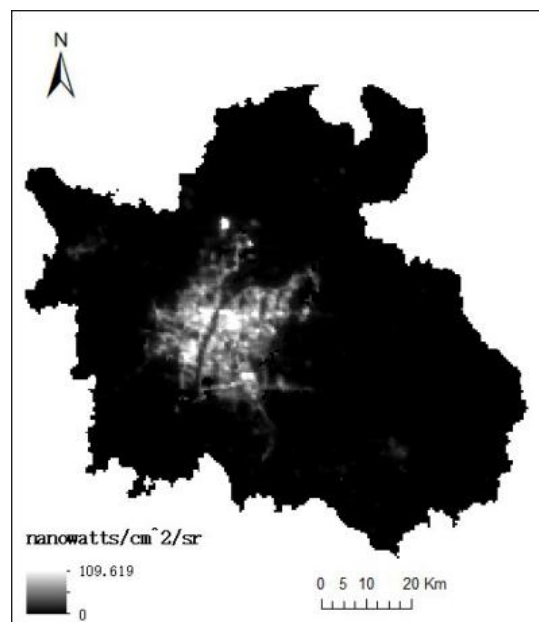


Figure 2. Annual average NTL data of 2019

3.1.2. The Kernel Density Estimation

Kernel density estimation (KDE) was used to calculate the unit density of the observed values of points and line elements in a specified neighborhood, which can directly reflect the distribution of discrete values in a continuous area. The principle of kernel density estimation was to assign weights according to the distance between the data point and the center point. The smaller the distance between the observation point and the center point, the greater the weight. The formula is as follows:

$$P_i = \frac{1}{n\pi R^2} \times \sum_{j=1}^n K_j \left(1 - \frac{D_{ij}^2}{R^2}\right)^2 \quad (1)$$

where P_i was the kernel density of the data point i , K_j was the weight of the data point j , D_{ij} was the distance between space point i and point j , and R was the distance threshold between the two points, that is, the bandwidth ($D_{ij} < R$), n was the number of point data in the bandwidth. A large number of studies had shown that the spatial weight function had little effect on the results of KDE. Reasonable choice of bandwidth R had a key influence on the result. Hinnerburg et al. studied the relationship between the bandwidth and the number of density centers and found that there were a certain number of bandwidth intervals to keep the density center stable[12]. It was reasonable to select the bandwidth in this interval.

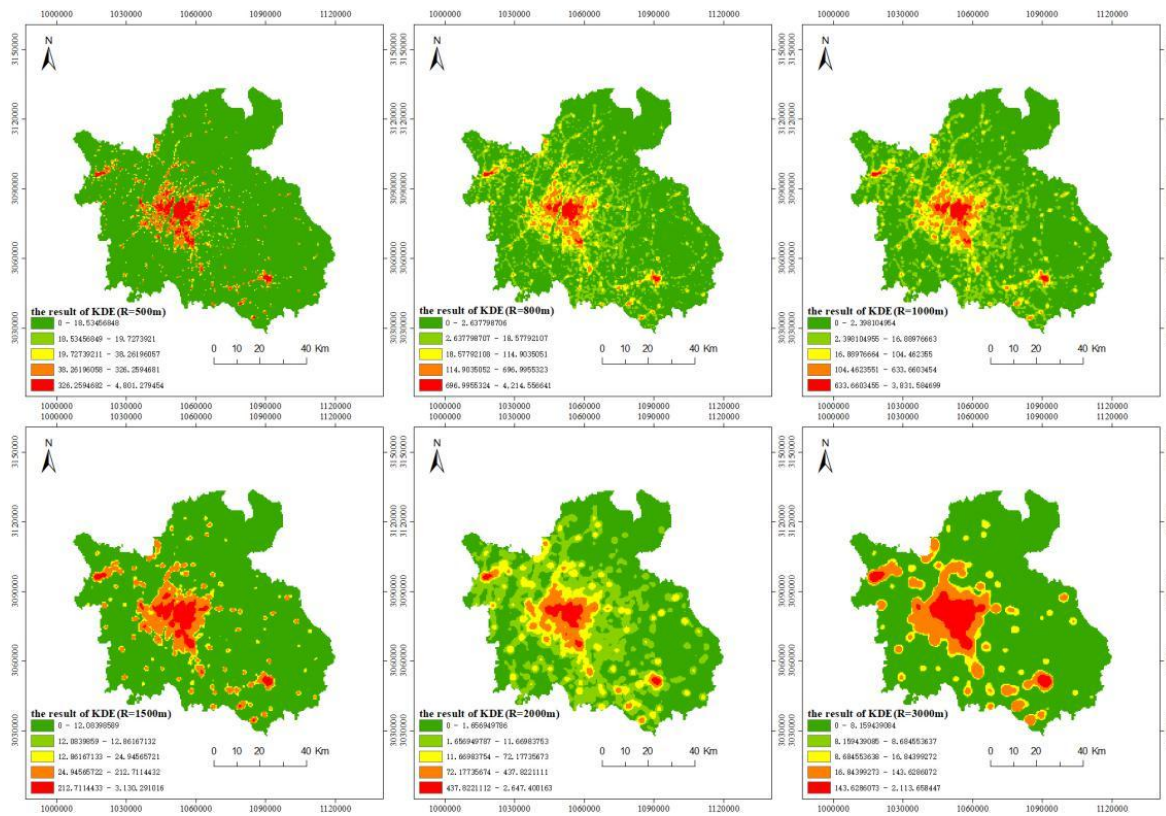


Figure 3. The result of kernel density estimate (R=500m-3000m)

As shown in Figure 3, the comparison showed that the density center was relatively stable between the bandwidth of 800m-1000m, so 1000m was selected as the bandwidth of KDE, and the pixel size was set to be 500m.

3.2. Regression Analysis between NPP-VIIRS NTL and Kernel Density of POIs

3.2.1. Logarithmic transformation

Logarithmic transformation was a common gray-scale transformation method in image processing. The formula is as follows:

$$f(i, j) = \ln[NTL(i, j)] \quad (2)$$

where $NTL(i, j)$ was the original NPP-VIIRS NTL intensity of the pixel at row i and column j . $f(i, j)$ was the result of the pixel corresponding to the NPP-VIIRS NTL data after logarithmic transformation. The formula is as follows:

$$g(m, n) = \ln[1 + k(m, n)] \quad (3)$$

where $k(m, n)$ was the original kernel density of the pixel at row m and column n . $g(m, n)$ was the result of the pixel corresponding to the POIs data after logarithmic transformation. In the logarithmic transformation values should be larger than 0, so pixels with 0 value in NTL or KDE were removed.

3.2.2. Pearson correlation coefficient

The Pearson correlation coefficient was one of the common methods to measure the linear correlation between two numerical variables. Its mathematical definition was:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (4)$$

where n was the number of samples; x_i and y_i were the variable values of the two variables respectively; $\bar{x}$ and $\bar{y}$ were the average values of the two variables respectively. The Pearson correlation coefficients of 2019 annual average NTL data and POIs KDE in every administrative division were respectively calculated.

4. Results and Discussion

As shown in table 1 and figure 4, generally speaking, the NTL intensity values were obviously positively correlated with the kernel density values of POIs in all districts and counties of Nanchang. It indicated NTL intensity was positively correlated with degree of urban building aggregation. Specifically, the correlation coefficients between NTL data and KDE of POIs data from high to low were Donghu, Honggutan, Qingshanhu, Qingyunpu, Xihu, Jinxian, Nanchang(county), Anyi and Xinjian. Donghu district as the oldest urban district of Nanchang city had the highest correlation coefficient of all. The correlation coefficients decreased from the center of the city to rural. Urban districts had remarkable high correlation coefficients than suburban counties. Xinjian district was an exception because it was rural before 2015 and had become urban district since 2016. There had been no substantial changes during a short period. The kernel density of the county centers in Anyi and Jinxian counties were higher, but the NTL intensity were lower. It may be due to the gathering of buildings in the center of the county and the infrequent human activities in the county at night.

Table 1. The correlation coefficient between NTL data and KDE of POIs data.

District	r
Donghu (urban)	0.8357
Honggutan (urban)	0.7427
Qingshanhu (urban)	0.7386
Qingyunpu (urban)	0.6575
Xihu (urban)	0.6561
Jinxian (rural)	0.6461
Nanchang (rural)	0.5931
Anyi (rural)	0.4990
Xinjian (urban)	0.4884

The NTL intensity indirectly expressed human social and economic activities intensity. The kernel density of POIs showed the concentration of places where human social and economic activities happened. The high consistency of the NTL intensity and the kernel density of POIs implied harmony in urban development planning and a mature urban structure. In our study of Nanchang city, urban districts had higher concentration of social and economic activities as well as high-intensity public and commercial lights than rural countries. Urban districts also had higher consistency of the NTL intensity and the kernel density of POIs than suburban counties. It implies that urban districts had better urban planning and more mature urban structure. Xinjian district as an exception had the lowest correlation coefficient between NTL data and KDE of POIs data. It implied the change of administrative attribute could not change the urban planning and urban structure in a short term.

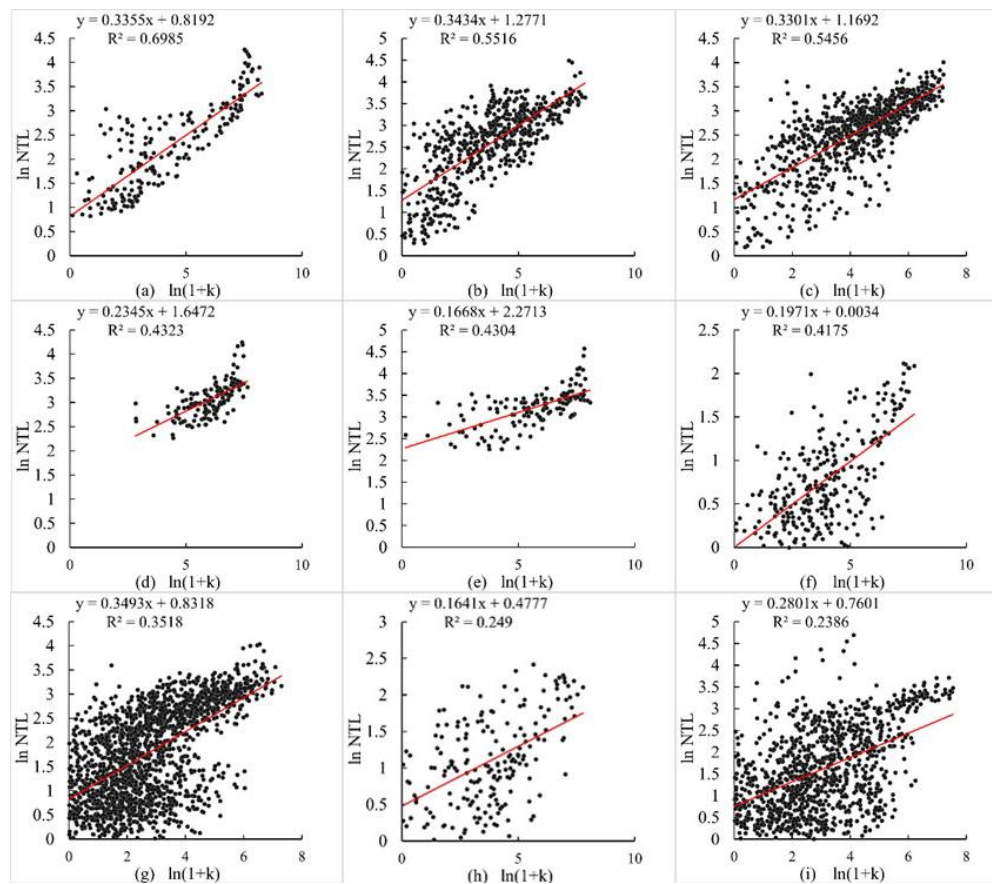


Figure 4. Scatter plots depicting the relationship between $\ln(NTL)$ and $\ln(1+k)$. Donghu (a), Honggutan (b), Qingshanhu (c), Qingyunpu (d), Xihu (e), Jinxian (f), Nanchang (g), Anyi (h), Xinjian (i).

5. Conclusions

To understand the spatial structure of a city, it is necessary to obtain accurate urban construction spatial distribution information at the pixel level. In this study, the linear regression analysis on the NTL intensity and POIs KDE quantitatively described the consistency between human social and economic activities intensity and urban construction. The specific conclusions were drawn as follows:

The correlation between NTL intensity and kernel density of POIs in Nanchang's urban districts was obviously higher than Nanchang's suburban country, especially the oldest urban district-Donghu district, which showed that the NTL intensity in the urban area had a strong consistency with the degree of building agglomeration. It could efficiently reflect the concentration of human social and economic activities.

The correlation between NTL intensity and kernel density of POIs in county-level suburban administrative areas was relatively weak. The POIs kernel density of the county center was high, but the NTL intensity was low. The non-consistence of NTL intensity and kernel density of POIs in rural area may imply that night time light remote sensing may not be accurate in characterizing the built-up spatial structure in rural areas.

The consistency of the NTL intensity and the kernel density of POIs could efficiently estimate harmony in urban development planning. High consistency could imply a mature urban structure and low consistency may imply that may be some problems in urban development planning.

6. Acknowledgments

This work was supported by the National Natural Science Foundation of China (41861052, 41701437), Natural Science Foundation of Jiangxi Province (20202BABL202045, 20202BABL213029) and Open

Foundation of Key Laboratory for Digital Land and Resources of Jiangxi Province, East China University of Technology (DLLJ201915)

7. References

- [1] Pan W B, Fu H M and Zheng P 2020 Regional Poverty and Inequality in the Xiamen-Zhangzhou-Quanzhou City Cluster in China Based on NPP/VIIRS Night-Time Light Imagery *Sustainability* **12** (06) 2547.
- [2] Zhao F, Ding J Y, Zhang S J, Luan G Z, Song L, Peng Z Y, Du Q Y and Xie Z Q 2020 Estimating Rural Electric Power Consumption Using NPP-VIIRS Night-Time Light, Toponym and POI Data in Ethnic Minority Areas of China *Remote Sens* **12** (17) 2836.
- [3] Zhao X Z, Yu B L, Liu Y, Yao S J, Lian T, Chen L J, Yang C S, Chen Z Q and Wu J P 2018 NPP-VIIRS DNB Daily Data in Natural Disaster Assessment: Evidence from Selected Case Studies *Remote Sens* **10** (10) 1526.
- [4] Ma M, Lang Q, Yang H, Shi K and Ge W 2020 Identification of Polycentric Cities in China Based on NPP-VIIRS Nighttime Light Data *Remote Sens*. **12** (19) 3248.
- [5] Shi K F, Chang Z J, Chen Z Q, Wu J P and Yu B L 2020 Identifying and evaluating poverty using multisource remote sensing and point of interest (POI) data: A case study of Chongqing, China *Journal of Cleaner Production* **255** 120245.
- [6] Tu W, Zhu T T, Xia J Z, Zhou Y L, Lai Y N, Jiang J C and Li Q Q 2020 Portraying the spatial dynamics of urban vibrancy using multisource urban big data *Computers, Environment and Urban Systems* **80** 101428.
- [7] Lu C Y, Pang M, Zhang Y, Li H, Lu C P, Tang X L and Cheng W 2020 Mapping Urban Spatial Structure Based on POI (Point of Interest) Data: A Case Study of the Central City of Lanzhou, China. *ISPRS Int. J. Geo-Inf* **9** (02) 92.
- [8] Lou G, Chen Q, He K, Zhou Y and Shi Z 2019 Using Nighttime Light Data and POI Big Data to Detect the Urban Centers of Hangzhou *Remote Sens* **11** (15) 1821.
- [9] Liu J, Deng Y, Wang Y, Huang H S, Du Q Y and Ren, F 2020 Urban Nighttime Leisure Space Mapping with Nighttime Light Images and POI Data *Remote Sens* **12** (03) 541.
- [10] Sun L, Tang L N, Shao G F, Qiu Q Y, Lan T and Shao J Y 2019 A Machine Learning-Based Classification System for Urban Built-Up Areas Using Multiple Classifiers and Data Sources *Remote Sens* **12** (01) 91.
- [11] Zhou Y, Chen Y, Liu Y, Tian F and Yi X 2019 Generation and Verification of NPP-VIIRS Annual Nighttime Light Data *Remote Sensing Information* **34** (2) 62.
- [12] Hinneburg A and Keim D A 1998 An Efficient Approach to Clustering in Large Multimedia Databases with Noise **98** 58-65.