IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 31, 2021, accepted February 9, 2021, date of publication February 16, 2021, date of current version March 4, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3059865

Estimating Socio-Economic Parameters via Machine Learning Methods Using Luojia1-01 Nighttime Light Remotely Sensed Images at Multiple Scales of China in 2018

BIN GUO[1], YI BIAN[1], DINGMING ZHANG[1], YI SU[1], XIAOXIA WANG[1], BO ZHANG[1], YAN WANG[1], QIUJI CHEN[1], YARUI WU[1], AND PINGPING LUO[2,3]

[1]College of Geomatics, Xi'an University of Science and Technology, Xi'an, 710054 China
[2]Key Laboratory of Subsurface Hydrology and Ecological Effects in Arid Region, Ministry of Education, Chang'an University, Xi'an 710054, China
[3]School of Water and Environment, Chang'an University, Xi'an 710054, China

Corresponding authors: Bin Guo (guobin12@xust.edu.cn) and Pingping Luo (lpp@chd.edu.cn)

This work was supported in part by the Fund Project of Shaanxi Key Laboratory of Land Consolidation under Grant 2019-JC11, and in part by the National Natural Science Foundation of China under Grant 41771576.

ABSTRACT Mapping socio-economic indicators with a raster format is still a great challenge. The nighttime light (NTL) datasets have been widely utilized to estimate the socio-economic parameters. However, the precision of the published datasets was too coarse to meet related issues such as flood losses assessment, urban planning, and epidemiological studies. The present study calibrated gross domestic product (GDP), population (POP), electric consumption (EC), and urban build-up area (B-A) at 100 m resolution for 45 cities of China in 2018 using Luojia1-01 NTL datasets via random forest (RF) as well as geographically weighted regression (GWR) model. The linear regression (LR), back propagation neural network (BPNN), and support vector machine (SVM) methods were selected for comparison with GWR and RF models. Besides, the Suomi National Polar-Orbiting Partnership-Visible Infrared Imaging Radiometer Suite (NPP-VIIRS) was chosen for comparison with Luojia1-01. The ten-folded cross-validation (CV) has been used for evaluating accuracy at county and city scales. Finally, the distribution maps of socio-economic parameters were illustrated and some findings were obtained. First, the validation results revealed that the calibration at the city-scale outperformed the county or district scale. Second, the precision of the Luojia1-01 NTL dataset surpassed the NPP-VIIRS NTL dataset on the same administrative scale except for some specific situations. Third, the precision of the simulation for the gross domestic product (GDP) is the highest than the others, followed by electric consumption (EC), build-up area (B-A), and population (POP). Fourth, the optimum model varied according to the socio-economic parameters. Fifth, the distribution of socio-economic parameters exhibited obvious spatial heterogeneity. This paper can supply scientific support for calibrating socio-economic parameters in other regions.

INDEX TERMS Luojia1-01, NPP/VIIRS, socio-economic parameters, GWR, machine learning, multiple scales, China.

I. INTRODUCTION

Socio-economic parameters are valuable data sources for governments making decisions and scientific researches [1], [2]. China, the world's second-largest economy, is undergoing massive infrastructure construction [3], [4]. Industrialization, urbanization [5] and human activities are causing severe disturbances to the land surface. The traditional statistical datasets with the political division scale are inadequate in describing socio-economic phenomena because the frequency of updates for statistical data is always yearly [6]. Also, the spatial heterogeneity of socio-economic phenomena within the administrative division can hardly be described for the data accuracy is restricted [7]–[12]. Meanwhile, the classic investigation methods for statistical data are time and finance-consuming [13]. Hence,

The associate editor coordinating the review of this manuscript and approving it for publication was Wentao Fan.

introducing new technologies and approaches to alter the traditional workflow in collecting socio-economic parameters is still a huge challenge [8].

Artificial lights are prevalent in most urban agglomeration areas due to urbanization and industrialization [14], [15]. Artificial lights can be detected by nighttime light (NTL) remote sensing at night, and the NTL has been confirmed to be a reliable data source for monitoring anthropic activities [16]. NTL data has been introduced in many fields recently, such as the evaluation of war losses [17], [18], the typhoon and flood disasters assessment [19], [20], the light pollution monitoring [21], the built-up area detection [22]–[24], impervious surface extracting [25], the cultural festival and protected area identifying [26], [27], the tourism activities [28], [29], the infrastructure monitoring [30], and the crime hot spots identifying [31]. Besides, the NTL data generally have a statistically significant relationship with socio-economic parameters including population [32], [33], gross domestic product [34], [35], poverty [36], [37], house vacancy [38], [39], carbon dioxide emissions [40], and electricity consumption [41], [42]. Previous studies revealed that the NTL remote sensing supplied a speedy, economical, and effective method to retrieve socio-economic activities in developed areas [43]–[46].

However, some deficiencies concerning spatial and radiation resolution of the existing NTL datasets are unavoidable. Specifically, although the Defense Meteorological Satellite Program's Operational Line scan System (DMSP-OLS) nighttime light imagery has exhibited excellent capacity in evaluating economic parameters, the over-saturation issues in urban centers can be hardly eliminated due to lack of onboard radiometric calibration and limited radiometric detection [47], [48], [49]. In October 2011, the Visible Infrared Imaging Radiometer Suite (VIIRS) sensor with a national polar-orbiting satellite (Suomi-National Polar-Orbiting Partnership, Suomi-NPP) was launched successfully [50]–[52]. Observational sensitivity and spatial resolution of NPP/VIIRS have been significantly promoted. The effectiveness of NPP/VIIRS in reflecting social and economic activities has been highly improved compared to DMSP/OLS [10], [53]. Whereas, gas flares, fires, volcanoes, or aurora have not been removed by the NPP-VIIRS images. The background noise has also not been deducted [10]. Besides, the capacity of NPP-VIIRS in capturing small cities was severely restricted owing to its limited spatial resolution [54].

The Luojia1-01 satellite, the first designed to collect high-resolution nighttime light images in the world, was launched by Wuhan University, China, on June 2, 2018 [21], [55], [56]. The spatial resolution of the Luojia1-01 nighttime light images has been greatly promoted with on-board calibration (TABLE 1). The brightness and details of Luojia1-01 images have also been improved compared with DMSP-OLS and NPP-VIIRS [57] (FIGURE 1). Meanwhile, the Luojia1-01 nighttime light images have demonstrated huge potentials in many fields. For example, Jiang *et al.* (2018) [21] detected artificial light pollution via Luojia1-01 and NPP-VIIRS data

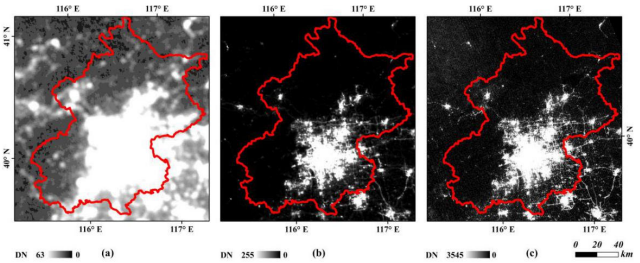


FIGURE 1. The nighttime imagery of Beijing, China extracted from DMSP-OLS (a), NPP-VIIRS (b), and Luojia1-01(c). Note: The images of DMSP-OLS, NPP-VIIRS, and Luojia1-01 were collected in 2013, December 2018, and November 2018, respectively. The DMSP-OLS image of Beijing used in Figure 1(a) was a yearly product.

TABLE 1. The comparison of parameters for DMSP-OLS, NPP-VIIRS, and Luojia1-01 [55].

Satellite	DMSP-OLS	NPP-VIIRS	Luojia1-01
Operator	U.S. Department of Defense	NASA/NOAA	Wuhan University
Available years	1992–2013	December 2011–present	June 2018–present
Wavelength range	400–1100 μm	505–890 μm	480–800 μm
Orbital altitude	830 km	830 km	645km
Orbit	Polar	Polar	Polar
Spatial resolution	2.7 km	742m	130m
Width	3000 km	3000 km	260km
Revisit time	12h	12h	15d
Pixel saturated	Saturated	No Saturated	No Saturated
On-board calibration	No	Yes	Yes

and concluded that the former dataset revealed better performance than the latter. Li *et al.* (2018) [23] found that Luojia1-01 achieved a higher precision in extracting an urban extent than NPP-VIIRS images. Chen *et al.* (2019) [25] reported that the Luojia1-01 dataset could produce a more precise map of impervious surfaces. Overall, Luojia1-01, with an improved NTL-intensity detection performance, enables a detailed look into artificial surface lighting distribution, which supplies a novel perspective in detecting human activities' dynamic on a small scale.

Socio-economic phenomena have a strong correlation with night-time light, which can be well depicted by the regression model [58]. Some previous studies used Luojia1-01 data to calibrate the socio-economic parameters via the linear regression method [2]. Nevertheless, as a single global model, the classical regression model requires the assumption of a prior probability distribution and cannot properly capture complex relationships [59]. So, the spatial autocorrelation feature of socio-economic data should be taken into consideration, and the previous methods need to be improved in precision. Some novel methods such as deep learning and machine learning give a new perspective to simulate the socio-economic parameters using nighttime light images. Although machine learning such as random forest (RF) has been introduced

to calibrate socio-economic parameters based on NTL data, the findings of comparing the feasibility of different models, the calibration precision at multiple scales, and the capacity of a variety of NTL datasets in estimating socio-economic parameters were scarce [54], [60], [61]. Therefore, evaluating the capacity of estimating socio-economic parameters using Luojia1-01 images via the classic regression model as well as machine learning methods at multiple scales is strongly desired.

The objectives of this study are: (i) to calibrate the socio-economic parameters including gross domestic product (GDP), population (POP), electricity consumption (EC), and urban build-up area (B-A) based on linear regression (LR), geographic weighted regression (GWR), Back Propagation (BP) neural network, support vector machine (SVM), and Random Forest (RF) using statistical, NPP-VIIRS, and Luojia1-01 data of China in 2018, (ii) to validate and compare the simulation results determined by different models and multi-source data at multiple scales, (iii) to evaluate the feasibility of calibration methods using NTL data to simulate socio-economic parameters.

II. STUDY AREA AND DATA SOURCES

A. STUDY AREA

China, with a land area of 9.6 million square kilometers accounting for about 1/15 of the total land area of the world, has a vast territory. Social and economic development demonstrated significant heterogeneity in China. The southeastern region has the most developed economy with speedy development. Whereas, the northwestern of China is undergoing massive infrastructure construction with relatively slow development. So, 45 cities, 352 districts (or counties) located in different regions have been chosen as the study areas to evaluate the robustness of the calibration model in a variety of economical structures. The selected study areas with different social and economic development statuses are distributed in eastern, southern, western, northern, and central China, respectively (TABLE 2, FIGURE 2). Meanwhile, we chose these 45 cities, 352 districts (or counties) as study areas due to the availability of NTL as well as socio-economic statistical datasets.

TABLE 2. The specific information of the study areas.

Regional name (The total districts)	Municipality or province	City name - code (The number of districts)
The Eastern region (62)	Zhejiang	Ningbo -19 (10), Quzhou -23 (6), Taizhou -20 (9), Jinhua -22 (8), Shaoxing -21 (6)
	Jiangsu	Suzhou -13 (6), Yangzhou -12 (5), Lianyungang -9 (6), Changzhou -14 (6)
	Fujian	Sanming -26 (12), Quanzhou -27 (10), Putian -25 (5)
The central region (94)	Jiangxi	Shangrao -24 (12)
	Henan	Anyang -5 (9)
	Anhui	Wuhu -16 (8), Bengbu -11 (7), Tongling -17 (4), Ma'anshan -15 (7), Anqing -18 (8), Huainan -10 (7)
	Hunan	Chenzhou -38 (11), Hengyang -36 (12), Zhuzhou -37 (9)
The Western Region (47)	Sichuan	Deyang -41 (6), Yibin -40 (10), Bazhong -42 (4)
	Shaanxi	Xi'an -44 (13), Ankang -43 (10), Tongchuan -45 (4)
The Southern Region (80)	Guangdong	Jiangmen -34 (7), Zhuhai -35 (3), Jieyang -31 (5), Guangzhou -32 (11), Shantou -30 (6), Shenzhen -33 (6), Meizhou -28 (8), Chaozhou -29 (3)
	Guizhou	Liupanshui -39 (4)
The Northern Region (69)	Beijing	Beijing -1 (16)
	Hebei	Handan -2 (18)
	Shandong	Taian -4 (6), Zaozhuang -7 (6), Rizhao -8 (4), Liaocheng -3 (8), Jining -6 (11)
Sum	14	45 (352)

Note: The code in Table 2 corresponds to Figure 2.

B. DATA SOURCES

Two kinds of nighttime light remotely sensed images were introduced to estimate socio-economic parameters including Luojia1-01 and the NPP data. Besides, the traditional statistic dataset on the counties' scale was chosen for calibrating and validating.

The Luojia1-01 NTL images can be downloaded free of charge at the website (http://www.hbeos.org.cn/). The phenomenon of data missing is existed due to the inconsistency of the shooting time of the Luojia1-01 images in various regions, so in this paper, the Luojia1-01 NTL images in December 2018 were chosen for the data coverage was relatively complete than other periods. Although the socio-economic phenomena changed all the time, the variations during the monthly scale were subtle that the influences for the modeling could be neglected. So, the counterpart images in November 2018 and October 2018 were chosen as substitutions if the data missing phenomenon existed in parts of the study areas in December 2018, and we assumed that the differences between the images in December 2018 and the images in November 2018 and October 2018 could be ignored.

Additionally, NPP/VIIRS nighttime light data, under the National Oceanic and Atmospheric Administration (NOAA) (https://www.ngdc.noaa.gov/ngdc.html), were obtained from the National Geophysical Data Center (NGDC). The images of NPP/VIIRS in December 2018 were selected in line with further comparisons with Luojia1-01.

The traditional statistic data in 2018 came from the statistical yearbooks published by local governments (Statistical Yearbook, 2019). The political boundaries originated from the Ministry of natural resources of China (http://bzdt.ch.mnr.gov.cn/). The following table is a detailed description of the data sources (TABLE 3).

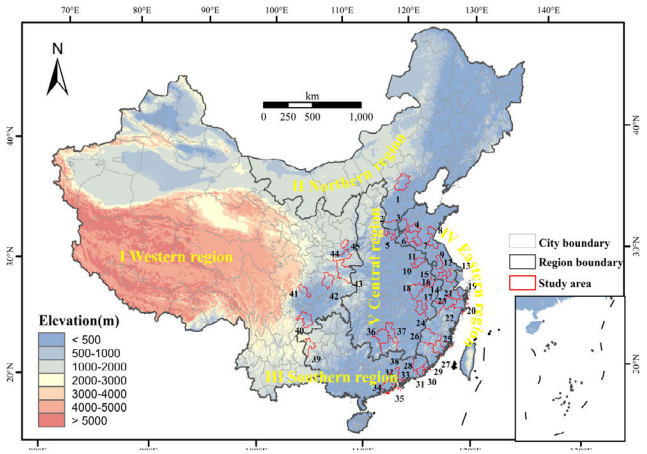


FIGURE 2. The distribution map of the study area of city scale, and the topographic of China in 2018. Note: the number in Figure 2 denotes the code of each city on the map, and the name of each city was presented in Table 2.

C. DATA PREPROCESSING

First, the Luojia1-01 and NPP-VIIRS images of the study areas were transformed into a CGCS2000 (China Geodetic Coordinate System 2000) reference system with an Albers projection after mosaicking and masking, then the Luojia1-01 was resampled to 100m and NPP-VIIRS was resampled to 500m, respectively.

Second, the positioning accuracy of Luojia1-01 NTL images without geometric control points (GCPs) is less than 650 m, which is almost five times the resolution of the image. Hence, the Luojia1-01 images have been geometrically corrected via the Google Map images before analysis. The intersection points of the road network have been selected as the GCPs benefit from the high spatial resolution of Luojia1-01 images. The geometric positioning error of the target image has been reduced through geometric correction. The temporary light values such as volcanic eruption, flare, exhaust gas combustion, and background noise were not excluded in the original NPP-VIIRS night light radiation data, which may result in negative and abnormal values in the original radiation value, large or sudden value. So, in the current study, outliers, and pixels with values larger than zero were removed [62] FIGURE 3. Besides, the NPP/VIIRS data had been corrected for absolute radiation before data releasing. Therefore, absolute radiation correction was only implemented for Luojia1-01 data in this study [63].

$$L = DN^{3/2} \times 10^{-10} \tag{1}$$

where L is the radiance value after absolute radiation correction in $w/m^2 \cdot \text{sr} \cdot \mu\text{m}$ The DN value is the image gray value for each pixel.

III. METHODS

A. WORKFLOW

The present study was conducted as follows (FIGURE 4). First, data preprocessing was performed on Luojia 1-01, NPP-VIIRS, traditional statistic data, and political boundaries.

TABLE 3. Description of the datasets.

Datasets	Names &Unit	Spatial-temporal coverage	Source	Data Pretreating
Statistical data	Gross Domestic Product (GDP) Unit: 10^{10} yuan Population (POP) Unit: 10^5 people Electricity Consumption (EC) Unit: 10^9 kw·h Built-up Area (B-A) Unit: 10^1 km^2		Statistical Yearbook	Standardizing the collected statistical data for unit consistency
Nighttime light images	Cloud-free average monthly product of NPP-VIIRS in December 2018.	45 cities, 352 districts (or counties), 2018	https://www.noaa.gov/	Resampling; Geometrical correction; Removing outliers; Projection transformation
	Cloud-free product of LuoJia1-01 in December 2018.		http://59.175.109.173:8888/app/login_en.html.	Resampling; Geometrical correction; Radiation correction; Projection transformation
Auxiliary data	Political Boundaries		http://sgic.geodata.gov.cn/	Projection transformation

Note: The population, GDP, electricity consumption, and the built-up area were chosen as the objectives of the current study because the population and GDP are the most important indicators for describing the development of the social economy in a country. Meanwhile, China is undergoing speedy urbanization and the amount of energy consumption is increasing. So, detecting the built-up area and the electricity consumption is valuable for policymaking. Besides, the availability of the related datasets had been investigated before the current study was conducted. We found that the spatial-temporal coverage for the population, GDP, electricity consumption, and the built-up area was the best. The study areas were distributed more in central, eastern, and southern China due to the availability of the related datasets.

Then the zonal statistics tool from ArcGIS was used to obtain the total NTL values of the two types of nighttime light remotely sensed data at counties and cities scale. Furthermore, the linear regression, the geographically weighted regression (GWR), the random forest (RF), the backpropagation neural network (BPNN), and the support vector machine (SVM) was implemented for calibrating and validating socio-economic parameters including GDP, EC, POP, and B-A at city and county scales. The performance of the above models was evaluated at multiple scales. Finally, the optimal model was selected for mapping the socio-economic parameters, and the spatial distribution characteristics of the four socio-economic parameters were analyzed.

B. CALIBRATION MODELS

a. LR models

$$SEP = a_1 \times TNL + b_1 \tag{2}$$

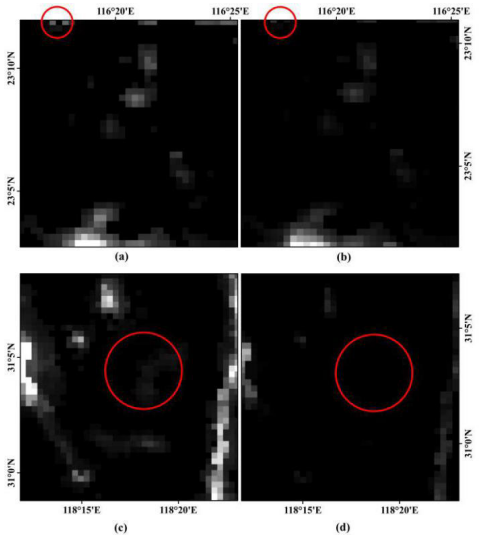


FIGURE 3. The NPP-VIIRS nighttime light images before and after noise eliminating of the examing area in 2018. Note: The above figures only demonstrate parts of the entire study areas due to the space of the article, and the processed methods of noise elimination for other areas are the same. Two regions bounded by red circles were originated from Shantou, Guangdong (a, b), and Wuhu, Anhui (c, d). The (a) and (c) represent the original images of the sampling areas, (b) and (d) denote the corrected images of the sampling areas. Two areas with noise are pointed through red circles.

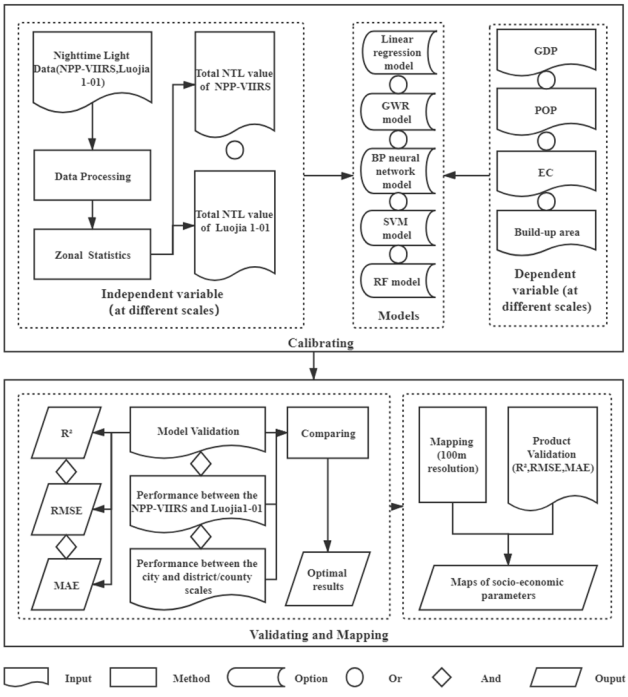


FIGURE 4. Research framework.

where SEP is the sum of socio-economic parameters in an administrative unit; TNL is the total nighttime light (TNL) in an administrative unit (i.e., the sum of all pixel values of nighttime light data in the administrative unit); and a_1 and b_1 are the regression coefficients and intercept, respectively.

b. GWR models

$$SEP_i = \beta_{1(u_i v_i)} + \beta_{2(u_i v)} \cdot TNL + \varepsilon_i \quad (3)$$

where the meanings of SEP and TNL are the same as equation (2); β_1 represents the intercepts at a specific location (u_i, v_i); and β_2 is the location-specific slope; The location (u_i, v_i) denotes the central coordinates of the monitoring site, ε_i is the error term for the county (city)$_i$ [64].

c. RF models

The Random forest (RF) algorithm [65] is a bagging method based on regression tree (CART) analysis and classification. The benefits of RF are the significance of each feature can be assessed with unbiased estimation during the classification process, and the issues with numerous missing data can be solved. Additionally, the efficiency of the RF model in processing big data without any dimensionality reduction outperforms traditional models. The classification trees were used to decide on choosing optimal trees in predicting. The number of classification trees in RF is large, and all variables have to be inputted into each tree with an independent feature for classing. Moreover, 99.9% of unrelated trees conduct predictions that cover all conditions. The basic theory of RF bagging is to choose the results of several weak classifiers and form a strong classifier [66]. In this study, the simulation result was optimal when one hundred trees (ntree = 100) were used for conducting the RF model. The importance values were not described because only the NTL data was selected as the independent variable.

d. BPNN models

The back-propagation neural network (BPNN) is one of the widely utilized artificial neural network methods. The BPNN consists of an input layer, one or more hidden layers, and an output layer and is relatively simple and easy-realized. There are two basic processes of BPNN: firstly, a signal-feed-forward process was conducted to input signals to the input layer computing the predicted values and forward to the output layer; secondly, an error-backward phase was implemented to compare the predicted and observed values and backward to the neurons in each layer, the weights, and thresholds of the neurons are modified. Determining the number of neurons in each layer, the maximum iterations, and the learning rate is necessary for designing BPNN [67]. In this study, epochs were 700, and the learning rate was 0.01.

e. SVM models

SVM, a supervised learning model and a prediction model based on error-based learning, can be used for pattern recognition and data analysis. So, SVM is widely utilized for classification and regression analysis. SVM is a binary classifier with a perceptron function-based classification model. The distances between the two groups can be measured by SVM to calculate the center between the two datasets and to divide the groups using the optimal discriminant boundary.

$$\omega \cdot x - b = 0 \quad (4)$$

where ω denotes the weight for the entity. b denotes the intercept. SVM detects an entity occurring at + 1 and

another occurring at − 1 based on the discriminant boundary and determines classification. The margin is the sum of the distances between the discriminant boundary and each entity [68].

C. MODEL VALIDATION

The performances of the calibration models were evaluated by the ten-folded cross-validation method [69], [70]. All samples were divided into 10 folds of the same size. Nine of the 10 folds were selected for model training and the remaining fold was used for model validation. Then, this process was repeated until all of the folds had been used once as the validation sample. In this study, the mean absolute error (MAE), root mean square error (RMSE), and determinate coefficients (R^2) were used to indicate the model accuracies.

$$R^2 = \frac{\sum_{i=1}^{n}(SEP_{e,i} - \overline{SEP})^2}{\sum_{i=1}^{n}(SEP_{t,i} - \overline{SEP})^2} \quad (5)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|SEP_{t,i} - SEPe, i\right| \quad (6)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(SEP_{t,i} - SEP_{e,i})^2} \quad (7)$$

where n is the number of administrative units, SEP represents socio-economic parameters including GDP, POP, EC, and B-A. $SEP_{t,i}$ denotes the real SEP of administrative unit i, $SEP_{t,i}$ represents the estimated SEP of administrative unit i, and SEP denotes the actual mean SEP of the administrative unit.

D. MAPPING SOCIO-ECONOMIC PARAMETERS

The optimal calibration model with the highest accuracy was selected to map the socio-economic parameters using 100 m resolution gridded variables. The predicted values of socio-economic parameters obtained from the optimal calibration model at the grid-scale were overestimated or underestimated. So, equation (8) was implemented for correcting the biases.

$$SEP'_{i,j} = SEP_{i,j} \times \left(\frac{SEP_{i,census}}{SEP_{i,estimated}}\right) \quad (8)$$

where $SEP'_{i,j}$, and $SEP_{i,j}$ are the corrected and estimated values of socio-economic parameters of administrative unit i at pixel j, respectively, $SEP_{i,census}$ is the socio-economic values of administrative unit i from the Statistical Yearbook, and $SEP_{i,estimated}$ is the estimated total socio-economic values of administrative unit i calculated with zonal statistic tool using ArcGIS.

IV. RESULTS

A. COMPARING THE CALIBRATED AND VALIDATED RESULTS OF SOCIO-ECONOMIC PARAMETERS USING LUOJIA1-01 AND NPP-VIIRS NTL DATASET BASED ON DIFFERENT MODELS AT MULTIPLE SCALES.

The socio-economic parameters in terms of GDP, POP, EC, and B-A were calibrated based on five different models including LR, GWR, SVM, BPNN, and RF using multiple NTL datasets on the city as well as county or district scales. Meanwhile, the ten-folded cross-validation method was implemented to evaluate the precision of the calibrated models. The results were shown in FIGURE 5, TABLE 5, and TABLE 6.

Clearly, no matter which socio-economic parameter, NTL dataset, and model we exam, the results of the calibration at the city-scale outperform the county or district scale. Also, for the same socio-economic parameter, the accuracy of the simulation is better at the city scale with relatively higher correlation coefficients (R^2) and lower RMSE as well as MAE than the county or district scale based on the same NTL dataset excluding some special situations. For example, the result of simulation for POP using the RF model at the county scale (R^2 = 0.83, RMSE = 2.25, MAE = 1.57)

TABLE 4. The R^2, RMSE, and MAE of optimal models for each socio-economic parameter.

socio-economic parameters	Optimal models	R^2	RMSE	MAE
GDP	GWR	0.77	40.28	22.28
EC	RF	0.70	16.36	9.29
POP	GWR	0.67	30.47	16.52
B-A	RF	0.69	23.49	15.07

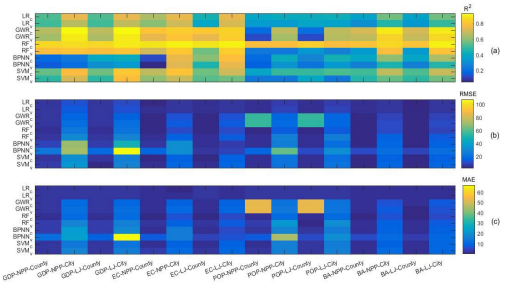


FIGURE 5. The R^2(a), RMSE (b), and MAE (c) of calibrated results for each socio-economic parameter including GDP, EC, POP, and B-A using Luojia1-01 and NPP-VIIRS NTL dataset based on LR, GWR, RF, and BP models at the county as well as city scales of China in 2018. Note: LR_C and LRv denote the correlation coefficient (R^2) of calibration and validation for the linear regression model, respectively. The meaning for other abbreviations of the vertical axis is similar to LR_C and LRv. GDP-NPP-County represents that the GDP was calibrated using NPP-VIIRS NTL dataset at the county scale, and GDP-LJ-City denotes that the GDP was calibrated using Luojia1-01 NTL dataset at the city scale. The definition for other abbreviations of the horizontal axis is similar to GDP-NPP-County and GDP-LJ-City.

TABLE 5. The r^2 of calibration and validation for evaluating the precision of calibrated models using npp-viirs and luojia1-01 ntl dataset at county and city scales of china in 2018. ($\times 10^{-2}$).

M	SEP	NPP-VIIRS				Luojia1-01			
		County		City		County		City	
		C	V	C	V	C	V	C	V
LR	GDP	52	51	74	66	53	51	75	67
	EC	65	59	74	70	44	38	74	71
	POP	38	35	41	35	42	39	43	37
	B-A	37	30	56	47	37	33	57	49
GWR	GDP	66	64	93	89	67	65	93	90
	EC	80	74	83	80	81	75	84	82
	POP	19	13	67	64	21	13	68	66
	B-A	71	68	90	84	72	69	90	85
RF	GDP	86	78	87	80	86	78	90	80
	EC	88	70	91	81	88	56	91	83
	POP	78	23	78	23	83	33	78	26
	B-A	80	34	85	78	80	35	91	80
BPNN	GDP	20	15	24	21	40	30	41	35
	EC	12	5	75	73	59	45	79	74
	POP	20	16	41	32	27	22	44	33
	B-A	38	26	61	45	39	26	65	47
SVM	GDP	55	48	78	71	56	48	81	79
	EC	68	61	80	69	57	54	70	56
	POP	39	23	49	24	50	39	52	24
	B-A	51	47	68	55	52	48	69	57

Note: M denotes the name of models, SEP denotes socio-economic parameters, C and V denote calibration and validation, respectively.

surpasses the city scale ($R^2 = 0.78$, RMSE = 14.02, MAE = 10.98).

The precision of the calibration based on the Luojia1-01 NTL dataset with relatively higher correlation coefficients (R^2) and lower RMSE as well as MAE surpasses the NPP-VIIRS NTL dataset at the same scale except for some specific situations. For example, the accuracy of simulation for EC using the NPP-VIIRS NTL dataset ($R^2 = 0.65$, RMSE = 3.25, MAE = 1.75) is better than the Luojia1-01 dataset ($R^2 = 0.44$, RMSE = 4.08, MAE = 2.37) at the county scale.

Overall, for each socio-economic parameter, no matter which NTL dataset, scale, and model we selected, the precision of the simulation for GDP are the most higher than others, followed by EC, B-A, and POP. No matter which socio-economic parameter, NTL dataset, the scale we choose to run the calibration model, the accuracy of the RF model is the best, followed by GWR, SVM, LR, and BPNN.

B. CHOOSING THE OPTIMAL MODELS FOR EACH SOCIOECONOMIC PARAMETER BASED ON COMPARISON RESULTS

The optimum models for each socio-economic parameter were selected based on section 4.1 comparison results. For all socio-economic parameters, the optimal models have been found when the Luojia1-01 was inputted as the NTL dataset and the calibration models were conducted at the city scale. Besides, the optimum model varied according to the socio-economic parameters. Specifically, for GDP, the optimal

TABLE 6. The rmse and mae of calibration and validation for evaluating the precision of calibrated models using npp-viirs and luojia1-01 ntl dataset at county and city scales of china in 2018.

M	SEP	NPP-VIIRS				Luojia1-01			
		County		City		County		City	
		C	V	C	V	C	V	C	V
LR	GDP	6.68	6.78	11.8	13.5	6.64	6.72	11.7	13.4
		3.07	3.13	2.61	2.90	3.03	3.08	2.47	2.73
	EC	3.25	3.52	5.22	6.01	4.08	4.32	5.72	6.44
		1.75	1.86	1.11	1.31	2.37	2.37	1.25	1.30
	POP	4.16	4.29	8.37	8.79	4.04	4.13	8.24	8.66
		2.91	3.03	2.24	2.37	2.83	2.87	2.24	2.38
	B-A	3.78	3.52	5.79	6.01	3.18	3.28	5.73	6.17
		1.91	1.86	1.44	1.31	1.90	1.97	1.49	1.48
GWR	GDP	5.98	6.23	16.9	20.6	5.91	6.07	16.	20.2
		2.91	2.99	10.6	12.4	2.89	2.91	10.1	11.8
	EC	2.45	2.79	12.8	14.5	2.47	2.77	12.0	13.8
		1.33	1.38	9.46	10.7	1.23	1.58	9.24	10.1
	POP	55.2	54.8	17.3	18.3	55.1	54.1	17.3	17.1
		52.9	52.5	13.1	14.2	52.8	52.3	12.9	12.8
	B-A	2.05	2.18	9.48	11.5	2.05	2.14	9.39	11.3
		1.26	1.28	6.90	7.97	1.21	1.11	7.62	6.51
RF	GDP	3.74	5.19	23.3	30.7	3.71	5.19	22.6	29.3
		1.79	3.45	11.5	16.9	1.72	3.45	10.7	16.0
	EC	2.20	4.61	10.7	15.6	2.06	4.26	10.4	14.0
		1.10	2.26	6.38	8.99	1.02	2.23	5.48	7.65
	POP	2.58	4.85	14.9	27.4	2.25	4.46	14.0	25.9
		1.77	3.38	11.3	21.0	1.57	3.11	10.9	19.6
	B-A	1.82	3.34	11.3	15.6	1.82	3.14	9.87	13.5
		1.00	1.94	6.73	9.47	1.13	1.24	6.33	8.23
BPNN	GDP	8.67	23.8	70.6	71.7	8.39	19.6	48.8	107
		3.70	12.4	22.9	23.6	3.31	8.61	18.2	66.8
	EC	5.26	17.5	32.4	32.5	3.50	5.52	8.43	9.54
		1.88	7.97	14.7	14.8	2.08	4.38	5.38	6.68
	POP	4.76	18.2	25.1	63.0	4.70	11.5	21.2	32.6
		3.07	7.79	17.7	45.3	2.84	6.16	15.5	19.7
	B-A	3.26	10.1	18.3	20.5	3.12	8.91	15.9	20.1
		1.75	5.97	9.08	10.6	1.59	4.06	8.58	10.0
SVM	GDP	6.63	7.19	33.0	36.2	6.61	3.14	27.4	28.8
		2.68	2.90	14.0	16.9	2.24	2.60	13.0	15.4
	EC	3.14	4.97	16.0	17.7	3.13	3.96	15.9	17.4
		1.63	3.26	9.20	11.0	1.52	2.53	9.09	10.5
	POP	4.32	4.80	22.1	22.4	3.85	4.27	21.0	22.1
		2.17	2.86	16.2	16.5	2.39	2.85	15.6	16.0
	B-A	2.97	3.03	16.4	17.4	2.85	3.01	16.2	16.9
		1.55	1.60	9.39	10.7	1.39	1.57	10.4	10.0

Note: M denotes the name of models, SEP denotes socio-economic parameters, C and V denote calibration and validation, respectively.

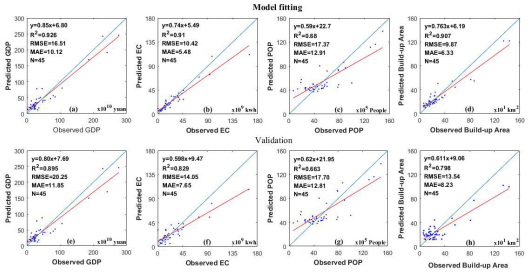


FIGURE 6. Scatter plots of observed and predicted values for GDP (a, e), EC (b, f), POP (c, g), and B-A (d, h) via GWR and RF models using the Luojia1-01 NTL dataset at the city scale of China in 2018. Note: the optimal calibration models for GDP, EC, POP, and B-A are GWR, RF, GWR, and RF, respectively.

model (R^2= 0.926, RMSE = 16.51, MAE = 10.12) was detected as we fitted the GWR model using the Luojia1-01 NTL dataset at the city scale. For EC, the best result (R^2=0.91, RMSE = 10.42, MAE = 5.48) was found when we ran the RF model using the Luojia1-01NTL dataset at the city scale. For POP, the optimum method (R^2=0.68, RMSE = 17.37, MAE = 12.91) was determined as we run the GWR model using the Luojia1-01NTL dataset at the city scale. For B-A, the optimal method (R^2=0.907, RMSE=9.87, MAE=6.33) was obtained when we conducted the RF model using the Luojia1-01 NTL dataset at the city scale (FIGURE 6).

C. MAPPING THE SOCIO-ECONOMIC PARAMETERS WITH 100M RESOLUTION BASED ON THE OPTIMAL MODELS

Each socio-economic parameter was mapped according to the optimum models determined by the R^2, RMSE, and MAE using the raster calculator tool from ArcGIS. The socio-economic parameters of typical cities located in different regions of China were illustrated due to the length of the article, and the maps of the remaining cities were demonstrated in the supplementary materials FIGURE 7 and Figure 9).

The distribution of GDP for Beijing exhibited obvious spatial heterogeneity, that is, the hotspots of GDP are mainly distributed inside the Sixth Ring Road and present a decreasing trend from inner to outside. There are about two regions distributed outside the Sixth Ring Road with a relatively higher GDP value than the neighborhood. One is the Yudu Mountain situated in Yanqing County, northwest of Beijing that is a famous tourism site, the other is the Miyun County, northeast of Beijing. Also, the distribution of GDP for Xi'an, Shaanxi Province demonstrated significant spatial features, that is, the concentrated GDP areas are mainly located inside the Third Ring Road and reveal a descending trend from inner to outside. The densest areas of GDP are situated inside the City Wall built in Ming Dynasty, in Xiaozhai commercial areas, Qujiang New District, and High Technology District. For the distribution of GDP in Shenzhen city, it represents a typical pattern that the most concentrated GDP areas except for the Pengshan, Yantian, and Dapeng District are mainly distributed along the South China Sea, and the

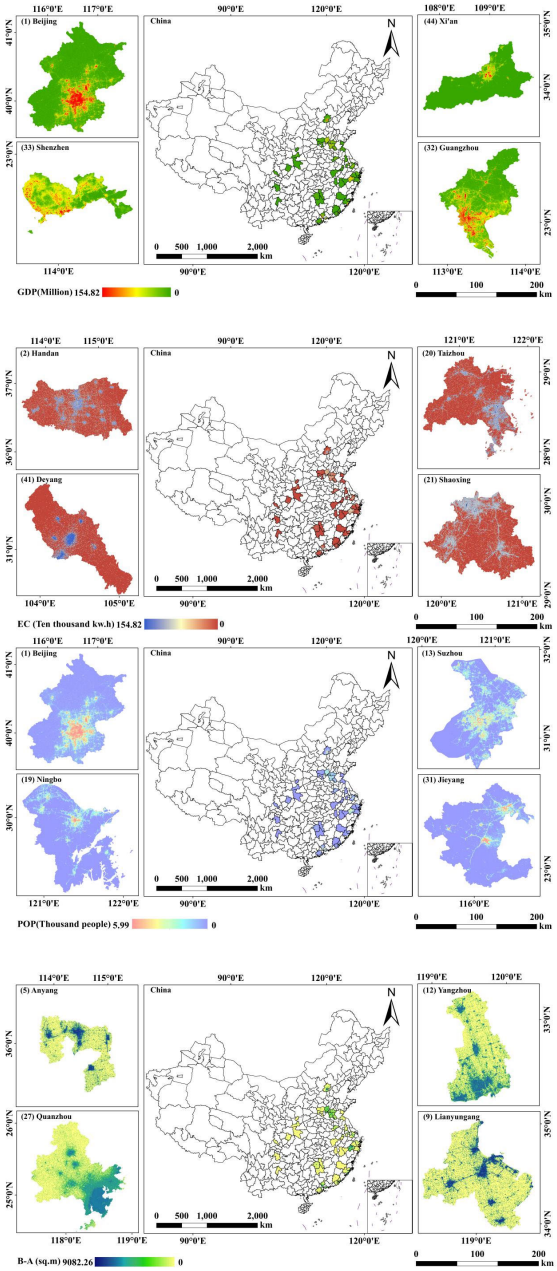


FIGURE 7. The maps with 100m resolution of four socio-economic parameters including GDP, POP, EC, and B-A for typical cities of China in 2018. Note: the maps of socio-economic parameters for other cities of China were presented in the Appendix section FIGURE 9 due to the limited space.

GDP of Longhua New District and Longgang District are also higher than adjacent regions. Meanwhile, the distribution characteristic of GDP for Guangzhou, Guangdong Province is that the concentrated GDP areas are mainly located in Old Seven District including the Yuexiu, Haizhu, Liwan, Tianhe, Baiyun, Huangpu, and Nansha District.

The distribution of EC for Handan, Hebei Province exhibited an obvious pattern that the most concentrated EC areas were mainly distributed in urban areas and the towns of

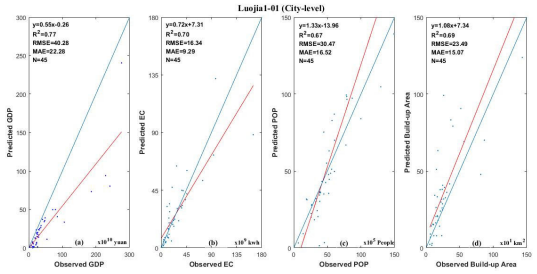


FIGURE 8. Scatter plots of observed and predicted values for GDP (a), EC (b), POP (c), and B-A (d) via GWR and RF models using the Luojia1-01 NTL dataset at the city scale of China in 2018. Note: The precision evaluation for each map of the socio-economic parameter is based on the actual statistical yearbook and predicted value obtained from the calibration models.

local governments. Also, the distribution of EC for Deyang, Sichuan Province demonstrated significant spatial features, that is, the concentrated EC areas are mainly located in Guanghan City and Jingyang District. The hotspots areas of EC for Taizhou City, Zhejiang Province are along the East China Sea. The hotspots areas of EC for Shaoxing City, Zhejiang Province are mainly distributed in the Heqiao and Yuexiu District and the towns of local governments.

The distribution feature of POP for Beijing was similar to GDP and the detailed description will not be stated in this section. The concentrated POP areas for Ningbo, Zhejiang Province are mainly located in the main urban areas, Cixi City, Ninghai County, and Yutao City. The hotspots areas of POP for Suzhou City, Jiangsu Province are located in Xiangcheng District, Kunshan District. The hotspots areas of POP for Jieyang City, Guangdong Province are mainly distributed in Puning City, Jiedong City, and Rongcheng City.

The B-A for Anyang, Henan Province is mainly located in the main urban areas, Linzhou City, Tangyin County, Neihuang City, and Hua County. The B-A for Quanzhou City, Fujian Province exhibits a decreasing trend from the southeast to the northwest because the main urban area is close to the South China Sea. The B-A for Yangzhou City, Jiangsu Province is mainly located in Hanjiang District, Yizheng City, and the majority area of Guangling District. The B-A for Lianyungang City, Jiangsu Province is mainly distributed in the Lianyun District, Xinpu District, and Haizhou District.

V. DISCUSSION

A. EVALUATING THE PRECISION OF THE MAPS FOR SOCIO-ECONOMIC PARAMETERS USING CROSS-VALIDATION METHOD BASED ON THE ACTUAL STATISTICAL DATASET AT THE CITY SCALE.

The precision of the 100m resolution maps for socio-economic parameters including GDP, POP, EC, and B-A was assessed based on the actual statistical dataset from local authorities at the city scale (TABLE 4 and FIGURE 8).

Although the precision of the GDP map was acceptable (R^2 =0.772, RMSE = 40.28, MAE = 22.28), the majority of cities' GDP except for Deyang City, Sichuan Province

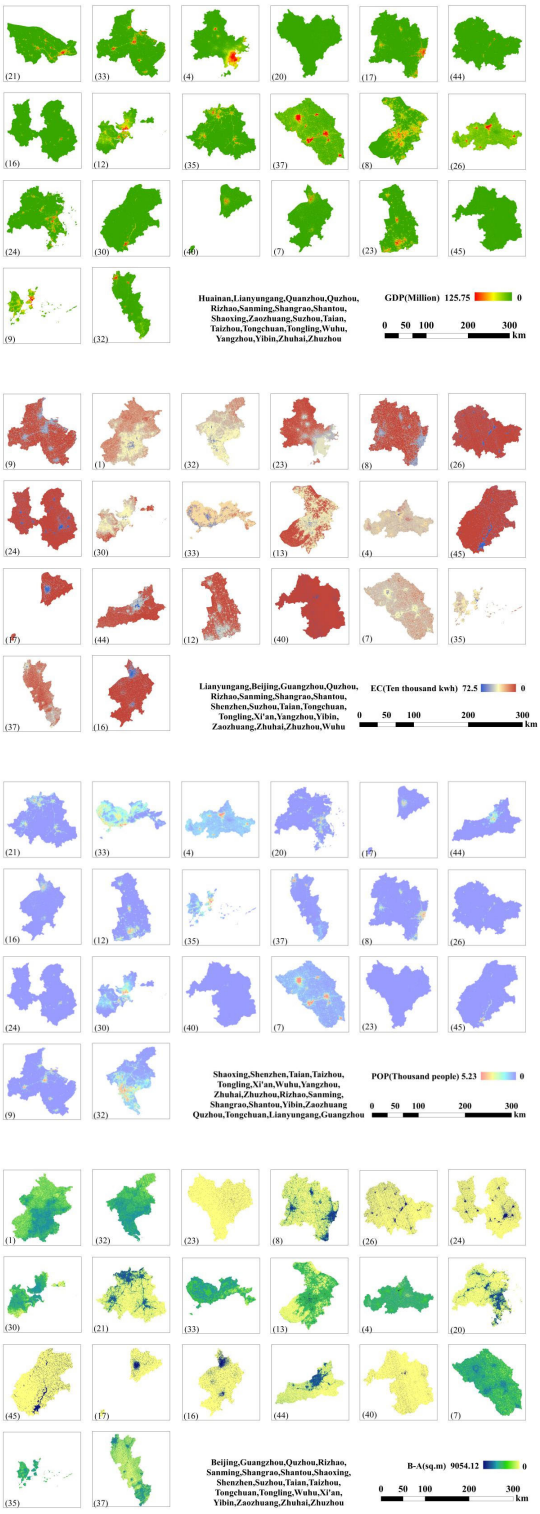


FIGURE 9. The maps with 100m resolution of four socio-economic parameters including GDP, POP, EC, and B-A for typical cities of China in 2018. Note: the number in Figure 9 denotes the code of each city on the map, and the name of each city was presented in Table 2. Figure 9 was presented here due to the limited space in the section results.

was underestimated (Figure 8(a)). The NTL has been proved that can be used to estimate socio-economic parameters in many previous studies. Whereas, there are still some existing

economic activities that have been neglected by NTL such as agriculture in rural areas, the industrial production without any lights at night. The capability of NTL for detecting primary industry and secondary industry was relatively weak. So, the predicted GDP from the current study was relatively lower than the actual statistical value.

Overall, the precision of the EC map was relatively reliable (R^2 =0.7, RMSE= 16.34, MAE = 9.29). However, Figure 8(b) demonstrated that thirteen cities' predicted EC was higher than the real value from the local government, and predicted EC for thirty-two cities was underestimated. For the overestimated situation, we speculated that some possible noise such as moonshine, the cloud may increase the unnecessary brightness of NTL images that may lead to overestimating the EC [54]. The most serious overestimated areas for EC are mainly distributed in the southeast of China close to the Sea such as Jiangmen that is adjacent to the South China Sea, Jinhua, Quanzhou. Contrarily, the EC was underestimated because some electric consumption can hardly be detected by nighttime light remotely sensed methods such as the secondary industry.

Meanwhile, the majority of cities' POP except for Beijing, Hengyang, Jieyang, Jining, Ningbo, Quanzhou, Suzhou, Taian, Xi'an, Zhuhai, and Zhuzhou were underestimated because the related dots were under the 1:1 line in Figure 8(c). The population data of the current study were obtained from the Statistical Yearbook of the local governments. As we know, the real population distribution exhibits spatial flow characteristics, that is, people may work or live in another place that the addresses are not the registered ones in the household register [72]. So the predicted POP of the present study showed some unavoidable errors. Lots of people from the current study areas entered into metropolitan such as Beijing, Suzhou, Xi'an for work. So, the actual population for places of outputting labors such as Deyang, Lianyungang, Putian, and Rizhao were lower than the number of population in the Statistical Yearbook, and the population for places of inputting labors was higher than the number of population in the Statistical Yearbook.

For the calibrated results of B-A, the R^2, RMSE, and MAE values for evaluating the simulation were 0.687, 23.49, and 15.07, respectively. The simulated B-A value for the majority of regions was overestimated except for nine places especially Guangzhou, Shenzhen, Wuhu, and Xi'an that the B-A was severely underestimated. The metropolitan area is always surrounded by suburbs in China. Although the suburban areas look like the urban areas at night through NTL images because the night light remotely sensed can hardly differentiate the suburb from the urban areas just depending on the night light. The most difference between suburban and urban is not only the building and roads but also the basic infrastructure such as water supply and drainage, natural gas supply. So, the suburban areas may be wrongly recognized as B-A by NTL in the current study. The overestimated areas for B-A are mainly located in Beijing, Hengshui, Ningbo, Quanzhou, Suzhou, Taian, Yangzhou, Zaozhuang, and Zhuzhou.

Obviously, the distribution of the socio-economic parameters demonstrated spatial heterogeneity owing to the different development levels. On the one hand, the POP, GDP, EC, and B-A exhibited higher values in the eastern and southern regions of China than in other areas because the majority of industries and population of China were accumulated in the eastern and southern regions of China. Besides, the regional advantage of eastern and southern China outperforms other regions due to the superior physical condition and adjacent to the ocean. On the other hand, as we know, the reform and opening-up policy of China was originally carried out in the eastern and southern regions of China in 1978. The policy had let the uneven distribution of socio-economic development.

B. ASSESSING THE CAPABILITY FOR LUOJIA1-01 AND NPP-VIIRS NTL DATASET IN SIMULATING SOCIO-ECONOMIC PARAMETERS

Obviously, the results from the precision evaluation demonstrated that the Luojia1-01 outperformed the NPP-VIIRS images in calibrating socio-economic parameters (Figure 5, Table 5, and Table 6). The possible reasons were analyzed below. First, the calibrated precision may be influenced by the resolution of the NTL images, and the spatial resolution of LuoJia1-01 images is better than the VIIRS images. Identifying the spatial heterogeneity of anthropic activities may benefit from the detailed resolution of NTL images due to a smaller overflow. Meanwhile, the actual area that may be detected accurately for the landscape patches can be easily identified by the higher-resolution NTL images. Furthermore, higher spatial resolution NTL images can help people to identify anthropic activities at a smaller scale and to detect more specific information about human activities. Second, as we know, human activities always occurred during 19:00 - 24:00 of the local time. Whereas, the overpass time of the NPP-VIIRS is about at 1:30 local time. So, some human activities such as the night market, shopping at night, and the igniting lights in tourism sites may be missed by the NPP-VIIRS, which may result in the precision of simulating socio-economic parameters are lower than the Luojia1-01 because the overpass time of Luojia1-01 is about at 22:30 local time that is in line with the local time of human activities. The light in the city tends to be brighter during 19:00 - 24:00 of local time than other times. Although the brightness of the light can be a benefit for extracting useful information from NTL, an unnecessary overflow generated some unuseful noise that increased bias errors for identifying socio-economic parameters. Furthermore, sunshine may not completely disappear at the overpass time of the Luojia1-01 in some western regions of China in summer such as Xinjiang Autonomous Region. So, some unavoidable noise from the Sun may produce negative effects for extracting socio-economic parameters.

Overall, though there are still some bias errors existing in extracting socio-economic parameters by Luojia1-01, the performance of Luojia1-01 is better than NPP-VIIRS.

C. COMPARING THE ROBUSTNESS FOR DIFFERENT MODELS IN FIXING SOCIO-ECONOMIC PARAMETERS

Clearly, the optimum model for calibrating GDP and POP based on the NTL dataset is the GWR (Figure 5, Table 5, and Table 6). The possible reason was analyzed below. The distribution of GDP and POP exhibits obvious spatial heterogeneity across China that the areas for dense population and economic activities of China are mainly accumulated in the southeast of China. The GDP and POP demonstrated significant spatial autocorrelation. The GWR model not only can determine the relations between dependent and explanatory variables but also can consider spatial heterogeneity of variables. So, it is no doubt that the performance of GWR outperforms other models in calibrating GDP and POP in the present study. For the EC and B-A, the optimal calibration model is RF. Random forest (RF) has been widely used in dealing with nonlinear relationships owing to its less sensitivity to noise and overfitting [65]. Also, RF does not need assumptions of a prior probability distribution, which is very suitable for complex simulations. The current study evaluated the outcome of the traditional regression model and RF model and found that the RF model was better than the others in simulating EC and B-A. This may be attributed to the multicollinearity of independent variables and the diversity of towns in study areas. The possible reasons were discussed below. The relationship between EC as well as B-A and NTL was very complicated, which can hardly be efficiently handled by the traditional regression model. Contrarily, the RF model is good at tackling multicollinear issues. So, the robustness of RF is more reliable than the regression model in simulating complex issues.

D. VERIFYING THE FEASIBILITY FOR DIFFERENT MODELS IN FIXING SOCIO-ECONOMIC PARAMETERS AT DIFFERENT SCALES

The results demonstrated that the precision of estimating the socio-economic parameters at the city scale was better than the county scale. The possible reasons were analyzed below. The structure of the industry was relatively complete and diverse at the city scale than the county scale, and the NTL dataset was suitable for retrieving socio-economic phenomena at the city scale. For example, some counties may have higher GDP owing to agriculture production, not service industries. The NTL remote sensing technology is different to detect agriculture production at night. However, the distribution of economic activities is relatively complete at the city scale. So, the accuracy for estimating GDP at the county scale is likely less than the city scale. Furthermore, it is a universal phenomenon that people born in the small county always entered into megacity for living in China during the rapid urbanization and industrialization process. The mobility of the population may much more significant at the county scale than the city scale because the population at the city scale is always larger than the county scale. Moreover, if the population flow occurred inner a city such as from a developed county to a developing county. The number of the population did not change at the city scale but varied dramatically at the county scale. Additionally, the brightness and extent of the NTL were always brighter and larger at the city scale than the county scale. There may exist plenty of feeble lights at night in a county but these lights were different to be detected through nightlight sensors. So, the calibration accuracy was influenced significantly at the county scale. On the contrary, feeble lights can be neglected for the impacts are relatively less at the city scale. So, the simulation precision of the population based on the NTL dataset at the city scale outperformed the county scale.

E. THE LIMITATION AND PLAN OF THE CURRENT STUDY

Though the current paper proved that Luojia1-01 can be implemented to retrieve the socio-economic parameters, there still exist some problems that need to be further addressed. Firstly, ''white'' LEDs that generate lots of radiation ranged from 450 nm to 480 nm are widely selected for igniting the street lights in China [71]. Whereas, parts of the radiation from the LEDs were not detected through the Luojia1-01 because the range of spectral for Luojia1-01 only from 480 μm to 800 μm [21]. This issue is still a challenge and will be tackled if the sensors of Luojia1-01 can be adjusted in the future. Secondly, the above results revealed that the accuracy of estimating socio-economic parameters by the NTL dataset was influenced by clouds and moonshine. Thirdly, although the synthetic Luojia1-01 NTL dataset of 2018 across China have been released in June 2019 by Wuhan University, China, the span of Luojia1-01 was scarce at spatial as well as the time scale. So, the Luojia1-01 dataset with larger spatial and longer time cover is strongly desired. Moreover, the availability of the Luojia1-01 seriously hampered the feasibility of calibrating socio-economic parameters on a large scale and long time series for the Luojia1-01 images that were missing during some periods in parts location of China. Fourthly, the population dataset used in the present study was the statistical data from the local government that was hard to illustrate the actual distribution of the population. Also, the electric consumption of the secondary industry was not be deleted which led to the outcome of the simulated EC was underestimated. Meanwhile, the GDP account for agriculture and industries was not be eliminated that the results of the calibrated GDP were also underestimated. Last but not the least, the present research was a local study, not a global finding. So, further data preprocessing concerning noise elimination including the deduction of the clouds and moonshine needs to be addressed. Additionally, the related studies on calibrating socio-economic parameters through the NTL dataset should be conducted on a global scale and in long time series. The novel dataset described actual human activities such as POI, Weibo sign-in data, and hotspots data from social media should be introduced to future study.

VI. CONCLUSION

This study estimated the socio-economic parameters based on multiple models using statistical data, NPP-VIIRS, and,

Luojia1-01 NTL data in 45 cities of China in 2018. Meanwhile, the ten-folded cross-validation was introduced to exam the precision of the LR, GWR, RF, BPNN, and SVM models on multiple scales. Moreover, the distribution of the socio-economic parameters was mapped using the optimum model in 2018. Finally, the evaluation of different models, scales, and NTL datasets was carried out for determining the potential feasibility of the present study. Some interesting results were achieved. The NTL remotely sensed images can be used to estimate the socio-economic parameters. The present study demonstrated that the accuracy at the city-scale was better than the county or district scale across China in 2018. The capability of NTL remotely sensed images in detecting primary industry, secondary industry, population flow, detailed urban features such as infrastructure was relatively weak. Overall, except for some special situations, the Luojia1-01 NTL dataset outperformed NPP-VIIRS NTL dataset in calibrating socio-economic parameters at the same scale because of high spatial and spectral resolution and the suitable overpass time. The GWR is good at simulating parameters with spatiotemporal features such as GDP and POP. The current study showed that the RF model was better than the others in simulating EC and B-A due to the efficiency of the RF in tackling multicollinear issues. This paper supplies a novel perspective on detecting socio-economic parameters with speedy and economic nighttime light remotely sensed images.

APPENDIX

See Fig. 9 and Tables 5 and 6.

ACKNOWLEDGMENT

(Bin Guo and Yi Bian contributed equally to this work.)

REFERENCES

[1] L. Deren and L. Xi, "An overview on data mining of nighttime light remote sensing," *Acta Geodaetica ET Cartographica Sinica*, vol. 44, pp. 591–601, Jun. 2015.

[2] D. Feldmeyer, C. Meisch, H. Sauter, and J. Birkmann, "Using OpenStreetMap data and machine learning to generate socio-economic indicators," *ISPRS Int. J. Geo-Inf.*, vol. 9, no. 9, p. 498, Aug. 2020.

[3] B. Guo, Y. Su, L. Pei, X. Wang, B. Zhang, D. Zhang, and X. Wang, "Ecological risk evaluation and source apportionment of heavy metals in park playgrounds: A case study in Xi'an, Shaanxi province, a northwest city of China," *Environ. Sci. Pollut. Res.*, vol. 27, no. 19, pp. 24400–24412, Jul. 2020.

[4] B. Guo, X. Wang, D. Zhang, L. Pei, D. Zhang, and X. Wang, "A land use regression application into simulating spatial distribution characteristics of particulate matter ($PM_{2.5}$) concentration in city of Xi'an, China," *Polish J. Environ. Stud.*, vol. 29, no. 6, pp. 4065–4076, Aug. 2020.

[5] B. Guo, Y. Su, L. Pei, X. Wang, X. Wei, B. Zhang, D. Zhang, and X. Wang, "Contamination, distribution and health risk assessment of risk elements in topsoil for amusement parks in Xi'an, China," *Polish J. Environ. Stud.*, vol. 30, pp. 601–617, Nov. 2020.

[6] C. D. Elvidge, B. Zhang, D. Zhang, and X. Wang, "Estimating population and energy consumption in Brazilian Amazonia using DMSP night-time satellite data," *Comput. Environ. Urban Syst.*, vol. 29, no. 2, pp. 179–195, 2005.

[7] C. D. Elvidge, K. E. Baugh, E. A. Kihn, H. W. Kroehl, E. R. Davis, and C. W. Davis, "Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption," *Int. J. Remote Sens.*, vol. 18, no. 6, pp. 1373–1379, Apr. 1997.

[8] X. Chen and W. D. Nordhaus, "Using luminosity data as a proxy for economic statistics," *Proc. Nat. Acad. Sci. USA*, vol. 108, no. 21, pp. 8589–8594, May 2011.

[9] J. Wu, Z. Wang, W. Li, and J. Peng, "Exploring factors affecting the relationship between light consumption and GDP based on DMSP/OLS nighttime satellite imagery," *Remote Sens. Environ.*, vol. 134, pp. 111–119, Jul. 2013.

[10] K. Shi, B. Yu, Y. Huang, Y. Hu, B. Yin, Z. Chen, L. Chen, and J. Wu, "Evaluating the ability of NPP-VIIRS nighttime light data to estimate the gross domestic product and the electric power consumption of China at multiple scales: A comparison with DMSP-OLS data," *Remote Sens.*, vol. 6, no. 2, pp. 1705–1724, Feb. 2014.

[11] H. Liang, Z. Guo, J. Wu, and Z. Chen, "GDP spatialization in ningbo city based on NPP/VIIRS night-time light and auxiliary data using random forest regression," *Adv. Space Res.*, vol. 65, no. 1, pp. 481–493, Jan. 2020.

[12] A. C. Townsend and D. A. Bruce, "The use of night-time lights satellite imagery as a measure of Australia's regional electricity consumption and population distribution," *Int. J. Remote Sens.*, vol. 31, no. 16, pp. 4459–4480, 2010.

[13] Z. Xiaolin, L. I. Qiang, S. MiaoGen, J. Chen, and J. Wu, "A methodology for multiple cropping index extraction based on NDVI time-series," *J. Natural Resour.*, vol. 23, no. 3, pp. 534–544, 2008.

[14] F. Falchi, P. Cinzano, C. D. Elvidge, D. M. Keith, and A. Haim, "Limiting the impact of light pollution on human health, environment and stellar visibility," *J. Environ. Manage.*, vol. 92, no. 10, pp. 2714–2722, Oct. 2011.

[15] K. J. Gaston, J. Bennie, T. W. Davies, and J. Hopkins, "The ecological impacts of nighttime light pollution: A mechanistic appraisal," *Biol. Rev.*, vol. 88, no. 4, pp. 912–927, Nov. 2013.

[16] C. D. Elvidge, K. E. Baugh, E. A. Kihn, H. W. Kroehl, E. R. Davis, "Mapping city lights with nighttime data from the DMSP operational Linescan system," *Photogramm. Eng. Remote Sens.*, vol. 63, no. 6, pp. 727–734, 1997.

[17] X. Li and D. Li, "Can night-time light images play a role in evaluating the syrian crisis?" *Int. J. Remote Sens.*, vol. 35, no. 18, pp. 6648–6661, Sep. 2014.

[18] X. Li, S. Liu, M. Jendryke, D. Li, and C. Wu, "Night-time light dynamics during the iraqi civil war," *Remote Sens.*, vol. 10, no. 6, p. 858, Jun. 2018.

[19] Liu, Du, Yi, Liang, Ma, and Pei, "Quantitative association between nighttime lights and geo-tagged human activity dynamics during typhoon mangkhut," *Remote Sens.*, vol. 11, no. 18, p. 2091, Sep. 2019.

[20] Y. Zheng, G. Shao, L. Tang, Y. He, X. Wang, Y. Wang, and H. Wang, "Rapid assessment of a typhoon disaster based on NPP-VIIRS DNB daily data: The case of an urban agglomeration along western taiwan straits, China," *Remote Sens.*, vol. 11, no. 14, p. 1709, Jul. 2019.

[21] W. Jiang, G. He, T. Long, H. Guo, R. Yin, W. Leng, H. Liu, and G. Wang, "Potentiality of using luojia 1-01 nighttime light imagery to investigate artificial light pollution," *Sensors*, vol. 18, no. 9, p. 2900, Sep. 2018.

[22] Z. L. Fleming, R. M. Doherty, E. von Schneidemesser, C. S. Malley, O. R. Cooper, J. P. Pinto, A. Colette, X. Xu, D. Simpson, M. G. Schultz, A. S. Lefohn, S. Hamad, R. Moolla, S. Solberg, and Z. Feng, "Tropospheric ozone assessment report: Present-day ozone distribution and trends relevant to human health," *Elementa, Sci. Anthropocene*, vol. 6, pp. 1–42, Jan. 2018.

[23] X. Li, L. Zhao, D. Li, and H. Xu, "Mapping urban extent using luojia 1-01 nighttime light imagery," *Sensors*, vol. 18, no. 11, p. 3665, Oct. 2018.

[24] X. Hu, Y. Qian, S. T. A. Pickett, and W. Zhou, "Urban mapping needs up-to-date approaches to provide diverse perspectives of current urbanization: A novel attempt to map urban areas with nighttime light data," *Landscape Urban Planning*, vol. 195, Mar. 2020, Art. no. 103709.

[25] X. Chen, X. Jia, and M. Pickering, "A nighttime lights adjusted impervious surface index (NAISI) with integration of landsat imagery and nighttime lights data from international space station," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 83, Nov. 2019, Art. no. 101889.

[26] S. Liu, X. Li, N. Levin, and M. Jendryke, "Tracing cultural festival patterns using time-series of VIIRS monthly products," *Remote Sens. Lett.*, vol. 10, no. 12, pp. 1172–1181, Dec. 2019.

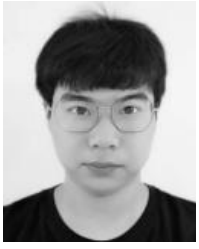
[27] K. J. Gaston, J. P. Duffy, and J. Bennie, "Quantifying the erosion of natural darkness in the global protected area system," *Conservation Biol.*, vol. 29, no. 4, pp. 1132–1141, Aug. 2015.

[28] C. Yang, Z. Jing, Y. Mingxiang, G. Binbin, L. Mei, Y. Liu, Q. Jiashen, and G. Peng, "Analysis of lighting changes in the tourist city Edogawa using nighttime light data," *J. Indian Soc. Remote Sens.*, vol. 46, no. 10, pp. 1617–1623, 2018.

[29] B. Devkota, H. Miyazaki, A. Witayangkurn, and S. M. Kim, "Using volunteered geographic information and nighttime light remote sensing data to identify tourism areas of interest," *Sustainability*, vol. 11, no. 17, p. 4718, Aug. 2019.
[30] E. C. Stokes and K. C. Seto, "Characterizing urban infrastructural transitions for the sustainable development goals using multi-temporal land, population, and nighttime light data," *Remote Sens. Environ.*, vol. 234, Dec. 2019, Art. no. 111430.
[31] H. Zhou, L. Liu, M. Lan, B. Yang, and Z. Wang, "Assessing the impact of nightlight gradients on street robbery and burglary in cincinnati of ohio state, USA," *Remote Sens.*, vol. 11, no. 17, p. 1958, Aug. 2019.
[32] T. Ye, N. Zhao, X. Yang, Z. Ouyang, X. Liu, Q. Chen, K. Hu, W. Yue, J. Qi, Z. Li, and P. Jia, "Improved population mapping for China using remotely sensed and points-of-interest data within a random forests model," *Sci. Total Environ.*, vol. 658, pp. 936–946, Mar. 2019.
[33] N. Levin and Y. Duke, "High spatial resolution night-time light images for demographic and socio-economic studies," *Remote Sens. Environ.*, vol. 119, pp. 1–10, Apr. 2012.
[34] C. P. Lo, "Urban indicators of China from radiance-calibrated digital DMSP-OLS nighttime images," *Ann. Assoc. Amer. Geographers*, vol. 92, no. 2, pp. 225–240, Jun. 2002.
[35] N. Zhao, Y. Liu, G. Cao, E. L. Samson, and J. Zhang, "Forecasting China's GDP at the pixel level using nighttime lights time series and population images," *GISci. Remote Sens.*, vol. 54, no. 3, pp. 407–425, May 2017.
[36] B. Yu, K. Shi, Y. Hu, C. Huang, Z. Chen, and J. Wu, "Poverty evaluation using NPP-VIIRS nighttime light composite data at the county level in China," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 3, pp. 1217–1219, Mar. 2015.
[37] K. Shi, Z. Chang, Z. Chen, J. Wu, and B. Yu, "Identifying and evaluating poverty using multisource remote sensing and point of interest (POI) data: A case study of Chongqing, China," *J. Cleaner Prod.*, vol. 255, May 2020, Art. no. 120245.
[38] Z. Chen, B. Yu, Y. Hu, C. Huang, K. Shi, and J. Wu, "Estimating house vacancy rate in metropolitan areas using NPP-VIIRS nighttime light composite data," *IEEE J.-Stars.*, vol. 8, no. 5, pp. 2188–2197, May 2015.
[39] A. Ingacheva and V. Kokhan, "Mapping China's ghost cities through the combination of nighttime satellite data and daytime satellite data," in *Proc. 32nd Eur. Conf. Modeling Simulation (ECMS)*, 2018, p. 1073.
[40] X. Liu, J. Ou, S. Wang, X. Li, Y. Yan, L. Jiao, and Y. Liu, "Estimating spatiotemporal variations of city-level energy-related CO2 emissions: An improved disaggregating model based on vegetation adjusted nighttime light data," *J. Cleaner Prod.*, vol. 177, pp. 101–114, Mar. 2018.
[41] G. Falchetta and M. Noussan, "Interannual variation in night-time light radiance predicts changes in national electricity consumption conditional on income-level and region," *Energies*, vol. 12, no. 3, p. 456, Jan. 2019.
[42] L. Lu, Q. Weng, Y. Xie, H. Guo, and Q. Li, "An assessment of global electric power consumption using the defense meteorological satellite program-operational linescan system nighttime light imagery," *Energy*, vol. 189, Dec. 2019, Art. no. 116351.
[43] C. D. Elvidge, P. Cinzano, D. R. Pettit, J. Arvesen, P. Sutton, C. Small, R. Nemani, T. Longcore, C. Rich, and J. Safran, "The nights at mission concept," *Int. J. Remote Sens.*, vol. 28, pp. 2645–2670, Jun. 2007.
[44] N. A. Rybnikova and B. A. Portnov, "Mapping geographical concentrations of economic activities in europe using light at night (LAN) satellite data," *Int. J. Remote Sens.*, vol. 35, no. 22, pp. 7706–7725, Nov. 2014.
[45] C. Li, G. Chen, J. Luo, S. Li, and J. Ye, "Port economics comprehensive scores for major cities in the yangtze valley, China using the DMSP-OLS night-time light imagery," *Int. J. Remote Sens.*, vol. 38, no. 21, pp. 6007–6029, Nov. 2017.
[46] C. D. Elvidge, J. Safran, B. Tuttle, P. Sutton, P. Cinzano, D. Pettit, J. Arvesen, and C. Small, "Potential for global mapping of development via a nightsat mission," *GeoJournal*, vol. 69, nos. 1–2, pp. 45–53, Oct. 2007.
[47] R. Paranunzio, S. Ceola, F. Laio, and A. Montanari, "Evaluating the effects of urbanization evolution on air temperature trends using nightlight satellite data," *Atmosphere*, vol. 10, no. 3, p. 117, Mar. 2019.
[48] K. Shi, Y. Chen, B. Yu, T. Xu, C. Yang, L. Li, C. Huang, Z. Chen, R. Liu, and J. Wu, "Detecting spatiotemporal dynamics of global electric power consumption using DMSP-OLS nighttime stable light data," *Appl. Energy*, vol. 184, pp. 450–463, Dec. 2016.
[49] N. Zhao, T. Ghosh, and E. L. Samson, "Mapping Spatio-temporal changes of Chinese electric power consumption using night-time imagery," *Int. J. Remote Sens.*, vol. 33, no. 20, pp. 6304–6320, 2012.
[50] X. Wu, J. Wen, Q. Xiao, Y. Yu, D. You, and A. Hueni, "Assessment of NPP VIIRS albedo over heterogeneous crop land in northern China," *J. Geophys. Res., Atmos.*, vol. 122, no. 24, pp. 13,138–13,154, Dec. 2017.
[51] C. D. Elvidge, F.-C. Hsu, M. Zhizhin, T. Ghosh, J. Taneja, and M. Bazilian, "Indicators of electric power instability from satellite observed nighttime lights," *Remote Sens.*, vol. 12, no. 19, p. 3194, Sep. 2020.
[52] X. Wang, M. Rafa, J. Moyer, J. Li, J. Scheer, and P. Sutton, "Estimation and mapping of sub-national GDP in uganda using NPP-VIIRS imagery," *Remote Sens.*, vol. 11, no. 2, p. 163, Jan. 2019.
[53] Z. Chen, B. Yu, W. Song, H. Liu, and J. Wu, "A new approach for detecting urban centers and their spatial structure with nighttime light remote sensing," *IEEE Trans. Geosci. Remote Sensing.*, vol. 55, no. 11, pp. 6305–6319, Nov. 2017.
[54] C. Wang, Z. Chen, C. Yang, Q. Li, Q. Wu, J. Wu, G. Zhang, and B. Yu, "Analyzing parcel-level relationships between luojia 1-01 nighttime light intensity and artificial surface features across Shanghai, China: A comparison with NPP-VIIRS data," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 85, Mar. 2020, Art. no. 101989.
[55] Wang, Zhong, and Su, "On-orbit signal-to-noise ratio test method for night-light camera in luojia 1-01 satellite based on time-sequence imagery," *Sensors*, vol. 19, no. 19, p. 4077, Sep. 2019.
[56] Li, Liu, Chen, and Sun, "Assessing the ability of luojia 1-01 imagery to detect feeble nighttime lights," *Sensors*, vol. 19, no. 17, p. 3708, Aug. 2019.
[57] Bu, Xu, Zhang, and Zhang, "Night-light image restoration method based on night scattering model for luojia 1-01 satellite," *Sensors*, vol. 19, no. 17, p. 3761, Aug. 2019.
[58] Y. Xu, A. Knudby, and H. C. Ho, "Estimating daily maximum air temperature from MODIS in British Columbia, Canada," *Int. J. Remote Sens.*, vol. 35, no. 24, pp. 8108–8121, 2014.
[59] L. Sun, L. Tang, G. Shao, Q. Qiu, T. Lan, and J. Shao, "A machine learning-based classification system for urban built-up areas using multiple classifiers and data sources," *Remote Sens.*, vol. 12, no. 1, p. 91, Dec. 2019.
[60] J. Holloway and K. Mengersen, "Statistical machine learning methods and remote sensing for sustainable development goals: A review," *Remote Sens.*, vol. 10, no. 9, p. 1365, Aug. 2018.
[61] B. Guo, D. Zhang, D. Zhang, Y. Su, X. Wang, and Y. Bian, "Detecting spatiotemporal dynamic of regional electric consumption using NPP-VIIRS nighttime stable light data—A case study of Xi'an, China," *IEEE Access*, vol. 8, pp. 171694–171702, 2020.
[62] J. Liu, Y. Deng, Y. Wang, H. Huang, Q. Du, and F. Ren, "Urban nighttime leisure space mapping with nighttime light images and POI data," *Remote Sens.*, vol. 12, no. 3, p. 541, Feb. 2020.
[63] B. Guo, X. Wang, L. Pei, Y. Su, D. Zhang, and Y. Wang, "Identifying the spatiotemporal dynamic of $PM_{2.5}$ concentrations at multiple scales using geographically and temporally weighted regression model across China during 2015–2018," *Sci. Total Environ.*, vol. 751, Jan. 2021, Art. no. 141765.
[64] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, Oct. 2001.
[65] S. Chen, M. P. Martin, N. P. A. Saby, C. Walter, D. A. Angers, and D. Arrouays, "Fine resolution map of top- and subsoil carbon sequestration potential in France," *Sci. Total Environ.*, vol. 630, pp. 389–400, Jul. 2018.
[66] R. Qiu, Y. Wang, D. Wang, W. Qiu, J. Wu, and Y. Tao, "Water temperature forecasting based on modified artificial neural network methods: Two cases of the yangtze river," *Sci. Total Environ.*, vol. 737, Oct. 2020, Art. no. 139729.
[67] Y. Ju, M. Kim, and J. Shin, "Detection of malicious code using the direct hashing and pruning and support vector machine," *Concurrency Comput., Pract. Exper.*, vol. 32, no. 18, Sep. 2020, Art. no. e5483.
[68] J. Wei, Z. Li, M. Cribb, W. Huang, W. Xue, L. Sun, J. Guo, Y. Peng, J. Li, A. Lyapustin, L. Liu, H. Wu, and Y. Song, "Improved 1 km resolution $PM_{2.5}$ estimates across China using enhanced space–time extremely randomized trees," *Atmos. Chem. Phys.*, vol. 20, no. 6, pp. 3273–3289, Mar. 2020.
[69] P. Luo, Y. Sun, S. Wang, S. Wang, J. Lyu, M. Zhou, K. Nakagami, K. Takara, and D. Nover, "Historical assessment and future sustainability challenges of Egyptian water resources management," *J. Cleaner Prod.*, vol. 263, Aug. 2020, Art. no. 121154.
[70] T. Kuester, A. S. De Miguel, K. Baugh, A. Jechow, F. Hölker, J. Bennie, C. D. Elvidge, K. J. Gaston, and L. Guanter, "Artificially lit surface of Earth at night increasing in radiance and extent," *Sci. Adv.*, vol. 3, no. 11, 2017, Art. no. e1701528.

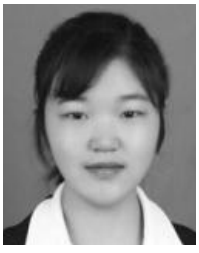
[71] F. J. Song, D. Q. Wang, and S. Q. Tian, "Governance mechanism of rural population hollowing under the background of balancing urban and rural development," *Adv. Mater. Res.*, vol. 989, pp. 5128–5131, Jul. 014.

BIN GUO received the Ph.D. degree in cartography and geographic information system from the School of Tourism and Environment, Shaanxi Normal University, Xi'an, China, in 2011. He is currently an Associate Professor with the College of Geomatics, Xi'an University of Science and Technology. His research interests include the application of nighttime light remote sensing, spatial analysis and modeling, and air and soil pollutants mapping.



YI BIAN received the bachelor's degree in surveying and mapping engineering from the Department of Geomatics, Shanxi Institute of Technology, Yangquan, China, in 2019. He is currently pursuing the master's degree with the College of Geomatics, Xi'an University of Science and Technology. His research interest is the application of nighttime light remote sensing.

DINGMING ZHANG received the bachelor's degree in geographic information science from the College of Geomatics, Xi'an University of Science and Technology, Xi'an, China, in 2018, where he is currently pursuing the master's degree. His research interest is the application of nighttime light remote sensing.



YI SU received the bachelor's degree in surveying and mapping engineering from the College of Geography and Environment, Baoji University of Arts and Sciences, Baoji, China, in 2018. She is currently pursuing the master's degree with the College of Geomatics, Xi'an University of Science and Technology. Her research interest is soil pollutants mapping.

XIAOXIA WANG received the bachelor's degree in surveying and mapping engineering from the College of Geography and Environment, Baoji University of Arts and Sciences, Baoji, China, in 2018. She is currently pursuing the master's degree with the College of Geomatics, Xi'an University of Science and Technology. Her research interest is air pollutants modeling and mapping.

BO ZHANG received the bachelor's degree in geographic information science from the College of Geomatics, Xi'an University of Science and Technology, Xi'an, China, in 2019, where he is currently pursuing the master's degree. His research interest is the application of hyperspectral remote sensing.

YAN WANG received the bachelor's degree in surveying and mapping engineering from the College of Geography and Environment, Baoji University of Arts and Sciences, Baoji, China, in 2019. She is currently pursuing the master's degree with the College of Geomatics, Xi'an University of Science and Technology. Her research interest is air pollutants and their effect on public health.

QIUJI CHEN received the Ph.D. degree in surveying and mapping from the School of Environment Science and Spatial Informatics, China University of Mining and Technology, Beijing, China, in 2005. He is currently a Full Professor with the College of Geomatics, Xi'an University of Science and Technology. His research interests include land reclamation and GIS application.

YARUI WU received the Ph.D. degree in environmental science from the School of Water and Environment, Chang'an University, Xi'an, China, in 2011. She is currently an Associate Professor with the College of Geomatics, Xi'an University of Science and Technology. Her research interest is environmental engineering.

PINGPING LUO received the Ph.D. degree in urban environmental engineering from the Graduate School of Engineering, Kyoto University, Japan, in 2012. He is currently a Professor with the School of Chang'an University. His research interests include hydrology and water resources, water conservancy and hydropower, urban flooding, soil and water conservation, environmental engineering, sustainable water environment, and geological disasters.