

Global Industrial Added Value 1 km Square Grid Dataset

Xue, Q.^{1,2} Song, W.^{1*} Zhu, H. Y.¹

1. Key Laboratory of Land Surface Pattern and Simulation, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China;
2. Institute of Architecture and Urban Planning of Chongqing Jiaotong University, Chongqing 400074, China

Abstract: The spatial attributes of industrial added value is an important indicator for assessing the impact of climate change on the industrial economic system, especially for assessing exposure of industrial economy in the context of climate change. This paper uses the 2010 global vegetation index data (EVI, from the MODIS monthly average data) and the 2010 night light remote sensing data (DMSP/OLS) to construct the night light adjustment index (EANTLI) to handle the problem of saturation of night light data, and obtain optimal global lighting data. Further, based on the optimal light data and the industrial added value data of various countries from the World Bank, a regression model was constructed to conduct spatial inversion of industrial added value, resulting in the generation of a global industrial added value 1 km square grid dataset. The industrial added value and the spatialized industrial added value of 178 provinces (states) globally, were randomly selected for correlation testing. The results show that the correlation coefficient between the industrial added value of the dataset and the statistical industrial added value is 0.93; taking the statistical data as the true value, the average precision of the industrial added value of the 178 regions is 80.14%. The dataset consists of two files: global 1 km industrial added value data (.tif), and verification data (.xlsx). The number of original data files is 6. The volume of data is 1.79 GB, and it is compressed into 1 file, of size 90.8 MB.

Keywords: Global; industrial added value; remote sensing inversion; 1 km square grid

1 Introduction

Since the 1980s, global change issues have become the focus of research by scholars in different disciplines from all over the world^[1-2]. As per the IPCC Fifth Assessment Report^[3], the global surface temperature has increased by 0.85 °C in the past 120 years. Fossil fuel combustion is a major factor in climate change. In recent decades, the increasing frequency of extreme climatic events^[4] has had a significant impact on socio-economic development.

Received: 31-01-2018; **Accepted:** 19-03-2018; **Published:** 25-03-2018

Foundation(s): Ministry of Science and Technology of P. R. China (2016YFA0602402)

***Corresponding Author:** Song, W. E-8333-2018, Institute of Geographic Sciences and Natural Resources Research, songw@igsnr.ac.cn

Data Citation: [1] Xue, Q., Song, W., Zhu, H. Y. Global industrial added value 1 km square grid dataset [J]. *Journal of Global Change Data & Discovery*, 2018, 2(1): 9–17. DOI: 10.3974/geodp.2018.01.03.

[2] Xue, Q., Song, W., Zhu, H. Y. Global industrial added value 1 km grid dataset [DB/OL]. Global Change Research Data Publishing & Repository, 2018. DOI: 10.3974/geodb.2018.01.14.V1.

The United States Meteorological Bureau's division, which is responsible for assessment of impacts on climate change, considers industry as a sensitive sector that is significantly affected by climate change^[5], particularly by mean value fluctuations^[6-7] and extreme climate events^[8]. Current research in climate change emphasizes the analysis of the trends of climate change, its likely negative impacts on different segments of the industrial economy, the development of appropriate adaptive measures and policies, and risk assessment.

According to the special report on extreme disaster risk^[9-10] issued by the IPCC, the risk hazards caused by climate change are mainly related to the climate events themselves and the exposure and vulnerability of the carrier. In climate risk assessment, the degree of exposure of vulnerable segments to climate events, especially extreme climate events, often directly affects the degree of damage and economic losses^[11]. However, studies on exposure degree are influenced by many factors such as population, policy, environment and industry. At present, most of the studies on exposure degree mostly adopt the construction of an index system^[12]. Within a certain administrative boundary, the mean value or total amount of an administrative area is used to express the exposure of a certain industry in that region. However, there are problems in the actual vulnerability assessment and risk hazard analysis, such as the inability to differentiate between internal features due to too coarse spatial resolution, and the inability to carry out evaluations of socio-economic systems based on climate raster data under multiple climate scenarios, or to simulate future exposures. Therefore, there is an urgent need for research on the spatial attributes of social and economic data at different scales to meet the basic data needs of climate change assessment. In particular, spatialized industry data for large area are still very scarce in risk assessment.

Based on the spatialization method for night light remote sensing inversion, this paper uses global Defense Meteorological Satellite Program Operational Linescan System (DMSP/OLS) night light remote sensing data, the industrial added value data of various countries from the World Bank, and moderate-resolution imaging spectroradiometer (MODIS) vegetation index products, to spatialize the global industrial economic system and build a kilometer grid dataset on global industrial added value. The grid dataset will help clarify the spatial-temporal characteristics of the grid distribution per kilometer of global industrial added value, provide scientific data for the assessment of exposure and vulnerability of industries at the scale of global change, and provide an important scientific basis for promoting industrial planning and layout, disaster prevention and mitigation of the global industrial sectors, as well as appropriate measures of response to global changes.

2 Metadata of Dataset

The metadata of global industrial added value 1 km square grid dataset is summarized in Table 1. It includes the dataset full name, short name, authors, year of the dataset, temporal resolution, spatial resolution, data format, data size, data files, data publisher, and data sharing policy, etc.

Table 1 Metadata summary of the Global Industrial Added Value 1 km Square Grid Dataset

Items	Description
Dataset full name	Global industrial added value 1 km square grid dataset ^[13]
Dataset short name	GlobIndusAddV1km
Authors	Xue, Q. E-8310-2018, Institute of Geographic Sciences and Natural Resources Research, 15123085462@163.com Song, W. E-8333-2018, Institute of Geographic Sciences and Natural Resources Research, songw@igsnr.ac.cn Zhu, H. Y. E-9420-2018, Institute of Geographic Sciences and Natural Resources Research, zhuhy@igsnr.ac.cn
Geographical region	Global
Year	2010
Temporal resolution	1 year
Spatial resolution	1 km square
Data format	.xlsx, .tif, .zip
Data size	1.79 GB (before compression), 90.8 MB (after compression)
Data files	Consists of 2 parts, including: 1) IndusAdd_Global.tif is the global industrial added value 1 km grid data, the volume of which is 1.79 GB 2) DataValidation.xls is the statistical table of precision validation on the global states and provinces, and the data volume is 29 KB
Foundation(s)	Ministry of Science and Technology of P. R. China (2016YFA0602402)
Data publisher	Global Change Research Data Publishing & Repository, http://www.geodoi.ac.cn
Address	No. 11A, Datun Road, Chaoyang District, Beijing 100101, China
Data sharing policy	Data from the Global Change Research Data Publishing & Repository includes metadata, datasets (data products), and publications (in this case, in the <i>Journal of Global Change Data & Discovery</i>). Data sharing policy includes: (1) Data are openly available and can be free downloaded via the Internet; (2) End users are encouraged to use Data subject to citation; (3) Users, who are by definition also value-added service providers, are welcome to redistribute Data subject to written permission from the GCdataPR Editorial Office and the issuance of a Data redistribution license, and; (4) If Data are used to compile new datasets, the ‘ten percent principal’ should be followed such that Data records utilized should not surpass 10% of the new dataset contents, while sources should be clearly noted in suitable places in the new dataset ^[5]

3 Methods

3.1 Basic Data

Source data used in this dataset include global night light data, MODIS data, and related statistical data. Specific sources are shown in Table 2.

The night light data was derived from non-radiatively calibrated average night lighting products produced by U.S. meteorological satellites through low-light-source imaging techniques^[19], which is widely applied in economic data inversion^[20], urban area extraction^[21], light pollution scope determination^[22] and other studies due to its large scale, high resolution and ability to reflect the characteristics of geographical entities through light intensity values. The light intensity of the night light data ranges from 0–63. When the light brightness reaches 63, the center light will not continue to increase with the light intensity. Instead, a phenomenon of saturation will occur^[23], resulting in the actual light value of the indicated area being much larger than the value of the light intensity represented by the image, and accurate

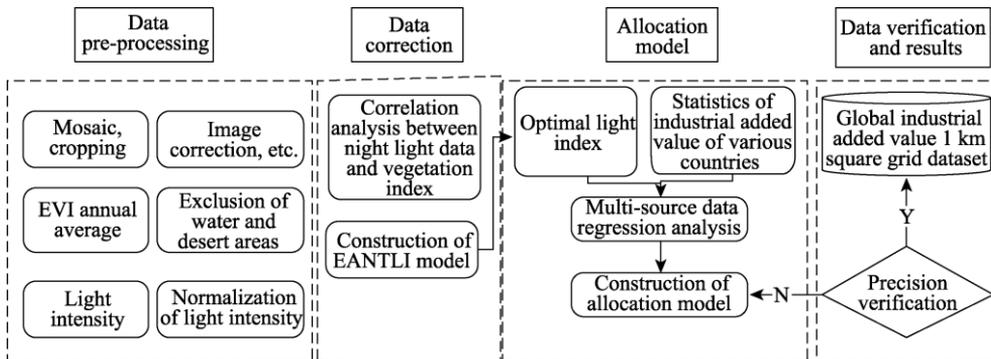
Table 2 Data sources and application in this study

Data name	Data source	Data Description	Data usage
Night light data	https://ngdc.noaa.gov/ ^[15]	The data used in this paper for 2010 global lighting data has a spatial resolution of 1,000 m and a brightness value range of 0–63	Obtain light intensity values and light normalization values; build EANTLI model and allocation model
MODIS Data	https://modis.gsfc.nasa.gov/ ^[16]	The data used in this paper is the global MOD13A3 product from January to December 2010, which contains 11 effective bands	Extract enhanced vegetation index (EVI) to construct an EANTLI model to correct night light data for spillover and saturation effects
Statistical data	http://data.worldbank.org.cn/ ^[17]	The data used in this paper comes from the World Bank's dataset of world development indicators and statistical offices of various countries	Used to build an allocation model and for precision verification
Vector data	http://www.gadm.org/ ^[18]	Administrative boundary	Mapping

remote sensing inversion cannot be performed due to the inability to process the tiny differences in the internal lighting of a city. Current methods for the correction of light data saturation include the use of radiometric calibration data published by the National Geophysical Data Center of the United States^[24] for correction, linear correction of power consumption and light intensity^[25], image correction of invariant area, image correction of vegetation index^[26] and other methods. Due to the large scale, taking into account the feasibility of data acquisition, vegetation index has been selected as the method for saturation correction of night light data for this dataset.

3.2 Data R&D Technical Route

The main technical steps of this dataset include data pre-processing, saturation correction of night light data, construction of allocation model, and precision verification. The specific technical route is as follows:

**Figure 1** Global 1 km industrial added value algorithm flow

3.3 Data R&D Algorithm

Based on the relationship between night light imagery, economic data, and enhanced vegetation index (EVI), the saturation of night light imagery is eliminated first, and a regression model of economic data and night light imagery is constructed to spatially process the industrial added value. The specific process is described as follows:

(1) Data pre-processing. Process the light data and MODIS vegetation index data by projection, trimming, mosaicing, binarization and normalization to obtain the annual average light intensity value (NTL). Calculate normalized light intensity index (NTLn) and annual average enhanced vegetation index (EVI), where the EVI data is the annual average data after deleting regions with value less than 0.01 (including water, bare rock, desert, etc.).

(2) Saturation correction of night light data. Construct a light index correction model enhanced vegetation index-adjusted NTL index (EANTLI) model to reduce light overflow and saturation, and use the NTL, NTLn and EVI data to build the EANTLI model. The calculation formula^[27] is as follows:

$$EANTLI = \left(\frac{2}{1 - NTL_n + EVI_i} - 1 \right) \times NTL \quad (1)$$

(3) Allocation model building. Based on the optimal light index EANTLI and statistical data of various countries, a regression allocation model is constructed to invert the industrial added value per kilometer. The formula is as follows:

$$I = \frac{I_i g EANTLI}{EANTLI_i} \quad (i = 1, 2, 3 \dots n) \quad (2)$$

Where I_i refers to the industrial added value of each country; $EANTLI_i$ refers to the optimal light value of each country. Through statistical analysis of ArcGIS software, the preliminary global industrial added value kilometer grid data is obtained by the regression allocation model.

(4) Precision verification and results. Randomly select the industrial added value data of a large number of state and provincial administrative units to verify the precision of data obtained in step 3. If the precision meets the requirements, the final global industrial added value 1 km square grid dataset will be obtained; otherwise, the allocation model should be rebuilt.

4 Results and Validation

4.1 Composition of Dataset Files

The global industrial added value kilometer grid dataset has a spatial resolution of 1 km and the data year is 2010. The dataset consists of two parts: data entity and verification data.

(1) Spatial distribution data

Data entities are stored in IndusAdd_Global.tif format, and can be read and written by ArcGIS software or Python and other mainstream programming languages by calling related libraries. The global industrial added value grid distribution for 2010 is shown in Figure 2. The global industrial added value ranges from 0–9,882.23 million USD/km², of which the high added value industries are mainly distributed on the east coast of North America, South America, Brazil along the Pacific coast, and many countries in Western Europe (Germany, Belgium, Netherlands, Denmark, Switzerland, France, Luxembourg, UK, Italy, Austria), China's Coastal Region, Japan, South Korea, and South and Southeast Asia.

(2) Validation data statistics

The DataValidation.xlsx data is the validated precision data of the 178 regions selected for

this dataset. The verification content and results are listed in Table 3.

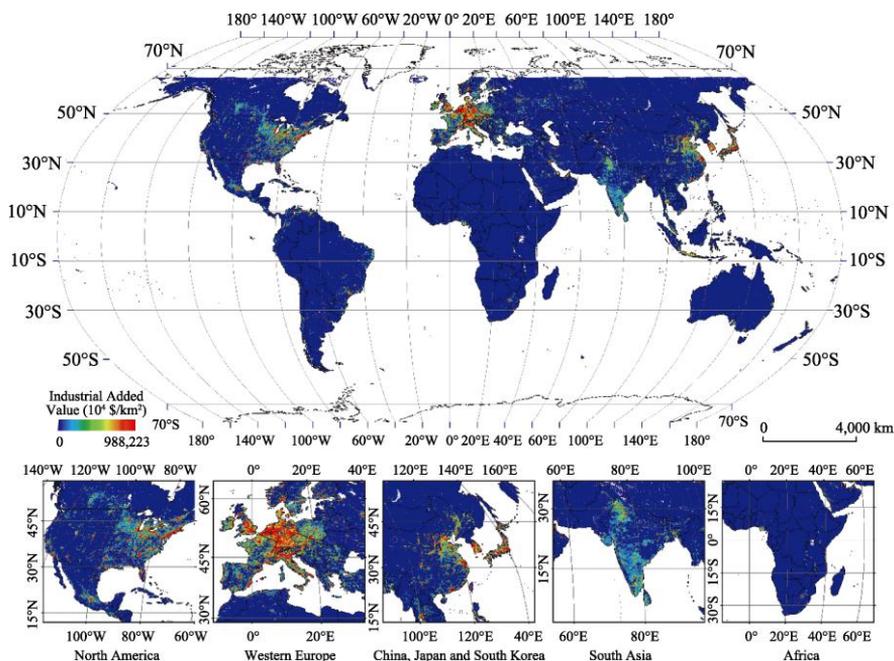


Figure 2 Global 1 km square GRID industrial added value spatial distribution

Table 3 Verification of global 1 km industrial added value dataset in 178 regions (excerpt)

Global Industrial Added Value (10^8 \$/km ²)				
Region names (EN)	Statistical data	GlobeIndus AddV1 km	Precision (%)	Statistical data source
Agusan del Norte	1.30	1.64	73.71	http://psa.gov.ph/
Anhui Province	942.75	669.17	70.98	http://www.stats.gov.cn/
Ehime Prefecture	110.50	116.87	94.24	http://www.stat.go.jp/english/index.htm
Alaska	213.89	213.45	99.80	https://www.bea.gov/
Ohio	1,043.95	1,229.37	82.24	https://www.bea.gov/
Amazonas	50.66	53.32	94.77	https://ww2.ibge.gov.br/english/default.php/
Antioquia	122.30	112.08	91.65	http://www.dane.gov.co/index.php/en/
Rostov region	61.88	61.41	99.24	http://www.gks.ru/

This dataset randomly selects 178 regions for precision verification (Table 3). The statistical data are the industrial added value data released by the statistics office of the selected verification regions. GlobeIndusAddV1km is the result of the spatialization of the selected verification region. The term precision refers to the relative precision between the two. The statistical data source is the website of the statistics office of the selected verification region. For example, the industrial added value of Agusan del Norte in 2010 was US\$130 million per square kilometer, and the statistical data came from <http://psa.gov.ph/>. This dataset had

acquired the industrial added value of the northern province of Agusan del Norte, in 2010, that is, US\$164 million per square kilometer, with a precision of 73.71%.

4.2 Data Result Validation

4.2.1 Night Light Remote Sensing Data Desaturation Test

Saturation of night light remote sensing data is the cause for low spatial precision of economic data. Therefore, this dataset uses the vegetation index to construct the EANTLI model to desaturate the night light remote sensing data. Figure 3 shows that the light intensity based upon the untreated NTL values (the left in each pair).

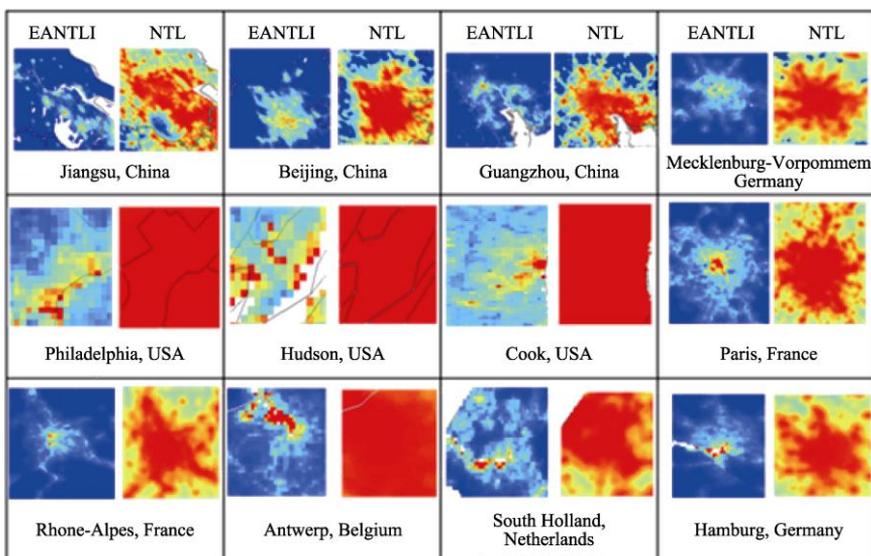


Figure 3 Comparison Chart of NTL and EANTLI

4.2.2 Spatial Precision Test and Error Analysis of Industrial Added Value at State and Provincial Levels

In order to verify the precision of the industrial added value 1 km square grid distribution dataset, statistical data on industrial added value in the statistical yearbooks of 178 provinces and states (Figure 4) were selected for this dataset, with the vector diagram of each province/state’s administrative region as the statistical layer, and the distribution data of global industrial added value as a target layer. A scatter diagram was prepared taking the statistical data as the x-coordinate and the spatial industrial added value data as the y-coordinate (Figure 5). From Figure 5, it can be seen that the spatially derived industry added value grid dataset shows a significant linear correlation with published statistical data. The correlation coefficient is 0.93 at a confidence level of 95%, which shows that the correlation is significant, and the average precision of the dataset reaches 80.14%, indicating that the spatial industrial added value datasets reliably represent the per kilometer distribution of industrial added value.

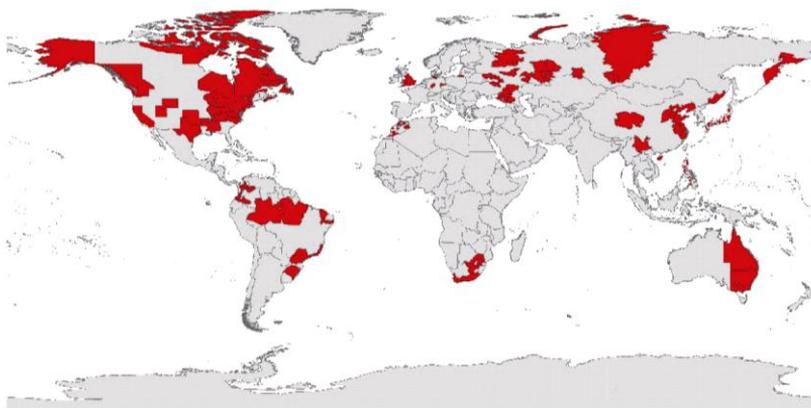


Figure 4 Location of 178 provinces and states for verification

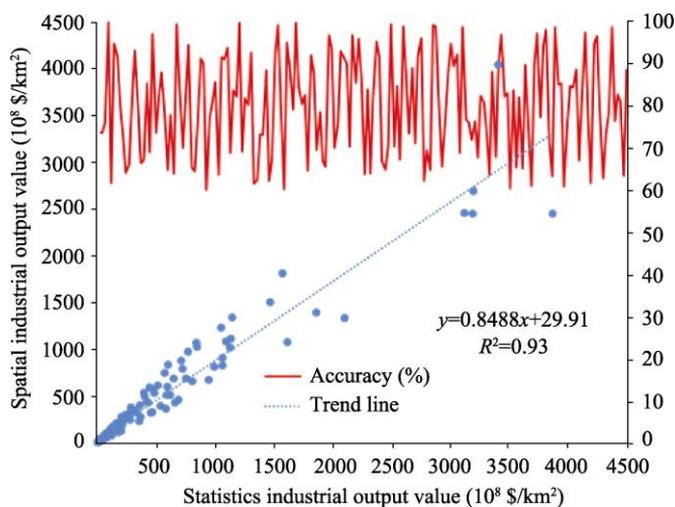


Figure 5 Precision verification of global industrial added value kilometer grid dataset

5 Discussion and Conclusion

The global industrial added value 1 km square grid dataset breaks the limits imposed by the administrative boundaries on statistical data, and achieves large-scale inversion of high spatial-temporal resolution of statistical data, with good results of precision verification. In addition, most of the spatial products at the global scale are spatial products of GDP and population. In particular, the current spatial datasets of GDP are mostly combined with secondary and tertiary industries, while separate secondary or tertiary spatial dataset products are few and far between. This dataset is based on industrial added value and night light data from countries around the world. This dataset has a high degree of precision, with average precision being 80.14%, which ensures precision in data use.

Author Contributions

Song, W., and Zhu, H. Y. carried out the overall design for the development of the dataset. Xue, Q. collected and processed global industrial added value data, designed models and algorithms, conducted data validation, and wrote the data paper.

References

- [1] Bryson, R. A. A perspective on climatic change [J]. *Science*, 1974, 184(4138): 753.
- [2] Zeng, N., Ding, Y., Pan, J., *et al.* Climate change: the Chinese challenge [J]. *Science*, 2008, 319(5864): 730–731.
- [3] IPCC, WGII. Climate change 2013: the physical science basis [J]. *Contribution of Working*, 2013, 43(22): 866–871.
- [4] Orłowsky, B., Seneviratne, S. I. Global changes in extreme events: regional and seasonal dimension [J]. *Climatic Change*, 2012, 110(3-4): 669–696.
- [5] Jager, J. Climate and Energy System [M]. Beijing: China Meteorological Press, 1988.
- [6] Wang, S. R. Impact of Climate Change on Sustainable Economic and Social Development in China and Its Response [M]. Beijing: Science China Press, 2011.
- [7] Li, D. H. W., Liu, Y., Lam, J. C. Impact of climate change on energy use in the built environment in different climate zones—a review [J]. *Energy*, 2012, 42(1): 103–112.
- [8] Rosenberg, E. A., Keys, P. W., Booth, D. B., *et al.* Precipitation extremes and the impacts of climate change on stormwater infrastructure in Washington State [J]. *Climatic Change*, 2010, 102(1-2): 319–349.
- [9] Aven, T., Renn, O. An evaluation of the treatment of risk and uncertainties in the IPCC reports on climate change [J]. *Risk Analysis*, 2015, 35(4): 701–712.
- [10] Mastrandrea, M. D., Mach, K. J. Treatment of uncertainties in IPCC assessment reports: past approaches and considerations for the fifth assessment report [J]. *Climatic Change*, 2011, 108(4): 659.
- [11] Fussler, H. M. Development and climate change: review and quantitative analysis of indices of climate change exposure, adaptive capacity, sensitivity, and impacts [R]. 2009.
- [12] Benzie, M., Hedlund, J., Carlsen, H. Introducing the transnational climate impacts index—indicators of country-level exposure [R]. *Stockholm Environment Institute*, 2016. DOI: 10.13140/RG.2.1.2839.7044.
- [13] Xue, Q., Song, W., Zhu, H. Y. Global industrial added value 1 km grid dataset [DB/OL]. Global Change Research Data Publishing & Repository, 2018. DOI: 10.3974/geodb.2018.01.14.V1.
- [14] GCdataPR Editorial Office. GCdataPR Data Sharing Policy [OL]. DOI: 10.3974/dp.policy.2014.05 (Updated 2017).
- [15] NOAA. Defense meteorological satellite program/operational linescan system (DMSP/OLS) [EB/OL]. 2013. <https://ngdc.noaa.gov>.
- [16] DAACNL. Vegetation indices monthly L3 global 1 km [DB/OL]. Version 5. <https://modis.gsfc.nasa.gov/>.
- [17] The World Bank. World Development Indicators [EB/OL]. <https://data.worldbank.org/>.
- [18] Areas, G. A. GADM database of global administrative Areas [EB/OL]. <http://www.gadm.org/>.
- [19] Cho, K., Ito, R., Shimoda, H., *et al.* Fishing fleet lights and sea surface temperature distribution observed by DMSP/OLS sensor [J]. *International Journal of Remote Sensing*, 2010, 20(1): 3–9.
- [20] Wu, J., Wang, Z., Li, W., *et al.* Exploring factors affecting the relationship between light consumption and GDP based on DMSP/OLS nighttime satellite imagery [J]. *Remote Sensing of Environment*, 2013, 134(7): 111–119.
- [21] Amaral, S., Cmara, G., Miguel, A., *et al.* Assessing nighttime DMSP/OLS data for detection of human settlements in the Brazilian Amazon [J]. *Nature*, 2001, (4): 895–903.
- [22] Yang, M., Wang, S., Zhou, Y., *et al.* Review on applications of DMSP/OLS night-time emissions data [J]. *Remote Sensing Technology & Application*, 2011, 26(1): 45–51.
- [23] Ma, L., Wu, J., Li, W., *et al.* Evaluating saturation correction methods for DMSP/OLS nighttime light data: a case Study from China's Cities [J]. *Remote Sensing*, 2014, 6(10): 9853–9872.
- [24] Jing, X., Shao, X., Cao, C., *et al.* Comparison between the suomi-NPP day-night band and DMSP-OLS for correlating socio-economic variables at the provincial level in China [J]. *Remote Sensing*, 2015, 8(1): 17.
- [25] Letu, H., Hara, M., Tana, G., *et al.* A Saturated Light Correction Method for DMSP/OLS Nighttime Satellite Imagery [J]. *IEEE Transactions on Geoscience & Remote Sensing*, 2012, 50(2): 389–396.
- [26] Zhang, Q., Seto, K. C. Mapping urbanization dynamics at regional and global scales using multi-temporal DMSP/OLS nighttime light data [J]. *Remote Sensing of Environment*, 2011, 115(9): 2320–2329.
- [27] Li, Z., Zhang, X., Jing, Z., *et al.* An EVI-based method to reduce saturation of DMSP/OLS nighttime light data [J]. *Acta Geographica Sinica*, 2015, 70(8): 1339–1350.