Research on GDP Forecasting Models of Heilongjiang Province in China Using Long Time Series Nighttime Light Data

Tian Xie

MSc Thesis

23 April 2019

TU Delft

1920 HIT

哈爾濱工業大學
HARBIN INSTITUTE OF TECHNOLOGY

Research on GDP Forecasting Models of Heilongjiang Province in China Using Long Time Series Nighttime Light Data

MSc Thesis

Student: Tian XIE
Student Number: 4619420
University: Delft University of Technology
Faculty: Technology, Policy and Management
Programme: Engineering and Policy Analysis

Supervisor:
Scott Cunningham
Delft University of Technology
Faculty of Technology, Policy and Management
Department of Multi-Actor Systems

Supervisor:
Xiaofeng HUI
Harbin Institute of Technology
School of Management
Department of Finance

Harbin, April 2019

Preface

Here it comes, the very last page of this paper. The mood at this moment is complicated. The paper, which should have been completed in 2018, has not ended until now. Despite the hard process, the results are satisfying.

First of all, I especially want to thank the two supervisors of my thesis. They are Professor Hui Xiaofeng from Harbin Institute of Technology, and Professor Scott Cunningham from Delft University of Technology. Professor Hui's rigorous academic attitude and the superb research ability have always been my role models. Professor Scott's professional teaching and research abilities and extensive professional involvement have inspired me to explore and deepen my research in the process of writing this thesis. I would like to thank the two supervisors for their support and help.

I would also like to thank Dr. Bhartendu Pandey of Yale University, Associate Professor Li Xi of Wuhan University, and Associate Professor Zhao Shuhe of Nanjing University for their inspiration in remote sensing. Thanks to Dr. Ryan Barnes, a data scientist, for his guidance on Python time series analysis.

In the process of completing the thesis, I have always recalled my wonderful times in the Netherlands, and the lovely classmates and friends. It is because of studying with you, working on projects with you, spending time with you, that I have mastered a lot of practical knowledge, got courage to explore my interests, better understood how to cooperate and share, better combined study and life. I will not list all the names of you here. But I will definitely remember every one of you. I wish you all have a bright future. I look forward to meeting you again.

Finally, I would like to thank my family, my friends and my sudden shown-up girlfriend who helped and accompanied me during my thesis time. It would never be finished without your support.

Tian Xie

Beijing, April 2019

Summary

Nighttime light remote sensing data is regarded as an important data that can be used to measure human socioeconomic activities. Because of the positive correlation between long time-series nighttime light data and economic development, long time-series nighttime light data can be used to construct GDP forecasting models to predict GDP. However, there are different temporal resolutions, measurement standards, and errors due to different sources of nighttime light data. How to eliminate the data errors and the differences between different data sources to construct a reliable long time-series nighttime light data is currently a major problem. In addition, different GDP forecasting models have different applicable conditions and forecasting accuracy. How to choose the optimal GDP forecasting model is another major problem. In this paper, data errors and differences are dealt with at first to obtain long time-series nighttime light data that can be quantitatively analyzed. Then, the advantages and disadvantages of different GDP forecasting models are analyzed and compared. The main contents of this paper are as follows.

Applying the methods of polynomial regression and RSR (ridgeline sampling regression), the intercalibration errors of DMSP/OLS (U.S. Air Force Defense Meteorological Satellite Program/Operational Linescan System) nighttime light data are calibrated. Three criteria for assessing calibration effects are proposed. DMSP/OLS nighttime light data obtained by combining two calibration methods is optimized.

The NPP/VIIRS (Suomi National Polar-orbiting Partnership/visible infrared imaging radiometer suite) nighttime light monthly data is summed and averaged to obtain NPP/VIIRS nighttime light annual data. For dealing with the differences between DMSP/OLS and NPP/VIIRS nighttime light data, two assumptions are proposed to unify the two nighttime light datasets.

Three kinds of GDP forecasting models are modeled and analyzed. They are linear regression model commonly used in the field of remote sensing, ARIMA model of univariate GDP time series analysis commonly used in the field of econometrics, and ARIMAX model of multivariate GDP time series analysis based on long time-series nighttime light data. ARIMAXe model with exogenous variables and ARIMAXs model with seasonal factors are constructed as sub-models of ARIMAX model. Nighttime light data is applied to time series analysis for the first time. Finally, this paper compares and assesses the prediction results of four GDP forecasting models on the GDP of Heilongjiang province in 2017.

List of figures

List of tables

Contents

1. Introduction

With the continuous development of space technology, satellite-based applications are also increasing. As one of the important applications of satellites, remote sensing can provide multi-spectral and multi-temporal geographic information as well as services for national conditions research, urban planning, land resource management, agricultural management and emergency disaster response. As a unique category of remote sensing, nighttime light remote sensing can provide a new perspective for the evaluation of many social and economic activities.

When observing the Earth at night from space, human activities are reflected in various forms of light. With the absence of clouds in the sky at night, remote sensing satellites can observe various kinds of light on the earth, such as city lights, fishing boat lights, oil and gas fields burning, natural fires like forest fire and volcanoes. These images of visible light from the Earth acquired at night without cloud covering are nighttime light remote sensing images (Li & Li, 2015a).

According to different spatial resolutions, nighttime light remote sensing images can be divided into high-resolution and low-resolution nighttime light remote sensing images. High-resolution nighttime light remote sensing images have resolutions ranging from a few meters to tens of meters. Their application scenarios are generally in urban or other small spatial scales, providing more detailed spatial information for urban planning and the like. This article will focus on low-resolution nighttime light remote sensing images with resolutions ranging from a few hundred meters to several kilometers. Due to the low resolution of images, the spatial scale that can be studied is large. In general, the original nighttime light remote sensing image taken by satellite needs a series of pre-processing such as splicing and system error elimination to obtain the image product available for users.

The nighttime light remote sensing images used in this paper are raster images, which means that each image is composed of many grids, each grid represents one pixel, and each pixel contains brightness information values. Therefore, each piece of nighttime light remote sensing image contains a large amount of data information. For the long time series study, a number of nighttime light remote sensing images over a period of time can reflect changes in the brightness and distribution of nighttime lights during this period of time in an area, thus providing more spatial and temporal information, which request a larger amount of data.

In the process of mining large-scale remote sensing data, it often encounters data reliability, system error, observation error, format mismatch and other data reliability and usability problems. For different research topics, the processing methods of night light remote sensing data are not the same, including the elimination of systematic errors and other processing method. This paper will also propose corresponding data processing methods for the regional economic studies in China.

The application of nighttime light remote sensing data is extremely extensive, including socio-economic statistical indicator estimation, urbanization and urban expansion research, oil and gas field monitoring, fishery observation, ecological environment and health effects research. Estimates of socio-economic statistical indicators include national economic indicator estimates, population estimates, electricity consumption estimates, and carbon emissions estimates. This paper will apply nighttime light remote sensing data modeling to predict and analyze the Gross Regional Product (GRP) of Heilongjiang Province, that is, the GDP statistics in the region.

According to the definition of the OECD (2003), Gross domestic product is an aggregate measure of production equal to the sum of the gross values added of all resident institutional units engaged in production (plus any taxes, and minus any subsidies, on products not included in the value of their outputs). The sum of the final uses of goods and services (all uses except intermediate consumption) measured in purchasers' prices, less the value of imports of goods and services, or the sum of primary incomes distributed by resident producer units. GDP is considered to be a most important macroeconomic measurement of the regional or countrywide economic development. Therefore, current and predicted GDP have been widely concerned.

The accounting data of GDP data comes from the national statistical survey data on the one hand, and from the records of the administrative department on the other hand. Due to its inherent economic investigation, statistics, and calculation methods, the following limitations exist. First, the economic census is difficult to cover all regions. The first-hand economic data in some remote areas is difficult to obtain or with low reliability. Second, there may be differences in statistical caliber and data processing methods in different administrative level units such as provinces, cities and counties. Third, in some regions, there may be phenomena such as “water injection” of GDP data. In terms of GDP forecasting, traditional forecasting methods are simple because of the existing simple model and single input. It is difficult to make accurate predictions.

Nighttime light remote sensing data can make up for the limitations mentioned above. Nighttime light remote sensing images are indistinguishable measurements of global nighttime lightness, with data covering the globe and without data differences due to administrative differences. The remote sensing data of nighttime light has irreplaceable value for the socio-economic activity statistics in war zones, remote areas and national borders (Henderson, Storeygard, & Weil, 2008). In terms of GDP data prediction, nighttime light remote sensing data can provide unique space-time input. This paper aims to establish a Heilongjiang GDP forecasting model using long time series nighttime light remote sensing data to predict the GDP of Heilongjiang Province in a timely and accurate manner.

2. Literature review

In the field of nighttime light remote sensing, the following researches have developed a macro framework, which is also a summary of the field. Henderson, Storeygard, and Weil (2012) introduced in detail the research directions in this field on the basis of different satellite platforms, processing and analysis methods of nighttime light remote sensing data, and applications of nighttime light remote sensing images in socio-economic activity analysis results till October 2016. Li and Li (2015a) introduced the future development trends of different satellite remote sensing observation platforms and sensors, various applications of nighttime light remote sensing and data mining methods. Li and Li (2015b), and Li, Yu, and Li (2017) also proposed the application of nighttime light remote sensing images to analyze the spatial-temporal pattern of urban development along the "Belt and Road" countries, and pointed out the important role of nighttime light remote sensing images in the analysis of the development of social and economic activities in the countries along the "Belt and Road".

2.1 Research Status of DMSP/OLS Nighttime Light Imagery

The DMSP/OLS is called the US Air Force Defense Meteorological Satellite Program/Operational Linescan System, which is a linear scanning operation system carried out by the US Air Force's Defense Meteorological Satellite Project. The product is global satellite remote sensing images released from 6 satellites from 1992 to 2013, hereinafter referred to as OLS data. Due to the lack of on-board calibration devices, there is no uniform standard for each satellite's annual brightness detection value, so that continuous time series studies are not possible. Data intercalibration is required. Meanwhile, observation oversaturation is likely to occur in high-luminance regions because of the limit of pixel digits. Intercalibration and error elimination are necessary processes for applying OLS data for continuous time series studies. Therefore, under the premise of ensuring data quality, researchers have proposed various OLS data processing methods.

At present, the widely used method for OLS data intercalibration was proposed by Elvidge et al. (2009), which is referred to as the Elvidge method. The paper has firstly extracted many human settlements as samples, found that the F12 satellite had the highest luminous value which is called digital number (DN) in 1999. F121999 data is selected as the reference data, and Sicily Island as the reference location. Because the

island has the smallest change in luminous brightness over decades. And the DN value is more evenly distributed between the minimum and maximum values. After determining the reference data and the reference location, the article draws the DN values of the reference location in the other annual images and the DN value of the reference year in Cartesian coordinates, and performs polynomial fitting on the data to convert the DN value of the non-reference year to the standard reference year DN value to complete the intercalibration work. The method has simple process, good correction effect and is widely accepted. Elvidge et al., in the report (Elvidge, Hsu, Baugh, & Ghosh, 2014) published in 2014, calculated the polynomial fit coefficients for the intercalibration of another OLS nighttime light remote sensing image product.

Another method of intercalibration of OLS data was proposed by Zhang et al. (2016) in 2016, which automatically identifies pseudo-invariant regions based on a new sampling strategy. In the density map generated by the reference image and the target image, the data is selected by polynomial fitting to generate a ridge sampling regression (RSR) curve, which is referred to as the RSR method. A calibration model is then generated using points on the ridgeline regression curve to minimize inconsistencies among annual OLS data. The results show that the method calibrates the OLS annual data to a large extent. And the effect is better than the classic Elvidge method. The method is simple, robust and easy to operate. It provides data consistency guaranteeing for continuous time series OLS data research.

Liu, He, Zhang, Huang, & Yang (2012) used China as the research object, Jixi City of Heilongjiang Province as the reference area, F162007 as the reference data, and applied the Elvidge method to calibrate the OLS data for year 1992-2008. Pandey, Joshi, and Seto (2013) selected the Lucknow region of India as the reference area, and used F162007 as the reference data. The Elvidge method was then applied to calibrate the OLS data from 1992 to 2009. Based on the principle of minimum ground change, Li et al. (2013) used Beijing as the research object, F152001 as the reference data, constructed a linear regression model, and proposed a calibration method that can automatically screen the reference pixels. Hsu, Baugh, Ghosh, Zhizhin, and Elvidge (2015) selected Los Angeles as the reference area, and the brightness-calibrated 2006 OLS data as reference data. Then the linear regression model was used to intercalibrate the OLS data from 1996 to 2010. Liu et al. (2015) proposed a simplified OLS nighttime light remote sensing data intercalibration method and Theil-Sen regression model when studying the relationship between urbanization and vegetation coverage in major cities around the world. Henderson, Storeygard, and Weil (2012) proposed an

intercalibration model that considers the annual correction effect and is widely adopted by economists. Pandey, Zhang, and Seto (2017) found that although there are many methods for calibrating OLS data, there is a lack of evaluation and comparison of the calibration effects of each method. So two metrics for the evaluation of intercalibration effect of each methods are proposed. The first is to minimize system errors. The second is to use the relationship between socio-economic data such as economic activity and electricity consumption and OLS data to assist judgment. The comparison evaluation draws the following conclusions. The global intercalibration effect is better than the regional intercalibration. The calibrated data still has systematic errors. The calibration is difficult to conduct in high latitude areas and weak light areas. The article points out that the accuracy of the intercalibration results needs to be improved from two aspects. One is to modify the engineering project itself or the original algorithm on the satellite itself or the original data. The other is to apply relative calibration based on the VIIRS satellite platform.

Then the development of global OLS data studies and some classic application scenarios will be introduced in chronological order. Since the OLS data was electronically archived in 1992, Elvidge, Baugh, Kihn, Kroehl, and Davis (1997b) published the first article in the field, systematically introduced DMSP/OLS, including methods for detecting and locating light sources. The results were displayed using Stable Lights images of the United States in 1997. Since then, researchers have found that OLS data is related to many socioeconomic parameters reflecting human activities. Elvidge et al. (1997a) found that time series analysis can eliminate the effects of transient light. Analysis of 52 countries shows that OLS data can be used as an indicator for spatial measurement of population size and greenhouse gas emissions. Sutton, Roberts, Elvidge, and Baugh (2001) conducted a study on urban population density in the United States based on OLS data. Imhoff, Lawrence, Stutzer, and Elvidge (1997) used the processed OLS city light data to map cities in the United States. Sutton, Roberts, Elvidge, and Baugh (2001) used OLS data to analyze the size of the global city bright areas, and estimated the global population based on the relationship between the size of the city's bright areas and the urban population. The results show that the global population estimated by remote sensing data in 1997 was 6.3 billion, which is close to the 5.9 billion widely accepted census data. Henderson, Yeh, Gong, Elvidge, and Baugh (2003) proposed a method for setting thresholds to identify urban boundaries, studied San Francisco, Beijing, and Lhasa, and compared the results with high-resolution Landsat TM images. Ebener, Murray, Tandon, and Elvidge (2005) used

OLS data and demographic data to spatially model income per capita as an indicator of regional affluence and to study its relationship with people's health at national and subnational levels. Elvidge et al. (2009) used the OLS data to map the global poverty map. Raupach, Rayner, and Paget (2010) used a combination of urban nighttime light remote sensing data and population density census data to study the spatial distribution of carbon dioxide emissions from fossil fuels. Zhang and Seto (2011) tested the ability of long time series OLS data in measuring regional and global urban growth through population, land use, and land cover data. In the study of urbanization in India, China, Japan, and the United States based on OLS data from 1992 to 2008, it was found that India's urban growth rate was higher than China every two year from 1992 to 2000, while India's and China's growth rates were much higher than the United States and Japan. Ma, Zhou, Pei, Haynie, and Fan (2012) applied OLS data to compare the urbanization of more than 200 prefecture-level administrative regions in China between 1994 and 2009. Liu, He, Zhang, Huang, and Yang (2012) studied the urban expansion of China from 1992 to 2008 with a new calibration method. Zhang and Seto (2013) randomly selected 240 locations in the world to generate urbanization types through OLS data, and compared the results with Google Earth images to study the extent to which OLS data can identify different types of urbanization. The results show that the recognition rate reaches 93%, but with an over-recognition phenomenon. Pandey, Joshi, and Seto (2013) studied the urbanization of India after calibrating the OLS data. Yi et al. (2014) proposed the Urban Light Index (ULI) to establish a concentric circle model, and carried out a study focusing on temporal and spatial changes of urbanization progress in 34 prefecture-level administrative regions in the three northeastern provinces of China, which are Harbin, Changchun, Shenyang and Dalian. Liu et al. (2015) applied the calibrated data to study the relationship between global urbanization and reduced vegetation coverage. Wang (2016) studied the urbanization of Southwest China based on OLS data. Baidu map and the Development Research Center of the State Council applied OLS data on urban agglomeration identification (Lan & Kan, 2018). Huang, Wang, and Xie (2016) used OLS data to analyze the urbanization process in Guangdong Province for 20 years. Pan (2017) conducted a coupling study on residential development and economic growth based on OLS data and city point of interest (POI).

In the study of various social and economic activities based on DMSP/OLS data, there is a specific focus on the study of GDP. Sutton and Costanza (2002) used OLS data to generate Light Energy (LE) indicators for measuring GDP changes on a smaller spatial

scale. At the same time, the paper studies the spatial and temporal distribution of Ecosystem Services Product (ESP) and Subtotal Ecological-economic Product (SEP) values. Doll, Muller, and Morley (2006) found a positive correlation between the nighttime light remote sensing data and the GRP through the study of the remote sensing data of 11 EU countries and the United States. The article also pointed out that in the economically developed areas, the luminous brightness detection value cannot fully reflect the economic development of the region due to brightness saturation. This over-saturation phenomenon needs further studies. Sutton, Elvidge, and Ghosh (2007) performed linear regression on OLS and GDP data to study the positive correlation between them in China, India, Turkey, and the United States. The results show that the coefficients of determination of the remote sensing data and GDP data obtained by applying the regression model in China, India, Turkey, and the United States are 0.96, 0.84, 0.95, and 0.72, respectively, meaning there is a strong positive correlation. Henderson, Storeygard, and Weil (2012) found the logarithmic regression relationship between OLS and GDP data with models, made predictions on GDP data in areas where GDP data is hard to obtain to with low quality. The article also pointed out that the nighttime light remote sensing data is more suitable for predicting the long-term growth trend of economic activities. Ghosh et al. (2010) performed regression analysis on the relationship between the sum of light intensity and economic activity in China, India, Mexico and the United States. After obtaining the correlation coefficient between the two variables, a spatial distribution map of global economic activities was generated with 1km^2 pixels. Chen and Nordhaus (2011) applied the panel regression model proposed by Henderson, Storeygard, and Weil (2012), found that there is a strong correlation between the nighttime light remote sensing data and the GDP data, which is of high value for regions lacking economic census data. Wu, Wang, Li, and Peng (2013) studied the panel data of 169 countries and regions for 15 consecutive years and found that the amount of light used is related to GDP per capita, spatial distribution of human activities and national savings rate. There is an inverted U-shaped relationship between light consumption per capita and GDP per capita. Keola, Andersson, and Hall (2015) applied nighttime light remote sensing data to assess income growth in farmland and woodland areas. The results show that a univariate nighttime light remote sensing data is unable to provide much information on the situation of agricultural and forestry land, but can be used to assess the economic activities of agricultural and forestry land combing with land cover data. Han, Zhou, Wang, Liu, and Yao (2012) studied the relationship between nighttime light remote

sensing data and GDP of primary, secondary and tertiary industries.

2.2 Research Status of NPP/VIIRS Nighttime Light Imagery

In addition to DMSP/OLS data, another type of data used in this paper is the successor to DMSP/OLS data, the Suomi National Polar-orbiting Partnership/Visible Infrared Imaging Radiometer Suite (NPP/VIIRS), hereinafter referred to as VIIRS data. This nighttime light remote sensing image products have been electronically archived since April 2012. Compared with OLS images, VIIRS images have higher resolution, equipped with calibration devices, which make it more and more popular among researchers.

Although the VIIRS is equipped with an on-board calibration device, its image data still has systematic errors and some errors caused by external factors. Several scholars have focused their research on the interpretation and elimination of VIIRS data errors. Levin (2017) analyzed the relationship between the monthly cloudless VIIRS data and the seasons in China and North America during the two years 2014 and 2015. The research found that VIIRS data has a high positive correlation with snow cover and star reflectance, and a high negative correlation with Normalized Difference Vegetation Index (NDVI). Therefore, when applying VIIRS data for annual time series analysis, seasonal changes in data should be considered. More than 30 researchers jointly published an article introducing NASA's black marble night light research project (Román et al., 2018). The Black Marble Project is the project that NASA comprehensively processes the VIIRS data and archives all corrected nighttime light data. The article mainly introduces the following four things in the VIIRS original data improvement algorithm: lunar radiation modeling to solve the nonlinear changes of phase and vibration; vector radiation transmission and lunar bidirectional surface anisotropic reflection simulation to correct the atmosphere and BRDF effects; modeling of Earth's optics and canopy radiation transmission to account for seasonal changes in NTL; filling gaps to reduce persistent data gaps. Wu, Yang, Dong, Zhang, and Xia (2018) analyzed the relationship between VIIRS data and snow cover.

Elvidge, Baugh, Zhizhin, and Hsu (2013) systematically compared the characteristics of VIIRS data and OLS data, including satellite parameters and image performance for the first time. Cao et al. (2013) focused on long-term performance in calibration, verification, radiometric calibration, and uncertainty after VIIRS satellites' launches. Chen (2017) and Shu (2013) used multi-source nighttime light remote sensing images to identify urban spatial morphological structures and urban cluster spatial patterns. Huang (2016) conducted population spatialization research based on nighttime light

remote sensing images and social perception data.

Similar to the OLS data, the VIIRS data can also be applied to the study of economic indicators like GDP. Li, Xu, Chen, and Li (2013) selected 31 provincial-level administrative regions and 393 counties in mainland China as the research objects, and compared the relationship between the total brightness of nighttime light of the VIIRS data and the corresponding GRP. The results show that the correlation between VIIRS data and GRP is higher than OLS data. Shi et al. (2014) also selected the China mainland as the research object. At first, a series of pretreatments were carried out on the original nighttime light remote sensing data. Then the linear regression was used to study the relationships between total brightness of VIIRS data and GDP and Electricity Power Consumption (EPC) respectively. The results of the paper show that the coefficient of determination R^2 between the VIIRS data and the parameters being studied is higher than the OLS data. Wu, Yang, Dong, Zhang, and Xia (2018) proposed a pre-processing method for VIIRS data, and selected the data from July to September of the VIIRS monthly data to form annual data, and studied the imbalance of economic development in various regions of China.

2.3 Research Status of High Resolution Nighttime Light Imagery

In addition to DMSP/OLS images and NPP/VIIRS images, there is a class of high-resolution nighttime light remote sensing images that are not the focus of this paper, but are still worthy of attention. Israeli scholar Levin and other researchers analyzed high-resolution images such as satellite images from the U.S. DMSP, Argentina SAC-C, Israeli EROS-B and ISS astronauts on the International Space Station (Levin & Duke, 2012; Levin, Johansen, Hacker, & Phinn, 2014; Levin & Phinn, 2016). The research found that these high-resolution images can study the nighttime light changes on smaller geographic scales, combined with ground information, providing more time and space information. The "Luojia-1" 01 satellite developed by Wuhan University with consideration of low-altitude nighttime light remote sensing and navigation signal enhancement functions was also successfully launched on June 2, 2018 at the Jiuquan Satellite Launch Center with the Long March II carrier rocket. The first image was received on the 4th of the month. These high-resolution nighttime light remote sensing images will become the focus of future research in the field of nighttime light remote sensing.

2.4 Research Status of GDP Forecasting Models

Scholars use various models to predict GDP. The specific models include exponential smoothing model, autoregressive moving average model ARMA, autoregressive integrated moving average model ARIMA, multivariate autoregressive integrated moving average model ARIMAX, autoregressive conditional heteroskedasticity model ARCH, data packet processing method GMDH, BP neural network, gray prediction, genetic programming GP model, etc.

The following studies have all applied a single predictive model to predict GDP. Yong and Bao (2008) applied ARMA to forecast the GDP of Inner Mongolia during the 11th Five-Year Plan period. Wang and Xu (2007) used the ARIMAX model to predict the tertiary industry's output value, capital expenditure and GDP. Yang (2008) applied the improved BP neural network algorithm to predict the GDP of Hunan Province. Chen (2009) discussed the application of genetic algorithm and BP neural network in GDP forecasting. Kong, Li, Shi, and Huang (2016) applied the grey prediction GM(1,1) model to predict the GDP of Hainan Province. Wang (2017) applied neural networks to predict GDP.

The following studies compare multiple GDP prediction models. Chen (2016) used the ARIMA model to predict the GDP of Shandong Province, and a multivariate time series prediction using ARIMAX model with the tertiary industry as an input. The results show that the accuracy of the prediction results obtained by ARIMAX is higher than that of the ARIMA model. Teng and He (2010) applied ARCH and ARIMA models respectively, and GMDH combined the first two models to predict China's quarterly GDP. The results show that the self-combined GMDH model has better prediction accuracy. He and He (2012) used the ARIMA and GMDH models to predict China's quarterly GDP, and introduced the Purchasing Manager Index (PMI) to apply the ARCH model for prediction. The results show that the ARCH model has better results. Shan (2014) used ARIMA model and BP neural network to predict GDP, and used genetic algorithm to optimize neural network. The results show that genetic algorithm optimized BP neural network is much better than ARIMA for GDP prediction. Xiong (2011) applied the model of ARIMA and neural network integration to predict GDP. The results show that the prediction accuracy of the integrated model is higher than that of the single model. Sun and Jiang (2008) constructed a non-linear GDP forecasting model using genetic programming GP method, predicted the GDP of Heilongjiang Province, and compared the results with the gray GM (1,1) model. The results show that the GP model is more ideal. The results show that the combined model has higher precision.

Andrews, Dean, Swain, and Cole (2013) systematically compared the prediction effects of ARIMA and ARIMAX on data prediction.

2.5 Literature Overview and Paper Structure

At present, when remote sensing scientists make GDP predictions using nighttime light data, the methods used have not much difference. With a fixed reference area and a certain time point, GDP data and nighttime light remote sensing data are regressed. The fit curve is used to predict GDP. The actual and predicted GDP are compared. The difference between this two values are used to evaluate the effect of GDP prediction. Another way of research is to take a time point in each of the OLS data and the VIIRS data for comparative study.

The ***Figure 2-1*** below shows the most of current researches in this field.

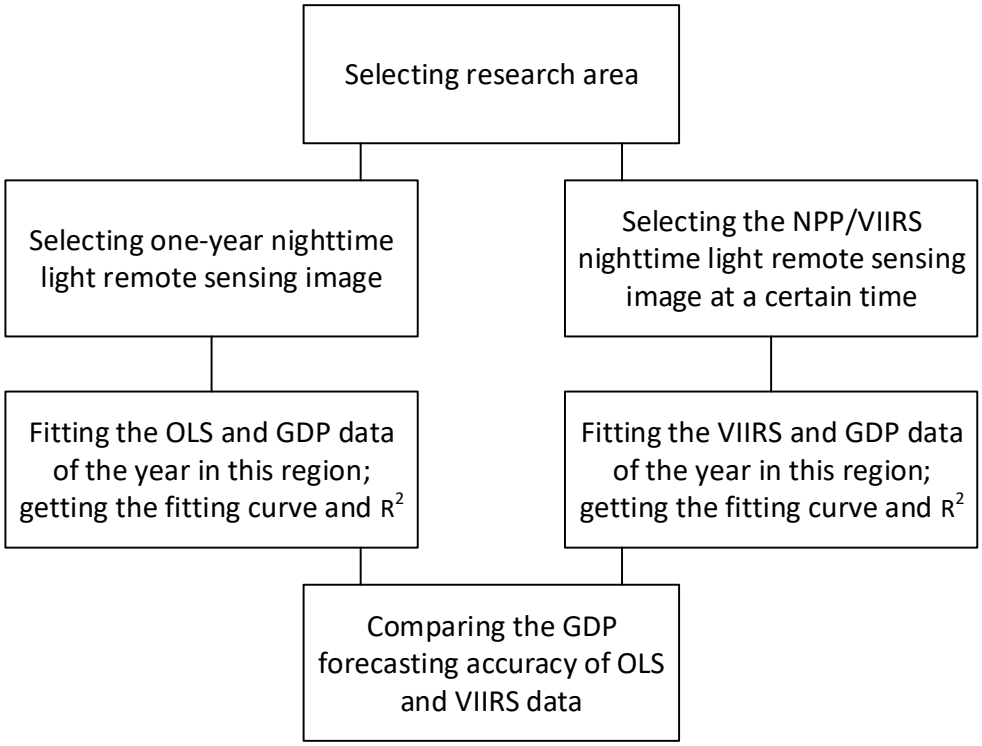


Figure 2-1 Current Research Status of GDP Forecasting Using Nighttime Light Data

The study of GDP forecasting can be divided into two aspects. On the one hand, it is the univariate time series analysis and prediction of GDP. This kind of research mainly analyzes the GDP of selected regions in the selected time using a certain GDP forecasting model, and obtains the predicted values calculated by the model. Based on the difference between the predicted value and the actual value, the prediction accuracy is evaluated. The applicability of the model to the GDP forecast in the situation is evaluated. Furthermore, multiple models, or a combination of multiple

models are studied under the same assumptions. The GDP prediction accuracy of different models are compared. Models commonly used in this type of research include exponential smoothing models, ARIMA, neural networks, or a combination of several models.

On the other hand, it is the multivariate time series analysis and prediction of GDP. The GDP time series is used as a response sequence. One or more other time series such as tertiary industry output value, purchasing manager index, etc. are added as input sequences. According to the characteristics of the response sequence, the input sequence, and the relationship between the input sequence and the response sequence are used to analyze and predict GDP. The models selected for this type of research include ARIMAX models, ARCH models, gray prediction models, and so on.

The ***Figure 2-2*** below shows the research contents and scale of GDP forecasting.

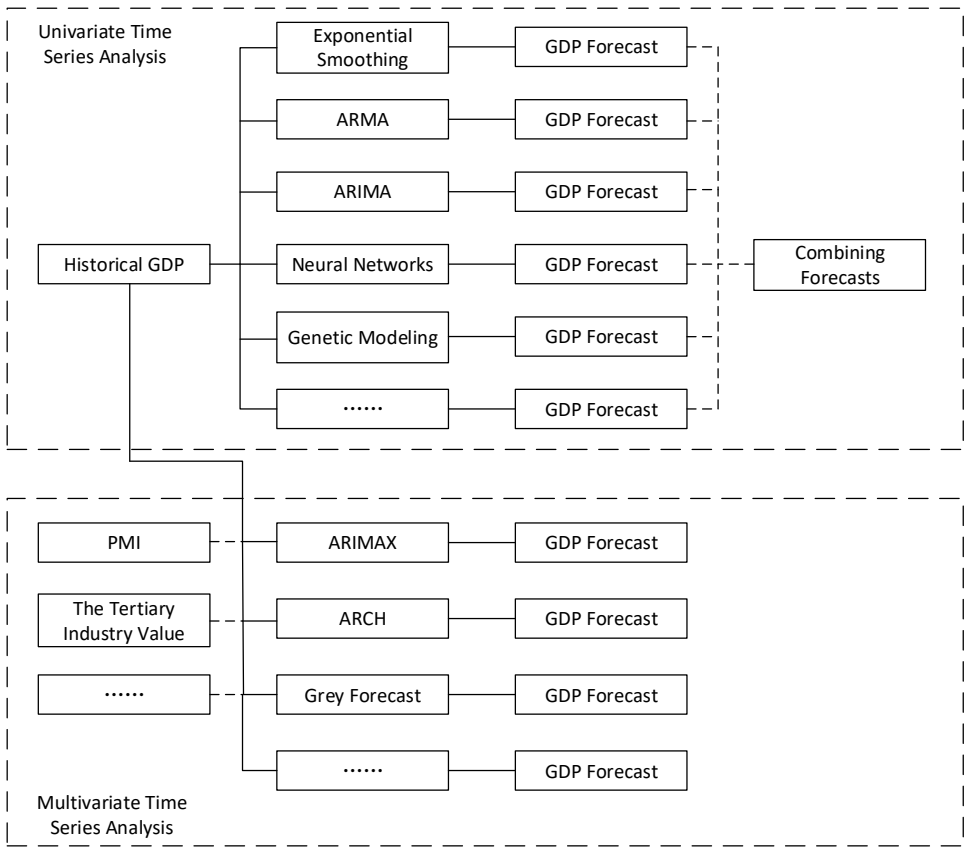


Figure 2-2 Current Research Status of GDP Forecasting Models

At present, the research on the application of GDP forecasting based on nighttime light remote sensing data is only applicable to the areas where the nighttime light remote sensing images and GDP data both exist. And the predicted value is a nighttime light fitting value of the existing GDP data. The significance of this method is to analyze the

correlation between the remote sensing data and the GDP data and calculate the coefficient of determination. The prediction of future GDP is only the value on the fitted curve, and the prediction error is large.

The application of predicting GDP using models is mostly focused on the univariate time series analysis and prediction of GDP data. In the few studies of multivariate time series analysis and prediction of GDP, the exogenous variables are merely social and economic parameters, such as the purchasing manager index, the tertiary industry output value.

In view of the shortcomings in the above researches, this paper aims to apply long time series nighttime light remote sensing data to the GDP forecasting model, improve the accuracy of GDP forecasting, and expand the application of nighttime light remote sensing data.

Here is the research framework of this paper.

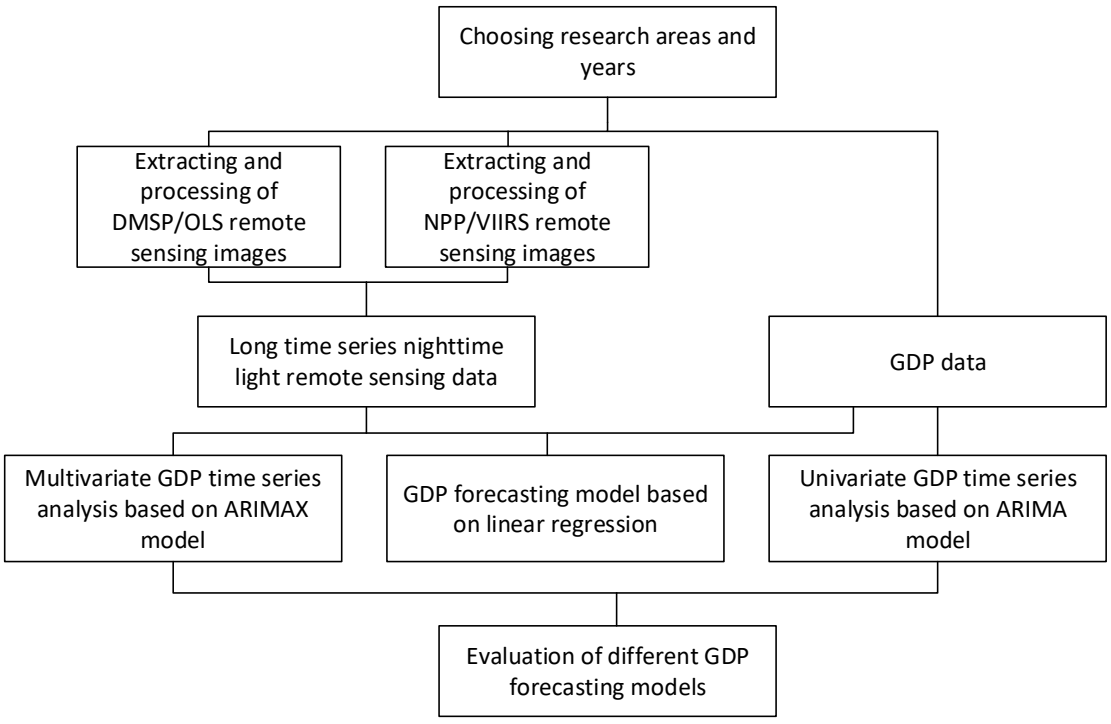


Figure 2-3 Paper Structure

The main research contents of this paper will include the following sections:
(1) The raster data model is built in ArcGIS software to generate output in a loop. The model is used to process the original nighttime light remote sensing images and export data in batches. The extracted nighttime light remote sensing data is then imported into Python. The data processing and visualization are completed in Python. The

characteristics of the nighttime light remote sensing data and the relationship between the nighttime light remote sensing data and economic activities are determined through visual images.

(2) Then the processes include determining the selection criteria of the pseudo-invariant region and the reference year, selecting the reference data and the reference region, using the polynomial regression method to intercalibrate the DMSP/OLS nighttime light remote sensing data, calculating the latest DMSP/OLS data, and obtaining the data after regression intercalibration. At the same time, the DMSP/OLS nighttime light remote sensing data will be intercalibrated by ridgeline sampling regression method. Then, the evaluation method of the two intercalibration is proposed. Two intercalibration methods are compared and evaluated to obtain the optimal long time series DMSP/OLS nighttime light remote sensing data. Next, the monthly NPP/VIIRS nighttime light remote sensing data is extracted, and synthesized into NPP/VIIRS annual data. Two hypotheses are proposed. The DMSP/OLS time series data is combined with the NPP/VIIRS time series data to form an optimal data series, so that it can serve long time series studies.

The paper is going to apply the commonly used linear regression model to predict GDP. Then the ARIMA model is used to analyze and predict the GDP data of Heilongjiang Province in a univariate time series. The prediction results and accuracy of the ARIMA model are obtained. At the same time, this paper will combine the nighttime light remote sensing data with the GDP data for the first time, propose a long time series nighttime light data as the input parameter of the ARIMAX model, use the nighttime light data to quantitatively predict the GDP of Heilongjiang Province, and obtain the prediction result and precision of the ARIMAX model. The effects of different models on Heilongjiang's GDP forecast will be compared and evaluated.

Part I: Nighttime Light Data Preparation

3. Satellite Imagery Processing

In order to apply the nighttime light remote sensing images to time series analysis, it is necessary to obtain long time series nighttime light remote sensing data which can be used as an input sequence. The next sessions will discuss the methods of extracting and processing nighttime light remote sensing data, and finally obtain long time series nighttime light remote sensing data.

This chapter will first introduce the characteristics of DMSP/OLS and NPP/VIIRS nighttime light remote sensing images, including their respective development history, satellite parameters, satellite payload, image product categories, error sources, and application scenarios. Then, the method of extracting the original nighttime light remote sensing image is proposed. The results will then be visualized.

3.1 Introduction of DMSP/OLS Nighttime Light Imagery

This section will introduce the development history, payload, product classification and error sources of DMSP/OLS.

As mentioned previously, DMSP/OLS is the U.S. Air Force Defense Meteorological Satellite Program/Operational Linescan System, which will be referred to as OLS. OLS was originally a constellation of military meteorological satellites used to detect night clouds. After being declassified in 1972, the researchers found that OLS sensors are highly capable of detecting light in urban lighting, gas combustion, fishing boat operations, and other visible light. Initially, there were fewer applications for nighttime lights (NTL) images. Until 1992, the National Oceanic and Atmospheric Administration's National Geoscience Data Center (NOAA/NGDC) released electronic archives of OLS images. Today, NOAA released OLS image data from 1992 to 2013. The subsequent image data is still collected by the military. But it is not publicly available.

The DMSP/OLS payload is an oscillating scanning radiometer with a scan width of 3000 km and two spectral bands for visible near-infrared and thermal infrared of the NTL. The OLS has two telescopes and a spot-amplifier tube for detecting radiation between 0.47μm and 0.95μm in the visible and near-infrared bands. The system has a 0.56km spatial resolution mode, but averaged by the on-orbit computer using 5×5 blocks, and transmitted to NOAA, resulting in an image with a resolution of 2.7km. OLS flies

around the globe for 14 rounds per day, covering the globe every day and night for 24 hours. Its nighttime overcast time is approximately between 20:30 and 21:30 every day. At night, the light radiation that OLS can detect from 1.54×10^{-9} lowest to $3.17\times10^{-7}\ \mathrm{W\cdot cm^{-2}\cdot sr^{-1}\cdot mm^{-1}}$ highest. The data value of each pixel is displayed with a 6-digit DN (digital number) value, from 0 without brightness to 63 with maximum brightness, which means OLS has a relative Instead of absolute radiation metrics. Since 1992, a total of nine solar-synchronous polar-orbiting satellites, named from F10 to F18, have provided OLS data. In most years, two satellites simultaneously provide OLS electronically archived data for intercalibration of OLS sensors. For long time series studies, the use of these overlapping data for calibration is often a necessary step because OLS lacks on-board calibration and inter-satellite calibration devices.

OLS imaging products include from high frequent daily images to NOAA synthesized low frequent annual images. Clouds, daylight, and moonlight will affect the detection of specific areas to a certain extent. These effects have been eliminated in the annual image. Since the cost of acquiring daily and monthly images is high, and the annual images are freely available, most of the research use global annual stable lights composites. Currently, the public data of OLS is the fourth edition. The previous editions of the data are no longer available.

OLS nighttime light remote sensing images have reduced systematic errors caused by light sources such as the sun, moon, and aurora. The sunlight is eliminated according to the solar elevation angle. The moonlight is eliminated according to the luminousness. The cloud layer is eliminated according to the OLS thermal infrared spectroscopy data and the National Centers for Environmental Prediction (NCEP) data. The original captured images were re-projected using Platte Carre projection covering from west 180 longitude to east 180 longitude, and from south 65 latitude to north 75 latitude. Every 30 arc second forms the minimal grid, approximately 1 km^2 around the equator. Each image is named after the satellite number and year. For example, F152000 represents the image of the F15 satellite in 2000, see ***Table 3-1***. On this basis, OLS nighttime light remote sensing image products are divided into 4 categories, all of which are available for download in geotiff format.

Year	F-10	F-12	F-14	F-15	F-16	F18
1992	F101992					
1993	F101993					

1994	F101994	F121994				
1995		F121995				
1996		F121996				
1997		F121997	F141997			
1998		F121998	F141998			
1999		F121999	F142000			
2000			F142001	F152000		
2001			F142002	F152001		
2002			F142003	F152002		
2003				F152003		
2004				F152004	F162004	
2005				F152005	F162005	
2006				F152006	F162006	
2007				F152007	F162007	
2008					F162008	
2009					F162009	
2010						F182010
2011						F182011
2012						F182012
2013						F182013

Table 3-1 Annual DMSP/OLS Remote Sensing Images

The first type of imaging products is cloud-free coverage images, which records all detected data. It can be used in low-light areas, but with limited quality. Observations are even zero in some years. The second type of image products is the average visible image, which records the observations obtained by the average visible spectrum with a DN value ranging from 0 to 63, and 255 for areas with no observations. The third type is stable lights product, which records stable lights in human settlements and industrial areas, excluding temporary lights such as natural flares. The background noise of the images is eliminated with DN value recorded as 0. Detected lights are recorded with DN value ranging from 1 to 63. The unobserved area is recorded with DN value equal to 255. This type of image is also the most widely used image data for studying human socioeconomic activities. The fourth type is the Average Lights Times Percent Image. This type of image uses the DN value of the detected visible light multiplied by the probability that the area is detected. The systematic error in the image is not eliminated.

There are three major errors in OLS stable lights data, namely intercalibration, saturation, and blooming. Intercalibration means that OLS lacks on-board calibration and inter-satellite calibration. When using long time series OLS data for analysis, images from different years or different satellites cannot be directly applied due to lack of consistency. Oversaturation means that since each pixel has only 6 bits, the DN value is up to 63. For some areas that are too bright, that is, areas where the DN value may exceed 63. The pixel points are only recorded at 63, which causes oversaturation. This phenomenon is very common in urban areas. Blooming means that for some areas covered with snow and water, light may be reflected to the adjacent area after being exposed to snow and water, causing bright light to be detected in the vicinity of the dull area. These light are detected with no human activities.

3.2 Introduction of NPP/VIIRS Nighttime Light Imagery

This section will introduce the development history of NPP/VIIRS, sensors, advantages and disadvantages of images, image sources, and image product classification.

As mentioned before, NPP/VIIRS is called Suomi National Polar-orbiting Partnership/visible infrared imaging radiometer suite, hereinafter referred to as VIIRS, which is the successor of OLS. Since December 1st, 2011, VIIRS nighttime light remote sensing images have been electronically archived and published by NOAA/NGDC. The VIIRS has a Day/Night Band (DNB) that can be used to detect bright light. VIIRS has higher spatial and radiometric resolution than OLS. It dose not have three critical errors which OLS has. On November 18, 2017, NPP/VIIRS' successor JPSS-1 (Joint Polar Satellite System-1) was successfully launched and renamed NOAA-20. The project will continue to provide updated images for nighttime light remote sensing.

Suomi NPP, launched on October 28th, 2011, carries five Earth observation sensors, including VIIRS. The VIIRS multispectral sensor has 22 bands, including the DNB for detecting nighttime light (NTL). The VIIRS scan width is kept within 3040km and the spatial resolution is 742m, which is 45 times the position of the OLS at nadir point and 88 times at the edge of the scan. The radiation resolution of VIIRS is 256 times higher than OLS. The radiation sensitivity is 10 times higher. VIIRS can solve the problem of low brightness with small detection interval. The measured spectral response is 505-890nm full width at half maximum, which is another significant improvement over OLS. Specifically, VIIRS can detect the change in brightness of a streetlight within 1km^2, so VIIRS images can be combined with ground-based nighttime measurements to obtain multi-dimensional luminous information.

However, VIIRS still cannot give higher spatial resolution and more spectral bands. In addition, VIIRS is not superior to OLS in all respects. For example, some cities have switched from using incandescent lamps to using more energy-efficient LED lamps. VIIRS will therefore detect a reduction in brightness due to its insensitivity to wavelengths below 500 nm. OLS is more likely to detect LED light because it has a slightly wider wavelength, from 0.4 to 1.1μm.

The VIIRS data pixels are available in 12 and 14 bits, depending on the noise level, and are ultimately measured in nanoWatts/(cm2•sr). Although the measurement range of VIIRS is from 3×10^{9} to $0.02\ \mathrm{W\cdot cm^{-2}\cdot sr^{-1}}$. The noise is actually limited within $5\times10^{-11}\ \mathrm{W\cdot cm^{-2}\cdot sr^{-1}}$, so that extremely weak brightness can be detected. Its three gain stages allow for the detection of larger dynamic radiation ranges during the day, dusk and night. Its over-the-top time is about 1:30 am every day at nadir point. Compared to the over-the-top time of OLS from 20:30 to 21:30, the VIIRS can detect less ground light.

Unlike the US Department of Defense, VIIRS is a satellite project jointly operated by NASA and NOAA. Its images are freely available, including monthly and daily nighttime light remote sensing images. The most original sensor data and geographic location information of the VIIRS is available from NOAA's Comprehensive Large Array-DATA Stewardship System (CLASS) and the University of Wisconsin NASA Atmospheric SIPS server. NOAA offers processed VIIRS global and regional imaging products, and the most widely used of which is the first version of the VIIRS DNB nighttime light remote sensing image.

In the VIIRS raw data, DNB filters out stray light and moonlight. The cloud cover is filtered by the VIIRS Cloud Mask product (VCM). Images taken at the edge of the scan are not used in the integrated image. The first version of the DNB data has undergone monthly and annual averages. The monthly image did not filter out aurora, flares, fishing boat operations and other temporary light. The annual image filtered out temporary light and background noise. But currently there is only annual VIIRS data for 2015. In monthly images, there are many places in the world that are difficult to obtain good data quality in some years. Cloud cover is part of the reason. The intensive light in the equatorial region and the whiteness around the Earth polar in summer are also important reasons. The first version of the DNB data product has a minimum grid of 15 arc seconds, covering a global area from 65 degrees south latitude to 75 degrees

north latitude. One image is divided into six tiles for release.

The first version of the DNB monthly data is available in two configurations. The first type excludes any data affected by stray light, referred to as VCM data. The second contains data corrected by stray light, referred to as VCMSL data. For VCMSL, there is more data coverage near the Polar Regions. But the data quality is reduced. The annual data only excludes the configuration of stray light, like VCM data. In this paper, the first type of data will be used for research. The VCM data in the DNB monthly data of the first edition of VIIRS is referred to as VIIRS data.

The ***Table 3-2*** below compares OLS and VIIRS in some key aspects.

Satellites	DMSP/OLS	NPP/VIIRS
Operator	DOD	NASA/NOAA
Electronic Data	From 1992 to 2013	From Dec. 2011 till now
Wavelength Range	0.4-1.1μm	505-890μm
Spatial Resolution	2.7km	742m
Phase Resolution	12h	12h
Coverage	From 65°S to 75°N From 180°E to 180°W	From 65°S to 75°N From 180°E to 180°W
Day/Night Over-the-top Time	Day: 8:30-9:30 Night: 20:30-21:30	Day: 13:30 Night: 1:30
Radiometric Resolution	6 digits	12 or 14 digits
Unit	DN value from 0 to 63 (relative measurement)	nanoWatts/(cm^2.sr) (absolute measurement)
Upper Limit	$3.17\times10^{-7}W\cdot cm^{-2}\cdot sr^{-1}\cdot\mu m^{-1}$	$0.02W\cdot cm^{-2}\cdot sr^{-1}\cdot\mu m^{-1}$
Lower Limit	$1.54\times10^{-7}W\cdot cm^{-2}\cdot sr^{-1}\cdot\mu m^{-1}$	$3\times10^{-9}W\cdot cm^{-2}\cdot sr^{-1}\cdot\mu m^{-1}$
On-orbit Calibration Devices Equipped	No	Yes

Table 3-2 Attributes Comparison of DMSP/OLS and NPP/VIIRS

3.3 Method of Nighttime Light Data Extraction

From the raw nighttime light remote sensing images to quantifiable data, a series of processing is needed. The ArcMap of ArcGIS 10.5 software is used to process nighttime light remote sensing images. Python is used to wash, analyze and visualize data. Both DMSP/OLS and NPP/VIIRS need to go through the processes as follows.

(1) Downloading and Importing Raw Nighttime Light Remote Sensing Images into ArcMap

As introduced, the brightness information of nighttime light remote sensing images is recorded as Digital Number (DN) in pixels. The higher the DN value, the brighter the pixel is. The essence of extracting nighttime light remote sensing data is to extract these DN values.

(2) Downloading and Importing China Administrative Vector Map into ArcMap

Because the study object of this paper is Heilongjiang province in China, the China administrative vector map should be imported into ArcMap. The vector maps used here is the Boundaries Map 4.3 from the Wambachers-OSM website (https://wambachers-osm.website/boundaries/). The base map is the OpenStreetMap. It allows users to choose districts at different administrative levels. The maps can be exported in various formats. This website is built and maintained by Walter Nordmann.

Figure 3-1 China Provincial-level Vector Map

The figure above is the China administrative vector map used in this paper, which includes 31 provincial district, Hong Kong, Macau and Taiwan. The Demjok area in Ngari Prefecture in Tibet Autonomous Region which is under territorial disputes between China and India is not exhibited here in the map. But the data in this area is included in Tibet Autonomous Region in the research.

(3) Building Circular Exporting Model in ArcMap, Obtaining Remote Sending Data Files

in DBF format.

Long time series nighttime light remote sensing images need to be processed for long time series studies. To increase the efficiency of image processing, a circular exporting model is built in ArcMap to extract images in batches. The figure below shows how this model looks like for dealing with DMSP/OLS images.

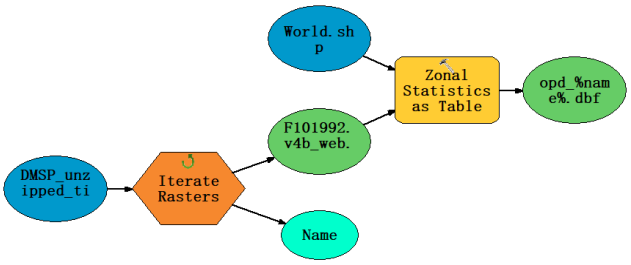


Figure 3-2 ArcMap Raster data output model

(4) Using Python to read, analyze and visualize DBF files

There are two quantitative research directions of nighttime light remote sensing data. One is to study the distribution of pixels in an area. The other is to study the changes of DN values. When using nighttime light data to study urbanization, the distribution of pixels is studied to analyze the change of cities. In order to study China regional economy, this paper will use the Total Nighttime Light (TNL) which is the sum of all DN values of the research areas, as the indicator for economic development.

3.4 DMSP/OLS Nighttime Light Data Extraction and Analysis

As can be seen from previous sections, the DMSP/OLS data used in this paper is the stable light image product. The background noise and the temporary light such as natural flare are eliminated. The stable light of human residential areas and industrial areas is recorded. The image pixel is recorded in 6 bits, that is, the DN of the pixel is 0 in the area where no bright light is observed, and 1 to 63 in the area where the light is detected according to the DN value, and 255 in the area where the light is not successfully observed. Its available electronic imaging products are archived from 1992 to 2013. Stable light images are also the most widely used image data for studying human socioeconomic activities. Figure below shows the global nighttime light remote sensing map of the F18 satellite in 2013.

Figure 3-3 F182013 Global Nighttime Light Remote Sensing Image

It can be intuitively sensed from the above picture that the better the economic development and the more frequent human activities, the higher the brightness on the OLS nighttime light remote sensing image, such as the US East and West coast, most of Europe, Japan and Beijing, Tianjin in China etc.. The weaker the economic development, the rarer the human activities, the lower the brightness on the OLS nighttime light remote sensing image is, such as the inland of Africa, parts of inland of South America, the Siberia and the Far East of Russia, the Mongolian, the Tibetan Plateau of China, etc.. Therefore, OLS nighttime light remote sensing images are considered to be more straightforward to reflect the level of economic development and human activities.

Figure 3-4 Re-projected F182013 China Nighttime Light Remote Sensing Image

As shown in ***Figure 3-4***, the original OLS global nighttime light remote sensing image is cropped according to the Chinese administrative division vector. The WGS 1972 BE

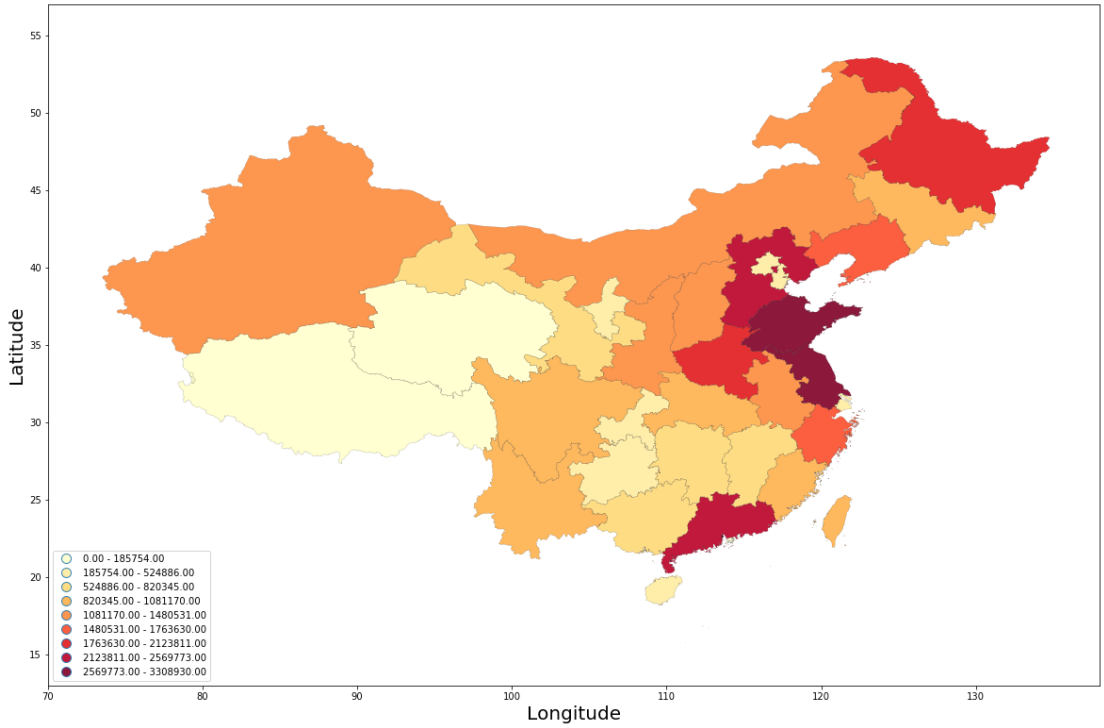
South China Sea Lambert projection coordinate system is constructed. The OLS China nighttime time light remote sensing image can be obtained after cropping and re-projection. Similarly, the brighter nights in the southeastern coastal regions and the provincial capitals of the inland provinces reflect the higher level of economic development and the more frequent human activities in these areas.

Then a raster data exporting model is built in ArcMap to extract nighttime light remote sensing data in loops. The model output is a DBF format file, which can be opened in tabular form. Since it contains much data, the following table only selects some parameters of some provinces for display.

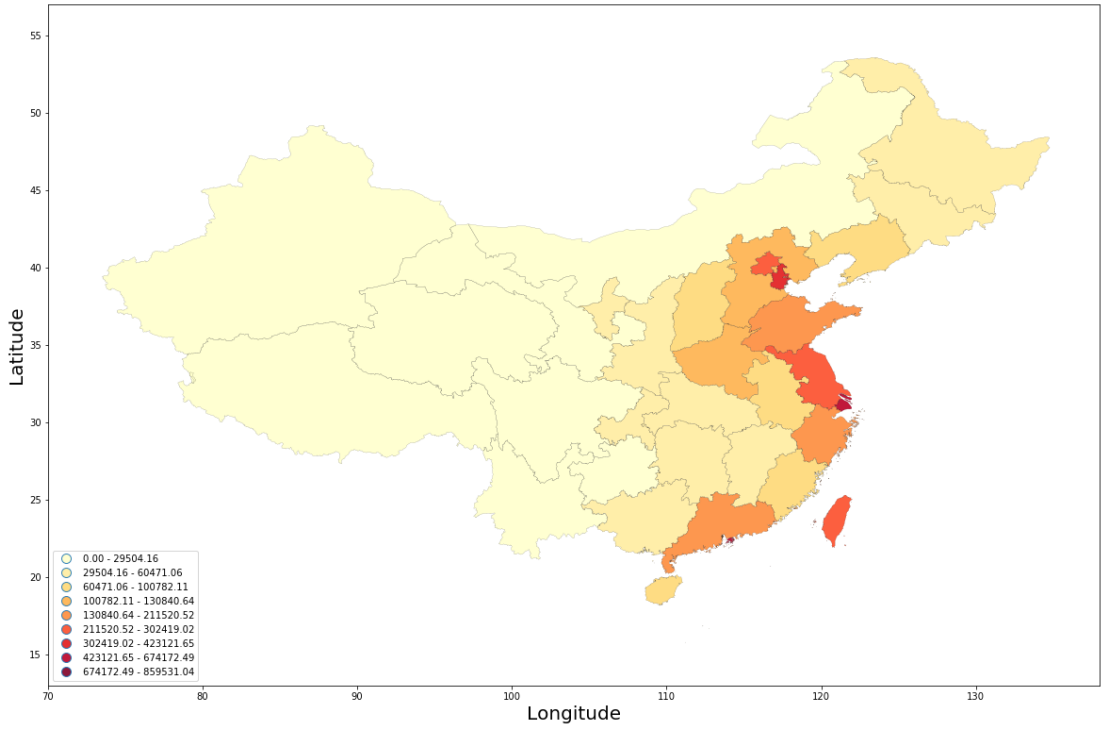
NAME	COUNT	AREA	MIN	MAX	SUM
Qinghai	999218	69.39013833380	0	63	185754.00000000000
Tibet Autonomous Region	1547855	107.48992969600	0	63	56582.00000000000
Xinjiang	2522731	175.18965137600	0	63	1383496.00000000000
Gansu	629102	43.68763853940	0	63	702958.00000000000
Inner Mongolia	1857445	128.98923507900	0	63	1480531.00000000000
Jilin	306838	21.30819427400	0	63	979905.00000000000
Heilongjiang	782442	54.33624956530	0	63	2123811.00000000000
Guangxi	301925	20.96701372120	0	63	820345.00000000000
Guizhou	230203	15.98631931660	0	63	471663.00000000000
Henan	232394	16.13847209310	0	63	2111568.00000000000
Taiwan	46472	3.22722219640	0	63	904889.00000000000
……	……	……	……	……	……

Table 3-3 F182013 Nighttime Light Output Data (partial)

In ***Table 3-3***, NAME is the name of the administrative district. COUNT is the number of pixels included in the administrative district. AREA is the size of the administrative area. MIN and MAX are the maximum and minimum values of the DN in the administrative district. SUM is the sum of DN values of pixels of the administrative district. According to the previous section, this paper will focus on extracting the SUM value to represent the economic development level of the administrative region. The DBF file can be read, processed, and visualized with Pandas in Python.



a) F182013 TNL Distribution Map of China Provincial Districts



b) F182013 Unit Area TNL Distribution Map of China Provincial Districts

Figure 3-5 F182013 China Nighttime Light Remote Sensing Visualization Map

The ***Figure 3-5*** above shows the results obtained by visualizing with Python. Some areas of the South China Sea are not shown here. But all data in the South China Sea are included in the subsequent analysis. In the ***Figure 3-5*** **a)**, the color depth of each

administrative area reflects the sum of the DN values of all the pixels contained in the corresponding area. The results show that the darkness of the colors may be related to the degree of economic development, the administrative area, and the nightlights associated with non-economic activities in the administrative district. The larger the administrative area, such as Xinjiang, Heilongjiang, etc., the more pixels it contains, so the deeper the color depth. While Beijing, Tianjin, Shanghai, Chongqing, etc., due to their relative smaller areas, contain fewer pixels. So the color depths are smaller. The annual forest fires in Heilongjiang Province and the burning of oil and gas fields in Xinjiang Autonomous Region have also added many light sources with little or no relationship with human activities. In addition, due to the oversaturation of the pixel points, the DN value detected in a place with a high degree of economic development is smaller than its actual nighttime light level. In order to study the relationship between nighttime light data and economic development, the area factor is excluded. The color depth of each administrative area in the ***Figure 3-5*** **b)** reflects the sum of the DN values of all the pixels contained in the administrative area, divided by the value of the administrative area. The results show that the color depths of Beijing, Tianjin, Jiangsu, Shanghai, Taiwan, Hong Kong and Macao are in the first echelon, and the color depth changes gradually from the southeast coast to the inland central and western regions, which is basically in line with the economic development level of various regions in China.

Consistency in data is a basic condition for long time series studies. According to the document description of the original nighttime light remote sensing image, the OLS sensor has no temporal consistency due to the lack of on-board and inter-satellite calibration devices. In order to visually reflect this time inconsistency and facilitate the later research, Python will be used to process and visualize long time series OLS data.

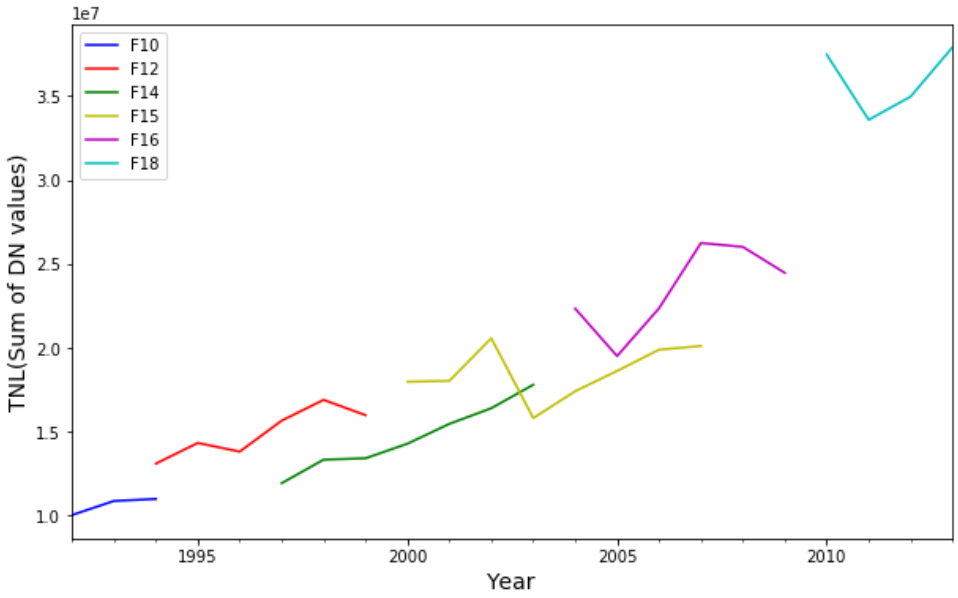


Figure 3-6 DMSP/OLS China Nighttime Light Data

The ***Figure 3-6*** above shows the data of all 34 images of DMSP/OLS from 1992 to 2013. Each point represents the sum of all DN values of the whole country taken in that year, namely TNL. If it is assumed that all of these data are correct and error-free, and that there is a strong positive correlation between the luminous remote sensing data and the degree of economic development, then the following conclusions can be drawn: (1) The DN value of each corresponding pixels should be the same by two different satellites. Therefore the sum of DN values should also be the same; (2) the long time series TNL should be gradually increasing. And the TNL of the later years should not have large drawdowns compared to the TNL of the previous years.

It can be seen from ***Figure 3-6*** that the above two points are not satisfied based on the conclusions assumed in the ideal case. For the error of the data, the following possible explanations can be drawn: (1) Due to the lack of inter-satellite calibration devices, images taken by different satellites do not have uniform standards in time, that is, lack of consistency among different satellites; (2) due to lack of on-board calibration device, the images captured by the same satellite in different years do not have a uniform standard in time, that is, the same satellite lacks temporal consistency; (3) the on-board load may attenuate with the increase of running time, resulting in a decrease in radiation detection sensitivity, and the detected brightness will therefore be lowered.

3.5 NPP/VIIRS Nighttime Light Data Extraction and Analysis

As can be seen from the previous chapter, the NPP/VIIRS products used in this paper

are the VIIRS first-version DNB (Day/Night Band) nighttime light remote sensing images. Due to the lack of a complete series of annual images, this article uses monthly image that is a VCM configuration which removes all stray light. The product has been updated since April 2012. This paper selects monthly VIIRS images from April 2012 to December 2017. Each pixel in the image has 12 or 14 digits, which is enough to detect all the light. And there is no oversaturation phenomenon which exists in the OLS image. Since the VIIRS image is divided into 6 tiles, the image used in this article is one of the 6 tiles, ranging from latitudes of 0-75 degrees north and longitudes of 60-180 degrees east, just covering the whole of China. The ***Figure 3-7*** shows the VIIRS data for December 2016.

Figure 3-7 VIIRS Nighttime Light Remote Sensing Image on Dec. 2016

The OLS and VIIRS images selected in this section are relatively brighter images. The VIIRS image was taken in December 2016 and the OLS images were taken in 2013. In theory, the brightness perception of VIIRS images should be much stronger than that of OLS images. However, the comparison of the two images shows that VIIRS has fewer bright regions, resulting in lower brightness perception. There are two reasons for this phenomenon. Firstly, because there is no saturation phenomenon in the VIIRS images, the recorded brightness is true brightness, so the range of DN value is much larger than the OLS. Therefore, the full brightness DN value 63 in the OLS may not be the brightest in the VIIRS. Secondly, the OLS annual image is an integration of the OLS daily image, covering one year. But the VIIRS image covers only one month. So the two have differences on the time length of observation.

Figure 3-8 Re-projected China Nighttime Light Image on Dec. 2016

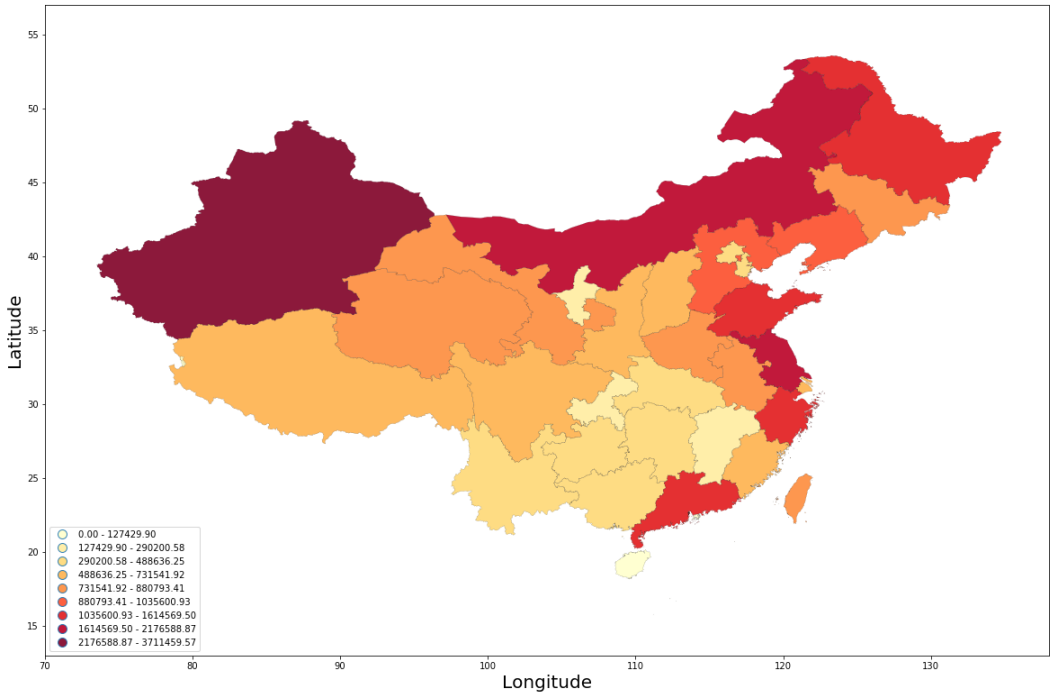
Similarly, the original VIIRS image was cropped using the China administrative division map and re-projected using the WGS 1972 BE South China Sea Lambert projection coordinate system to obtain the image shown in the Figure above. Applying a similar raster data circular exporting model in ArcMap, the VIIRS nighttime light remote sensing data can be extracted and shown below.

NAME	MIN	MAX	SUM
Qinghai	-0.17000000179	288.35998535200	765688.27027500000
Tibet Autonomous Region	-0.38999998570	120.62000274700	624003.73976800000
Xinjiang	-0.41999998689	2051.15991211000	3711459.56945000000
Gansu	-0.18000000715	263.77999877900	760907.11005000000
Inner Mongolia	-0.28000000119	4153.08984375000	2176588.87136000000
Jilin	0.00999999978	219.80999755900	833054.15005300000
Heilongjiang	-0.00999999978	355.82998657200	1419876.32033000000
Guangxi	-0.37000000477	316.79998779300	412536.23002000000
Guizhou	-0.77999997139	497.17999267600	367017.72002300000
Henan	-0.14000000060	360.33999633800	880793.40997000000
Taiwan	-0.23999999464	369.89001464800	868470.08985700000
……	……	……	……

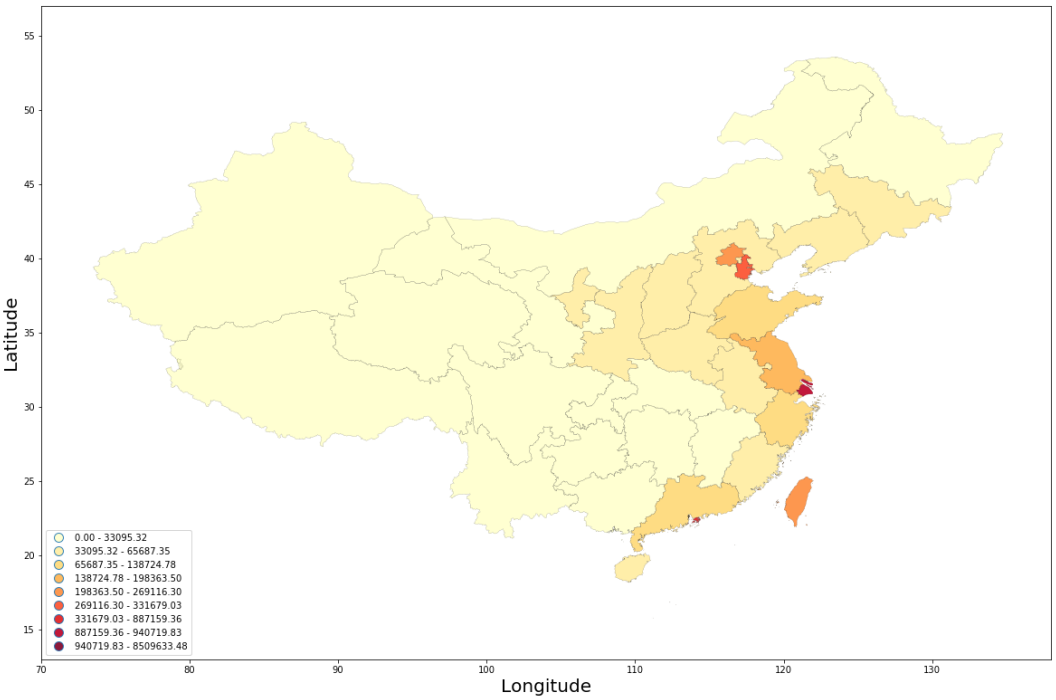
Table 3-4 NPP/VIIRS Nighttime Light Output Data on Dec. 2016 (partial)

The ***Table 3-4*** is not a complete output. Several parameters and administrative areas

are omitted. Comparing the OLS output data with the VIIRS output data, the VIIRS DN value can completely reflect the nighttime light brightness. And its detection range is much larger than the 0-63 range of the OLS. This article will then visualize the VIIRS monthly nighttime light remote sensing data for December 2016.



a) VIIRS TNL Distribution Map of China Provincial Districts on Dec. 2016



b) VIIRS Unit Area TNL Distribution Map of China Provincial Districts on Dec. 2016

Figure 3-9 NPP/VIIRS China Nighttime Light Visualization Map on Dec. 2016

By looking at the figure above and comparing it with visualized OLS image, similar conclusions can be drawn. The Python will be applied to visualized long time series VIIRS data.

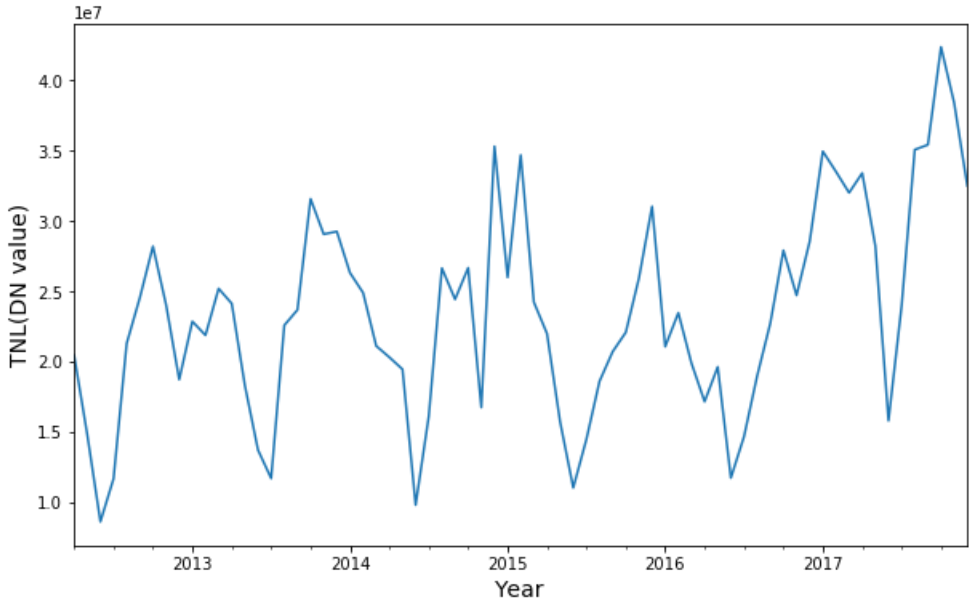
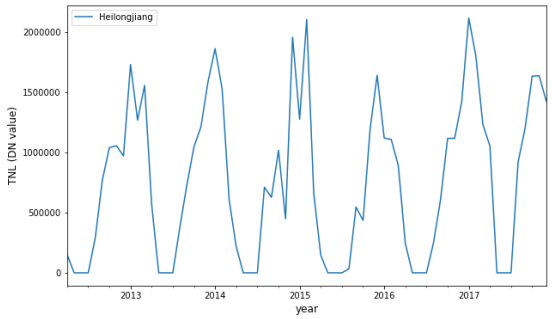
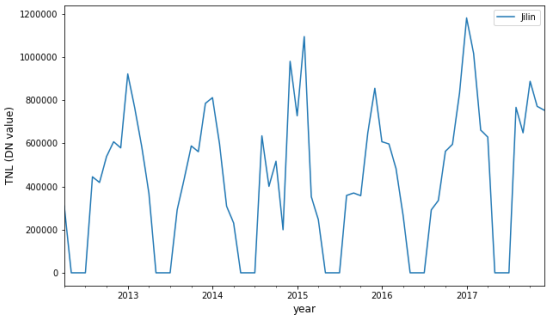


Figure 3-10 NPP/VIIRS China Nighttime Light Data

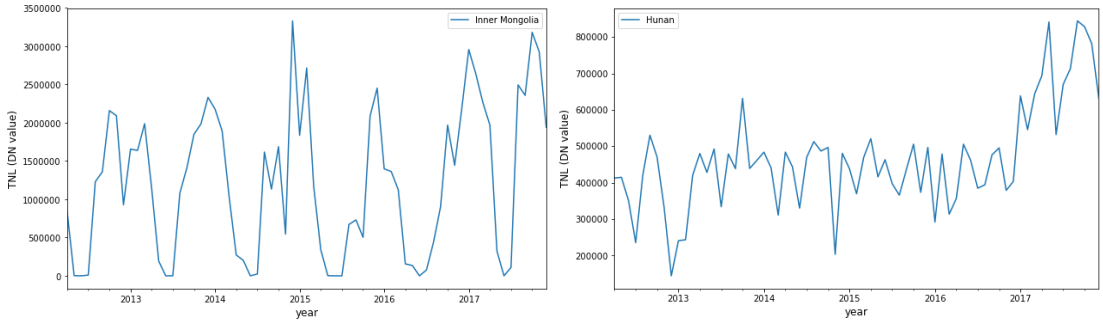
The above ***Figure 3-10*** is the visualization of monthly NPP/VIIRS nighttime light remote sensing data. Each point in the graph represents the sum of DN values in China, namely the national TNL. There are two obvious characteristics of the data in the graph. Firstly, the overall data rises gradually with time. Secondly, the data has seasonal volatility, with lowest value on May, June and July each year, and highest around December.



a) NPP/VIIRS Data in Heilongjiang Province

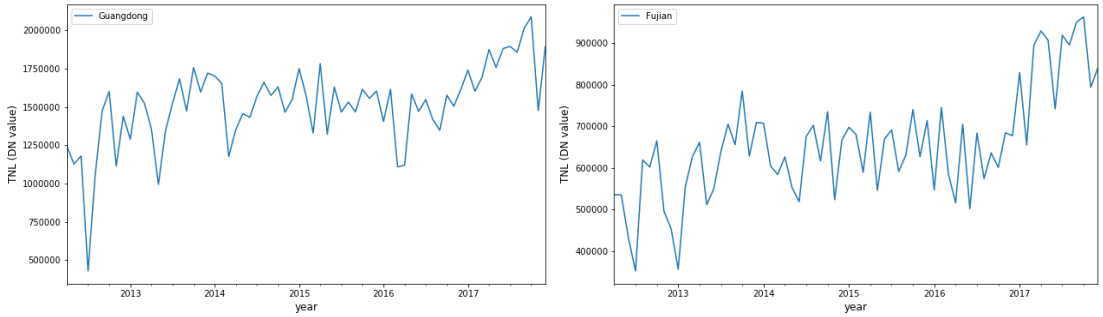


b) NPP/VIIRS Data in Jilin Province



c) NPP/VIIRS Data in Inner Mongolia

d) NPP/VIIRS Data in Hunan Province



e) NPP/VIIRS Data in Guangdong Province

f) NPP/VIIRS Data in Fujian Province

Figure 3-11 NPP/VIIRS Nighttime Light Data of Six Provincial Districts in China

In order to analyze the seasonal volatility of the national monthly VIIRS nighttime light remote sensing data in Figure 3, this paper plots the monthly data of VIIRS in all 34 provincial administrative regions. Six representative regions are selected and shown in ***Figure 3-11***. It is observed that in the northern part of China, such as Heilongjiang, Jilin, Inner Mongolia, there are cases where TNL is zero in all or some months in May, June and July. And the seasonal fluctuation characteristics are particularly obvious. In southern China, such as Hunan, Guangdong, and Fujian, though TNL still has fluctuations, the seasonal correlation is weak. And there is no case where TNL is zero. It is described in VIIRS data manual that in June at the north hemisphere and December at the south hemisphere, the Polar Regions are in summer. Due to the intense sunlight, VIIRS sensors cannot obtain high quality data. Therefore, the data detected during these periods are deemed as stray light. Because the VCM data used in this paper has removed all stray light, the TNL in north China in summer is zero. Since the national TNL is obtained by adding TNLs from various regions, the long-time series national TNL also exhibits strong seasonal fluctuations, with smaller summer TNL and larger winter TNL.

3.6 A Summary of the Chapter

This chapter introduced DMSP/OLS and NPP/VIIRS projects in general. The chapter also

pointed out the reason why the fourth version of DMSP/OLS stable lights product and the first version DNB VCM of NPP/VIIRS were chosen, and analyzed both products. After obtaining the raw nighttime light remote sensing images, this chapter proposed the procedure and methods of preprocessing, analyzing by ArcGIS and dealing with, analyzing, visualizing by Python. Then, the paper applied this process and method to deal with and visualize DMSP/OLS and NPP/VIIRS respectively. By reviewing data outcome and visualized figures, this chapter described and analyzed the inconsistency both in a single satellite and different satellites. Then the monthly variation and seasonal volatility has been depicted and analyzed. By dividing the outcome into several subsections, the reason behind the seasonal volatility of NPP/VIIRS has been explained, which paves the following study.

4. Nighttime Light Data Processing

In DMSP/OLS long time series nighttime light remote sensing, image data has two types of errors: systematic error and random noise. The main causes of these errors are the following: lack of on-board calibration device, small dynamic range of radiometer, saturation of signal acquisition et cetera. At the same time, changes in atmospheric conditions, satellite orbital deviation, on-board sensor attenuation, satellite over-the-top time differences, etc., also cause different DN values in the same region detected by different sensors. In addition, the algorithms introduce errors when processing annual nighttime light remote sensing images.

In this chapter, section 4.1 will apply the two most widely used methods to intercalibrate the latest OLS image data. In section 4.2, the data processing will be performed on VIIRS images. In section 4.3, this paper will combine the long time series OLS and VIIRS data to obtain the longest time series data ever since the release of the electronically archived data.

4.1 Methods for DMSP/OLS Nighttime Light Data Intercalibration

As introduced before, in order to study the global oil and gas combustion situation over a long period of time, Elvidge et al. (2009) proposed a polynomial regression intercalibration method of OLS, which is referred to as the Elvidge method. The method became one of the most widely used methods in the field of intercalibration of DMSP/OLS nighttime light data. The method will be introduced in detail below. The latest intercalibrated data will be obtained by applying this method.

The method was first applied in DMSP/OLS Average Lights X Product in 2009. In the paper, Elvidge firstly selected a large number of human settlements around the world. The total nighttime light (TNL) was then calculated in these areas separately. It turned out that the F12 satellite nighttime light remote sensing image in 1999 (F121999) had the highest TNL. Therefore, the F121999 was chosen as the reference image. All other annual images were calibrated to the F121999 standard. The reason why the image with the highest TNL is chosen is because it has the most of saturated pixels, namely pixels with the DN value 63. It will prevent data that does not appear saturated from becoming saturated.

The reference area needs to be selected when the reference image has settled. There are two standards for selecting reference area.

First, the variation of nighttime light in reference area needs to be as stable as possible. Theoretically, the nighttime light in this area needs to remain unchanged during long time period. So that when the nighttime light data of different years are intercalibrated, there is a unified scale. However, there is almost no place in the world where the nighttime light remains unchanged from 1992 till now. Therefore, it is only possible to find a reference area that is relatively stable over a long period of time.

Second, the DN values of pixels in the reference area need to be evenly distributed in the range of 0-63. DN values in most regions are usually concentrated at both ends of the DN value range, that is, close to 63 or 0. When conducting intercalibration using regression method, the evenly distributed DN values can remarkably increase the fitting precision. That is the reason why it is better to be an even distribution.

Based on the two criteria above, Elvidge chose Sicily in Italy as the reference area for OLS data intercalibration. Then each year's data was paired with F121999. The pair of data was drawn in a coordinate, with F121999 data in y axis, and the other year data in x axis. The formula below was used to fit the scatter plot.

$$DN_{adjusted} = C_0 + C_1 \times DN + C_2 \times DN^2$$

When the coefficients of the binary quadratic polynomial were calculated by regression, the above formula were then applied to intercalibrate all OLS images. The long time series OLS nighttime light data was then obtained.

After Elvidge proposed this method in 2009, most of the subsequent OLS long time series nighttime light remote sensing studies were intercalibrated using the Elvidge method, though the choices of reference images and reference areas varied in accordance with the corresponding study areas. Elvidge, Hsu, Baugh, and Ghosh (2014) reapplied this method to perform an intercalibration of OLS data, which has two differences from 2009. Firstly, the 2014 intercalibration is for the Stable Light Product, not for the Average Lights X Product. Secondly, the 2014 paper has intercalibrated a total of 33 OLS images from 1992 to 2012, while the 2009 article only intercalibrated the 2008 image. The intercalibration results in the 2014 paper are shown in ***Table 4-1*** below.

Satellite	Year	C0	C1	C2	R2
F10	1992	-2.0570	1.5903	-0.0090	0.9075
F10	1993	-1.0582	1.5983	-0.0093	0.9360
F10	1994	-0.3458	1.4864	-0.0079	0.9243

F12	1994	-0.6890	1.1770	-0.0025	0.9071
F12	1995	-0.0515	1.2293	-0.0038	0.9178
F12	1996	-0.0959	1.2727	-0.0040	0.9319
F12	1997	-0.3321	1.1782	-0.0026	0.9245
F12	1998	-0.0608	1.0648	-0.0013	0.9536
F12	1999	0.0000	1.0000	0.0000	1.0000
F14	1997	-1.1323	1.7696	-0.0122	0.9101
F14	1998	-0.1917	1.6321	-0.0101	0.9723
F14	1999	-0.1557	1.5055	-0.0078	0.9717
F14	2000	1.0988	1.3155	-0.0053	0.9278
F14	2001	0.1943	1.3219	-0.0051	0.9448
F14	2002	1.0517	1.1905	-0.0036	0.9203
F14	2003	0.7390	1.2416	-0.0040	0.9432
F15	2000	0.1254	1.0452	-0.0010	0.9320
F15	2001	-0.7024	1.1081	-0.0012	0.9593
F15	2002	0.0491	0.9568	0.0010	0.9658
F15	2003	0.2217	1.5122	-0.0080	0.9314
F15	2004	0.5751	1.3335	-0.0051	0.9479
F15	2005	0.6367	1.2838	-0.0041	0.9335
F15	2006	0.8261	1.2790	-0.0041	0.9387
F15	2007	1.3606	1.2974	-0.0045	0.9013
F16	2004	0.2853	1.1955	-0.0034	0.9039
F16	2005	-0.0001	1.4159	-0.0063	0.9390
F16	2006	0.1065	1.1371	-0.0016	0.9199
F16	2007	0.6394	0.9114	0.0014	0.9511
F16	2008	0.5564	0.9931	0.0000	0.9450
F16	2009	0.9492	1.0683	-0.0016	0.8918
F18	2010	2.3430	0.5102	0.0065	0.8462
F18	2011	1.8956	0.7345	0.0030	0.9095
F18	2012	1.8750	0.6203	0.0052	0.9392

Table 4-1 Polynomial Intercalibraion Coefficients by Elvidge in 2014

Based on the Elvidge method of 2014, this paper will intercalibrate the OLS data of F182013, and obtain the long time series OLS nighttime light remote sensing data from 1993 to 2013.

At first, the vector maps of Sicily were still downloaded from Boundaries Map 4.3 on the Wanbachers-OSM website. After cropping the global OLS image through ArcMap, the OLS images of F121999 and F182013 in Sicily were obtained. Then the two raster images were read into Matlab, each pair of pixels were plotted into the two-dimensional coordinate. The previous formula were used to regress the scatter plot. The result is shown in ***Figure 4-1*** below.

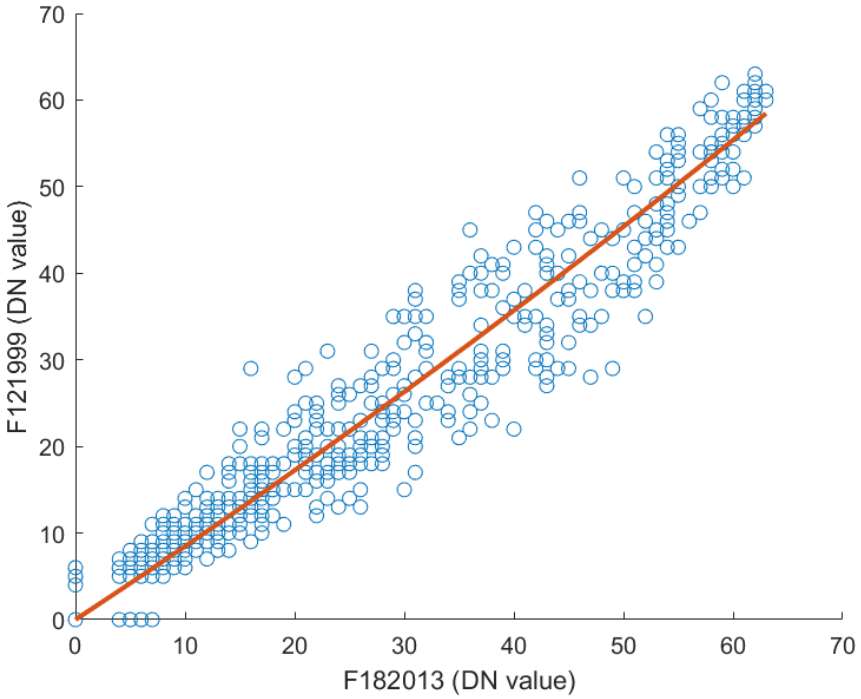


Figure 4-1 F182013 Intercalibration Fitted Curve by Elvidge Method

The fitted curve is as follows.

$$DN_{F182013adjusted} = 0.0023 + 0.8312 \times DN_{F182013} + 0.0015 \times DN_{F182013}{}^{2}$$

The coefficients are here.

Satellite	Year	C_0	C_1	C_2	R^2
F18	2013	0.0023	0.8312	0.0015	0.9780

Table 4-2 F182013 Intercalibration Coefficients of Elvidge Method

Using the binary quadratic polynomial fitting coefficients of ***Table 4-1*** and ***Table 4-2*** to calibrate the OLS nighttime light remote sensing data from 1992 to 2013, the long time series OLS data calibrated using Elvidge method can be obtained as shown below. Comparing ***Figure 3-6*** and ***Figure 4-2***, it can be found that the single satellite error and

inter-satellite error are significantly reduced.

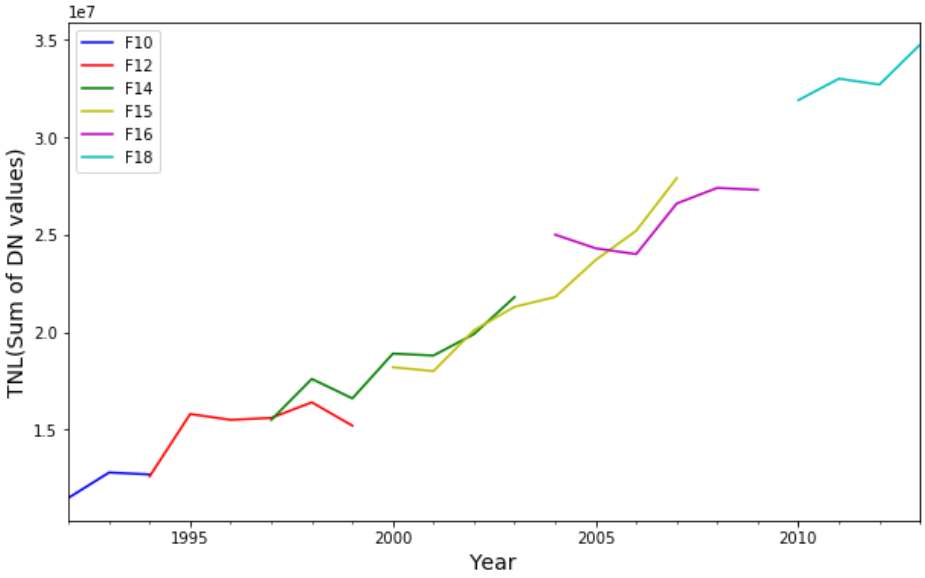


Figure 4-2 DMSP/OLS China Nighttime Light Data Calibrated by Elvidge Method

There is another DMSP/OLS intercalibration method called Ridgeline Sampling Regression proposed by Zhang, Pandey, and Seto (2016), hereinafter referred to as RSR method.

Due to the lack of observation stations on the ground that can be used for calibration, all intercalibration methods are using remote sensing images to find pseudo-invariant regions where the ground nighttime light remains stable over long periods of time. Most researches choose pseudo-invariant regions manually, with subjective interpretations of the researchers. At the same time, their studies are all based on the following two assumptions. First, the ground nighttime light basically remains unchanged in the selected reference area. Second, the error is globally consistent, namely the intercalibration of reference area applies to all regions of the world.

The purpose of intercalibration is to have consistency and continuity of OLS data in time series. In this article, Zhang proposed four criteria that must be met in order to achieve this goal: first, the reliability of pixel identification for pseudo-invariant region; second, the efficiency of intercalibration calculations; third, the universal applicability of intercalibration method out of the reference area; and fourth, the minimization of errors.

Zhang's RSR method proposed in the article meets the above four criteria. The method is mainly based on the assumptions that, in a short period of time, the changes of

nighttime light mainly come from a small number of pixels while the majority remains stable. Therefore, when plotting the density or frequency of DN values of the reference and target images, the densest places are where the most stable parts of two images. Changes mainly come from less dense places.

All OLS intercalibration methods use relative image intercalibration instead of absolute image calibration. That is, the absolute value of the OLS nighttime light remote sensing image measurement cannot be obtained. Therefore, the only applicable method is to calibrate target images to the reference image to make images in the same standard. The RSR method is also based on this principle.

The modeling process of the RSR method is as follow. It is assumed that the changes between the target image and reference image are small enough that most of pixels are stable worldwide. Therefore, most of the pixels can be considered as pseudo-invariant regions for constructing intercalibration model. This theory only applies to global images. Because sometimes regional pixels can vary dramatically. The theory also points out that the change between the target image and reference image with longer time interval may be more significant. Since the F152000 image is located in the middle of the long time series, it is selected as the reference image.

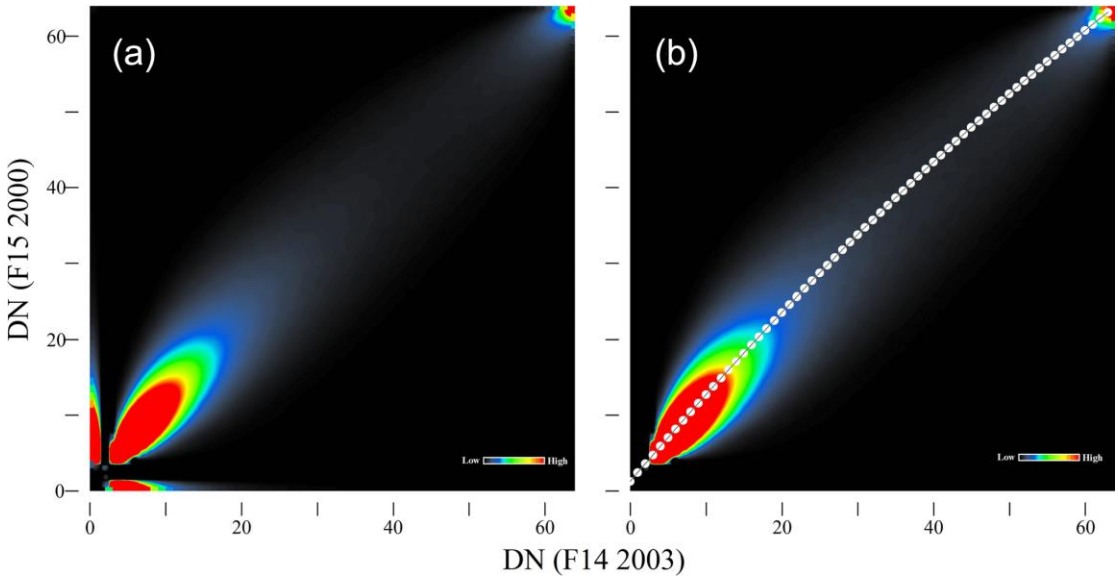


Figure 4-3 Dense Map and Ridgeline Regression Curve between F142003 and F152000 (Zhang, Pandey, & Seto, 2016)

The figure above is retrieved from Zhang's article. In the ***Figure 4-3***, the DN values of all pixels of F152000 and F142003 are in y axis and x axis respectively. Then the density plot can be drawn. The densest places are at two ends where DN values are 0 and 63.

It can also be seen from ***Figure 4-3*** **a)** that some of the pixels are scattered on the x-axis or y-axis. It means that those pixels have value in an image but zero in the other one. Those pixels mostly reflect the sudden changes of nighttime light on the ground, which may include the construction of dams and highways, natural flares and so on. It may also be caused by geographic mismatch of images or processing errors of annual images. Those pixels are not deemed as systematic errors and therefore rejected as random noise, see ***Figure 4-3*** **b)**.

The places with the highest density are the ridges. So the method is called the Ridgeline Sampling Regression. In theory, if the two images are identical, the ridgeline regression curve is a straight line. The curvature of the ridgeline indicated that the image needs to be calibrated. It can be seen from ***Figure 4-3*** **b)** that the RSR method only selects a total of 64 points with DN values from 0 to 63 for regression analysis. It does not apply regression analysis for all pixels for two reasons. First, the ridgeline regression analysis of all pixels is likely to cause the weight of the regression curve to be biased to both ends because the pixels are mainly concentrated at the lower and higher DN value. Second, the global image has a total of 43200 times 21600 pixels. Regression analysis of all pixels would make the calculation too heavy. Therefore, only 64 points were selected for ridgeline sampling regression. After selecting the points which are the densest, the following formula is applied for regression. The calculation of coefficients is based on ordinary least squares.

$$y = ax + bx^2 + c$$

In the article, Zhang collated the OLS data from 1992 to 2012. The calibrated images are publicly available on the Yale website (https://urban.yale.edu/data). This paper will conduct follow-up studies in accordance with this data. The ***Figure 4-4*** shows the results after extracting and intercalibrating using RSR method. Comparison of ***Figure 3-6*** and ***Figure 4-4*** shows that the OLS data error after RSR intercalibration is significantly smaller than the original raw OLS data.

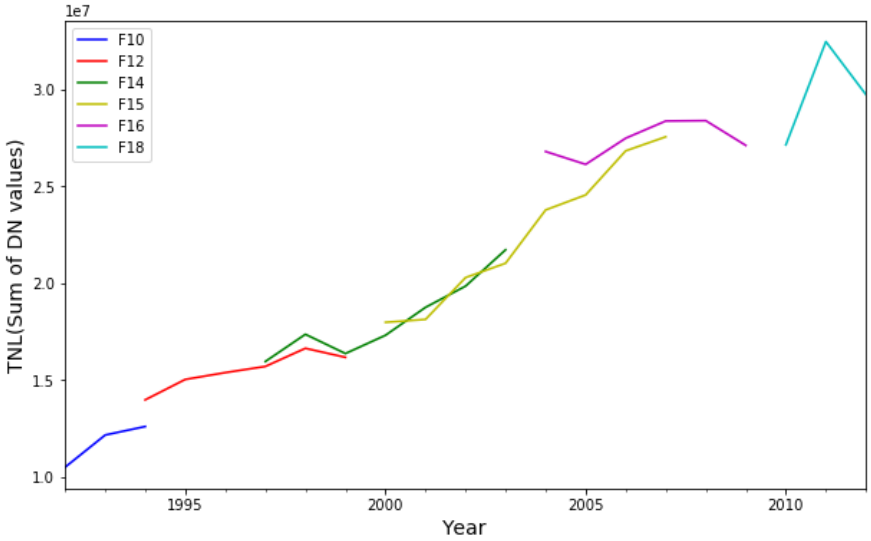


Figure 4-4 DMSP/OLS China Nighttime Light Data Calibrated by RSR Method

4.2 Assessment of DMSP/OLS Intercalibration Methods

How to obtain reliable long time series OLS nighttime light remote sensing data from 1992 to 2013 is the key to this study. Therefore, a comparative evaluation of the Elvidge method and the RSR method is required to obtain optimal calibrated data.

The evaluation of the DMSP/OLS data intercalibration will be based on the following three criteria. First, whether there are theoretical deficiencies in both intercalibration method. Second, for images taken by different satellites in the same year, the error is minimal after intercalibration. Third, the long time series nighttime light remote sensing data after intercalibration should show a steady growth trend as a whole.

Therefore, the intercalibration results will be evaluated in the following three aspects:
(1) Evaluation Based on Intercalibration Result

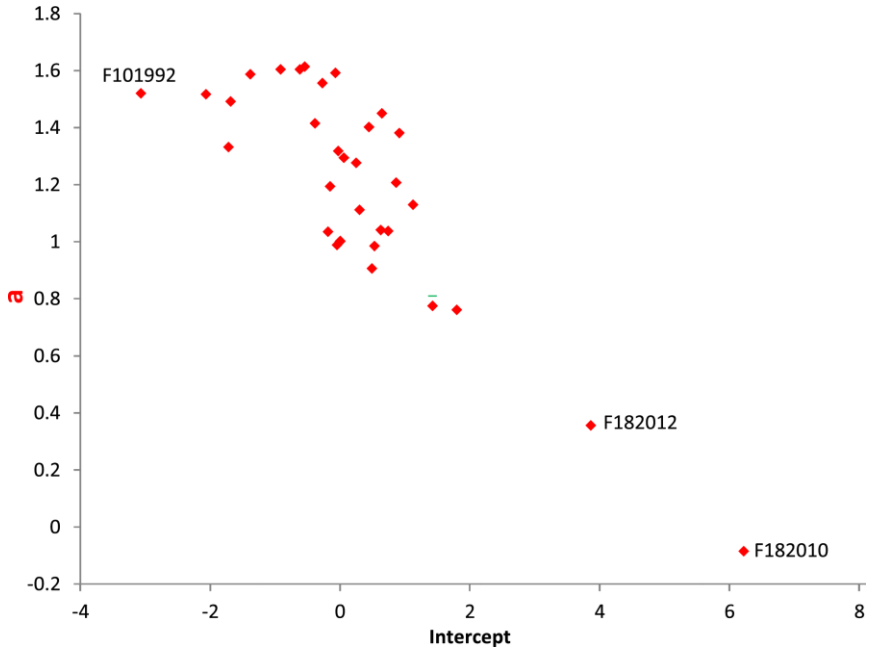


Figure 4-5 Coefficients Intercepts between the F152000 and Other Images after Intercalibration (Zhang, Pandey, & Seto, 2016)

Zhang stated in the article that the RSR method has very bad performance for F18 satellites, especially F182020 and F182012 (Zhang, Pandey, & Seto, 2016). The ***Figure 4-5*** is retrieved from the article. The horizontal axis represents the interval between the regression intercepts c and the target data. The vertical axis represents the interval between regression coefficient a and the target data. It can be found that the regression coefficients of F182010 and F182012 are quite different from others. The reason for this may be that F182010 and F182012 are far apart in time from the reference data F152000. However, F182011 does not have a large intercalibration error. So the errors of F182010 and F182012 may due to their own image quality. The conclusion can be drawn that the reliability of F182010 and F182012 calibrated by RSR method is low.

(2) Evaluation Based on Sum of Normalized Deviation Index

This paper believes that, theoretically, the data of the same year taken by different satellites should be identical after intercalibration, which is also a commonly accepted method for evaluating the intercalibration results. The ***Table 4-3*** below shows the years which have two satellite images.

Year	Satellite 1	Satellite 2
1994	F10	F12
1997	F12	F14
1998	F12	F14

1999	F12	F14
2000	F14	F15
2001	F14	F15
2002	F14	F15
2003	F14	F15
2004	F15	F16
2005	F15	F16
2006	F15	F16
2007	F15	F16

Table 4-3 Remote Sensing Images Shot by Two Satellites in the Same Year

By calculating the Total Sum of Lights (TSOL), the image differences in the same year could be compared. The formula is as follow:

$$SNDI = \sum NDI_t$$

$$NDI_t = \frac{|TSOL_{1t} - TSOL_{2t}|}{TSOL_{1t} + TSOL_{2t}}$$

$TSOL_{1t}$ and $TSOL_{2t}$ in formula above represent the TSOL values of two images in the same year in the reference area, which is China in this paper. NDI_t is the normalized index that guarantee the value ranging from 0 to 1. $SNDI$ is the sum of NDI_t. The smaller the $SNDI$ value, the smaller the difference between two images, the better the intercalibration results.

Based on this principle, the results of $SNDI_{Elvidge}$ and $SNDI_{RSR}$ is 0.2728 and 0.2678 respectively. The $SNDI_{raw}$ of the original OLS data is 1.1291. The results show that both intercalibration methods can calibrate the error of the original OLS data to a large extent. But the RSR method is slightly better than the Elvidge method.

(3) Evaluation Based on Steady Growth of Data after Intercalibration

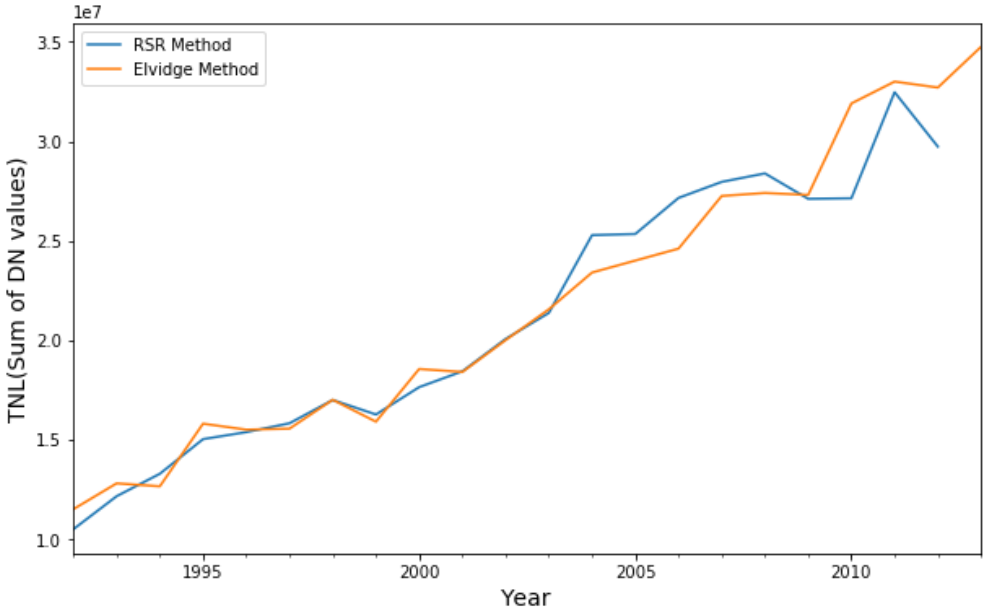


Figure 4-6 DMSP/OLS China Nighttime Light Data Calibrated by Two Methods

For the years which have two images, the average value of the two images after intercalibraion is used for the year. The long time series OLS nighttime light remote sensing data intercalibrated by the Elvidge method and RSR method are shown in ***Figure 4-6*** above.

China's economy has maintained steady and rapid growth over the past few decades. The nighttime light remote sensing data should reflect this trend to a certain extent. So the actual nighttime light data should follow this steady growth trend since 1992.

This paper proposes that the steady growth of the two intercalibration curves can be evaluated by calculating the number of drawdowns and the maximum drawdown rates. The formulas are as follows:

$$DD = \frac{OLS_t - OLS_{t+1}}{OLS_t} \times 100\%, OLS_{t+1} < OLS_t, DD > 0$$

$$DD_{\max} = \max\left\{\frac{OLS_t - OLS_{t+1}}{OLS_t} \times 100\%\right\}, OLS_{t+1} < OLS_t, DD_{\max} > 0$$

DD and $DD_{\max}$ represent the drawdown rate and maximum drawdown rate respectively. The smaller the better in both cases. OLS represents the calibrated OLS data. After calculation, the results are shown in the table below.

	Elvidge Method	RSR Method
Maximum Drawdown	0.0647	0.0842
Times of Drawdowns	6	3

Table 4-4 The Maximum Drawdown Rate and Times Comparison of two Methods

Although the drawdown times of RSR are smaller, the maximum drawdown rate is higher than Elvidge method. So the intercalibration effects cannot be judged.

From ***Figure 4-6***, the RSR method is better than the Elvidge method between 1992 and 2008, and inferior to the Elvidge method between 2008 and 2013. Therefore, if only the maximum drawdown rate and the number of drawdowns between 1992 and 2008 are considered, the results can be shown in the table below.

	Elvidge Method	RSR Method
Maximum Drawdown	0.0647	0.0430
Times of Drawdowns	5	1

Table 4-5 The Maximum Drawdown Rate and Times of two Methods from 1992 to 2007

From ***Table 4-5***, the RSR method is better in both maximum drawdown rate and number of drawdowns than Elvidge method from year 1992 to 2007.

In combination with the above three intercalibration quality evaluation criteria, this paper chose to use the RSR method to intercalibrate OLS data from 1992 to 2007, and the Elvidge method from 2008 to 2013. Since the Elvidge method and RSR method use F121999 and F152000 as reference images respectively, it is assumed that the difference between these two intercalibration methods can be approximated by the differences between the original F152000 and F121999 images. Therefore, the long time series OLS nighttime light remote sensing data shown in the following figure can be obtained. The overall trend is a steady upward trend.

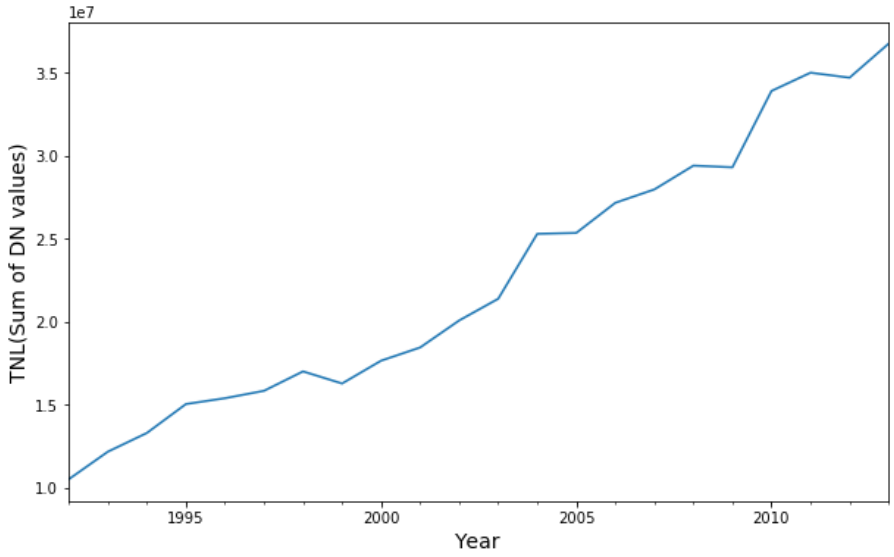


Figure 4-7 Long Time Series DMSP/OLS China Nighttime Light Data after Intercalibration

At the same time, the OLS data from 1992 to 2013 in Heilongjiang Province can be obtained as shown below in ***Figure 4-8***.

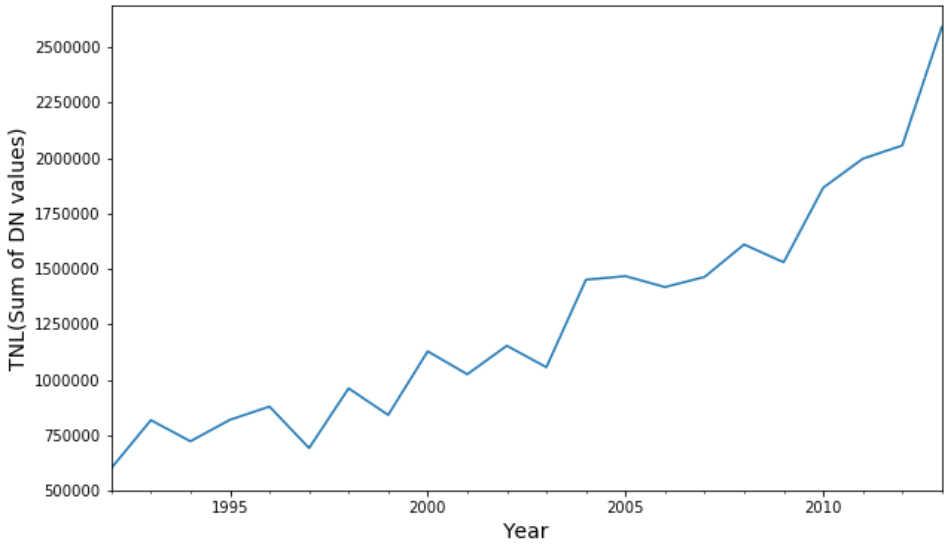


Figure 4-8 Long Time Series DMSP/OLS Nighttime Light Data after Intercalibration in Heilongjiang Province

4.3 NPP/VIIRS Nighttime Light Data Processing

The NPP/VIIRS nighttime light remote sensing data is the monthly data from April 2012 to December 2017. The ***Figure 4-9*** has shown the extracted China VIIRS data which is strongly periodic. Currently, only 2015 has one VIIRS annual image. In order to keep consistent with the OLS data in time resolution, this paper simply averages the monthly data to obtain the roughly annual data. The result is shown below.

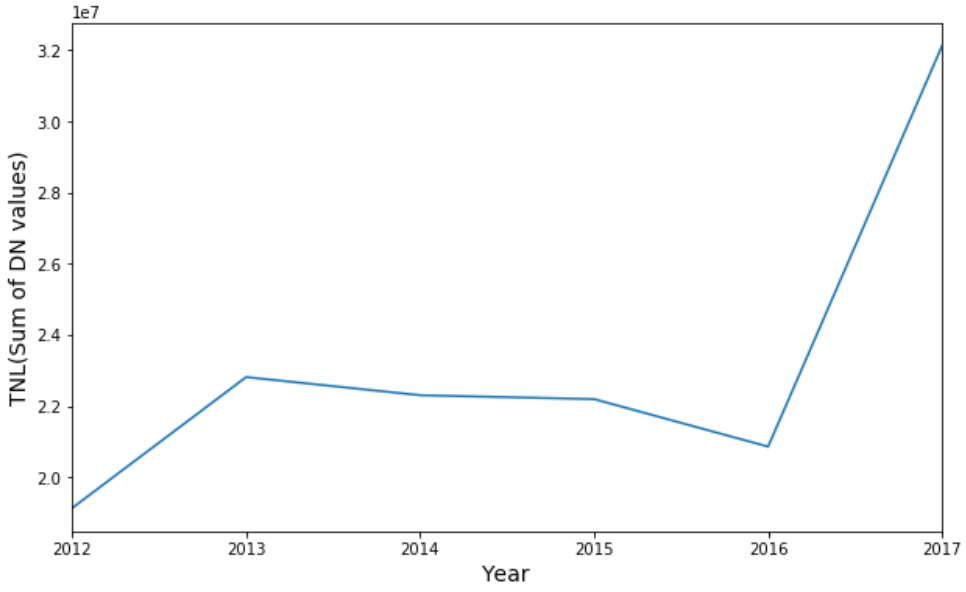


Figure 4-9 Annual NPP/VIIRS China Nighttime Light Data

As can be seen from ***Figure 4-9***, the simple calculated VIIRS data demonstrated a continuous slow downward trend from 2013 to 2016, and rose sharply in 2017.

This phenomenon is not in line with common sense. This occurrence may be caused by some VIIRS data errors. Based on studies from predecessors, the sources of errors might come from the aspects stated below.

Firstly, the VCM product of VIIRS DNB images has removed stray light. These stray light mainly concentrates at the equatorial region and the northern and southern hemispheres in their corresponding summer. Because the light intensity is not the same every year. So the light filtered by VCM is different, thus introducing errors.

Secondly, due to the serious haze in northern and some southern regions in China in recent years, the emergence of haze greatly affected the detection of bright radiation by the VIIRS sensor. The occurrence, duration and the concentration of haze in different regions in China vary from year to year, which leads to inaccurate detection of nighttime light in haze-covered areas.

Thirdly, in places where there are water, ice and snow, if there is nighttime light around, the water, ice and snow surfaces will reflect the nighttime light to the places where there is actually no light nearby, resulting in an increase in nighttime light. But the increases of nighttime light do not reflect any economic or human activities of the location and should therefore be considered as errors in the study.

Fourthly, according to energy conservation and environmental protection

requirements, China has been replacing incandescent lamps to more energy-saving and environmentally friendly LED lamps for a large scale in the past few years. The detectable brightness of LED is much lower than incandescent lamps. So the overall luminous brightness is lower over the past few years.

However, these potential errors are difficult to correct. NOAA is now trying to generate annual VIIRS image other than 2015. NASA is also developing new algorithms to correct errors in VIIRS images.

Since the errors of VIIRS annual data is smaller than the monthly and quarterly data, this paper will ignore these errors for now and use the sum of average monthly data to represent annual data. ***Figure 3-11*** **a)** demonstrates the VIIRS monthly data in Heilongjiang Province. The ***Figure 4-10*** below shows the roughly calculated VIIRS annual data, which has a similar trend with ***Figure 4-9***.

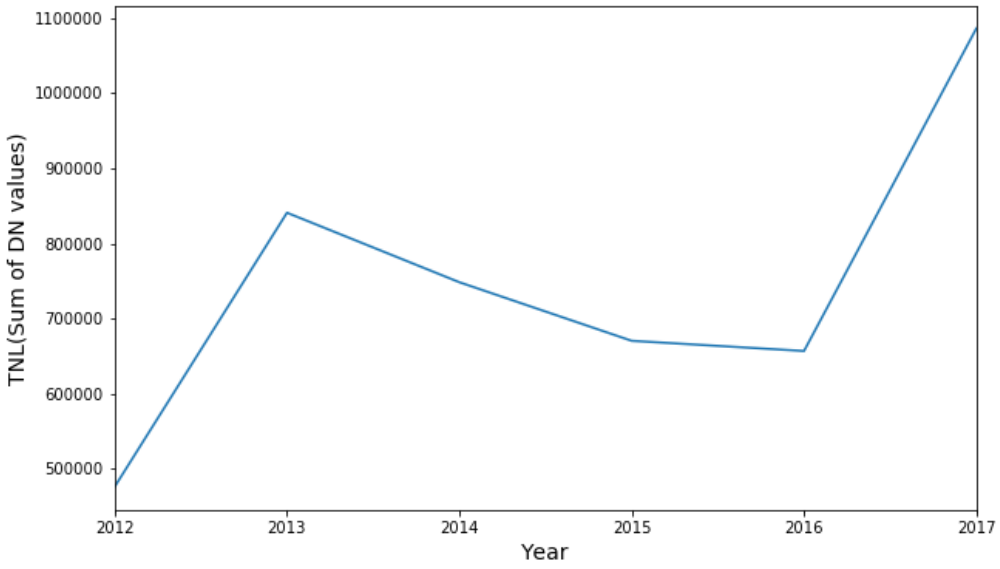


Figure 4-10 Annual NPP/VIIRS Nighttime Light Data in Heilongjiang Province

4.4 Unified Long Time-series Nighttime Light Data

Since the long time series DMSP/OLS and NPP/VIIRS remote sensing data are not uniform in metrics, how to connect the two together into a unified long time series data remains a major problem for researchers. Previous studies have focused on the comparison of data attributes and data quality, while few have combined two datasets. This paper will innovatively propose a simple and easy-to-use method to link OLS and VIIRS data into a longer time series remote sensing data. The method is based on two assumptions.

(1) Assumption one: OLS and VIIRS data are equal in TNL in 2013.

The OLS data ended in 2013 and the VIIRS data began in April 2012. The overlap between the two was from April 2012 to December 2013. However, the periodicity is not taken into account while calculating VIIRS annual data. The 2012 annual data lacks comparability due to the lack of the first three months' data. Therefore, this paper only considers that the OLS data and VIIRS data have the same TNL in 2013. In theory, except for the oversaturated pixels in OLS, the two datasets should be consistent in each pixels in 2013. But this cannot be the reality because the satellites, sensors, orbits, filming time are all different. So the assumption remains that the TNL of these two datasets are the same.

(2) Assumption two: The annual trends of OLS and VIIRS data are consistent with the annual trend of actual ground nighttime light.

Nighttime light remote sensing data is considered to have a strong positive correlation with human socioeconomic activities. In fact, as human activities become more and more frequent, the brightness and the range of nighttime light are also increasing. Therefore, the annual nighttime light DN value growth rate should be in line with the actual nighttime light growth rate even if it is detected by different satellites. That is, the annual growth rate of OLS and VIIRS data can both reflect the actual situation. No further process is needed.

Based on the above two assumptions, the VIIRS value in 2013 can be move to the point where OLS value is in 2013. The long time series nighttime light remote sensing data of China and Heilongjiang Province can then be drawn in ***Figure 4-11*** and ***Figure 4-12***.

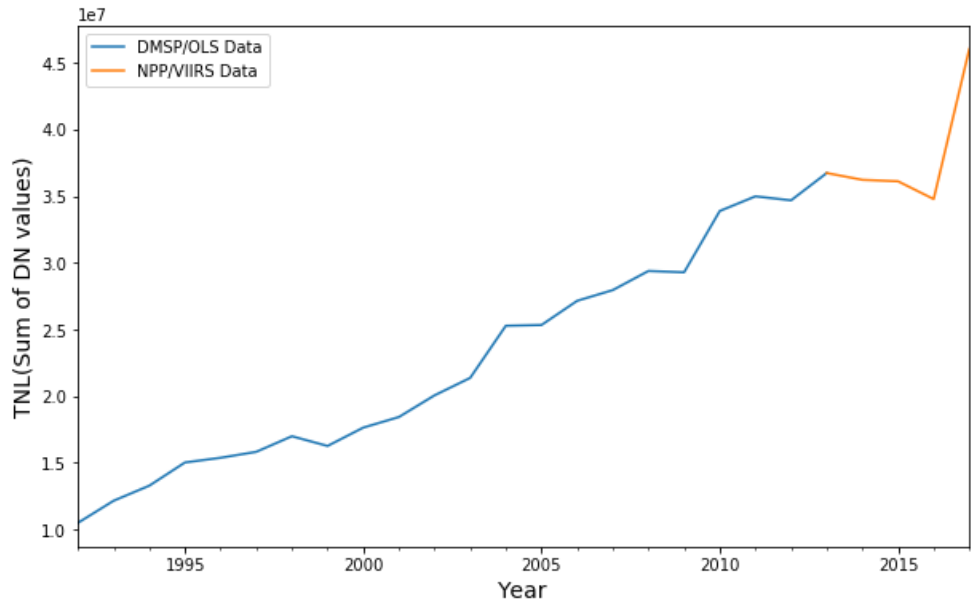


Figure 4-11 Long Time Series China Nighttime Light Data

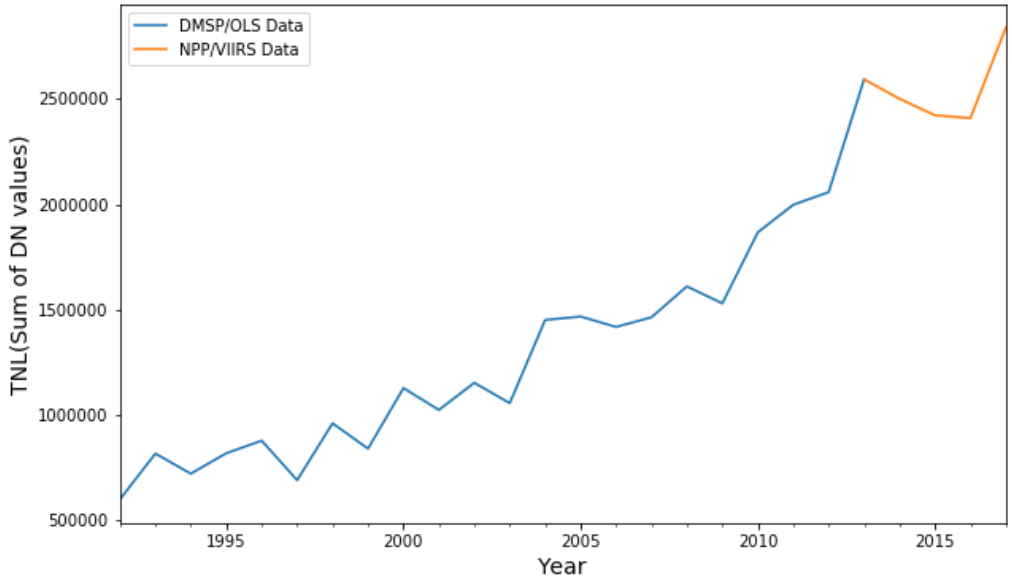


Figure 4-12 Long Time Series Nighttime Light Data in Heilongjiang Province

The ***Table 4-6*** below exhibits the time range of unified long time series data.

	Starting Time	Ending Time	Time Resolution	Terminated
DMSP/OLS	1992	2013	Annual	Yes
NPP/VIIRS	April 2012	December 2017	Monthly	Still Collecting
Unified Data	1992	2017	Annual	Still Collecting

Table 4-6 Timespans of Different Databases

4.5 A Summary of the Chapter

In this chapter, the DMSP/OLS nighttime light remote sensing data is intercalibrated, using the most widely accepted Elvidge polynomial regression method and RSR ridgeline sampling regression method. Three criteria were used to evaluate the performance of two intercalibration methods, including calibration results, normalized error index and the growth stability. The results show that the optimal strategy is to intercalibrate OLS from 1992 to 2007 using RSR method, and from 2008 to 2013 using Elvidge method.

Next, NPP/VIIRS annual data is calculate by averaging the sum of monthly data to keep the same time resolution as the OLS data. This chapter also discusses the inconsistency between the VIIRS annual data and the actual situation. Four potential reasons are analyzed, including the characteristics of VCM product, the haze distribution in China, the reflectivity of water, ice and snow, and the use of LED lights.

Finally, in order to have the longest nighttime light remote sensing data since the archival of data, two assumptions are proposed to unify DMSP/OLS and NPP/VIIRS data. The unified long time series nighttime light data exhibits steady growth trend.

Part II: Model Building and Application

In Part I of the paper, the nighttime light remote sensing images have been extracted and dealt with to obtain the quantifiable long time series nighttime light data. This part will be building various GDP forecasting models based on this study.

The chapter 5 will firstly use the traditional linear regression method to construct the GDP forecasting model. The chapter 6 will then use only GDP data to construct a univariate time series analysis model. The chapter 7 will use both nighttime light data and GDP data to construct a multivariate time series analysis model. Eventually, chapter 8 will compare the forecasting results to evaluate the pros and cons of different models.

5. Linear Regression GDP Forecasting Model for Heilongjiang Province

As introduced in chapter 2, most GDP analysis and researches based on nighttime light remote sensing data adopt a simple linear regression model, hereinafter referred to as LR model. This chapter will use the long time series nighttime light data to establish a linear regression model of the remote sensing data and GDP data of Heilongjiang Province. The model is used to predict GDP and analyze the prediction accuracy according to the results.

The GDP data of Heilongjiang Province is taken from the National Bureau of Statistics (NBS). This paper extracts the annual GDP data of Heilongjiang Province from 1992 to 2017, as shown in the table below.

Year	GDP	Year	GDP	Year	GDP
1992	959.7	2001	3390.1	2010	10368.6
1993	1198.3	2002	3637.2	2011	12582
1994	1604.9	2003	4057.4	2012	13691.58
1995	1991.4	2004	4750.6	2013	14454.91
1996	2370.5	2005	5513.7	2014	15039.38
1997	2667.5	2006	6211.8	2015	15083.67
1998	2774.4	2007	7104	2016	15386.09
1999	2866.3	2008	8314.37	2017	16199.88
2000	3151.4	2009	8587	—	—

Table 5-1 GDP of Heilongjiang Province (in 0.1 Billion Yuan)

Next, the nighttime light data and GDP data from 1992 to 2016 are extracted as observation data. The following formula is used to perform least squares fitting.

$$y_{GDP} = ax_{TNL} + b$$

Stats in Scipy library from Python is used to fit and plot. The result is in the figure below.

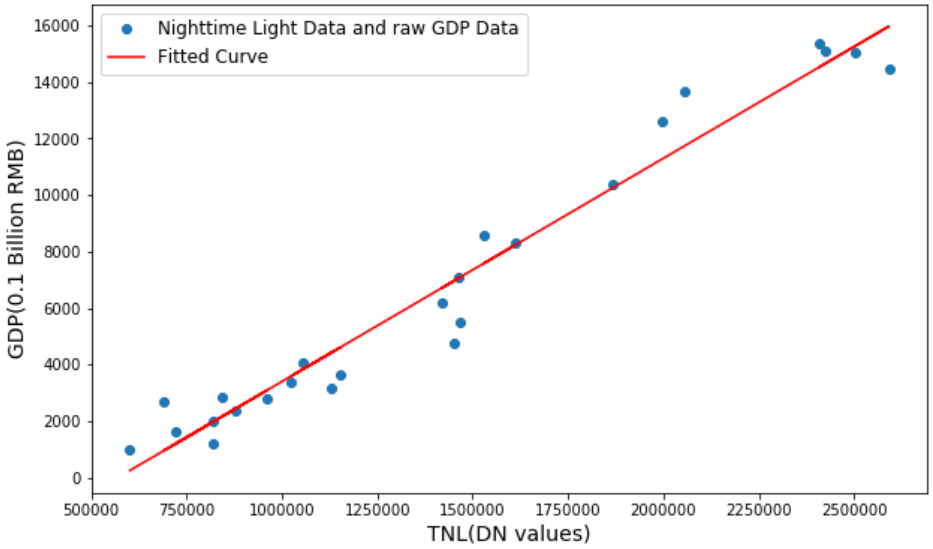


Figure 5-1 Nighttime Light Data and GDP of Heilongjiang Province

The fitted curve expression is as follows.

$$y_{GDP} = 0.0079x_{TNL} - 4511.7932$$

The fitting determination coefficient R^2 is 0.9586, which indicates that the fitting effect of the remote sensing data and GDP data of Heilongjiang Province is acceptable.

The nighttime light TNL value of Heilongjiang Province is substituted into the formula above. Then the forecast value of 2017 using LR model is calculated. In this paper, the 2017 GDP forecast of Heilongjiang Province obtained by LR model is recorded as GDP_{pred}. The actual 2017 GDP of Heilongjiang Province is recorded as GDP_{real}. The deviation of the predicted value from the actual value is recorded as the error r_{LR}, which is calculated in formula below.

$$r_{LR} = \frac{\left|GDP_{pred} - GDP_{real}\right|}{GDP_{real}} \times 100\%$$

The 2017 GDP_{real} of Heilongjiang Province is 1619.988 billion in RMB. The GDP_{pred} 2017 of Heilongjiang Province is 1793.367 billion in RMB. As can be seen from the

results, though the coefficient of determination of LR model in Heilongjiang Province is high, the prediction accuracy is low.

6. Univariate GDP Forecasting Model for Heilongjiang Province

6.1 Data Differencing and Stationarity test

The visualization figure can be drawn by importing GDP data in the ***Table 5-1*** into Python.

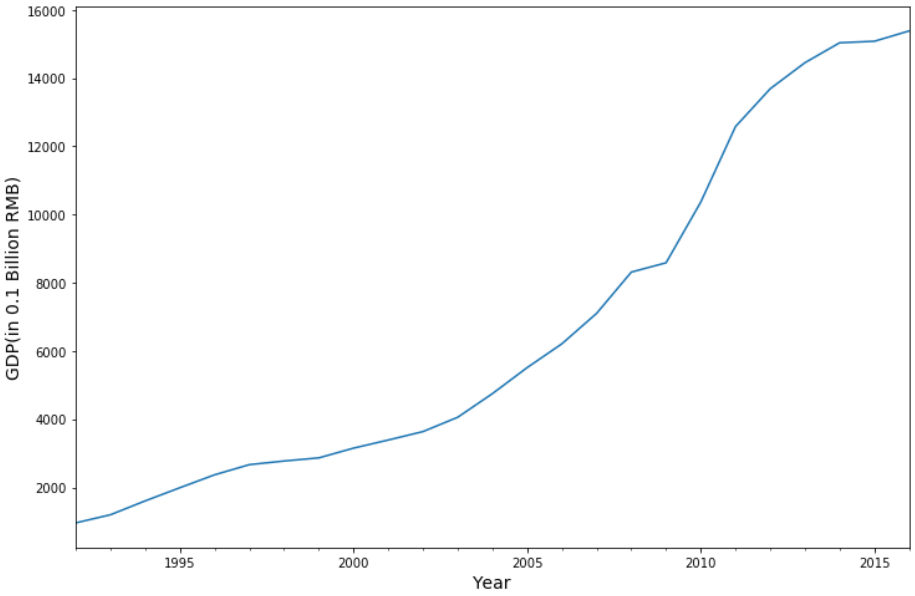


Figure 6-1 Heilongjiang GDP from 1992 to 2016

As can be seen from the figure above, the GDP data series of Heilongjiang Province is not stationary. Therefore, the data differencing is needed to make the series stationary. The ***Figure 6-2*** **a)** and ***Figure 6-2*** **b)** below demonstrate the data series after first and second order differencing.

a) GDP Series after 1st Order Difference **b) GDP after 2nd Order Difference**

Figure 6-2 Heilongjiang GDP Time Series Difference Plots

Through visual observation, the sequence after the 1st order difference is still not stable. The sequence after the 2nd order difference basically conforms to the condition of the wide stationary sequence. It is can be preliminarily determined to be a stationary sequence. The ADF test is then used to test whether it is a stationary sequence or not. From ***Table 6-1***, the P value is 0.167412. When the P value is less than the significance level, the null hypothesis is rejected and the sequence is considered stable. In general, the level of significance α is 0.05. However, in this case, there are only 26 observations in the sequence. And when applying the ADF test, there are only 14 actual test observations due to the calculation of the unit root. In the case where the sequence observations are small, the significance level can be relaxed to some extent. Therefore, in this example, the P value of 0.167412 can also be regarded as stationary.

Year Range	1992-2016
Test Statistic	-2.314254
P-value	0.167412
Lags Used	8
Numbers of Observations Used	14

Table 6-1 ADF Test Results of GDP Series after Difference

6.2 ARIMA Model Construction and Test

The autocorrelation function ACF and the partial autocorrelation function PACF of the sequence are calculated below. These two calculations can be done directly using the StatsModels library in Python. The results are shown in the following figure.

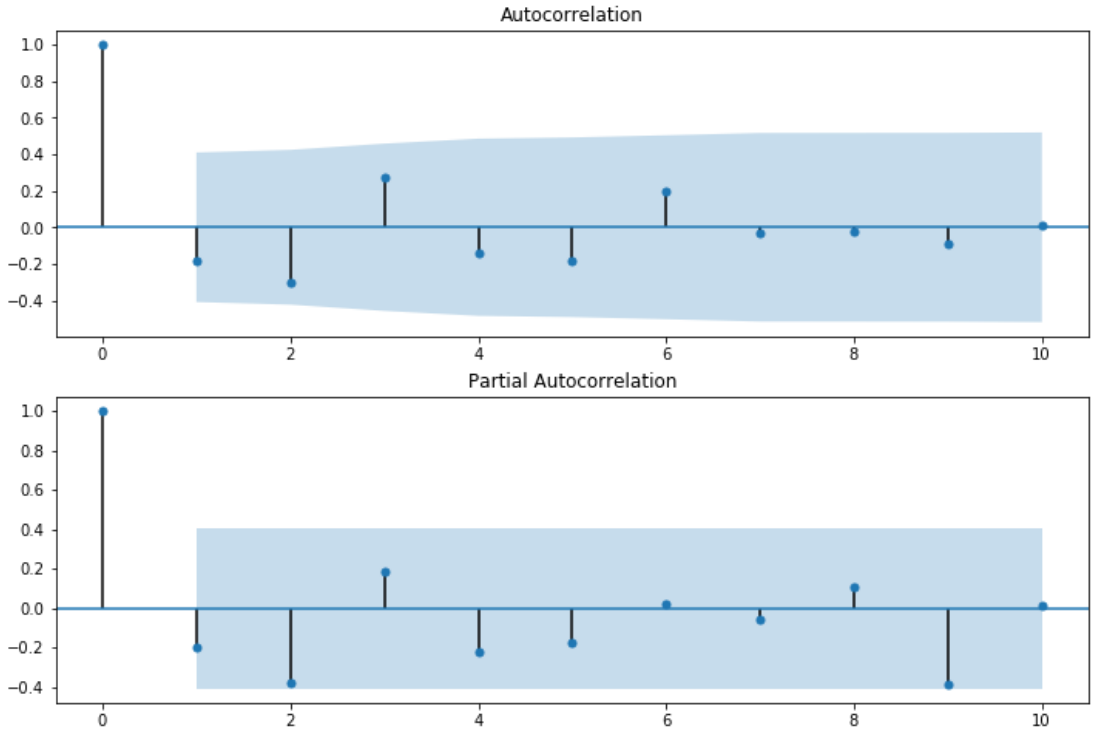


Figure 6-3 ACF and PACF of GDP Series after Difference

By looking at the figure above, both ACF and PACF fall within the range of two standard deviations from the 1st order. That is, the sequence after the differencing may be a white noise sequence. But due to the existence of the 2nd order differencing, the ARIMA(0,2,0) model can still be tested for analysis. In addition, it falls into 2 standard deviation after 2nd order. According to the judging standard, the applicability of the ARIMA(2,2,0) model can also be tested.

Then the information criteria are used to compare the ARIMA(0,2,0) model and ARIMA(2,2,0) model. In the Python library StatsModels, the criteria used to judge the pros and cons of the models are Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Hannan Quinn Information Criterion (HQIC). All three consider the number of independent variables in the residual fitting, and use different formulas or weights to "punish". The formulas are as follows.

$$\mathrm{AIC} = -2\ln(L) + 2k$$

$$\mathrm{BIC} = -2\ln(L) + \ln(n) * k$$

$$\mathrm{HQIC} = -2\ln(L) + \ln(\ln(n)) * k$$

In the formulas, L represents the maximum likelihood value. And k represents the number of unknown variables. It should be noted that comparing the calculated values

of different models can only indicate the relative merits of the models. The optimal result does not mean that the fitting accuracy can meet the requirements of using the model.

	ARIMA(0,2,0)	ARIMA(2,2,0)
AIC	353.85	354.20
BIC	356.12	358.74
HQIC	354.42	355.34

Table 6-2 Information Criteria of Two ARIMA Model

The results calculated by the AIC, BIC and HQIC criteria for the ARIMA(0,2,0) model and ARIMA(2,2,0) model are shown in the table above. The results show that the ARIMA(0,2,0) model is slightly better than the ARIMA(2,2,0) model under the three criteria. So the ARIMA(0,2,0) model will be used for analysis.

Similar to the ARIMA(0,1,0) random walk model, the ARIMA(0,2,0) is also a special model. As shown in ***Figure 6-4***, the residual approximates the constant mean white noise, but not white noise exactly. Therefore, the model ARIMA(0,2,0) still has practical significance and can be used for prediction.

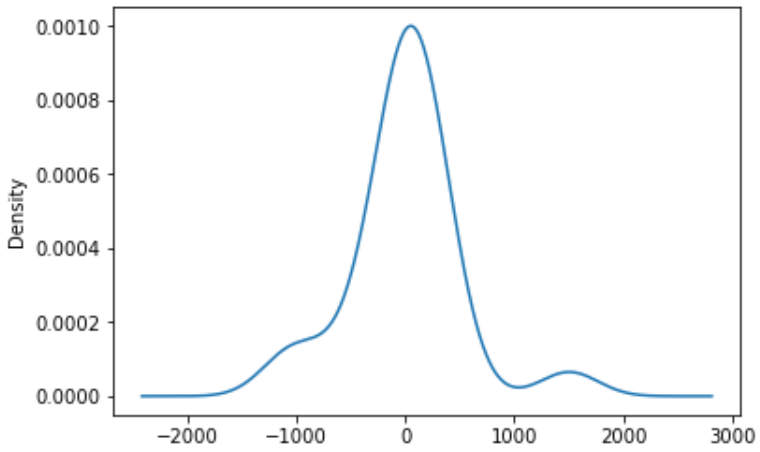


Figure 6-4 Residual Density Plot of ARIMA(0,2,0) Model

6.3 GDP Forecasting Using ARIMA Model

Python's StatsModels can directly calculate the predicted value of the sequence based on the principle of minimum variance prediction, and automatically conduct inverse difference to obtain the predicted curve from original time series. In the figure below, the orange line at the end indicates the forecast values in 2016, 2017 and 2018 by ARIMA(0,2,0) model.

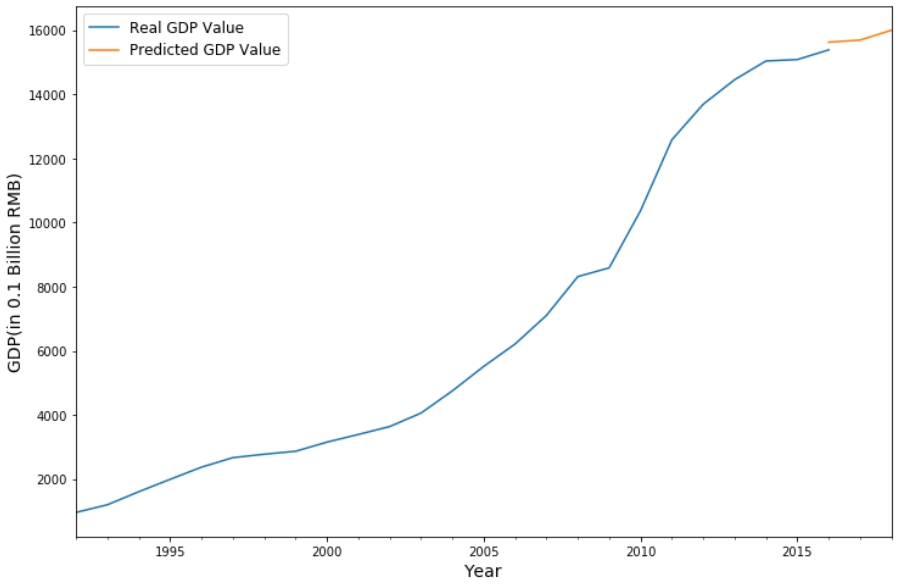


Figure 6-5 GDP Forecasting Value of Heilongjiang Province

The GDP forecast value and actual value will be expressed as GDP_{pred} and GDP_{real} here and after. The deviation of forecast value from ARIMA model and the actual value is recorded as the univariate time series error r_{ARIMA}. The GDP_{real} and GDP_{pred} in 2017 in Heilongjiang Province is 1619.988 billion and 1569.128 billion in RMB respectively. The r_{ARIMA} is 3.14%. The univariate time series model gives much better forecast result than LR model.

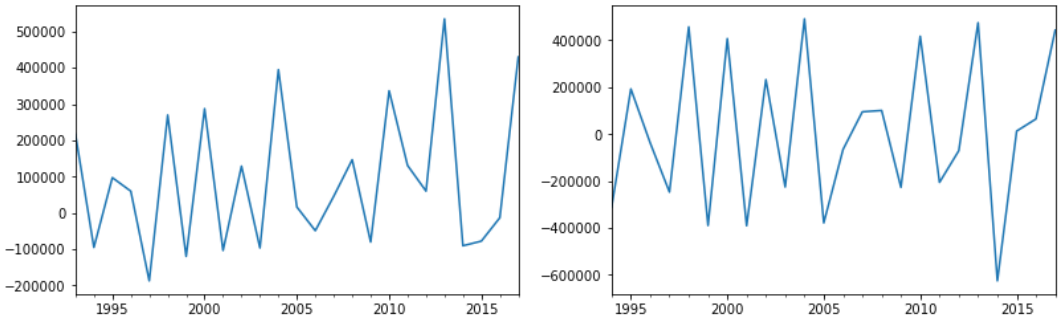
7. Multivariate GDP Forecasting Model for Heilongjiang Province Based on Nighttime Light Data

7.1 Data Differencing and Stationarity Test

By plotting the Heilongjiang GDP data and the long time series nighttime light remote sensing data in a single graph, the two curves can be seen in ***Figure 7-1***. The GDP data curve is smoother, while the nighttime light data curve has large fluctuations. But the overall trends of the two curves are similar.

Figure 7-1 Long Time Series Nighttime Light and GDP Data of Heilongjiang Province

As can be seen from ***Figure 7-1***, the long time series nighttime light remote sensing data is also a non-stationary time series, which needs to be differentiate. The ***Figure 7-2*** **a)** and ***Figure 7-2*** **b)** are the graphs obtained after the time series has gone through the 1st and 2nd order differences. The ADF stationarity test is performed after the 2nd order difference. The results can be seen in ***Table 7-1***.



a) **b)**

Figure 7-2 Nighttime Light of Heilongjiang Province after Difference (a) Nighttime Light Data of Heilongjiang Province after 1st Order Difference / b) Nighttime Light Data of Heilongjiang Province after 2nd Order Difference)

Year Range	1992-2017
Test Statistic	-3.438084
P value	0.009727
Lags Used	8
Number of Observations Used	15

Table 7-1 ADF Test Results of Nighttime Light Data after Difference

The results show that the P value of the sequence after the 2nd order difference is less than 0.05, so that the null hypothesis can be rejected. The sequence is regarded as a stationary time series.

7.2 ARIMAX Model Construction and Test

The autocorrelation function ACF and the partial autocorrelation function PACF of the long time series nighttime light remote sensing data after the difference are calculated. The results can be seen below.

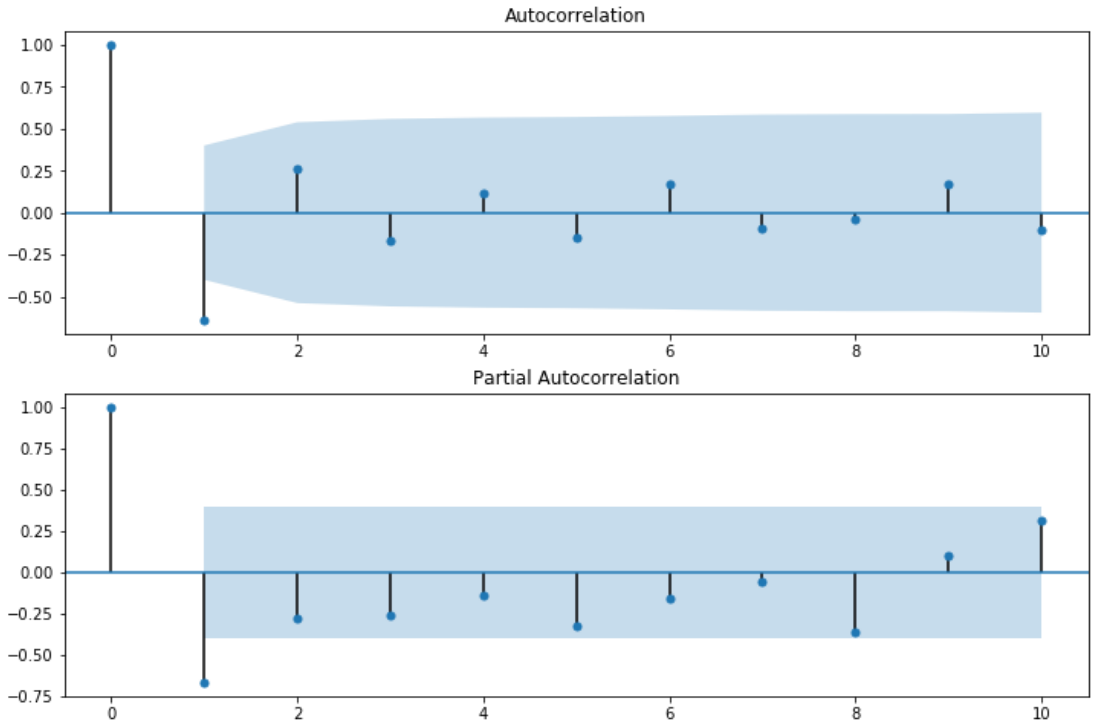


Figure 7-3 ACF and PACF Plots of Nighttime Light Data after Difference

As can be seen from ***Figure 7-3***, the ACF is truncated after a delay of 1st order. So the MA(1) model is selected. The PACF falls within the 2 standard deviation after a delay

of 1st order. But it falls also just a little within 2 standard deviation in the 5th and 8th order. So the AR(1), AR(5), are AR(8) are selected. But the ARIMA(1,2,1) model has a singular value decomposition SVD error. And the ARIMA(8,2,1) has only 16 actual observations when performing 8th order auto-regression, which affects the information criteria calculation because of instability of the sequence. Therefore, the model ARIMA(5,2,1) is selected for the subsequent analysis. The residual after fitting the ARIMA(5,2,1) model is shown in ***Figure 7-4* a)**. The residual probability density function is shown in ***Figure 7-4* b)**. It can be seen from ***Figure 7-4* b)** that the residual model is close to the constant mean white noise, that is, the model fitting effect is remarkable.

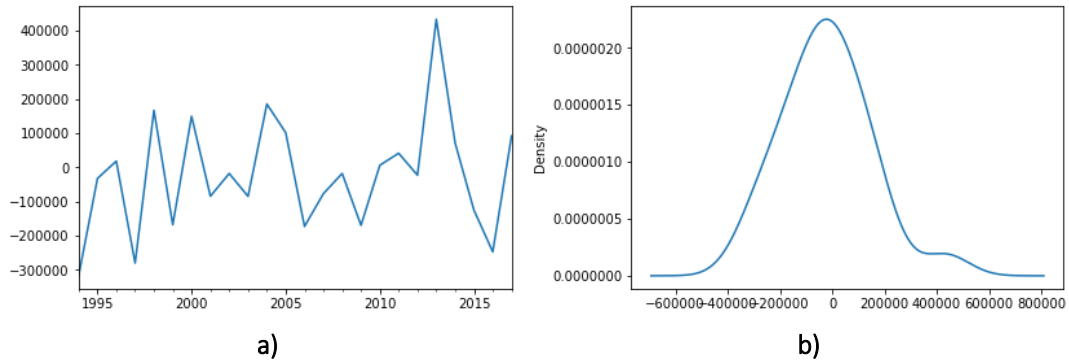


Figure 7-4 Residual Features of Nighttime Light Data in ARIMA(5,2,1) Model after Difference (a) Residual Sequence Fitted by ARIMA(5,2,1) / b) Density Plot of Residual Sequence)

7.3 GDP Forecasting Using ARIMAX Model

This article uses Python to implement the entire process of time series analysis. The StatsModels library used in this article is currently the most comprehensive mathematical and time series analytical library in Python. However, there is no specific analytical tool for ARIMAX in the library. The construction, analysis and prediction of the ARIMAX model can only be realized indirectly through other means. This article will use StatsModels to implement ARIMAX in two ways. One is to use ARIMA model with exogenous variables. The other is to use SARIMAX model with seasonal factors.

(1) ARIMAX model based on ARIMA tool with exogenous variables

In the StatsModels library, the exogenous variables can be added to the ARIMA analytical tool, meaning adding input sequence into the response sequence. Then the ARIMAX model can be constructed, which is described as the ARIMAXe model. However, in practice, the premise of using this method is that the response sequence and the input sequence have the same autoregressive order, moving average order,

and difference order. If they are inconsistent, the fitting is performed according to the order of the response sequence.

According to the results in the previous section, the Heilongjiang GDP time series model is $\mathrm{ARIMA}_{GDP}(0,2,0)$. However, the nighttime light remote sensing data series model is $\mathrm{ARIMA}_{NTL}(5,2,1)$. These two only have the same differential order. The autoregressive order and moving average order are different. Therefore, only the ARIMAe(0,2,0) model with exogenous variables can be used for analysis and prediction. The exogenous variable is nighttime light data series. At this time, only the data from 1992 to 2016 is used to fit the model. The AIC, BIC and HQIC results are 355.56, 358.97 and 356.42, respectively.

In order to obtain the 2017 GDP forecast of Heilongjiang Province, it is necessary to bring in the nighttime light remote sensing data in 2017 when applying the ARIMAXe(0,2,0) model prediction. The predicted GDP data is shown in the figure below.

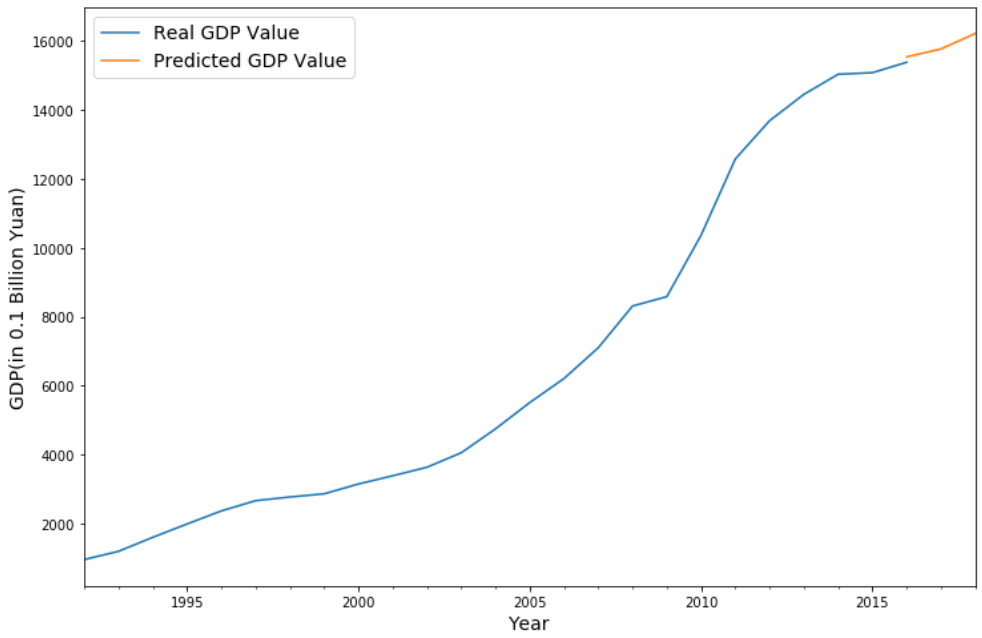


Figure 7-5 Heilongjiang GDP Forecasting Using ARIMAXe Model with Exogenous Variables

Using the ARIMAXe(0,2,0) model with exogenous variables based on long time series nighttime light remote sensing data, the predicted GDP of Heilongjiang Province in 2017 GDP_{pred} is 1571.175 billion in RMB. The real value GDP_{real} is known to be 1619.88 billion in RMB. The deviation of the predicted value from the real value $r_{ARIMAXe}$ is 2.64%. The prediction effect is good.

(2) ARIMAX model based on SARIMAX tool with seasonal factors

In the StatsModels library, there is also a class of tools called SARIMAX, which is one type of the ARIMAX model, hereinafter referred to as ARIMAXs. The S represents the seasonal effect, meaning adding seasonal factors to the ARIMAX model. The application method is stated below. Firstly the order of the response sequence $\{y\}$ is p_y, d_y, q_y. Then the order of the input sequence $\{x\}$ is p_x, d_x, q_x, and the seasonal factor is s. Normally the seasonal factor s is 12 or 4, indicating that the model has monthly or quarterly fluctuation effects. In this example, the response sequence is the Heilongjiang GDP data series. The orders p_y, d_y, q_y are 0, 2 and 0. The input sequence is the nighttime light data series. The orders p_x, d_x, q_x are 5, 2, and 1. The seasonal factor s is set to be 1 because the sequence is an annual sequence. By setting parameters above, the model results AIC, BIC and HQIC are 280.66, 287.33 and 281.33, respectively.

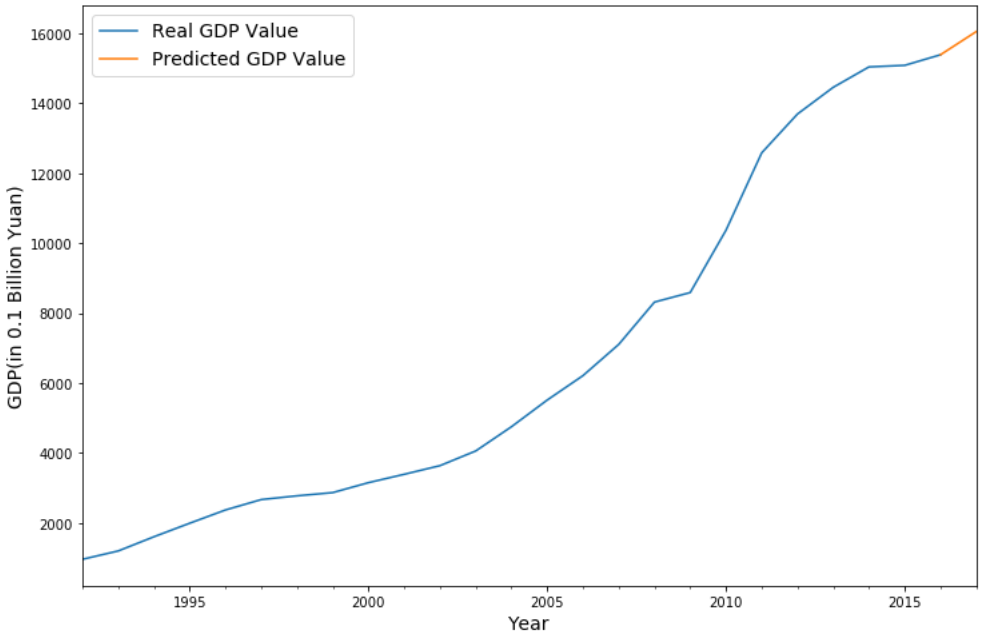


Figure 7-6 Heilongjiang GDP Forecasting Using ARIMAXs Model with the Seasonal Factor

The above figure shows the results of using the ARIMAXs model with the seasonal factor to predict the 2017 GDP forecast for Heilongjiang Province. The predicted value GDP_{pred} is 1604.802 billion in RMB, while the real value GDP_{real} is 1619.988 billion in RMB. The deviation $r_{ARIMAXs}$ is 0.94%. The results show that the predicted value is very close to the actual value. The prediction effect is very good.

8. Evaluation of GDP Forecasting Models for Heilongjiang Province

The ***Figure 8-1*** below visually shows the difference between the GDP forecast and the actual GDP value obtained from the four different GDP forecasting models in this part. For the sake of demonstration, only GDP values between 2014 and 2017 are shown. The ***Table 8-1*** below shows that values of the forecast GDP and actual GDP obtained by different models, and the deviations.

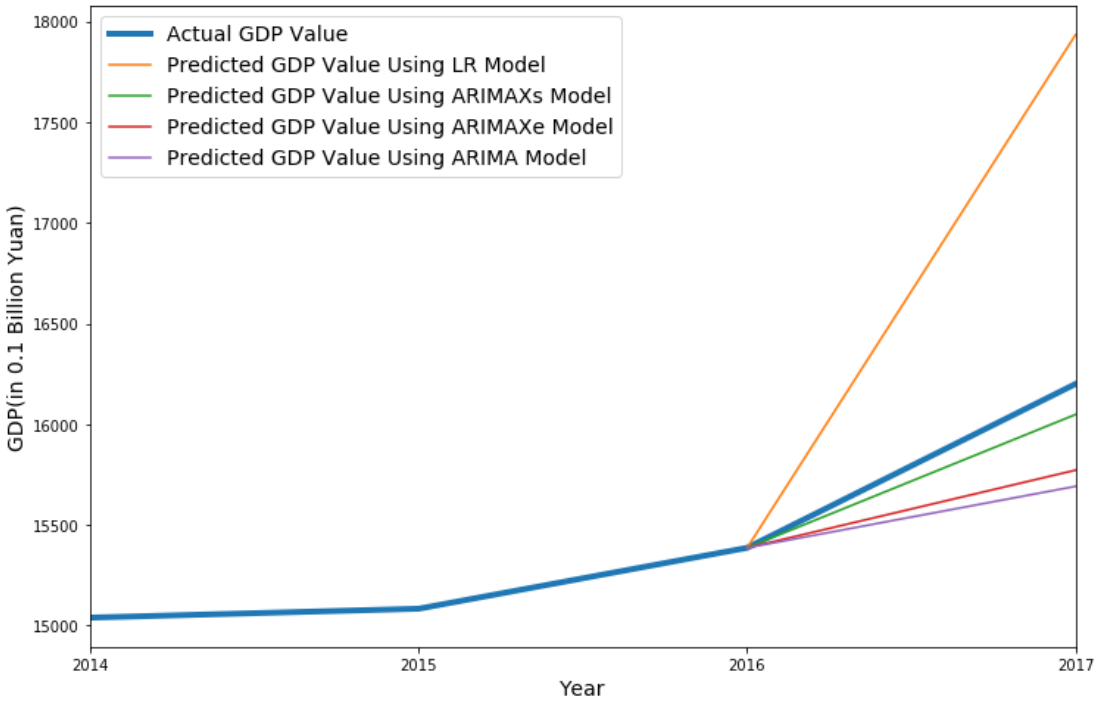


Figure 8-1 Comparison of Heilongjiang GDP Forecasting Using Different Models in 2017

The results show that the three GDP forecast models can all generate good forecasting results. Among them, the ARIMAXs model with seasonal factors has the highest prediction accuracy. The deviation error is 0.94%. The prediction accuracy of the ARIMAXe model with exogenous variables is second. The deviation error is 2.64%. The univariate time series ARIMA model is a bit less. The deviation error is 3.14%. The LR model based on linear regression has the lowest prediction accuracy, which has a deviation error of 10.70%.

One reason why the accuracy of the LR model is much lower than other time series analysis models is that the LR model only uses the nighttime light data and GDP data without considering the time variation, meaning the time information has not been

taken into account in these two datasets. Therefore, the results should theoretically be more error-prone than time series analysis.

	LR	ARIMA	ARIMAXe	ARIMAXs
GDP_{pred} (in 0.1 Billion Yuan)	17933.67	15691.28	15771.75	16048.02
GDP_{real} (in 0.1 Billion Yuan)		16199.88		
r	10.70%	3.14%	2.64%	0.94%

Table 8-1 Results Comparison of Different GDP Forecasting Models

The ***Table 8-2*** below shows the results of three time series based GDP prediction models under three information criteria AIC, BIC and HQIC. The results show that the ARIMAXs model, ARIMA model and ARIMAXe model are from superior to inferior.

	ARIMA	ARIMAXe	ARIMAXs
AIC	353.85	355.56	280.66
BIC	356.12	358.97	287.33
HQIC	354.42	356.42	281.33

Table 8-2 Information Criteria Test Results Comparison of Different GDP Forecasting Models

The model prediction accuracy is the most important index for evaluating models. And the ARIMA model and ARIMAXe model are both small in information criteria. So taking both prediction accuracy and information criteria into account, the ARIMAs model with seasonal factors is the best. The ARIMAXe model with exogenous variables is the second. And the univariate time series analysis ARIMA model is the third.

Conclusions

In this paper, the DMSP/OLS and NPP/VIIRS nighttime light remote sensing data have been extracted, processed and visualized. The causes of the errors from nighttime light remote sensing data have been analyzed. The errors have been eliminated. The long time series nighttime light remote sensing data sequence has been constructed. Based on the sequence, the GDP forecast models have been built and analyzed. The pros and cons of different GDP forecast models have been examined. The main research conclusions are as follow:

(1) The nighttime light remote sensing data has been visually demonstrated via Python visualization. At first, the raster data output circular model is built in ArcGIS to process the DMSP/OLS and NPP/VIRRS nighttime light images in batches. The output data of nighttime light images is read, processed, analyzed and visualized in Python. The visualized images visually reflect the distribution of nighttime light brightness in all provincial-level regions, indicating that the nighttime light remote sensing data can effectively reflect the development of regional economy.

(2) The optimized DMSP/OLS nighttime light remote sensing data after intercalibration has been obtained. Aiming at the error of DMSP/OLS nighttime light remote sensing data due to the lack of calibration devices, the error calibration of DMSP/OLS nighttime light data has been conducted by using the Elvidge method based on polynomial regression in Matlab. At the same time, the intercalibrated DMSP/OLS nighttime light data based on RSR method of ridgeline sampling regression has been collected. Three criteria for evaluating intercalibration effect are proposed to compare Elvidge method and RSR method. The results show that the DMSP/OLS remote sensing data is better intercalibrated by RSR method from 1992 to 2007, and by Elvidge method from 2008 to 2013. This results in the intercalibrated DMSP/OLS data from 1992 to 2013.

(3) The NPP/VIIRS nighttime light remote sensing data has been processed and combined with calibrated DMSP/OLS nighttime light data to obtain a quantifiable long time series nighttime light data. In order to maintain the consistency of the time resolution of nighttime light data, the NPP/VIIRS monthly data are summed and averaged for the lack of annual data to get the calculated annual data from 2012 to 2017. The errors between the monthly and annual NPP/VIIRS data have been distinguished. The causes of these errors have been discussed. Two assumptions have been proposed to unify DMSP/OLS and NPP/VIIRS data to obtain a long time series nighttime light data from 1992 to 2017 for the first time. This is the longest sequence

of nighttime light data since the electronic archiving and disclosure of nighttime light remote sensing images. It is available for quantitative analysis.

(4) Three GDP forecasting models have been constructed for Heilongjiang Province from 1992 to 2016 in Python. And the GDP of Heilongjiang Province in 2017 was forecast and compared with the actual value to calculate the forecasting error. The first model is LR model based on linear regression. The regression curve and coefficient of determination of Heilongjiang GDP data and nighttime light data are calculated using this model. The second model is the univariate time series analysis ARIMA model with only GDP data. The third is the multivariate time series analysis ARIMAX model based on long time series nighttime light data. Due to the characteristics of Python library, the model is divided into ARIMAXe model with exogenous variables and ARIMAXs model with seasonal factors. This is the first time that nighttime light remote sensing data has been applied to GDP time series analysis. The results show that the prediction errors of LR model, ARIMA model, ARIMAXe model and ARIMAXs model for Heilongjiang's 2017 GDP are 10.70%, 3.14%, 2.64%, 0.94%, respectively. That is, when the nighttime light data is added to the GDP forecasting model, the LR model's prediction of GDP is not better than the GDP univariate time series forecasting model. But ARIMAX model has higher prediction accuracy than the GDP univariate time series forecasting model. Therefore, GDP forecasting based on long time series nighttime light remote sensing data can improve the prediction accuracy of GDP.

Discussions

Limited to the quality of nighttime light remote sensing data, the imperfection of the Python library and the author's capability, this research still has the following shortcomings:

(1) The NPP/VIIRS first version DNB nighttime light remote sensing image product has only annual image in 2015. The annual images of other years are not available. The annual data calculated from monthly data in this paper is very rough. If the NOAA releases the annual images subsequently, the annual image published by NOAA should be used as the annual data of NPP/VIIRS nighttime light remote sensing. In addition, the paper found out that there are large systematic errors in the NPP/VIIRS data, analyzed the causes of the errors, and mitigated the errors. NASA is currently known to work on the correction of NPP/VIIRS data. If the NASA issues corrected data subsequently, the data published by NASA should then be used as the NPP/VIIRS nighttime light remote sensing data.

(2) The time series analysis in this article is implemented using the StatsModel library in Python. The StatsModel library is currently the most comprehensive library for time series analysis in Python, but is still not perfect. While the author was conducting the research, the library was still being updated. For the purposes of this paper, the StatsModel library lacks the processing of sequence heteroscedasticity, lacks model parameter significance testing, and lacks specific ARIMAX model analysis tools. If the library is updated later, or a better time series analysis library appears, the currently used model can be updated and improved.

(3) This paper only analyzes and evaluate the GDP forecasting of Heilongjiang Province. And the modeling process is only applied for Heilongjiang Province. Whether the model is applicable to other regions and countries remains to be studied.

(4) The study of GDP is generally divided into two parts: the total GDP and the GDP structure. In this paper, the GDP forecast based on long time series nighttime light remote sensing data only predicted the total GDP value. It does not provides any insights on the GDP structure. Whether the nighttime light remote sensing data can be used for the analysis of GDP structure remains to be discussed.

Bibliography

Andrews, B. H., Dean, M. D., Swain, R., & Cole, C. (2013). Building ARIMA and ARIMAX models for predicting long-term disability benefit application rates in the public/private sectors. *Society of Actuaries*, 1-54.

Bennett, M. M., & Smith, L. C. (2017). Advances in using multitemporal night-time lights satellite imagery to detect, estimate, and monitor socioeconomic dynamics. *Remote Sensing of Environment*, *192*, 176-197.

Cao, C., Xiong, J., Blonski, S., Liu, Q., Uprety, S., Shao, X., ... & Weng, F. (2013). Suomi NPP VIIRS sensor data record verification, validation, and long‐term performance monitoring. *Journal of Geophysical Research: Atmospheres*, *118(20)*, 11-664.

Chen, C. 陈聪聪. (2016). 基于 ARIMA 模型和 ARIMAX 模型的山东省 GDP 的预测与分析. 山东大学.

Chen, X., & Nordhaus, W. D. (2011). Using luminosity data as a proxy for economic statistics. *Proceedings of the National Academy of Sciences*, *108(21)*, 8589-8594.

Chen, Z. 陈佐旗. (2017). 基于多源夜间灯光遥感影像的多尺度城市空间形态结构分析. 华东师范大学.

Chen, Z. 陈志高. (2009). 遗传算法和 BP 神经网络在 GDP 预测中的应用. *计算机与数字工程*, 37(09), 172-175.

Doll, C. N., Muller, J. P., & Morley, J. G. (2006). Mapping regional economic activity from night-time light satellite imagery. *Ecological Economics*, *57(1)*, 75-92.

Ebener, S., Murray, C., Tandon, A., & Elvidge, C. C. (2005). From wealth to health: modelling the distribution of income per capita at the sub-national level using night-time light imagery. *International Journal of health geographics*, *4(1)*, 5.

Elvidge, C., Baugh, K., Hobson, V., Kihn, E., Kroehl, H., Davis, E., & Cocero, D. (1997a). Satellite inventory of human settlements using nocturnal radiation emissions: a contribution for the global toolchest. *Global Change Biology*, *3(5)*, 387-395.

Elvidge, C. D., Baugh, K. E., Kihn, E. A., Kroehl, H. W., & Davis, E. R. (1997b). Mapping city lights with nighttime data from the DMSP Operational Linescan System. *Photogrammetric Engineering and Remote Sensing*, *63(6)*, 727-734.

Elvidge, C. D., Baugh, K. E., Zhizhin, M., & Hsu, F. C. (2013, April). Why VIIRS data are superior to DMSP for mapping nighttime lights. In *Proceedings of the Asia-Pacific Advanced Network* (Vol. 35, No. 62).

Elvidge, C. D., Hsu, F. C., Baugh, K. E., & Ghosh, T. (2014). National trends in satellite-observed lighting. *Global urban monitoring and assessment through earth observation*, *23*, 97-118.

Elvidge, C. D., Sutton, P. C., Ghosh, T., Tuttle, B. T., Baugh, K. E., Bhaduri, B., & Bright, E.

(2009). A global poverty map derived from satellite data. *Computers & Geosciences*, *35(8)*, 1652-1660.

Elvidge, C., Ziskin, D., Baugh, K., Tuttle, B., Ghosh, T., Pack, D., ... & Zhizhin, M. (2009). A fifteen year record of global natural gas flaring derived from satellite data. *Energies*, *2(3)*, 595-622.

Ghosh, T., L Powell, R., D Elvidge, C., E Baugh, K., C Sutton, P., & Anderson, S. (2010). Shedding light on the global distribution of economic activity. *The Open Geography Journal*, 3(1).

Han, X., Zhou, Y., Wang, S., Liu, R., & Yao, Y. 韩向娣, 周艺, 王世新, 刘瑞, & 姚尧. (2012). 夜间灯光遥感数据的 GDP 空间化处理方法. *地球信息科学学报*, *14(01)*, 128-136

He, L., & He, Y. 何黎, & 何跃. (2012). 结合 PMI 的中国 GDP 预测模型. *统计与决策*, *2010(01)*, 84-86.

Henderson, J. V., Storeygard, A., & Weil, D. N. (2012). Measuring economic growth from outer space. *American economic review*, *102(2)*, 994-1028.

Henderson, M., Yeh, E. T., Gong, P., Elvidge, C., & Baugh, K. (2003). Validation of urban boundaries derived from global night-time satellite imagery. *International Journal of Remote Sensing*, *24(3)*, 595-609.

Hsu, F. C., Baugh, K., Ghosh, T., Zhizhin, M., & Elvidge, C. (2015). DMSP-OLS radiance calibrated nighttime lights time series with intercalibration. *Remote Sensing*, *7(2)*, 1855-1876.

Huang, H., Wang, F., & Xie, X. 黄鸿健, 王芳, & 解学通. (2016) 基于灯光遥感数据的广东省 20 年城市化进程研究. *城市学刊*, *37(01)*, 1-8.

Huang, Y. 黄益修. (2016). 基于夜间灯光遥感影像和社会感知数据的人口空间化研究. 华东师范大学.

Imhoff, M. L., Lawrence, W. T., Stutzer, D. C., & Elvidge, C. D. (1997). A technique for using composite DMSP/OLS “city lights” satellite data to map urban area. *Remote Sensing of Environment*, *61(3)*, 361-370.

Keola, S., Andersson, M., & Hall, O. (2015). Monitoring economic development from space: using nighttime light and land cover data to measure economic growth. *World Development*, *66*, 322-334.

Kong, Z., Li, G., Shi, M., & Huang, M. 孔朝莉, 李国徽, 石明, 黄美婷. (2016). 基于 GM(1,1)与主成分回归的海南 GDP 预测及其影响因素分析. *数学的实践与认识*, *46(17)*, 66-80.

Lan, Z. & Kan, C. 兰宗敏, & 阚长城. (2018) 基于大数据的城市群识别与空间特征研究. Retrieved from http://www.drc.gov.cn/yjlyyyjbm/8.html

Levin, N. (2017). The impact of seasonal changes on observed nighttime brightness

from 2014 to 2015 monthly VIIRS DNB composites. *Remote Sensing of Environment*, *193*, 150-164.

Levin, N., & Duke, Y. (2012). High spatial resolution night-time light images for demographic and socio-economic studies. *Remote Sensing of Environment*, *119*, 1-10.

Levin, N., Johansen, K., Hacker, J. M., & Phinn, S. (2014). A new source for high spatial resolution night time images—The EROS-B commercial satellite. *Remote Sensing of Environment*, *149*, 1-12.

Levin, N., & Phinn, S. (2016). Illuminating the capabilities of Landsat 8 for mapping night lights. Remote Sensing of Environment, 182, 27-38.

Li, D. & Li, X. 李德仁, & 李熙. (2015a). 论夜光遥感数据挖掘. *测绘学报*, 591-601.

Li, D. & Li, X. 李德仁, & 李熙. (2015b). 夜光遥感技术在评估经济社会发展中的应用——兼论其对“一带一路”建设质量的保障. *宏观质量研究*, *3(4)*, 1-8.

Li, D., Yu, H. & Li, X. 李德仁, 余涵若, & 李熙. (2017). 基于夜光遥感影像的“一带一路”沿线国家城市发展时空格局分析. *武汉大学学报• 信息科学版*, *42(6)*, 711-720.

Li, X., Chen, X., Zhao, Y., Xu, J., Chen, F., & Li, H. (2013). Automatic intercalibration of night-time light imagery using robust regression. *Remote sensing letters*, *4(1)*, 45-54.

Li, X., Xu, H., Chen, X., & Li, C. (2013). Potential of NPP-VIIRS nighttime light imagery for modeling the regional economy of China. *Remote Sensing*, *5(6)*, 3057-3081.

Liu, Y., Wang, Y., Peng, J., Du, Y., Liu, X., Li, S., & Zhang, D. (2015). Correlations between urbanization and vegetation degradation across the world’s metropolises using DMSP/OLS nighttime light data. *Remote Sensing*, *7(2)*, 2067-2088.

Liu, Z., He, C., Zhang, Q., Huang, Q., & Yang, Y. (2012). Extracting the dynamics of urban expansion in China using DMSP-OLS nighttime light data from 1992 to 2008. *Landscape and Urban Planning*, *106(1)*, 62-72.

Ma, T., Zhou, C., Pei, T., Haynie, S., & Fan, J. (2012). Quantitative estimation of urbanization dynamics using time series of DMSP/OLS nighttime light data: A comparative case study from China's cities. *Remote Sensing of Environment*, *124*, 99-107.

OECD. (2003). GROSS DOMESTIC PRODUCT (GDP) – CONSTANT PRICES. Retrieved February 23, 2019, from https://stats.oecd.org/glossary/detail.asp?ID=1164

Pan, S. 潘思东. (2017). 基于夜光遥感和小区 POI 的住宅发展与经济增长的空间耦合研究. *地球信息科学学报*, *19(05)*, 646-652.

Pandey, B., Joshi, P. K., & Seto, K. C. (2013). Monitoring urbanization dynamics in India using DMSP/OLS night time lights and SPOT-VGT data. *International Journal of*

Applied Earth Observation and Geoinformation, *23*, 49-61.

Pandey, B., Zhang, Q., & Seto, K. C. (2017). Comparative evaluation of relative calibration methods for DMSP/OLS nighttime lights. *Remote Sensing of Environment*, *195*, 67-78.

Raupach, M. R., Rayner, P. J., & Paget, M. (2010). Regional variations in spatial structure of nightlights, population density and fossil-fuel CO2 emissions. *Energy Policy*, *38(9)*, 4756-4764.

Román, M. O., Wang, Z., Sun, Q., Kalb, V., Miller, S. D., Molthan, A., ... & Seto, K. C. (2018). NASA's Black Marble nighttime lights product suite. *Remote Sensing of Environment*, *210*, 113-143.

Shan, Y. 单玉隆. (2014). ARIMA 模型与遗传算法优化神经网络在 GDP 预测中的应用. 兰州大学.

Shi, K., Yu, B., Huang, Y., Hu, Y., Yin, B., Chen, Z., ... & Wu, J. (2014). 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 Sensing*, *6(2)*, 1705-1724.

Shu, S. 舒松. (2013). 基于稳定夜间灯光遥感数据的城市群空间模式识别方法研究. 华东师范大学.

Sun, C., & Jiang, M. 孙彩, & 姜明辉. (2008). 基于 GP 的非线性 GDP 预测模型的构建与应用. *哈尔滨工业大学学报(社会科学版)*, *2008(01)*, 133-138.

Sutton, P. (1997). Modeling population density with night-time satellite imagery and GIS. *Computers, Environment and Urban Systems*, *21(3-4)*, 227-244.

Sutton, P. C., & Costanza, R. (2002). Global estimates of market and non-market values derived from nighttime satellite imagery, land cover, and ecosystem service valuation. *Ecological Economics*, *41(3)*, 509-527.

Sutton, P. C., Elvidge, C. D., & Ghosh, T. (2007). Estimation of gross domestic product at sub-national scales using nighttime satellite imagery. *International Journal of Ecological Economics & Statistics*, *8(S07)*, 5-21.

Sutton, P., Roberts, D., Elvidge, C., & Baugh, K. (2001). Census from Heaven: An estimate of the global human population using night-time satellite imagery. *International Journal of Remote Sensing*, *22(16)*, 3061-3076.

Teng, G., & He, Y. 腾格尔, & 何跃. (2010). 基于 GMDH 组合的中国 GDP 预测模型研究. *统计与决策*, *2010(07)*, 17-19.

Wang, Y. 王永杰. (2017). 人工神经网络算法在 GDP 和 CPI 中的预测应用. 中北大学.

Wang, Y. 王云菲. (2016). 基于 DMSP/OLS 夜间灯光遥感数据的中国西南地区城市化研究，云南大学.

Wang, Y., & Xu, J. 汪远征, & 徐雅静. (2007). 多元平稳时间序列 ARIMAX 模型的应用. *统计与决策*, *2007(18)*, 132-135.

Wu, J., Wang, Z., Li, W., & Peng, J. (2013). Exploring factors affecting the relationship between light consumption and GDP based on DMSP/OLS nighttime satellite imagery. *Remote Sensing of Environment*, *134*, 111-119.

Wu, R., Yang, D., Dong, J., Zhang, L., & Xia, F. (2018). Regional inequality in China based on NPP-VIIRS night-time light imagery. *Remote Sensing*, *10(2)*, 240.

Xiong, Z. 熊志斌. (2011). 基于 ARIMA 与神经网络集成的 GDP 时间序列预测研究. *数理统计与管理*, *30(02)*, 306-314.

Yang, T. 杨挺. (2008). 基于 BP 神经网络改进算法的湖南省 GDP 预测研究. 中南大学.

Yi, K., Tani, H., Li, Q., Zhang, J., Guo, M., Bao, Y., ... & Li, J. (2014). Mapping and evaluating the urbanization process in northeast China using DMSP/OLS nighttime light data. *Sensors*, *14(2)*, 3207-3226.

Yong, H., & Bao, G. 雍红月, 包桂兰. (2008) 组合时间序列 ARMA 模型在经济预测中的应用——内蒙古十一五期间 GDP 预测. *数学的实践与认识*, *2008(21)*, 19-23.

Zhang, Q., Pandey, B., & Seto, K. C. (2016). A robust method to generate a consistent time series from DMSP/OLS nighttime light data. *IEEE Transactions on Geoscience and Remote Sensing*, *54(10)*, 5821-5831.

Zhang, Q., & Seto, K. (2013). Can night-time light data identify typologies of urbanization? A global assessment of successes and failures. *Remote Sensing*, *5(7)*, 3476-3494.

Zhang, Q., & Seto, K. C. (2011). Mapping urbanization dynamics at regional and global scales using multi-temporal DMSP/OLS nighttime light data. *Remote Sensing of Environment*, *115(9)*, 2320-2329.