

## ANALYSIS OF VISIBLE INFRARED IMAGING RADIOMETER SUITE CAPABILITY FOR POPULATION ESTIMATION ON JAVA ISLAND

Mohamad Dimas Henru Nurhadi<sup>a</sup>, Ari Cahyono<sup>b,\*</sup>

<sup>a</sup>Undergraduate Program in Cartography and Remote Sensing, Department of Geographic Information Science, Faculty of Geography, Universitas Gadjah Mada, Yogyakarta, Indonesia; mohamad.dimas.h.n@mail.ugm.ac.id

<sup>b</sup>Department of Geographic Information Science, Faculty of Geography, Universitas Gadjah Mada, Yogyakarta, Indonesia; aricahyono@ugm.ac.id

Commission 4, WG 7

**KEY WORDS:** VIIRS, nighttime light, population, linear regression, remote sensing, estimation, Java Island, Indonesia.

### ABSTRACT

Population data, despite their significance, are often missing or difficult to access, especially in cities/regencies not belonging to the metropolitan areas or centers of various human activities. This hinders practices that are contingent on their availability. In this study, population estimation was carried out using nighttime light imagery generated by the Visible Infrared Imaging Radiometer Suite (VIIRS) instrument. The variable illuminated area was integrated with the population data using linear regression based on an allometric formula so as to produce a regression value, correlation coefficient ( $r$ ), and coefficient of determination ( $r^2$ ). The average  $r^2$  between the illuminated area and the total population was 0.86, indicating a strong correlation between the two variables. Validation using samples of population estimates from three different years yielded an average error of 73% for each city and 7% for the entire study area. The estimation results for the number of residents per city/regency cannot be used as population data due to the high percent error, but for the population on a larger regional scale, in this case, the island of Java, they have a much smaller percent error and can be used as an initial picture of the total population.

### 1. INTRODUCTION

*Badan Pusat Statistik* (BPS-Statistics Indonesia) (2020) stated that the population in Indonesia reached 268,074.6 million in 2019 with a growth rate of 1.31% from 2010. Indonesia has a land area of 1,916,906.77 km<sup>2</sup> (BPS, 2020), while its waters reach 3,257 million km<sup>2</sup> or almost 60% of the total area (BPS-Statistics Indonesia, 2019). The population density is up to 140 people per km<sup>2</sup> (BPS-Statistics Indonesia, 2020). The increase in population from year to year has resulted in challenges such as increasing food needs, water needs, especially clean water, and energy consumption, both renewable and non-renewable (Daily et al., 1998). Furthermore, if the rapid population growth remains unchecked, one of the consequences in the next few years will include urban development, which primarily affects physical features in cities like built-up land and land functions (Kulshrestha, 2007). Therefore, population data are needed to make the latest regulations and policies to control population size and growth rates and, ultimately, minimize the resultant adverse effects (Chowdhury et al., 2012).

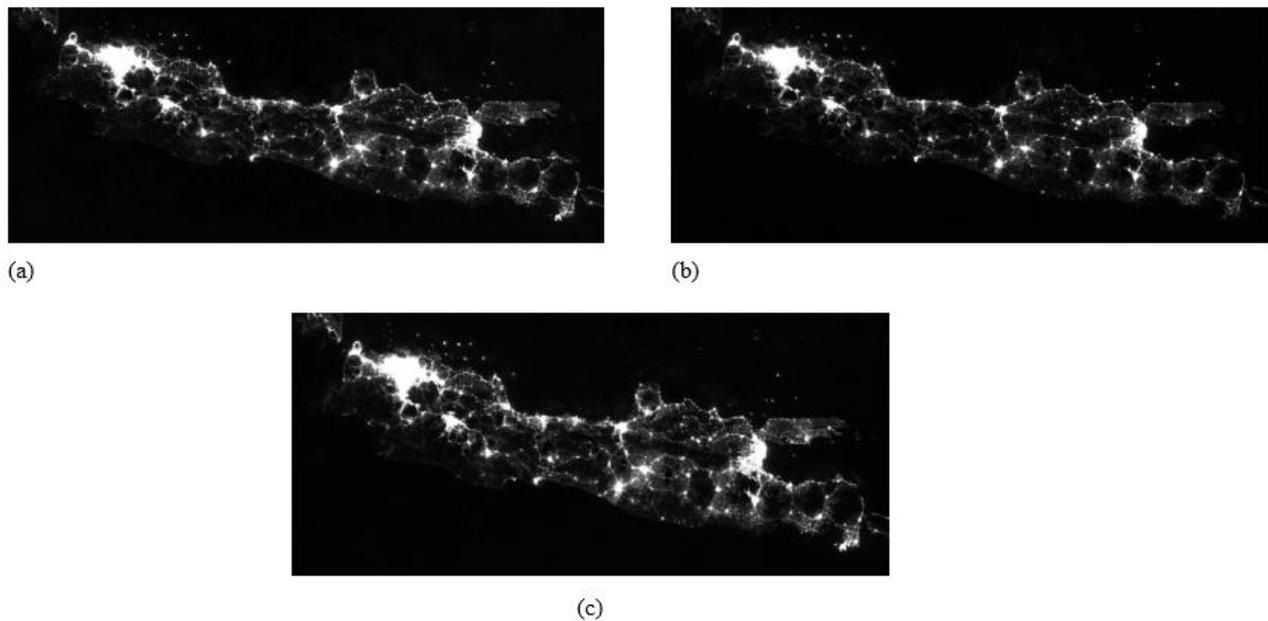
Fitria (2014) describes the three methods used by BPS-Statistics Indonesia to collect population data, namely population census, population survey, and population registration. The population survey occurring every ten years comprises a series of activities ranging from acquiring data to publishing demographic, economic, and social data from each region (Fitria, 2014). Ajie (2008) defines population survey as population enumeration with a technique of taking samples in an area to be used as a reference for the condition of its population and population registration as

the recording of population data which the authorities carry out regularly to the lowest level such as neighborhood and urban village. However, the population census, survey, and registration require a lot of effort, energy, cost, and time in their implementation (Priyono, 1992). To overcome these limitations, remote sensing offers image products that can be used to generate time-sensitive data.

One form of remote sensing-derived data is satellite imagery. Each satellite has a different sensor to capture the spectral energy reflected by objects on the earth's surface. For example, the Visible Infrared Imaging Radiometer Suite (VIIRS) is one of the sensors aboard the Suomi NPP satellite that has a Day-Night Band (DNB) to detect human presence, urban settlements, and other activities at night that require lighting (Amaral et al., 2006). Sutton et al. (1997) also explain that remote sensing products, especially nighttime light (NTL) data, are an alternative source of information for identifying urban settlements that can be used indirectly for further analysis, such as the spatial distribution and estimation of populations. Previous studies have proved a good correlation between the light-emitting or lit area and the number of residents. Therefore, linear regression is used to estimate population numbers from the logarithmic form of the allometric formula that mathematically defines population counts as a function of lit areas to find the allometric coefficient and exponent. This method requires that the regression formula be back-transformed to an allometric formula. This study aimed to analyze the application of nighttime light imagery in estimating population size.

---

\* Corresponding Author



**Figure 1.** Java Island at night as captured by VIIRS from year to year: (a) 2015 (b) 2017 (c) 2019

## 2. DATA AND METHODS

### 2.1 Data

The data used were the yearly composite VIIRS DNB images in 2015, 2017, and 2019 with a median data type acquired from EOG Data Mines ([HTTP://eogdata.mines.edu/products/vnl/](http://eogdata.mines.edu/products/vnl/)) and Google Earth Engine. The NTL imagery used in this study has a spatial resolution of 500 meters and a spectral resolution of 0.5–0.9  $\mu\text{m}$  according to the DNB sensor. Data on the number of residents per city/regency in 2015, 2017, and 2019 were obtained from BPS-Statistics Indonesia. Each city/regency in question is part of the island of Java that administratively belongs to the provinces of DKI Jakarta, Banten, Jawa Barat, Jawa Tengah, Jawa Timur, and DI Yogyakarta.

### 2.2 Methods

#### 2.2.1 Threshold Determination

Chowdhury et al. (2012) said that NTL data need to be corrected to remove the blooming effect resulting from a fault in the sensor, which can overestimate light emissions in an urban region so that the lit area looks more extended than it should be. This effect can be corrected using a threshold, which, according to Liu et al. (2016), is determined by setting the pixel value of the lowest light intensity in the least developed cities as the minimum value and the highest intensity in the most developed ones as the maximum value of the threshold. In addition, this threshold approach is a trial-and-error method, meaning that it requires several attempts to find the optimal minimum and maximum values.

#### 2.2.2 Lit Area Extraction

The extracted value is not the light intensity but the light-emitting or lit area, which is considered a settlement. The lit area was used as a variable in population estimation because many scholars have proved its strong linear correlation with population number (e.g., Amaral et al., 2006; Sutton et al., 2001; 1997; Zhuo et al., 2009).

#### 2.2.3 Samples

The samples in this study were population data per city/regency acquired from BPS-Statistics Indonesia. They were divided into modeling samples to create a population estimation model and validation samples to assess the accuracy of the model generated using the modeling samples. This division followed the ratio of 7:3, i.e., seven modeling samples for every three validation samples.

#### 2.2.4 Linear Regression Analysis

Chowdhury et al. (2012) found a strong linear relationship between the log of the lit area and the log of the population number from previous studies. Therefore, the following logarithmic model was applied (Lo, 2001):

$$\log P = \log a + b \log A \quad (1)$$

where  $P$  = population  
 $A$  = lit area  
 $a$  = coefficient  
 $b$  = exponent

The logarithmic model is the result of the transformation of the allometric formula between the built-up area and the total population based on research conducted by Tobler (1969). The allometric formula in question is as follows:

$$P = aA^b \quad (2)$$

where  $P$  = population  
 $A$  = lit area  
 $a$  = coefficient  
 $b$  = exponent

The logarithmic model was used for linear regression analysis because the transformed allometric formula has the basic form of a regression formula, as presented below:

$$Y = a + bX \quad (3)$$

where  
 Y = dependent variable  
 X = independent variable  
 a = intercept  
 b = coefficient

Harvey (2002) further explained that the coefficient of determination of the regression results does not determine the accuracy of the model; therefore, the above regression formula must be back-transformed to the allometric formula used for calculating the estimated population. The back-transformation from regression to allometric is shown in the formula below (Chowdhury et al., 2012):

$$Y = e^a \cdot (X)^b \quad (4)$$

where  
 Y = estimated population  
 X = lit area  
 a = coefficient  
 b = exponent  
 e = Euler's number (2.71828)

### 2.2.5 Percent Error Calculation

Linear regression analysis can produce a correlation coefficient, a coefficient of determination, and a standard error of estimate to determine the strength of intervariable correlation in 2015, 2017, and 2019. However, the coefficient of determination generated by the regression analysis does not show the accuracy of the model, warranting the need to calculate percent error (% error) as a better indicator for the overall accuracy (Chowdhury et al., 2012). The percent error was calculated using the following formula:

$$RE = \frac{(X)-(Y)*100}{Y} \quad (5)$$

where  
 RE = relative error (% error)  
 X = estimated or expected population  
 Y = actual population observed

## 3. RESULTS

### 3.1 Linear Regression Analysis Results

The linear regression analysis used NTL data in 2015, 2017, and 2019 with a divided modeling sample. It produced the correlation coefficient (r), coefficient of determination (r<sup>2</sup>), and standard error of estimate to define the correlation between lit area and population. The linear regression results of each year can be seen in table 1.

Year	r	r <sup>2</sup>	Std. Error of estimate
2015	0.929	0.863	0.34037
2017	0.930	0.865	0.33204
2019	0.931	0.866	0.33359

**Table 1.** Results of the regression analysis models in 2015, 2017, and 2019.

Table 1 shows that each year had a high r<sup>2</sup> value, which can be concluded beforehand that the lit area as an independent variable has a significant influence on the dependent variable, population.

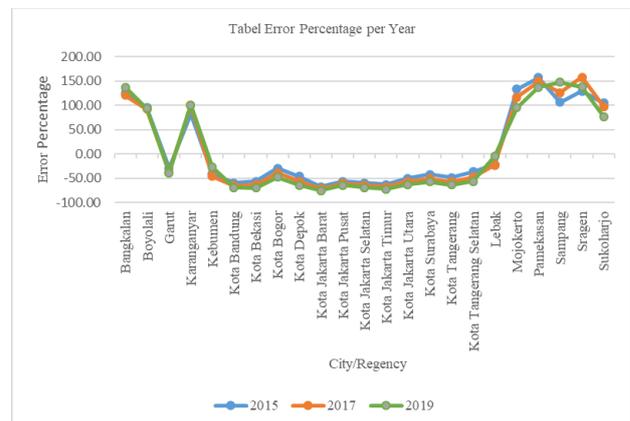
Moreover, the average r<sup>2</sup> was 0.865, indicating that the lit area significantly influences 86% of the entire modeling samples. In addition, the 2017 and 2019 data generated the best models, as indicated by their has highest r<sup>2</sup> and smallest standard error of estimate with very slight differences. The similarity shared by both models produced allometric formulas with coefficients and exponents that were not much different (Table 2). This allows the 2017 model to estimate the population in 2019 using only the 2019 lit area data without substantial deviation from the estimation results of the 2019 model itself, which then enables predictions for the coming years using the 2019 model.

Year	Coefficient	Exponent
2015	9.8626	0.7249
2017	9.9002	0.6927
2019	9.8471	0.678

**Table 2.** Coefficients and exponents of the allometric formulas back-transformed from the regression formulas.

### 3.2 Population Estimation

Estimation was carried out using the allometric formulas that had been back-transformed from the regression formulas. Using the allometric formulas, the lit area was integrated with the validation samples to see the accuracy of the model. A % error close to 0 indicates that the population estimated using the model represents the actual population. Figure 2 shows the % error of each validation sample in 2015, 2017, and 2019 for several cities and regencies. The % errors of the cities showed that the expected population was underestimated or smaller than the actual population observed. On the contrary, the population estimates in the regencies were mostly overestimated or much larger than the actual population. The validation results also showed that the highest overestimate was in 2015 in Pamekasan Regency with 156.69% error, while the largest error for the underestimated population, -75.59%, was in 2015 in the City of Jakarta Barat.



**Figure 2.** Graphic of percent errors of the population estimation models per year by city/regency.

Table 3 groups the estimation results, both underestimated and overestimated, of the cities/regencies into their respective provinces. At the provincial level, the % errors were generally smaller than at the city/regency level because the samples used were at the city/regency level. Given that a province covers a larger area than a city/regency, these results suggest that predictions at a higher regional level also produce population estimates with higher accuracy. Also, based on Table 3, the capital city of the country, DKI Jakarta, was the only province in 2015 with underestimated population, indicating that most of its cities have an areal extent that is not proportional to the

population or, in other words, high population density. On the other hand, in Jawa Timur and DI Yogyakarta, the majority of

their regencies with fewer populations than the lit area, resulting in an overestimate.

Error Percentage 2015					
Province	Actual Population	Estimated Population	Error Percentage	Median Error Percentage	Average Error Percentage
DKI Jakarta	10177924	4091485	-59.80	60.79	68.85
DIY Yogyakarta	3679176	5277136	43.43	47.66	46.42
Banten	11955243	12914932	8.03	42.51	57.78
Jawa Barat	46709569	47222749	1.10	29.24	39.78
Jawa Tengah	33168959	42291266	27.50	40.17	47.99
Jawa Timur	39363743	58124741	47.66	53.37	63.09
Total	145054614	169922310	17.14	42.13	52.44

**Table 3.** Percent errors of the 2015 population estimates per province.

Percent Errors 2017					
Province	Actual Population	Estimated Population	% Error	Median % Error	Mean % Error
DKI Jakarta	10,374,235	3,635,179	-64.96	66.03	70.99
DI Yogyakarta	3,762,167	5,242,018	39.34	53.27	45.71
Banten	12,448,160	11,369,452	-8.67	52.10	48.81
Jawa Barat	48,037,827	40,676,763	-15.32	34.11	34.67
Jawa Tengah	33,641,781	43,116,554	28.16	36.45	46.28
Jawa Timur	39,909,055	59,119,618	48.14	81.73	63.07
Total	148,173,225	163,159,584	10.11	47.42	50.44

**Table 4.** Percent errors of the 2017 population estimates per province.

Percent Errors 2019					
Province	Actual Population	Estimated Population	% Error	Median % Error	Mean % Error
DKI Jakarta	10,557,810	3,218,153	-69.52	70.63	80.19
DI Yogyakarta	3,842,932	5,758,213	49.84	53.44	49.63
Banten	12,927,316	11,089,819	-14.21	51.21	44.68
Jawa Barat	49,316,712	40,965,792	-16.93	35.11	38.41
Jawa Tengah	34,091,497	45,120,481	32.35	38.27	43.54
Jawa Timur	40,325,338	60,259,106	49.43	57.61	65.20
Total	151,061,605	166,411,564	10.16	47.42	52.24

**Table 5.** Percent errors of the 2019 population estimates per province.

For the entire island of Java, the estimated population in 2015 had a 17.14% error, with a median of 42.13% and a mean of 50.44%. Compared with this, the 2017 developed model was better in that it showed a higher correlation between lit area and population and a smaller standard error of estimate, meaning that the resulting estimates were more accurate. Table 4 shows that the estimates in 2017 were substantially different from those in 2015. There were three provinces with underestimated population numbers in 2017, but this case was only found in one province in 2015. Jawa Barat is one of the provinces with very different estimation results, namely, overestimate in 2015 (1.10% error) but underestimate in 2017 (-15.32% error), because the regression formulas (coefficients and exponents) derived for the two years were fairly significantly different. However, most cities in the province tended to have underestimated results with

large % errors. This is due to the fact that many cities were densely populated and clustered.

Figure 3 compares the estimates with the actual populations of the cities/regencies of Jawa Barat Province in 2017. It shows that most cities/regencies with underestimated results tended to have large differences between the estimated and actual populations. Although several cities/regencies showed overestimated results, the estimates were not much different from the actual numbers. Most of the provinces with underestimated results formed a graphic pattern, as shown in Figure 3.

Figure 4 compares the estimates with the actual populations of the cities/regencies of Jawa Tengah Province in 2019. It shows that the modeled populations in almost all cities/regencies were overestimated because this province has regencies with larger lit

areas (an independent variable) than the cities. This is especially because the cities tend to have narrow areas administratively. In addition, cities are more developed than their surroundings and experience population growth every year because of the impact of urban economic and physical developments that attract influxes of migrants.

Figure 5 shows a graph comparing the lit areas in 2015, 2017, and 2019 in several cities/regencies. The lit area had covered the entire cities (100%) since 2015; thus, there would not be an increase in 2017 and 2019. On the contrary, the lit area in the regencies continued to increase during the observation years. For these reasons, the population estimates in the cities were less accurate than in the regencies, as evidenced by the increasing % errors of the estimated populations in the cities in, for instance, the DKI Jakarta Province (see Tables 3, 4, and 5).

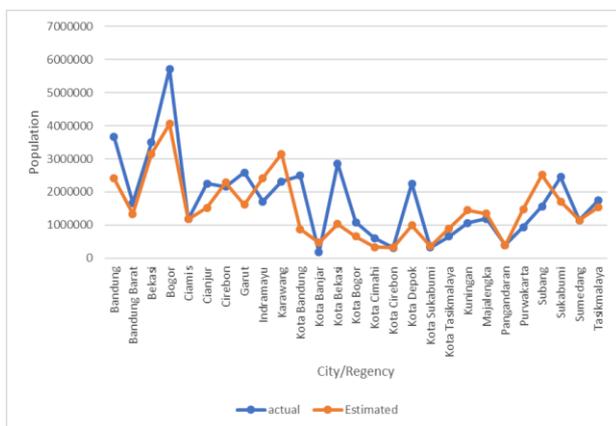


Figure 3. Comparison of estimated populations to actual populations in Jawa Barat Province in 2017 by city/regency.

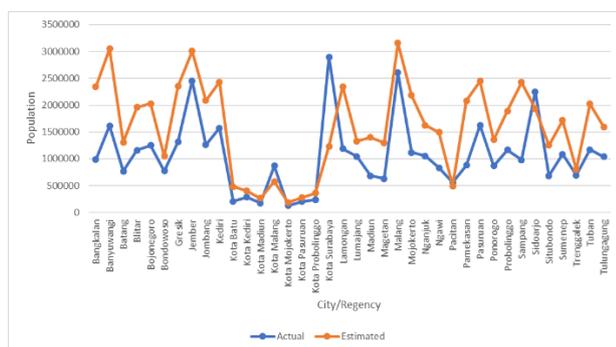


Figure 4. Comparison of estimated population to actual population in Jawa Tengah Province in 2019 by city/regency.

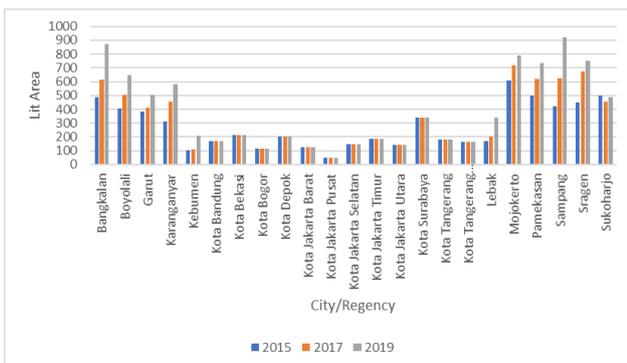


Figure 5. Lit areas in 2015, 2017, and 2019 per city/regency.

#### 4. SUMMARY

Nighttime light (NTL) images like VIIRS products can assist in providing population data by estimating the population based on the lit area. However, the derived population estimates have fewer percent errors for areas with one regional level above that of the samples used to avoid larger errors. In this case, the sample used is at the regency/city level, and the estimated populations of the provinces and Java Island have much smaller % errors; therefore, it is better to use the estimation results of the province or the island to give an initial picture of the population estimates in the city/regency level. The disadvantage of population estimation using lit area from NTL data as the only independent variable is that, in the city, it does not increase from year to year, resulting in a bias that will increase the % error of the estimates. To overcome this weakness, it is necessary to add ancillary variables such as light intensity that increases from year to year in each city/regency. Aside from adding or using more variables, another possible way to decrease errors is also to apply a correction factor, i.e., a constant value, to make the estimates closer to the actual population. However, the correction factor applied to each city/regency, province, and Java Island differs, and estimations at the city/regency level will require diverse correction factors.

#### ACKNOWLEDGMENTS

The authors greatly appreciate invaluable inputs from fellow researchers in the preparation of this paper. Gratitude is extended to Universitas Gadjah Mada, Yogyakarta, Indonesia, for funding the research and result publications through the Final Student Project Recognition (*Rekognisi Tugas Akhir*, RTA) in the fiscal year 2021 with the grant number 3143/UN1.P.III/DIT-LIT/PT/2021.

#### REFERENCES

Ajie, M. U. (2008). *Pengantar Studi Kependudukan*. Banda Aceh: Yayasan Pena Banda Aceh.

Amaral, S., Monteiro, A. M. V., Camara, G., & Quintanilha, J. A. (2006). DMS/OLS night-time light imagery for urban population estimates in the Brazilian Amazon. *International Journal of Remote Sensing*, 27(5), 855–870. <https://doi.org/10.1080/01431160500181861>.

BPS-Statistics Indonesia. (2019). *Statistik Sumber Daya Laut dan Pesisir 2019*. Jakarta: Badan Pusat Statistik.

BPS-Statistics Indonesia. (2020). *Statistik Indonesia 2020*. In *Statistical Yearbook of Indonesia*. Jakarta: Badan Pusat Statistik.

Chowdhury, P. K. R., Maithani, S., & Dadhwal, V. K. (2012). Estimation of urban population in Indo-Gangetic Plains using night-time OLS data. *International Journal of Remote Sensing*, 33(8), 2498–2515. <https://doi.org/10.1080/01431161.2011.616550>.

Daily, G., P. Dasgupta, B. Bolin, P. Crosson, J. du Guerny, P. Ehrlich, C. Folke, A. M. Jansson, B.-O. Jansson, N. Kautsky, A. Kinzig, S. Levin, K.-G. Mäler, P. Pinstrop-Andersen, D. Siniscalco, and B. W. (1998). Food Production, Population Growth, and Environmental Security by Gretchen Daily. *Science*, 281, 1291–1292. <https://doi.org/10.1126/science.281.5381.1291>.

Fitria, P. (2014). *Kamus Geografi: Istilah dan Penjabarannya*. Bandung: Nuansa Cendekia.

Harvey, J. T. (2002). Estimating census district populations from satellite imagery: Some approaches and limitations. *International Journal of Remote Sensing*, 23(10), 2071–2095. <https://doi.org/10.1080/01431160110075901>.

Kulshrestha, S. K. (2007). Peri-urban areas: the concept, concern, and planning and development policies. *56th National Town & Country Planners Congress*, 31–37.

Liu, Y., Delahunty, T., Zhao, N., & Cao, G. (2016). These lit areas are undeveloped: Delimiting China's urban extents from thresholded nighttime light imagery. *International Journal of Applied Earth Observation and Geoinformation*, 50, 39–50. <https://doi.org/10.1016/j.jag.2016.02.011>.

Lo, C. P. (2001). Modeling the population of China using DMSP operational linescan system nighttime data. *Photogrammetric Engineering and Remote Sensing*, 67(9), 1037–1047.

Priyono. (1992). Seluk beluk registrasi penduduk dan peranannya dalam perencanaan pembangunan kependudukan. *Forum Geografi No.10, VI, 33–45*. <https://doi.org/https://doi.org/10.23917/forgeo.v6i1.4691>.

Sutton, P., Roberts, D., Elvidge, C., & Meij, H. (1997). A comparison of nighttime satellite imagery and population density for the continental United States. *Photogrammetric Engineering and Remote Sensing*, Vol. 63, pp. 1303–1313.

Tobler, W. R. (1969). Satellite Confirmation of Settlement Size Coefficients. *Area*, 1(3), 30–34. Retrieved from <http://www.jstor.org/stable/20000359>.