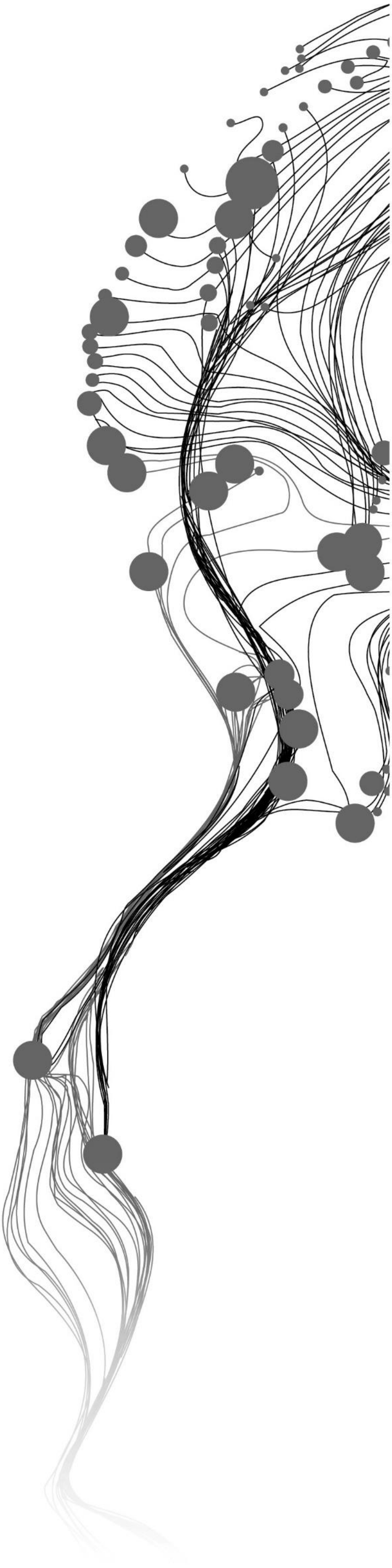
TOWARDS POPULATION DENSITY RETRIEVAL USING GAS-FLARE CORRECTED DMSP-OLS NIGH-TIME LIGHT OBSERVATIONS

ANDI MAYA PURNAMASARI
February, 2017

SUPERVISORS:
Valentijn Venus, M.Sc.Ing
Dr. A.G Toxopeus



TOWARDS POPULATION DENSITY RETRIEVAL USING GAS-FALRE CORRECTED DMSP-OLS NIGHT-TIME LIGHT OBSERVATIONS

ANDI MAYA PURNAMASARI
Enschede, The Netherlands, February, 2017

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.
Specialization: Natural Resources Management

SUPERVISORS:
Valentijn Venus, M.Sc.Ing
Dr. A.G Toxopeus

THESIS ASSESSMENT BOARD:
Dr. Ir. C.A.J.M. de Bie (Chair)
Dr. Ir. B.G.J.S. Sonneveld (External Examiner, SOW-VU, Centre for World Food Studies, The Netherlands)

ABSTRACT

Many socioeconomic and environmental problems could be explained with proper information of the distribution human population. Night-time light satellite images from DMSP-OLS is one of the most useful data to simply explain the population distribution. However, its saturation and gas flare contamination may mislead the information. As the oil and gas producer, Indonesia ranks 10th globally in generating gas flare. Thus, this research aims to examine the gas flare contamination in DMSP-OLS night-light observation in Indonesia by optimizing the available data of gas flare observation to improve the vectors that define the geographic regions containing gas flare, as a preliminary attempt to remove the light contamination from the stable night-light image. Whereas, the saturation and background noise in the nigh-light intensity image were resolved by setting a threshold for lit and unlit area. The gas flare removal demonstrated in this study intents to improve the variance of population density from recent census data at various administrative unit (province, regency, district and villages) with light observation. This work has revealed that an adequate procedure in correcting light contamination would significantly improve the coefficient of determination of light intensity in explaining variance of population density at various administrative unit (province, regency, district and villages). This study also reported the impact of gas flare removal in improving application of night-light data to assess population distribution, particularly in Indonesia.

Keywords:
DMSP-OLS stable night-light data, Night-light intensity, Population density, Gas flare

ACKNOWLEDGEMENTS

I gratefully acknowledge my supervisors Valentijn Venus, M.Sc. Ing and Dr. A.G Toxopeus for the valuable remarks, that were always helpful to come up with a good solution.

I would like to thank The Chairman of the thesis assessment board, Dr. Ir. C.A.J.M. de Bie, for the comments and recommendations.

I would also to thank drs. R.G. Nijmeijer for the support as Course Director NRM. In general, thanks to all lectures at the ITC for their interesting courses.

I would also extend my gratitude to my colleague in the Ministry of Marine and Fisheries for the help on data collection, sharing their thoughts and time

I would like to thank the NRM and GEM students in ITC for their support and encouragement, which contributed to the completion of my study.

Finally, I would give my very profound gratitude to my parents, husband, my sisters and best friends, who have encouraged and inspired me. I owe you big time.

Enschede, February 2017

Andi Maya Purnamasari

ABBREVIATIONS

DMSP	Defense Meteorological Satellite Program
DN	Digital Number
GPW	Gridded Population of the World
GRUMP	Global Rural Urban Mapping Project
NTL	Night-time Light
NPP	National Polar Partnership
OLS	Operational Linescan System

TABLE OF CONTENTS

Abstract i
Acknowledgements ii
Abbreviations iii
Table of contents iv
List of figures vi
List of tables vii
1. Introduction 1
1.1. Background 1
1.2. Research Problem 2
1.3. Research Objectives 3
1.4. Research Questions 3
1.5. Hypothesis 4
1.6. Research Assumption 4
1.7. Structure of the Research 4
2. Literature Review 5
2.1. Population Distribution 5
2.1.1. Gridded Population of the World 6
2.1.2. Global Rural Urban Mapping Project (GRUMP) 7
2.1.3. LandScan 8
2.2. Night-time Light as a Proxy for Population Density 9
2.2.1. Stable Night-Light 10
2.2.2. Light Contamination 12
3. Methodology 14
3.1. Population Distribution 14
3.1.1. Census Data 14
3.1.2. Noise 15
3.1.3. Consistency 15
3.2. Night-time Light 15
3.2.1. DMSP-OLS stable night-light observation 16
3.2.2. Removing Light Contamination 17
3.3. Night-Light Intensity and Population Density 20
3.3.1. Incorporating Stable night-light and Population Data 20
3.3.2. Night-Light Intensity as a Proxy for Population Density 21
4. Result and Discussion 24

4.1. Population Distribution 24
4.1.1. Removing Potential Noise 24
4.1.2. Internal Consistency Check 25
4.2. Removing Light Contamination 25
4.3. Night-time Light as a Proxy of Population Density 27
4.3.1. Normal Distribution 28
4.3.2. Correlation Coefficient 29
4.3.3. Regression Analysis 30
4.4. Discussion 34
4.4.1. Gas flare correction of DMSP-OLS night-time light improve the variance of population density 34
4.4.2. Incorporating Population Distribution and Marine Debris Case 35
5. Conclusion and recommendation 39
5.1. Conclusion 39
5.2. Limitations and Improvements for Future Assessment 39
Reference 41
Appendix 1 List of Data 46
APPENDIX 2 pROVINCE DATA 47
aPPENDIX 3 rEGENCY 49
Appendix 4 District 51
Appendix 5 village 52
AppendiX 6 Pearson's table 53
appendix 7 z table 54

LIST OF FIGURES

Figure 1 Population Distribution in Census 2010 5
Figure 2 Comparison of GPWv3 and GPWv4 Population Density Estimation 6
Figure 3 GRUMP Population Density 2000 8
Figure 4 Sources of Night-time Light (Elvidge, et.al., 2015) 9
Figure 5 Night-time light dataset process 11
Figure 6 Cloud free coverage, Raw average visible band, Cleaned up average visible band 11
Figure 7 Stable night-light image 12
Figure 8 Population of Indonesia 14
Figure 9 DMSP-OLS night-light observation 16
Figure 10 Gas Flares Observation 17
Figure 11 Gas flare removal process 18
Figure 12 Circular pattern of gas flare in the stable night-light 19
Figure 13 Circular and ellipticity of gas flare could be affected by in-situ wind direction 19
Figure 14 incorporating the population data and the light data 21
Figure 15 Linear Regression of Night-light Intensity and Population Density 22
Figure 16 Village administrative unit with sliver polygon 24
Figure 17 Population data of village level compared with tabular data of district level for Jakarta Utara... 25
Figure 18 Gas Flare Estimation 25
Figure 19 Identified Gas Flares 26
Figure 20 Aggregation of Administrative Unit 27
Figure 21 Distribution of Night-Light Intensity in Province level 29
Figure 22 Scatter diagram showing the relationship between night-light intensity and population density with and without gas flare removal 31
Figure 23 Bootstrap Distribution of Mean of Ln Light Intensity 32
Figure 24 Significant correlation between night-light intensity and population density in lit area 33
Figure 25 Simplified diagram of lifecycle of plastic and ways it becomes marine debris 35
Figure 26 Significant Correlation between Night-Light Intensity and Population Density in Pixel Level... 37

LIST OF TABLES

Table 1 Comparison of Gridded Population of the World Version 7
Table 2 Summary of Global Population Distribution Maps 8
Table 3 Comparison of DMSP-OLS and NPP-VIIRS in Light Observation 10
Table 4 Summary of Census 2010 and Administrative Boundary of Indonesia 15
Table 5 Area with sliver polygon 24
Table 6 Identified Gas Flares 26
Table 7 Summary of Administrative unit aggregation 27
Table 8 Summary of Light data in Administrative Unit 28
Table 9 Correlation Coefficient of Light Intensity and Population Density 30
Table 10 Regression Analysis of Light Intensity and Population Density 31
Table 11 Regressiont Coefficient of Light Intensity and Population Density 31
Table 12 Correlation Coefficient of Light Intensity and Population Density in Pixel Level 37
Table 13 Regression Coefficient and Bootsrap for Coefficient of Light Intensity and Population Density in Pixel Level 38

1. INTRODUCTION

This chapter describes the background of the study, justification of the research problem, research objectives and research questions, followed by hypotheses to address these research objectives.

1.1. Background

The reasons for environmental decline are complex, but population factors play a significant role (Creel, 2003). Many global environmental problems are human induced or effect of human activities. One reason for the small number of studies that incorporate human factors in global change modelling is the lack of suitable population related data (Tobler, Deichmann, Gottsegen, & Maloy, 1997). The available information on population distribution are usually from censuses that release statistical data and associated variables in a spatially aggregated form, typically at a given administrative unit (Khomarudin, Strunz, Post, Zoßeder, & Ludwig, 2009). However, administrative borderlines are not usually drawn to represent geographical phenomena. Furthermore, human population density in an administrative region does not provide the spatially explicit details (Elvidge, Baught, Kihn, Kroehl, & Davis, 1997), accurate and spatially detailed population distribution data were considered more valuable than distribution obtained with aggregated population data (Stevens, Gaughan, Linard, & Tatem, 2015).

Information of the size and distribution population is essential for understanding and responding to many socioeconomic and environmental problems. There is demand for population products that accurately identify the spatial distribution and density of the population to meet the growing demand for immediate and well-informed decision making (Maantay, Maroko, & Herrmann, 2007). High resolution, contemporary data on human population distributions are vital for measuring impacts of population growth, monitoring human-environment interactions and for planning and policy development (Stevens et al., 2015). Yet, population input data are inevitably highly variable in terms of quality, resolution, and accuracy, in ways that are not quantifiable (Balk et al., 2006), to make the most of these data available, improvement methods are needed to more accurately estimate population distributions.

Population estimation is an uncertain science (Nelson, 2004). When dealing with modelling of population distribution, a wide range of issues need to be taken into consideration, such as type and sources of population and boundary data, attribute and spatial accuracy, and modelling techniques (Deichmann, 1996). Population modelling methods essentially involve some form of re-distribution of aggregate census counts using ancillary datasets at finer spatial detail that are known to influence human population distribution (Linard, Alegana, Noor, Snow, & Tatem, 2010). While statistical model aims to identify the relationship between variables that determine the population distribution, i.e. socio-economic variables, remote sensing image, and population census data.

Remote sensing and geographic information systems (GISs) have long been used to estimate population, particularly for large areas (Wang & Wu, 2010). Current global population datasets that adjust the effectiveness of census data such as Gridded Population of the World (GPW), Global Rural Urban Mapping Project (GRUMP) and LandScan. GPW models the distribution of the human population on a continuous raster surface. The series started in 1995 and now it is in fourth version (GWPv4), that published in 2015. The gridded population dataset was created using uniform distribution or proportional allocation, where population was allocated into grid cells through the simple assumption that the population of a grid cell is an exclusive function of the land area within that pixel. Another high-quality population product which already incorporate various ancillary data sets, such as night time lights, transportation network, and various landmark, to generate final population data sets was the disaggregated population density by Global Rural Urban Mapping Project (GRUMP). Yet, the underlying modelling frameworks of these product only rely on coarser-scale and more generalizable input data that provide more flexibility for incorporating regional and global scale data sets (Jia, Qiu, & Gaughan, 2014; Stevens et al., 2015).

1.2. Research Problem

One of the most intuitive uses of night-time light satellite images from Defense Meteorological Satellite Program–Operational Linescan System (DMSP-OLS) is to use it as a proxy for the location of human population (Doll, 2010). The night-time light images can detect the artificial lights from cities, towns, industrial sites and other human activities at night (Elvidge et al., 1997). The brightness of nocturnal light reveals the density of human activities (Bergs, 2007).

The night-time light data have been used in many application of remote sensing techniques, such as economic activity (Elvidge, Sutton, et al., 2009; Mellander, 2013), greenhouse gas emissions (Elvidge et al., 1997), light pollution (Gallaway, Olsen, & Mitchell, 2009), disaster management (Kohiyama et al., 2004; Liu & Clarke, 2003), social analysis (Roychowdhury, Jones, Arrowsmith, & Reinke, 2011), urban extents (Henderson, Yeh, Gong, Elvidge, & Baugh, 2003; P. C. Sutton, 2003), and population studies (Anderson, Tuttle, Powell, & Sutton, 2010; Gao, Huang, He, Sun, & Zhang, 2016; Ghosh, Anderson, Elvidge, & Sutton, 2013; Pozzi, Small, & Yetman, 2002).

These remotely sensed images cannot indicate population density directly but can describe the urban morphology of built-up and non-developed areas (Mennis, 2003). Night-time light images record the artificial lights from cities at night, thereby closely representing anthropogenic activities at the pixel level (Roychowdhury et al., 2011). Given its nature and spatial resolution, the DMSP-OLS night-time light images are the most suitable data source to represent urban concentration and expansion on continental and global scale (Henderson et al., 2003). Night-time lights provide a versatile and user friendly data source for the scientist, whether it is used simply to define an urban area or used more intensively to model population, economic activity or some other socio-economic parameter (Doll, 2008).

Aside from the limited spatial extent, pixel saturation and insufficient detection for infrequently lit area, the light from gas flare still exist in the stable night-light product and extremely bright, which often saturate the DMSP-OLS visible band (Doll, 2008; Elvidge et al., 2010). Gas flare is a combustion device to burn associated, unwanted or excess gases and liquids released during normal operation in many industrial processes, such as oil-gas extraction, refineries and coal industry. Refinery and gas production generates associated gases that are usually flared and vented. It is not related with population activities and human settlement, however the brightness of gas flares recorded in stable night-light often misinterpreted as urban extent.

Despite the extensive application of the night-time light data, only few researchers have addressed the issues of gas flares in the stable night-light product, most of the research have only focus on gas flare in global scale. Previous work by Bhandari and Roychowdhury (2011) and Pestalozzi (2012) eliminate the gas flare by obtaining mask from a map featuring location and extension of DMSP-OLS gas flare observation, which converted in a binary raster so that the gas flares locations had value zero whereas all others pixels had value one. While Lowe (2014) eliminate the gas flares using Overlay and Erase feature in ESRI ArcGIS. However, the polygons that encircle the gas flares are quite large. Consequently, it is inevitable that certain areas in proximity of gas flare with settlement and man-made light are cancelled out. While instead of removing the gas flare explicitly in monitoring socioeconomic parameter, Propastin and Kappas (2012) implemented different threshold values for pixels' digital numbers (DN) to delineate the lit urban areas of individual settlements and removing noisy emission sources such as fires and conspicuous gas flares. Even with this interest, no one to the best of our knowledge has reported the impact of gas flare removal in improving application of night-light data to assess population distribution, particularly in Indonesia.

Indonesia ranks as the world's 17th oil and 6th gas producer (Gustya, 2005), which is linked to gas flare combustion, rank 10th globally in generating gas flare. Thus, this research aims to examine the gas flare contamination in DMSP-OLS night-light observation, in Indonesia by optimizing the available data of gas flare observation to improve the vectors that define the geographic regions containing gas flare, as a preliminary attempt to remove the light contamination from the stable night-light image. The gas flare removal demonstrated in this study intents to improve the variance of population density from recent census data at various administrative unit (province, regency and villages) with light observation. In other words, if the gas flare correction procedure as proposed in this study is adequate, the correlation will improve significantly.

1.3. Research Objectives

The aim of this research is to improve data input quality of population distribution using gas flare corrected DMSP-OLS night-light data. In order to achieve the aim of this research, the specific objectives were formulated:

1. To identify the initial correlation between night-light intensity and population density
2. To enhance the coefficient determination between light intensity and population density by correcting the gas flare contamination.

1.4. Research Questions

Research questions were made to address the objectives aforementioned, as follows:

1. Is there a difference in the coefficient correlation of night-light intensity and population density by correcting the gas flare contamination?
2. How well does the variance of night-light intensity predicts the population density by correcting the gas flare contamination?

1.5. Hypothesis

The hypothesis to be tested, as follow:

1. H_0 : Gas flare contamination removal in a night-light dataset will not give any positive impact in correlation coefficient of night-light intensity and population density ($r_1 \geq r_2$, $z \geq 0$)
   H_1 : Gas flare contamination removal in a night-light dataset will not give any positive impact in correlation coefficient of night-light intensity and population density ($r_1 < r_2$, $z < 0$)

2. H_0 : Gas flare contamination removal in a night-light intensity dataset will not give any positive impact in explaining variance of population density ($\bar{R}^2$ with gas flare removal $\leq$ $\bar{R}^2$ without gas flare removal)
   H_1 : Gas flare contamination removal in a night-light intensity dataset will give a positive impact in explaining variance of population density ($\bar{R}^2$ with gas flare removal > $\bar{R}^2$ without gas flare removal)

1.6. Research Assumption

Despite the difference in observation time, the NPP-VIIRS gas flare estimation recorded after 2012 is still sufficient as independent observation of gas flare, for the reason that the gas flares occur at permanently fixed locations and are mostly continuously active for a period of years. The second measurement could narrow down the gas flare position which is useful for differential correction to reduce the errors.

1.7. Structure of the Research

The thesis will be presented in five chapters, and will be organized as follows:

Chapter 1 : Introduction, this section will present the general overview abbot the research. It will describe the main idea and justification of the topic, research problem, research objectives and research question.

Chapter 2 : Literature Review, this chapter will cover the theoretical background of the study.

Chapter 3 : Methodology, this chapter will describe the study area, data preparation and workflow to achieved the research objectives.

Chapter 4 : Result and Discussion, this chapter will show and discuss the result of the research

Chapter 5 : Conclusion and Recommendation, this chapter will present the answer of the research questions and the limitations with the conclusions and recommendations for further study.

2. LITERATURE REVIEW

This chapter cover the theoretical background of the study related to population distribution, night-time light as a proxy for population density and how population distribution linked with marine debris estimation.

2.1. Population Distribution

Available information on the human population distribution are usually collected through censuses. Governmental agencies generally release this as statistical data in a spatially aggregated form, typically only at a given administrative unit-level (Khomarudin et al., 2009). Census data is costly to collect, and updating of this data is therefore generally carried out only every 10 years or so.

Population density, or number of people living in an administrative unit is therefore commonly a derived variable since the raw census data is not published. However, the human population density in an administrative region does not necessarily provide enough spatially-explicit information to describe the distribution of pressures on the natural resources in a region (Elvidge et al., 1997). While conceptually straightforward, population density may be misleading particularly when population within a region is unevenly distributed (Deichmann, 1996).

A distinction in census enumeration is made between de jure and de facto population. The former is the population usually resident at a place, excluding visitors and including those residents that are temporarily absent. De facto population refers to the number of people present in an enumeration area at the time of the census. The most basic of population indicators is the size of the total population within a clearly defined geographic region (Deichmann, 1996). The number of people is assumed to be distributed homogenously within each unit area, even in the part of uninhabited areas e.g. lakes, forest, swamps and areas with high slopes (Bielecka, 2015). In fact, these spatial units do not reflect the spatial patterns of the built environment and the population distribution when analysed at fine spatial scales (Tenerelli, Gallego, & Ehrlich, 2015). Therefore, on vector data based choropleth population maps, the populations seem to be homogenously distributed over the area of the administrative unit, despite possibly significant variations in real population densities (Schneiderbauer & Ehrlich, 2007).

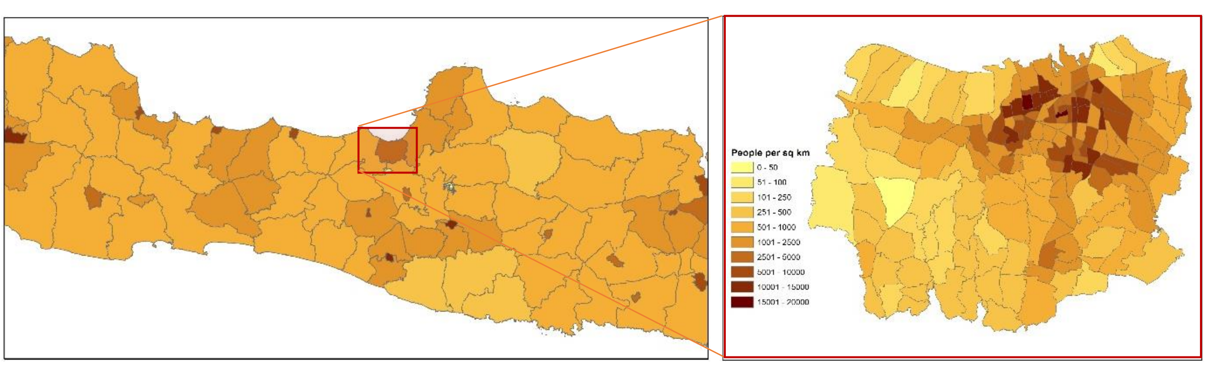


Figure 1 Population Distribution in Census 2010

Census data are used as benchmark data for studying population changes, and are key input for making projections of future population pressure, households changes, labour force and employment needs. Even though administrative borderlines used in census are not usually drawn to represent geographical phenomena, current global population datasets such as Gridded Population of the World (GPW), Global Rural Urban Mapping Project (GRUMP) and LandScan attempt to improve upon the effectiveness and usefulness of census data.

The fact that boundary and population data often did not come from the same source, derives the synchronicity problem. Administrative boundaries change frequently, and in some cases, boundary was only available for a previous census, and did not match the units for which population figures were defined (Tobler et al., 1997). In the few cases where a mismatch occurred, the population data had to be aggregated or disaggregated using simple proportional areal weighting, although more complex areal interpolation techniques could be applied (Goodchild, Anselin & Deichmann, 1993).

2.1.1. Gridded Population of the World

GPW models the distribution of the human population on a continuous raster surface. The purpose of GPW is to provide a spatially disaggregated population layer that is compatible with data sets from social, economic, and earth science disciplines, and remote sensing data. The series started in 1995 and now it is in fourth version (GWPv4), that published in 2015. GPWv4 is a minimally modelled data set that uses uniform distribution to disaggregate census data from their native input units into a thirty arc-second global grid (Doxsey-Whitfield et al., 2015).

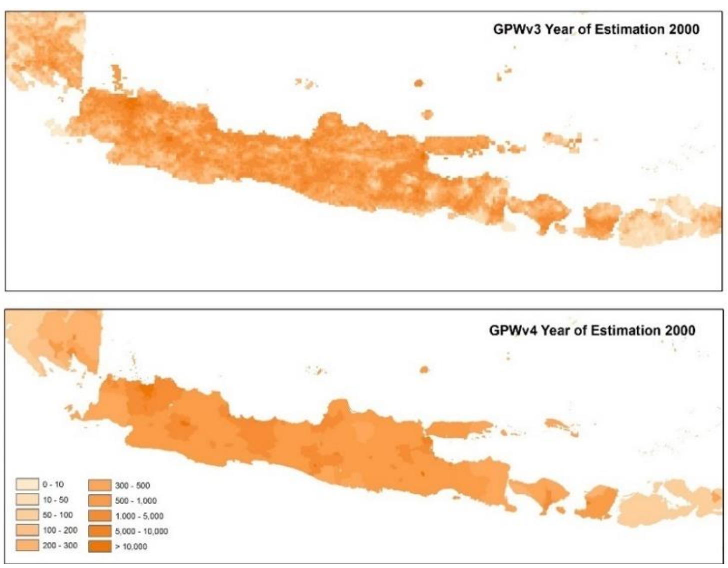


Figure 2 Comparison of GPWv3 and GPWv4 Population Density Estimation

The GPWv4 is a new release improving upon the previous version (GPWv3), using the most recent census data (round 2010). The two basic inputs of GPW are non-spatial population data (tabular counts of population from the recent census) and spatially-explicit administrative boundary data (Global Administrative Areas version 2; www.gadm.org). The GPWv4 provides a finer horizontal resolution and estimated the population data for 2000, 2005, 2010, 2015, and 2020. Overall the improvement of GPWv4 from the previous version as listed by Doxsey-Whitfield et al. (2015) as repeated in the table below.

Table 1 Comparison of Gridded Population of the World Version

Summary Information	GPWv1	GPWv2	GPWv3	GPWv4
Publication Year	1995	2000	2005	2015
Years of Estimation	1994	1990, 1995	1990, 1995, 2000	2000, 2005, 2010, 2015, 2020
Grid Resolution	5 arc-minute (~10 km)	2.5 arc-minute (~5 km)	2.5 arc-minute (~5 km)	30 arc-second (~1 km)
Number of Input units (subnational geographic units)	19,000	127,000	~400,000	~12,500,000
Census Variables	Total Population	Total Population	Total Population	Total Population, Sex, Age, Urban/Rural Designation

Source: CIESIN, 2015

The gridded population dataset was created using uniform distribution or proportional allocation, where population was allocated into grid cells through the simple assumption that the population of a grid cell is an exclusive function of the land area within that pixel. GPW proportionally allocated total population to grid cells based on the assumption that population is distributed evenly over administrative units (Tobler et al., 1997). Apart of applying the water mask, the method used does not include other geographic data as parameters to spatially disaggregate the population density. The areal-weighting method does not incorporate ancillary geographic data (e.g. land cover, urban extent, etc.) to allocate the population within a grid cell (Doxsey-Whitfield et al., 2015; Stevens et al., 2015).

2.1.2. Global Rural Urban Mapping Project (GRUMP)

While GPW is a simple redistribution across census unit, Global Rural Urban Mapping Project (GRUMP) builds on GPW by incorporating urban-rural designations in the spatial reallocation of population for each census unit, primarily derived from satellite nightlights (Balk et al., 2006; Stevens et al., 2015). GRUMP comprises three data products. First, GRUMP provides a higher resolution gridded population data product at 30 arc-seconds, or ~1km at the equator. Second, GRUMP's urban extents data set delineates urban areas based on NOAA's night-time lights data set and buffered settlement centroids (where night lights are not sufficiently bright). Third, GRUMP provides a points data set of all urban areas with populations of greater than 1,000 persons (SEDAC, 2005).

The objective of GRUMP is to disaggregate the urban area populations from the total population of the administrative unit into which the urban area falls (Balk et al., 2006). The most important source are night-time satellite images that show areas lit by streetlights and other permanent light sources that are concentrated in urban settlements (Elvidge et al., 1997). GRUMP is already integrating various ancillary data sets such as night-time lights, transportation network, various landmarks, etc. to generate final population data sets. This allows us to allocate urban and rural population separately, which effectively increases the number of input units and thus the effective resolution of the population grid (Balk et al., 2006). Yet, the underlying modelling frameworks of these products often rely on coarser-scale and more generalizable input data that provide more flexibility for incorporating regional and global scale data sets (Jia et al., 2014; Stevens et al., 2015).

2.1.3. LandScan

The LandScan dataset was first produced in 1998 as an improved resolution global population distribution database for estimating populations at risk. The original LandScan algorithms integrated globally consistent, but relatively coarse, spatial data. The basic concept of the LandScan data sets is to perform a spatial allocation of census reported population numbers based on models developed with spatially disaggregated data. The term population count is used instead of population density - that is based on residence, where people are likely to be during the day (Doll, 2008; Elvidge, Sutton, et al., 2009).

Figure 3 GRUMP Population Density 2000

The LandScan population distribution used a multi-layered, dasymetric, spatial modelling approach that is also referred to as a "smart interpolation" technique. In LandScan models, the typical dasymetric modelling is improved by incorporating and employing multiple ancillary or indicator data layers. The modelling process uses sub-national level census data for each country and ancillary datasets, including land cover, roads, slope, urban areas, village locations, and high resolution imagery analysis (Oak Ridge National Laboratory, 2014).

Table 2 Summary of Global Population Distribution Maps

	GPW	GRUMP	LandScan
Authors/ Developers	Center for International Earth Science Information Network (CIESIN), Columbia University	Center for International Earth Science Information Network (CIESIN), Columbia University	Oak Ridge National Laboratory (ORNL)
Resolution	30 arc second (~1km)	30 arc second (~1km)	30 arc second (~1km)
Scope	Global	Global	Global
Input data	- Census data - Administrative boundaries - Coastlines	- Census data - Administrative boundaries - Coastlines - Satellite night time light-derived urban extents	- Census data - Administrative boundaries - Land Cover - Coastlines - High resolution imagery - Elevation and slope - Roads - Populated areas (urban boundaries) and populated points (towns and villages)
Website	http://sedac.ciesin.columbia.edu/data/collection/gpw-v4	http://sedac.ciesin.columbia.edu/data/collection/grump-v1	http://web.ornl.gov/sci/landscan/

Source: Caribbean Handbook for Risk Information Management (CHARIM)

Table 2 summarize the global population maps. All the maps have similar resolution and global scope. All the maps generated using census and administrative data. However, the major difference is that GRUMP used satellite night-time light derived from urban extents. The use of night-time light imagery in LandScan was subsequently dropped as a model input due to the "overt effect of economic development on the brightness and intensity of lighting" (Elvidge et al., 2007 in Doll, 2008)

2.2. Night-time Light as a Proxy for Population Density

Night-time lights offer a consistent global measure of the spatial extent of human habitation, development and intensity of economic activity (Small, Elvidge, Balk, & Montgomery, 2011). Night-time lights data are a class of urban remote sensing products derived from satellite sensors with specialized low light imaging capabilities (Elvidge, Hsu, Baugh, & Ghosh, 2014). These images were collected by DMSP-Operational Linescan System (OLS) and NPP-Visible Infrared Imaging Radiometer Suite (VIIRS) sensors, which both have unique capability for global mapping of artificial lighting present at the Earth's surface (Elvidge, Baugh, Zhizhin, & Hsu, 2013) capable of detecting shortwave outgoing radiances down to 5E-10 Watts/cm^{2}sr and 2E-11 Watts/cm^{2}sr respectively. Both sensor use optical and near infrared data to produce the night light imagery which comprises of light emission from both man-made and the natural light sources in the earth surface (Elvidge et al., 2013). Thus, it is possible to detect artificial sky brightness surrounding cities and gas flares (Doll, 2008).

Both the DMSP-OLS and NPP-VIIRS visible band were designed to detect moonlit clouds at night. However,

Figure 4 Sources of Night-time Light (Elvidge, et.al., 2015)

not only it is possible to detect clouds illuminated by moonlight, but also it detects lights from cities, towns, industrial sites, gas flares, as well as ephemeral events such as fires and lightning-illuminated clouds.

Throughout this thesis, we will use the term stable night-light to refer to stable night-light data recorded using the DMSP-OLS sensor, whereas NPP-VIIRS light observation refers to the light data recorded using Visible Infrared Imaging Radiometer Suite carried by the Suomi National Polar-orbiting Partnership satellite.

Table 3 Comparison of DMSP-OLS and NPP-VIIRS in Light Observation

	DMSP-OLS	NPP-VIIRS
Platform	U.S. DMSP	NASA-NOAA Joint Polar Satellite System
Band	VIS an TIR	22 Spectral bands
Spatial Resolution	2.7 km Ground Sample Distance	742 m Ground Sample Distance
	5+ km Ground Instantaneous field of view	750 m Ground Instantaneous field of view
DN Values	6 bit quantization (VIS)	14 bit quantization in DNB
Calibration	No inflight calibration	Inflight calibration
Saturation	Saturation on bright lights	No saturation
Launch	Flown since 1972	Launched in 2011
Operation	Stop operation in 2012	Expected to continue for several decades

Adapted from Elvidge (2012)

Brightness and spatial extent of emitted light are often correlated to population density. The brightest pixels observed generally correspond to fully developed mixed use urban areas with near total outdoor illumination. Less brightly lit pixels may correspond to lower built area density with dimmer outdoor lighting throughout or to a small number of discrete lighted areas with somewhat brighter lighting (Small et al., 2011). Models of population derived from DMSP-OLS data have the potential to provide an inexpensive, easy to update means of mapping the size and spatial distribution of the human population at a global scale (Sutton, Roberts, Elvidge, & Baugh, 2001).

The drawback of the DMSP-OLS as identified by Elvidge et al. (2007) in Huang et al. (2014) were the blooming effect (i.e., overestimation of lit area) due to the coarse spatial resolution of data and reflectance of light from adjacent areas (e.g., water bodies). It shows the DMSP-OLS lights larger than sources on the ground, which over-represent built-up area and under-represent small settlements that are either poorly or infrequently lit due to insufficient detection by the sensor (Balk et al., 2006; Elvidge, Baugh, Ziskin, Anderson, & Ghosh, 2010). NPP-VIIRS data employ on board calibration, which is not available for the DMSP-OLS data (Elvidge, Baugh, Chi Hsu, Zhizhin, & Ghosh, 2015).

2.2.1. Stable Night-Light

The DMSP-OLS data screened to exclude the constraints i.e. sunlit and moonlit data, glare, observation with clouds and lighting features from the aurora (Michalopoulos & Papaioannou, 2012). By analysing the location, frequency, and appearance of lights observed in an image times series, it is possible to distinguish four primary types of lights present at the earth's surface, mostly man made light: human settlements, fires, gas flares, and fishing boats (Pozzi et al., 2002).

A night-time lights composite is made to serve as a baseline of persistent light sources. It made as an average of the highest lights imagery annually (Elvidge, Baugh, Chi Hsu, Zhizhin, & Ghosh, 2015). The resulting annual composite images of time-stable night-lights are created by overlaying all images captured during a calendar year, dropping images where lights are shrouded by cloud or overpowered by the aurora or solar glare (near the poles), and removing ephemeral lights like fires and lightning (Michalopoulos & Papaioannou, 2012). The final output was a geo-referenced composite of nighttime stable night-light (NSL) images with a spatial resolution of 1 km. These images record the percent frequency at which lights were detected, normalized by the number of cloud-free observations (Huang et al., 2014).

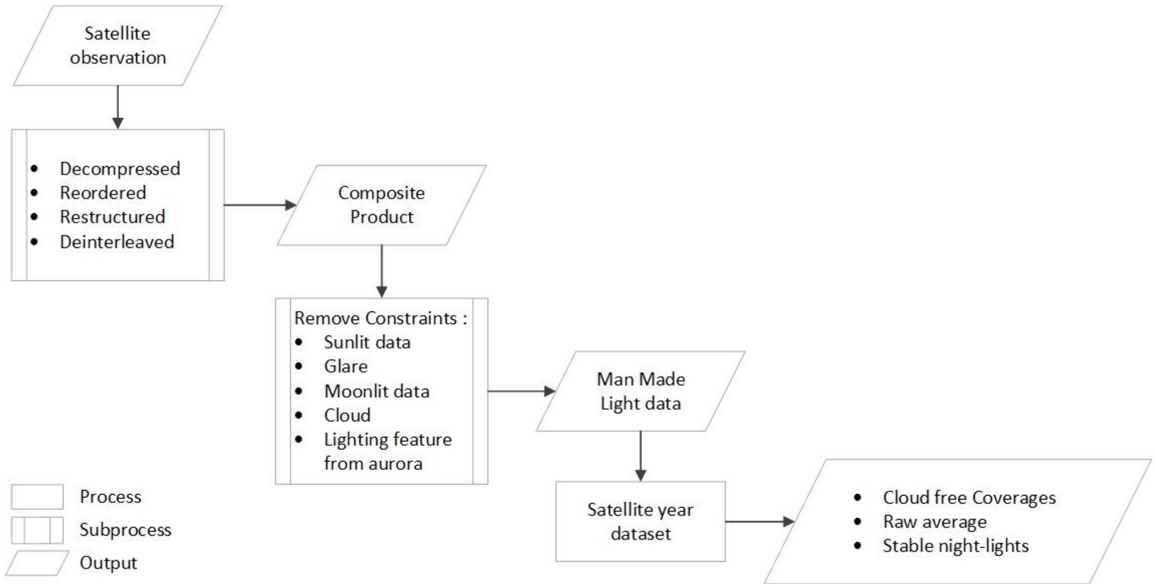


Figure 5 Night-time light dataset process

The stable night-light (Figure 6) contains of the lights from cities, towns, and other sites with persistent lighting, including gas flares. Ephemeral events, such as fires have been discarded. The background noise was identified and replaced with values of zero. Areas with zero cloud-free observations are represented by the value 255 (Lowe, 2014). Although the DMSP-OLS data are produced in byte format (0–255), the pixels values of the stable night-lights product are 6-bit digital numbers ranging from 0 to 63, due to the fact that the digital numbers in the image represent the percent frequency at which lights were detected, normalized by the number of cloud-free observations (Huang et al., 2014; Propastin & Kappas, 2012). However, with six bit quantization and a limited dynamic range, the recorded data are saturated in the bright cores of urban centers, in which the night-time light may be brighter, but the DN values are all 63 (Ma, Wu, Li, Peng, & Liu, 2014). Thus, Pixels with DN value of 63 are defined as saturation pixels (Doll, 2008; Elvidge, Sutton, et al., 2009).

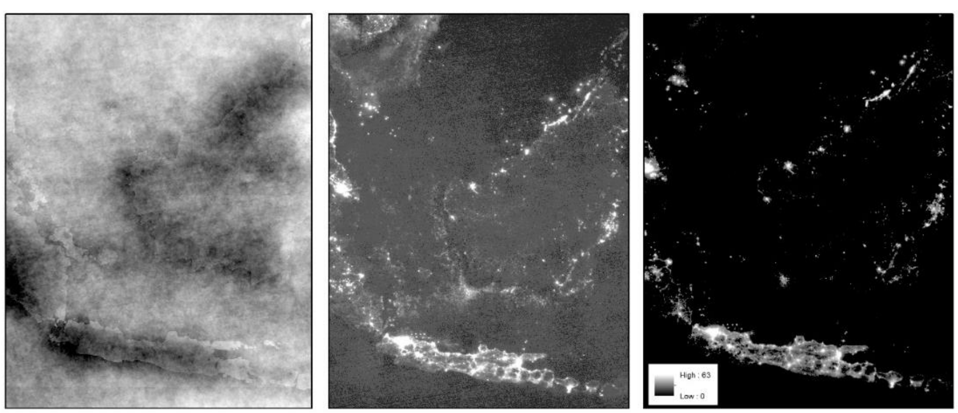


Figure 6 a) Cloud free coverage, b) Raw average visible band, c) Cleaned up average visible band

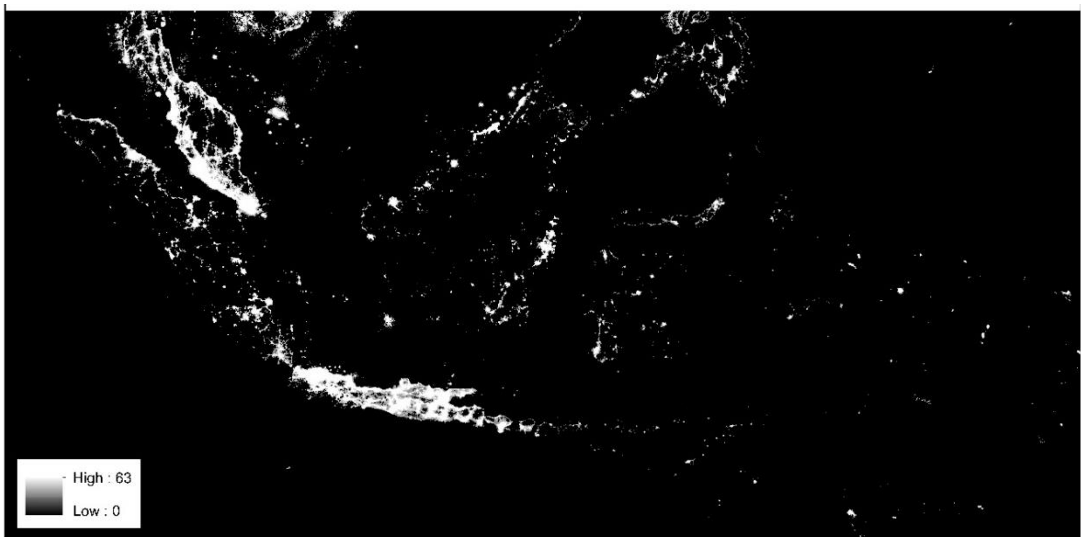


Figure 7 Stable night-light image

2.2.2. Light Contamination

Despite of the limited spatial extent, the light from gas flare still exist in the stable night-light product and extremely bright, which often saturate the DMSP-OLS visible band (Doll, 2008; Elvidge et al., 2010). Gas flaring is a combustion device to burn associated, unwanted or excess gases and liquids released during normal or unplanned over-pressuring operation in many industrial processes, such as oil-gas extraction, refineries, chemical plants, coal industry and landfills. Gas flares are smaller in size than biomass fires, occur at flare stacks and pits permanently fixed to a particular location and are mostly continuously active (Small et al., 2011) It is widely recognized as a waste of energy and an added load of carbon emissions to the atmosphere (Elvidge et al., 2009). It also generates noise, heat and provided large areas uninhabitable (Emam, 2015).

Gas flares typically very bright, show up as circular shape of saturated pixels surrounded by a sort of glowing of light, bright in the center and dim at the outer edges (Pestalozzi, 2012), yet it is impossible to identify gas flares inside of brightly lit urban centers (Elvidge et al., 2009). Moreover, most of the flare located in the remote area, thus satellite sensor has the potential for global systematic observation of flares and estimation of flared gas volume/CO2 emission. Remote sensing offers a potential solution to this through the detection, continuous monitoring, and mapping of flare locations over extended periods (Obinna Chukwubuikem Diony Anejionu, Blackburn, & Whyatt, 2014). However, none of the existing sensors have been designed specifically for detection and monitoring of gas flares (Elvidge, Zhizhin, Baugh, Hsu, & Ghosh, 2016).

The DMSP-OLS instrument is a sensor with an infrared telescope, a visible telescope, and a photomultiplier tube to enhance visible band signal at night. The photomultiplier tube is crucial for flare monitoring because it intensifies the visible and near-infrared bandpass by approximately one million times, allowing detection of surface lights one million times fainter than a satellite not equipped with similar technology. In addition, DMSP-OLS instrument include accurate geolocation.

The most extensive time series of global gas flaring, with national estimates of flared gas volumes, comes from low light imaging night-time data acquired by DMSP-OLS (Elvidge, Baugh, Zhizhin, & Chi Hsu, 2012), where gas flare identified in time series of annual light from 1994-2012. Elvidge, et al. (2009) defined that DMSP-OLS identify the gas flare based on three general characteristics as follow:

- Gas flares tend to form circular lighting features with a bright center and wide rims;
- Most gas flares are active for a period of years, but there are few gas flares that persist with little change in intensity over a full decade
- Gas flares tend to be in remote locations, outside of urban centers. In addition, LandScan global population density grid and NASA MODIS satellite hot spot data were used to clarify the identified gas flares on land.

Gas flare can easily be identified when it located offshore or in isolated area which not impacted by urban lighting (Elvidge et al., 2009). Furthermore, in order to improve the DMSP-OLS gas flare observation, Google Earth were used for visual confirmation of individual flares, label false detections and to eliminate the inclusion of human settlements features (Elvidge et al., 2009). The last year where DMSP-OLS estimated gas flare was 2012. DMSP-OLS orbit degradation from 2013 onward, resulted in solar contamination that made it impossible to produce global data for estimating flared gas volumes (Elvidge et al., 2016). NPP-VIIRS took over the functions of the DMSP-OLS since 2012, including gas flare monitoring. NPP-VIIRS is an improvement over DMSP-OLS in both spectral and spatial resolution. NPP-VIIRS observation works best at night and resulting in equally high quality fire monitoring, but subject to problems with spectral confusion of clouds and gas flares.

Few researchers have addressed the gas flares issues in stable night-light data. Previous work by Bhandari and Roychowdhury (2011) and Pestalozzi (2012) eliminate the gas flare by obtaining mask from a map featuring location and extension of DMSP-OLS gas flare observation, which converted in a binary raster so that the gas flares locations had value zero whereas all others pixels had value one. Whereas Lowe (2014) eliminate the gas flares using Overlay and Erase feature in ESRI ArcGIS. However, the polygons that encircle the gas flares are quite large. Consequently, it is inevitable that certain areas in proximity of gas flare with settlement and man-made light are cancelled out. While instead of removing the gas flare explicitly in monitoring socioeconomic parameter, Propastin and Kappas (2012) implemented different threshold values for pixels' Digital Numbers (DN) to delineate the lit urban areas of individual settlements and removing noisy emission sources such as fires and conspicuous gas flares.

3. METHODOLOGY

This third chapter will explain about the population distribution of the study area, the potential noise and how to overcome the signal noise in the data, as well as methodology to answer the research questions.

3.1. Population Distribution

Indonesia named as the 15th largest country. As an archipelagic country, Indonesia consists of 13.466 islands with total of 1.922.570 km² land area and 3,257,483 km² sea area. Indonesia stretches from 6°04'30" north latitude to 11°00'36" south latitude, and from 94°58'21" to 141°01'10" east longitude and lies on equator line. Based on the recent census (2010), the population of Indonesia in 2010 is 237.641.326 inhabitants, which makes it as the fourth most populous country on earth after China, India, and the United States. The density is 124 people s per square kilometres. The national population growth 1.49% per year.

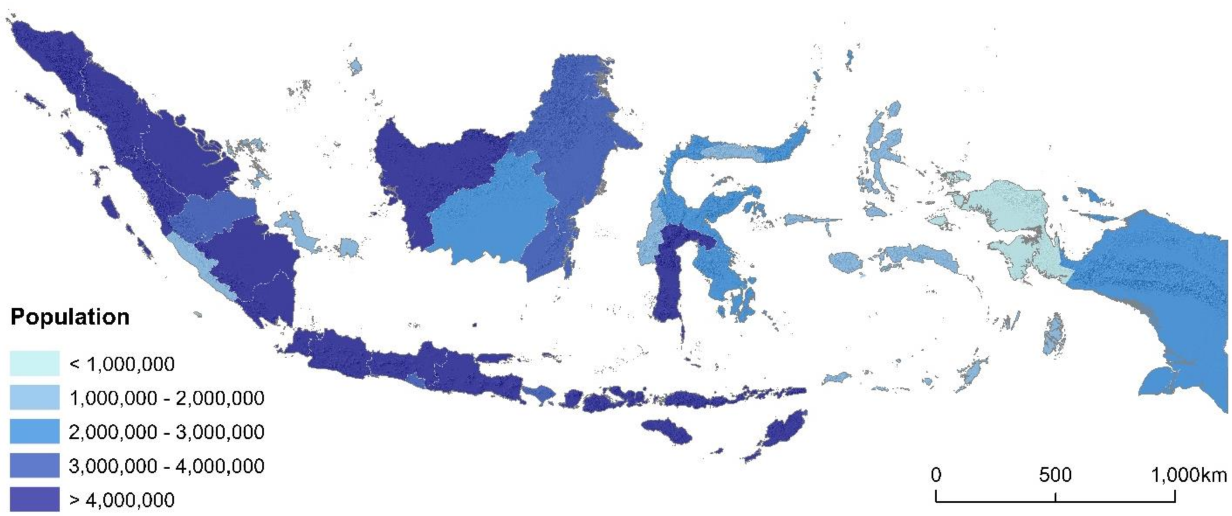


Figure 8 Population of Indonesia

3.1.1. Census Data

The census data is the only consistent source for demographic data with a national scope. It is the most reliable and detailed information in describing the finest administrative units and used as the basis of government development programs. The census data retrieved from National Census 2010 produce by Indonesia Central Bureau of Statistics, enumerated at the smallest governmental administrative region equivalent to village. This data relationally linked with the administrative boundary.

The most basic of population density, indicate by the total population within defined administrative boundary. The administrative unit has a unique identifier, which used to associate the polygon with the census data, and serves as the main key to linked the spatial features with the external tables. Changes in the administrative boundaries can affect the population density calculation, thus the administrative polygons at the village level data, which obtained from Central Bureau of Statistics, linked with the census data using unit identifier in the attribute. The identifier for the administrative unit used in this research for village unit data composed of: province code + regency code + district code + village code.

Table 4 Summary of Census 2010 and Administrative Boundary of Indonesia

Observed	Census 2010	Administrative Boundary
Smallest unit	Villages	Villages
Villages Count	77,064	77,474 Polygon
Population Count	237,641,326	-
Aggregated to level 4	6652 Districts	6652 Districts
Aggregated to level 3	504 Regency	504 Regency
Format	Tabular	Shapefile
Data Published	2010	2010

3.1.2. Noise

To reduce noise, the census data were checked for quality. Once this done, potential noise in census data can be removed. The author noticed several 'sliver polygon', particularly in the administrative 'village' units. Sliver polygon are here defined as a small, narrow, polygon features that typically appears following the overlay of two or more geographic datasets (Esri, n.d.). The issue of sliver polygon presence was checked by adjusting the administrative boundary with the census data in village level using the unique identifier. First the extent of the area was sorted from the smallest to the largest area. If there is a doubt about the extent of the area, the population density was checked. Areas with triangular shaped polygons and unreasonably high population densities were removed.

3.1.3. Consistency

If one process can be measured well, it can be used to make reasonable estimates of others. As a preventive measure to reduce error and improve data quality, the data should be examined for its quality to ensure the consistency and usability of the data. The internal consistency is done by compare relationship of the data within the same census data. In this case, the aggregated census data of village level compared with the tabular data of the census data in the same regency with coarser level (district). Any substantive tabulation at the observed administration level will have the same aggregation value with the adjusted population at coarser level. This is important in relation to improve data input quality, since the census data will be used as the basis of population distribution modelling and statistic purpose. The internal consistency check could detect the possibility of error, where if there is inconsistency in the census data, it will produce an additional error component in the model.

3.2. Night-time Light

On account of the DMSP-OLS sensor capability in detecting artificial lights from cities, towns, industrial sites and other human activities at night, the night-light data may be useful as a proxy for population density. However, the light saturation and contamination that still exist in the stable night-light data have to be corrected before examined the relationship of the dataset.

3.2.1. DMSP-OLS stable night-light observation

The night-time light (NTL) data obtained from the Defense Meteorological Satellite Program's satellite with Operational Linescan System-imager, abbreviated to DMSP-OLS, with ~1 km cell resolution. The file named with the satellite number and the year, this research will use F182010 file (F18 stand for the satellite number and 2010 is the year the data was recorded). There are 3 type of data that restored in each file, as mentioned by Lowe (2014), i.e.:

F182010_vdc_cf_cvg.tif : Cloud-free coverages tally the total number of observations that went into each 30 arc second grid cell. This image can be used to identify areas with low numbers of observations where the quality is reduced. In some years, there are areas with zero cloud-free observations in certain locations

F182010_4c_avg_vis.tif : Raw avg_vis contains the average of the visible band digital number values with no further filtering. Data values range from 0-63. Areas with zero cloud-free observations are represented by the value 255.

F182010_v4c_stable_lights.avg_vis.tif : The cleaned up avg_vis contains the lights from cities, towns, and other sites with persistent lighting, including gas flares. Ephemeral events, such as fires have been discarded. Then the background noise was identified and replaced with values of zero. Data values range from 1-63. Areas with zero cloud-free observations are represented by the value 255.

Specifically, stable night-light is the file that will be used in this analysis. However, there is contamination of light from saturation, background noise and gas flare that still exist in the Stable night-light product. Thus, the light contamination should be excluded from the analysis before the Stable night-light used to model the relationships between population density and the night-light intensity.

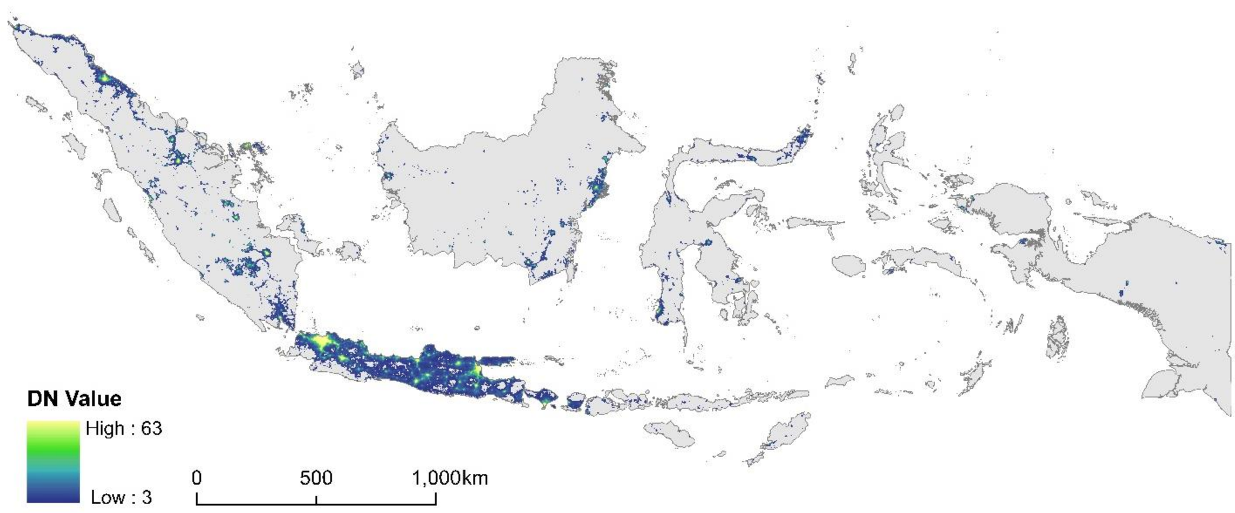


Figure 9 DMSP-OLS night-light observation

3.2.2. Removing Light Contamination

As mentioned before, in the stable night-light image, the background noise identified and replaced with values of zero. Therefore, this value should be excluded from the analysis, so the data values will range between 1-63. However, since the Stable night-light use six-bit quantization and has limited range, the light recorded in bright urban area might be saturated. In which the area with brighter light has limited DN values to 63. Thus, Pixels with DN value of 63 are defined as saturation pixels (Doll, 2008) and should be excluded from the analysis.

In contrast to Propastin and Kappas (2012), given the varying environmental context of flares in study area a simple threshold method alone was unsuitable for flare detection (Anejionu, Blackburn, & Whyatt, 2015). Of a few studies that had performed gas flare removal, the DMSP-OLS gas flare observation were most commonly used. However, the polygons that encircle the gas flares estimated by DMSP-OLS are quite large. Consequently, certain areas which is settlement or urban extent that located closed to the gas flare feature might have been cancelled out.

Therefore, the gas flare correction of the DMSP-OLS will be demonstrated by optimizing the available data of gas flare estimation. These data include:

1. Polygon that define the geographic regions containing gas flares[1] produced by DMSP 18 years' record of gas flare (1994-2012), hereafter refer as DMSP-OLS gas flare observation;
2. Identity of flares clarified with Google Earth and visualized with placemark[2], hereafter refer as identified individual gas flares; and
3. The estimated gas flare by the NPP-VIIRS[3] which record gas flare activity from 2013 onward, to refer as NPP-VIIRS gas flare observation.

.

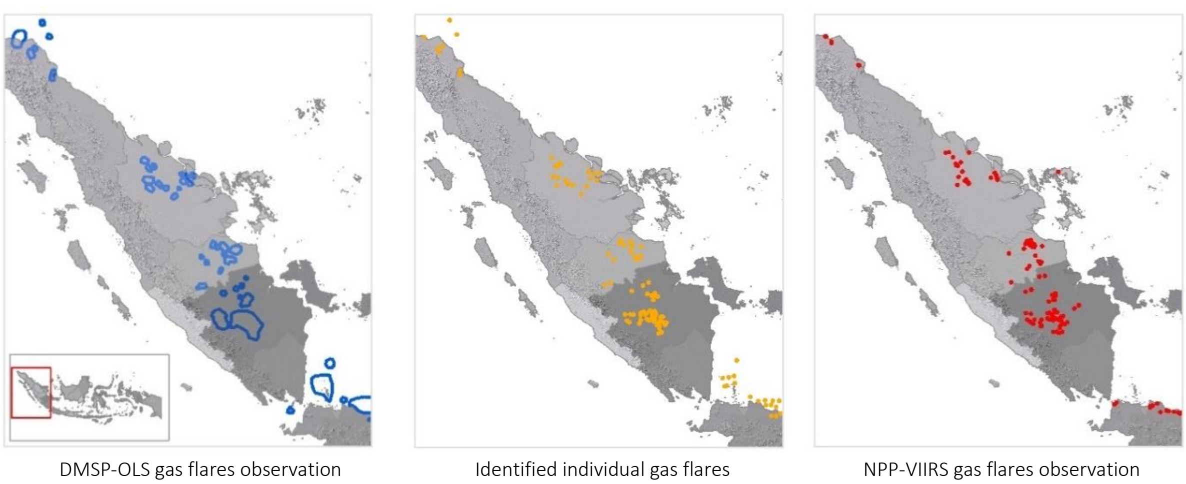


Figure 10 Gas Flares Observation

[1] Retrieved from https://ngdc.noaa.gov/eog/interest/gas_ flares_countries_shapefiles.html) on January 7th 2017
[2] Retrieved from https://ngdc.noaa.gov/eog/interest/gas_flares_countries_kmz.html on January 7th 2017
[3] Retrieved from https://ngdc.noaa.gov/eog/viirs/download_global_flare.html on January 19th 2017

These data will be used to improve the vectors that define the geographic regions containing gas flare estimated by DMSP-OLS gas flare observation. The process of removing the gas flare as shown in Figure 11. In order to confirm the absence or presence and the extend of gas flares, the following will be performed:

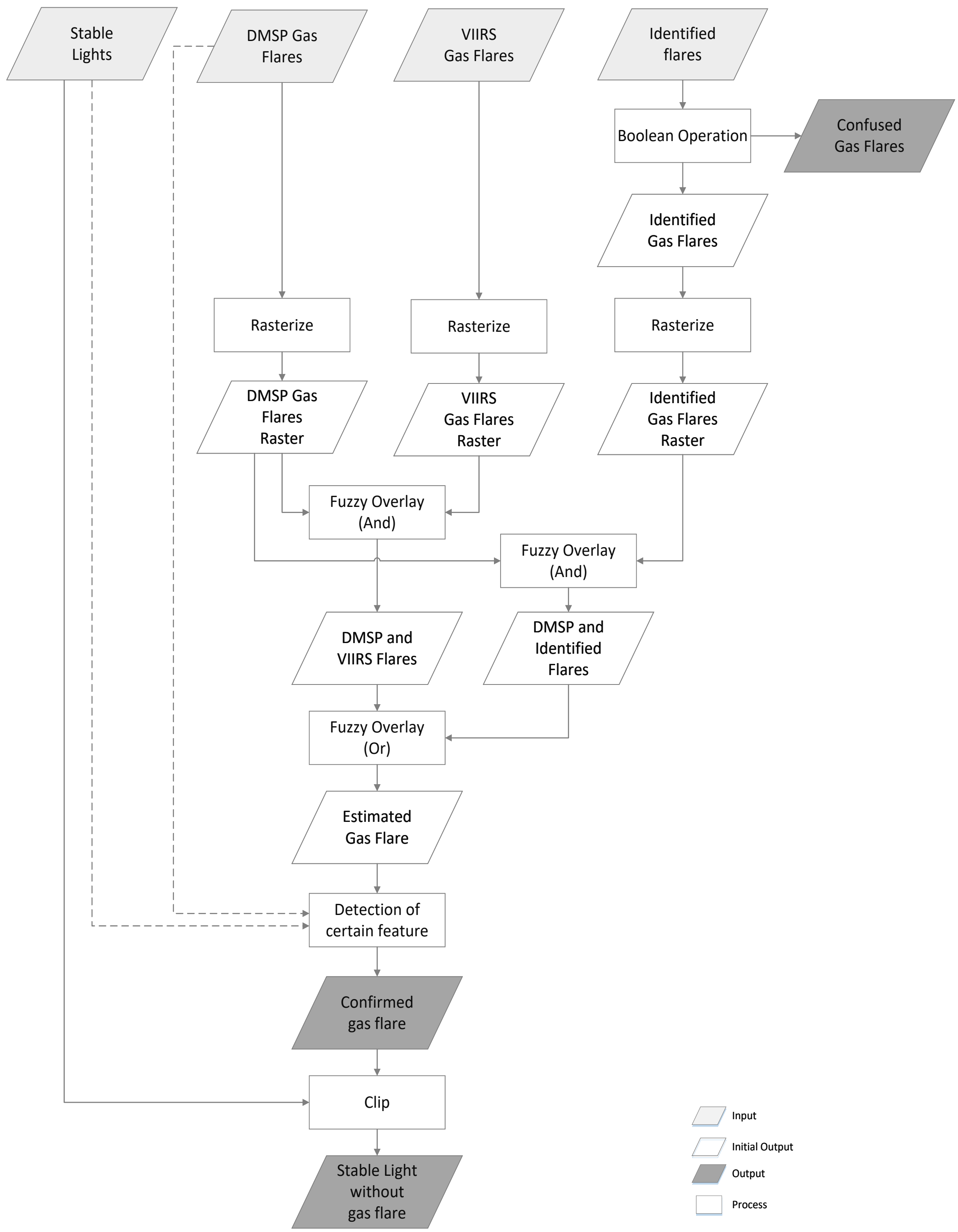


Figure 11 Gas flare removal process

1. To improve the gas flare estimation, NPP-VIIRS gas flare observation should be included because of its higher sensitivity compare to DMSP-OLS. Despite the difference in observation time, most gas flares are permanently fixed to location and are mostly continuously active for a period of years. Moreover, the second measurement could narrow down the gas flare position which is useful for to reduce the errors.
2. A buffer of 1 to 3 pixels was applied around each flares pixel from NPP-VIIRS and DMSP-OLS to accommodate the irregularity shape and variance of gas flare.
3. The polygon feature of DMSP-OLS gas flare observation and the NPP-VIIRS gas flare observation will be converted into raster, and spatial analysis will be performed using 'Fuzzy overlay' feature to combine the DMSP-OLS gas flare observation and NPP-VIIRS gas flares observation using 'AND' as overlay type. This process will produce raster map combination of gas flares overlays.

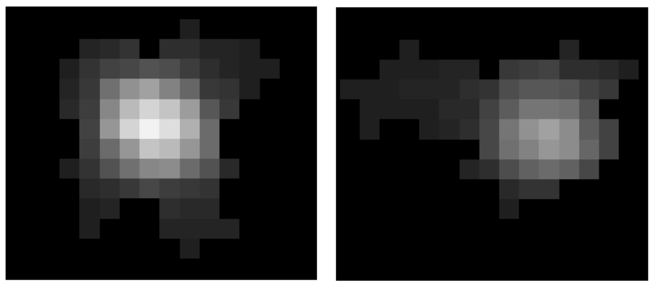


Figure 12 Circular pattern of gas flare in the stable night-light

4. The Placemarks in Google Earth used to confirm the DMSP-OLS gas flare observation, converted to point feature. The red placemarks for features that were either confirmed or consistent with gas flares, while the green placemarks were created for cities, towns, airports, industrial sites, or mines that might be confused as gas flares. These placemarks will also overlay with the polygon feature of DMSP-OLS gas flare observation, but first the confused gas flares (cities, towns, airports, industrial sites, or mines), were removed using Boolean operation, 1 for the gas flare, 0 for the confused gas flare. This should be done to label false detections and to eliminate the inclusion of human settlements features. The individual gas flare then converted to raster, to perform the Fuzzy overlay using the same overlay type as the previous step, and resulting the raster map combination of DMSP-OLS gas flare observation and identified individual gas flares.

Figure 13 Circular and ellipticity of gas flare could be affected by in-situ wind direction

5. Large uncertainties were expected when dealing with gas flare estimation using different sensor and platforms, therefore Fuzzy overlay techniques are used, to analyse the relationship and interaction between the two set of raster maps generated in the previous process, a fuzzy overlay-procedure with 'OR' operator will be performed. The fuzzy overlay describes the interaction of the inaccuracies in the membership of the sets. This type of operator was chosen to include the maximum intersection of the input raster datas. The result of this process is the estimated gas flares.
6. To analyse the relationship and interaction between, which will be used for image object detection. Since it is impossible to identify gas flares inside of brightly lit urban centers, the characteristic of gas flare that show up with circular or ellipticity of light, bright in the centre and dim at the outer edges as shown in Figure 12, were considered. This were optimized by change the visualization of stable night-light image was stretched to show subtle difference between black to white.
7. The estimated map of gas flare will then be used to detect certain feature of gas flare in stable night-light imagery, such as presence of gas flare, the shape and its variance. The presence of the gas flare can be seen in the result both of fuzzy overlay map of gas flare, while for the shape of gas flare, it has to noted that the gas flare can be seen as circular or ellipticity, since the gas flare can be influenced by the direction of wind. The result of this image object detection is the confirmed gas flare, that will be used as a clip for the stable night-light to generate stable night-light image without gas flare.

3.3. Night-Light Intensity and Population Density

In an attempt to achieve the research objectives, the night-light intensity will be used as a proxy for population density by performing normality test, coefficient correlation and regression analysis.

3.3.1. Incorporating Stable night-light and Population Data

As shown in Figure 13, to identify the correlation between night-time light and population density, first the administrative geometry and corresponding census data at finest administrative unit (village level) needs to be established. The census data at village level will then be examined for the possible noise and internal consistency. This data will then be aggregated to coarser level (district, regency and province).

The process is illustrated below for the provincial level. 'Zonal statistic' feature will be performed to summarize statistics on the raster values of a light data within the census data in each level of administration, the statistic to be used is the sum of light value. Aside from that, the polygon of census data in each administrative level were transform from feature to point. This points used to extract the value of light data for the centroid of each administrative data where the point is on and recorded in the attribute table of census data. The sum of light data will be used to calculate the night-light intensity by divide the value with the extend of the area in each administrative level. This data will later be used for cross validation, in order to examine the relationship between night-light intensity and population density.

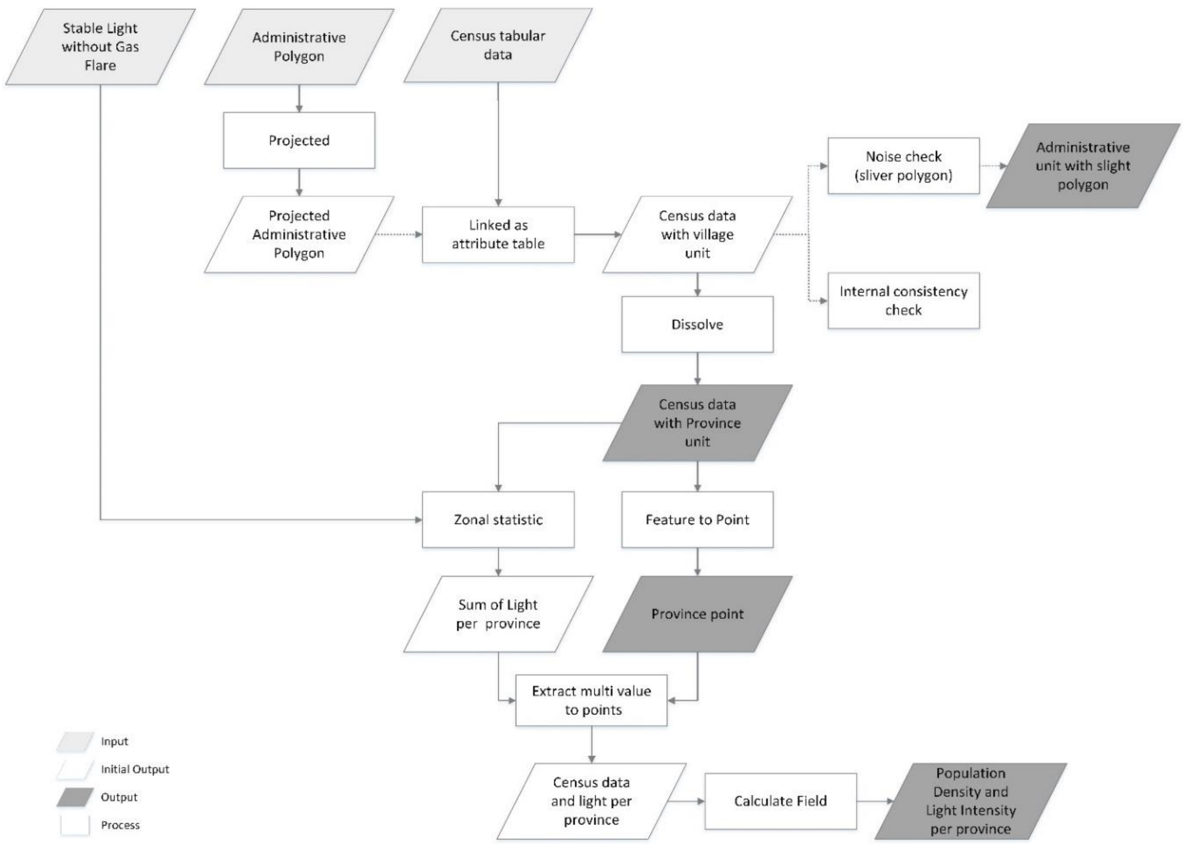


Figure 14 incorporating the population data and the light data

3.3.2. Night-Light Intensity as a Proxy for Population Density

With the intention to predict value from another measured variable, between the night-light intensity and the population density, statistical significance was analysed by using IBM SPSS Statistic. First the frequency distribution of each variable was visualized. If it has significant fraction of the observation on the value of zero and the observations are not normally distributed, the data has to be transformed, however before the transformation, a small constant need to be added to the original scores avoid the undefined zero value, since there is no logarithm of the value 0. The transformation used as follows

$$\text{Light Intensity } y_{(i)} = \ln(0.01 + \textit{Light Intensity}).$$

Outliers might cause non-normality or other problem in the regression model, thus corrective action of outliers, must be considered carefully. To assure that all the observations were considered, the problem of outliers were minimized by using a transformation. This is done using a similar procedure as indicated by Michalopoulos & Papaioannou (2012). The transformation itself is a replacement that changes the shape of a distribution or relationship of the observed data. Once the transformed data meet normality assumption, a regression analysis may be performed.

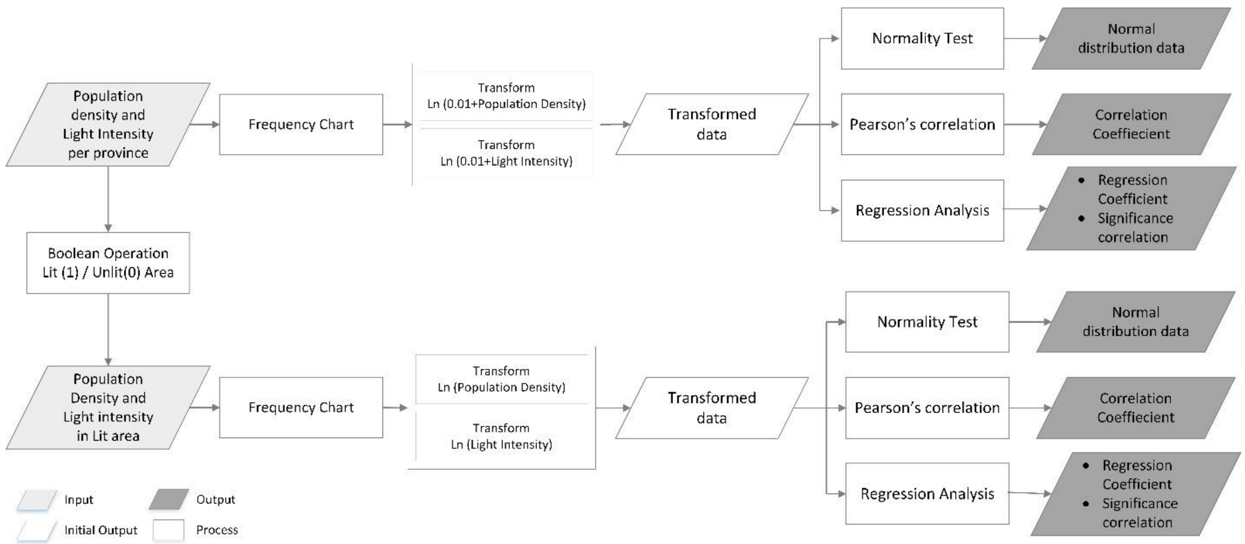


Figure 15 Linear Regression of Night-light Intensity and Population Density

The second observation started by creating dummy variable to distinguished the lit (1) and unlit (0) area, where later only lit area that will be counted in. Due to insufficient detection by the sensor, the night-light intensity might over-represent dense area and under-represent small settlements, thus the unlit area could be excluded in the process. The unlit area defines by the DN value within a pixel, the threshold of unlit area subject to the extent and development of the area. If the data to be processed, the lit area, is not normally distributed it transformed into Light Intensity *y(i) = ln(light intensity)*.

Simple linear regression used to model the relationship between the population density and the night-light intensity as shown in Figure 15. The term 'simple' refers to single predictor variable. The idea is that the probability distribution of a random variable of population density depend on the value of night-light intensity of some predictor variable. This step aim to fine the equation that describes the relationship between two variables. Linear regression estimates the regression coefficients of β_0 and $\beta_1$in the following equation.

$$y = \beta_0 + \beta_1 x$$

Y is the population density, it is a continuous response variable (dependent variable), while *x* is the night-light intensity as an explanatory or predictor variable (independent variable). The value of β_0 is the *Y*-intercept, the mean of *Y* for population with x =0, while the value β_1 is the slope, it is the increase in the mean of *Y* for a unit increase in *x*. A scatter plot of the population density as *y* and night-light intensity as *x* could support to interpreted the relationship.

In an effort to investigate the reliability of the estimated intercept and slope, the correlation coefficient (r) were studied. The correlation coefficient ranges from -1 to 1. This value indicates the linear relationship between two variables. As the value gets closer to $\pm$ 1, the relation between the variables is stronger. In contrary, when the value closer to zero, there is no linear relationship between the variables.

R^2 is the square of correlation coefficient. It explains how well the proportion of population density variation can be explained by the variation of the night-light intensity. A coefficient determination, R^2 indicates in this study to what extent the night-light intensity can able to explain the variance in population distribution. Because of the various aggregation levels used in this study, an adjusted R^2 is preferred correct for the varying number of observations at village, district, regency and provincial level. Also, a scatter plot of the population density as dependent (y-axis) and night-light intensity as independent (x-axis) could confirm the existence of expected relationship.

Bootstrapping as one of the primary method in validating statistical findings, developed to provide standard errors and confidence intervals for regression coefficients and predicted values in situations in which the standard assumptions are not valid. The bootstrap relying on random sampling with replacement selected from the original observation, thus bootstrap take thousands of random samples to yield 95% confidence intervals for statistical differences. Each replacement may be selected more than once, by creating thousands of alternate versions of a data set to describe closer to true observation of the population. Bootstrap also reduce the impact of outliers which ensure the stability and reliability of the models.

Bootstrapping will be performed to ensure that statistic models are reliable and will produce accurate results. In addition, bootstrap can be used to control and check the stability of the results. For linear regression, bootstrap can be performed when sample size is small. Thus, bootstrap may be useful to evaluate a study in estimating the effect size for more generalizable study.

4. RESULT AND DISCUSSION

4.1. Population Distribution

The potential noise of population distribution data was identified and corrected. While the internal consistencies of the population distribution data were examined. The result reported in this following section.

4.1.1. Removing Potential Noise

The potential noise in the census data has to be identified, since the data itself consist of two different type of data, one is tabular data, and the other is geometry of the administrative unit. As the tabular data associated with the administrative geometry using the unique identifier, the area then calculated and sorted. There are many area with small extend, yet the population density was calculated, and the small polygon were detected by sorting the population density from the highest population density. As a result, Table 5 listed the slight extent of area in village level that indicated as 'sliver polygon'. This phenomenon drive the unreasonably high population density.

Table 5 Area with sliver polygon

Village	Census ID	Population (people)	Area (km^2)	Density (people/km^2)
Tanjung Benoa	5103010005	6,767	0.0175	386,026
Kesuben	3328060010	9,218	0.0420	219,303
Sambimulyo	3510020004	8,035	0.0174	460,786
Jambu	1709031002	1,132	0.0029	391,335
Latta	8171040007	1,658	0.0157	105,774
Bonan Dolok I	1206030011	315	0.0001	5,221,658

Figure 16 shows the areas with triangular shaped polygons. However, this area also contains information about population number. Therefore, the geometry shape of the village was examined and merged with the closest village in the same district to avoid losing information about the population number.

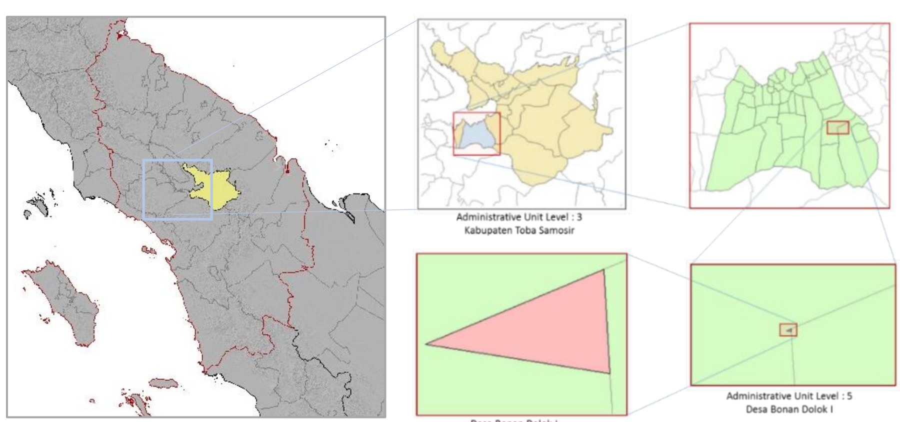
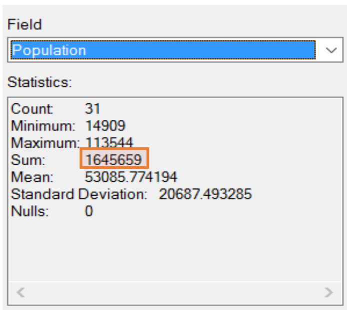


Figure 16 Village administrative unit with sliver polygon

4.1.2. Internal Consistency Check

Internal consistency examined by compare relationship of the data within the same census data. In this study, several areas were observed, using the aggregated census data of village level compared with the tabular data of the census data in the same regency with coarser level (district). One of the result presented in Figure 17, where the census data proves the internal consistency. Any substantive tabulation at the observed administration level have the same aggregation value with the adjusted population at coarser level. This is check is important in relation to improve data input quality and statistic purpose.



Regency : Jakarta Utara

District		Penduduk
010	Penjaringan	306,456
020	Pademangan	149,809
030	Tanjung Priok	375,276
040	Koja	288,091
050	Kelapa Gading	154,692
060	Cilincing	371,335
Sum of Population		**1,645,659**

Figure 17 Population data of village level compared with tabular data of district level for Jakarta Utara

4.2. Removing Light Contamination

This research aims to improve data input quality of population distribution using stable night-light data. As a preliminary attempt, the gas flare contamination had to be detached carefully to avoid eliminating the presence of human settlement in the stable night-light, because the gas flare circular shape has saturated pixel of glowing light that can be misled as bright lit urban area.

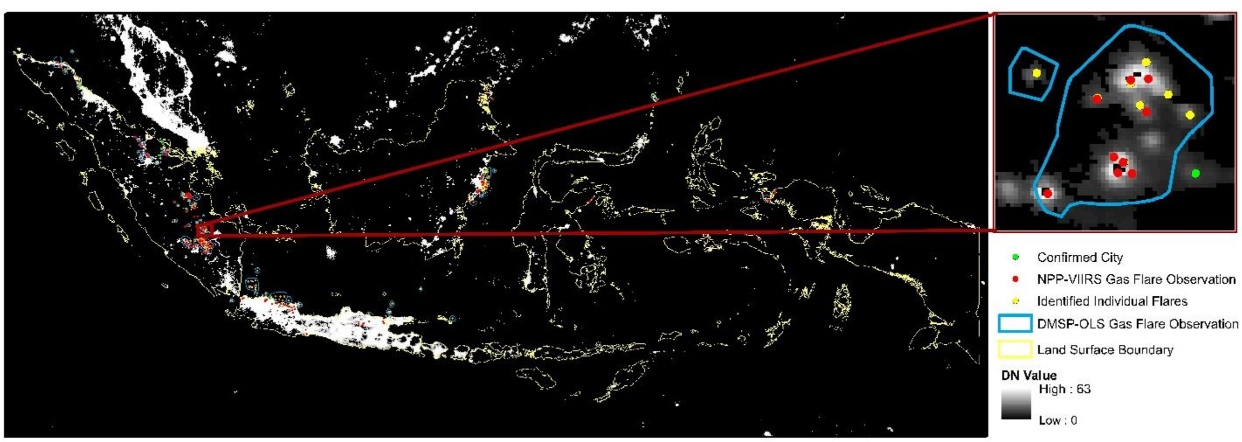


Figure 18 Gas Flare Estimation

To examine the gas flare contamination in the stable night-light image, the NPP-VIIRS gas flare observation and DMSP-OLS gas flare observation with polygon feature were used, together with the identity of individual flares that has been clarified in Google Earth. The method represents alternatives for gas flares removal in stable night-light image and enhancement of previous method that had been demonstrated by Bhandari and

Roychowdhury (2011) and Pestalozzi, (2012) where both study obtaining mask from a map featuring location and extension of DMSP-OLS gas flare observation. In addition, the suggested method by Propastin and Kappas (2012) that implemented different threshold values for pixels' Digital Numbers (DN) to delineate the night-lit urban areas of individual settlements and removing noisy emission sources such as fires and conspicuous gas flares were considered to define the lit and unlit area.

To assess presence and absence of gas flares inside a polygon of the DMSP-OLS gas flare observation, the 'fuzzy overlay' feature was used in ESRI ArcGIS. As summarized in the table below, of a total of 196 flares estimated by NPP-VIIRS located on land surface (onshore), some 102 flares are located inside DMSP-OLS polygon. While for 154 identified individual flares (107 flares onshore and 19 flares offshore), there are 135 flares that located inside the polygon of DMSP-OLS gas flare observation. As highlighted, there are 11 estimated flares that confirmed as cities or densely-populated towns and were therefore discarded from the gas flare dataset.

Table 6 Identified Gas Flares

	Inside of DMSP-OLS polygon	Outside DMSP-OLS Polygon	Land	Offshore
NPP-VIIRS Estimation	102	94	196	-
DMSP-OLS Identified	135	19	107	47

The presence of estimated gas flare inside and outside the polygon of DMSP-OLS gas flare observation, were detected for certain feature, such as shape and brightness. The confirmed feature of gas flares was used calculate the extent area of gas flares and to clean stable night-light image from the gas flare contamination.

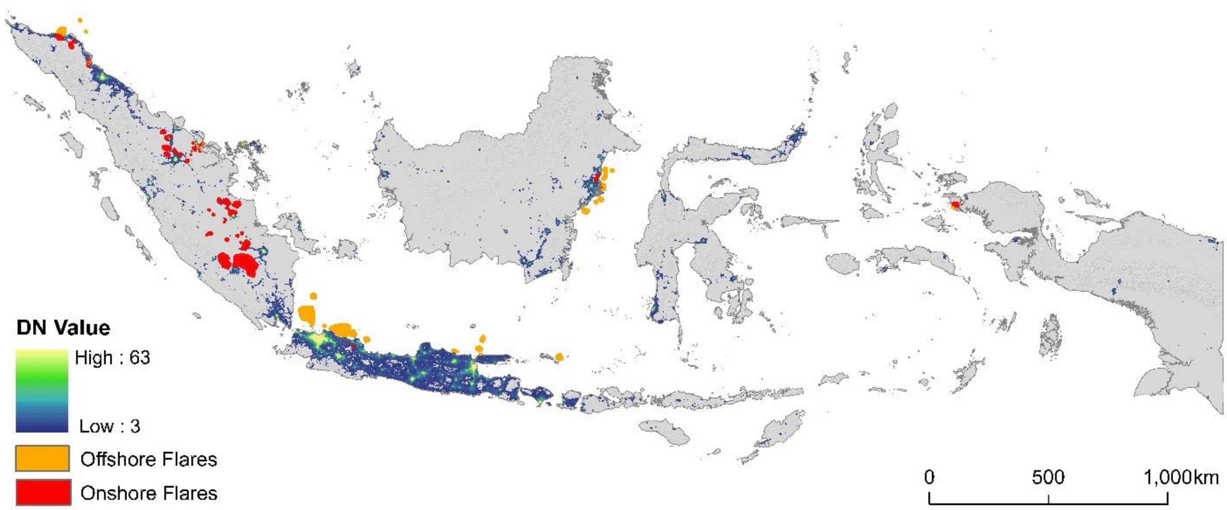


Figure 19 Identified Gas Flares

As might have been expected, the total extent of gas flare is 39641 km² or only 0.76% of total extent Indonesia. Sixty percent (23959 km²) of the gas flares extend in the land area, while approximately 40% or 15682 km² of the gas flares located off shore. This extent is slightly lower than the value we anticipated. Given that the findings are based on limited data of gas flares estimation, the result from such analysis can be improved by considering other satellites based earth observing systems that has a capability to detect gas flares as data sources.

4.3. Night-time Light as a Proxy of Population Density

The ESRI ArcMap dissolve function was used to aggregate the census data based for the different administrative levels. Summary statistic of the aggregation process are listed in the table below.

Table 7 Summary of Administrative unit aggregation

Admin. Level	Number of Admin. Unit	Population		Area (km^2)		Average Population Density (people/km^2)	Total Population
		Min	Max	Min	Max		
Village	77,063	1	160,083	0.00289	4968.64	1252.94	237,641,326
District	6652	111	513,920	0.657653	11,908.96	1099.88	237,641,326
Regency	497	6144	4,771,932	10.83	44200.51	1012.19	237,641,326
Province	33	460,422	43,053,732	661.21	316436.87	683.77	237,641,326

Table 7 shows the consistency of the total population in every aggregation of administration unit, which is 237,641,326 people. The reduced number in administrative unit as well as in average population density, and the increasing value of minimum and maximum population confirmed the soundness of the aggregation process.

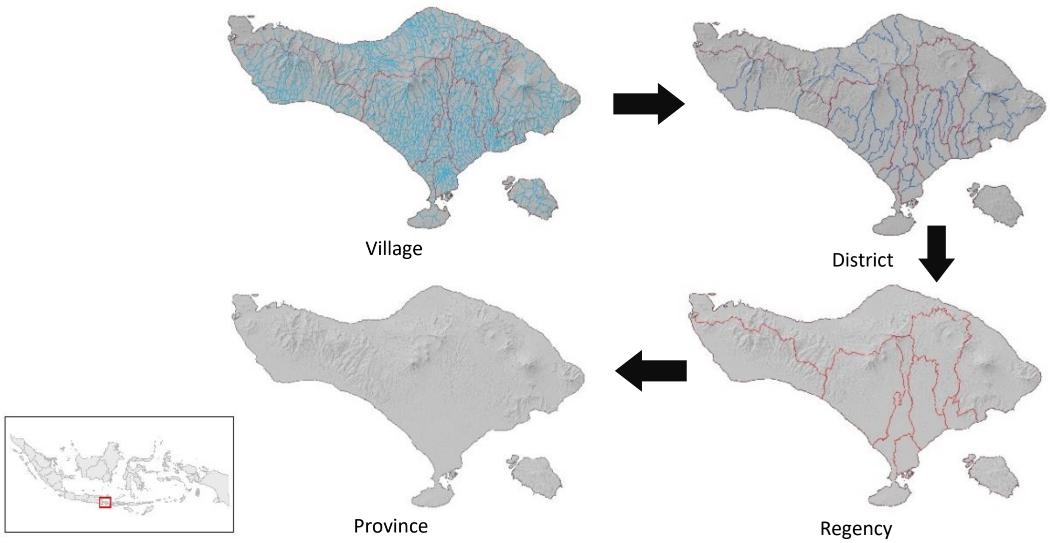


Figure 20 Aggregation of Administrative Unit

Further examination on each administrative unit, as a result of 'zonal statistic' summarized in the table below. The table presented the pixel count per administrative unit, luminosity, the variance of night-light intensity and the frequency missing data in each administrative unit. All the information of population distribution and night-light intensity were put together and summed up in provincial level. However, any aggregation to coarser level of administrative unit tend to reduce the signal noise. These table highlighted that the finer the administrative level, the frequency of missing data increase. The DN value '0' is not appeared in the table, since it replaced the identified background noise. These results provide the considerable insight of the stable night-light data used in this study.

Table 8 Summary of Light data in Administrative Unit

Admin. level	Pixel Count		Luminosity		Light Intensity		Missing Value
	Min	Max	Min	Max	Min	Max	
Village	1	5812	4	4550	0.0073	8515.54	335
District	1	13914	5	16597	0.002043	3805.84	4
Regency	14	51917	5	65306	0.00083	81.26	1
Province	759	368767	2747	558940	0.076	70.73	0

4.3.1. Normal Distribution

The next set of analyses investigate the distribution of the data in the dependent variable. Figure 19 briefly showed the distribution of the night-light intensity in stable night-light image in province level, without and with gas flares removal. The normality tested were examined with *Saphiro-Wilk W test*, since number of the observed data is small, N=33 for the observation without gas removal and N=23 for the observation with gas flare removal. Given the a =0.05, the reported distribution of night-light intensity is not normally distributed with p-value < 0.05. Thus, the distribution of the data was needed to reduce the impact of unequal variances. Using log normal transformation, the right skew distribution reduced and the problem of unequal variances was overcome, resulting in p-value > 0.05, meaning that the data has normal distribution.

Table 9 Saphiro-Wilk Normality Test

Variable	*N*	Mean	*W*	*p*-value
Without gas flare removal				
Light Intensity	*33*	5.321	*0.416*	*< 0.0001*
Ln Light Intensity	*33*	0.274	*0.960*	*0.256 (two tails)*
With gas flare removal				
Light Intensity	*23*	3.119	*0.658*	*< 0.0001*
Ln Light Intensity	*23*	-0.022	*0.947*	*0.192 (two tails)*

α= 0.05

In this provincial level of aggregation, the number of observation and variance of the data is smaller and the signal noise in finer level of administration tend to be reduced. This set of analysis also performed for the other administrative level as presented in Appendix 3-5. In finer level of administration (village, district and regency level), the distributions of night-light intensity were reported have right skewed distribution. This result derived by the significant fraction of zero in the data, while in province level, this value already summed up and have no zero value for night-light intensity. Thus, transformation of the data was needed to reduce the impact of these values and unequal variances. The log normal transformation, could reduce the right skew distribution. However, the significant fraction in the data that were derived by zero value, made the log normal transformation is not possible. Therefore, to avoid numerical problems with a log of zero, a small value of 0.01 were add before the transformation. This were done to make the smallest number in the data positive, and log normal transformation possible to performed. Also this allow us to include all the observations, so that true outliers' can be detected as suggested by Michalopoulos and Papaioannou (2012).

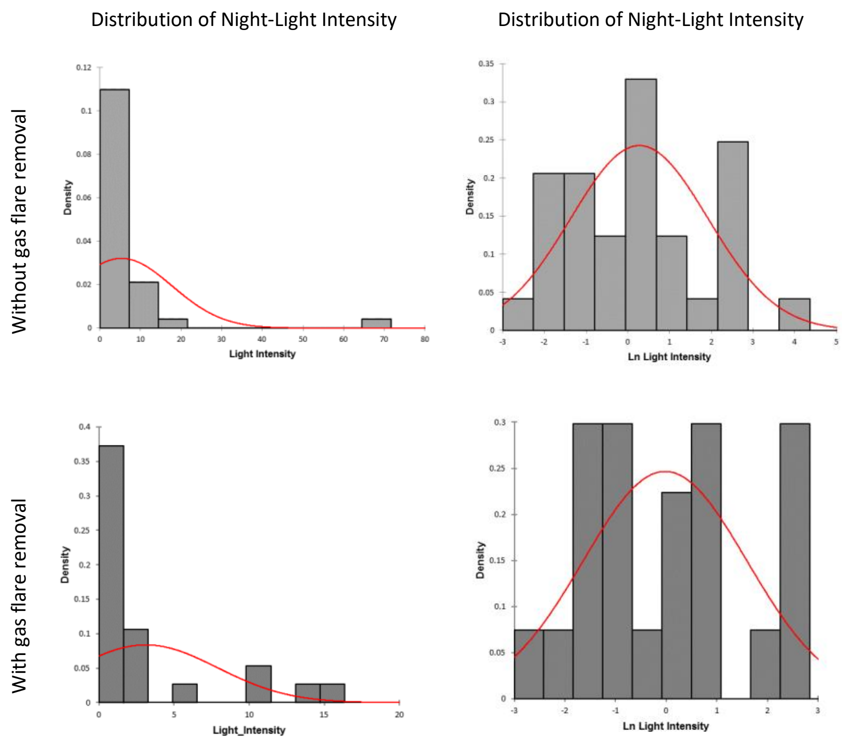


Figure 21 Distribution of Night-Light Intensity in Province level

Another way to deal with the outliers and reduce the signal noise in the finer administrative level is by focusing only on the observation that affect the investigating factors, in this case, only the observation of lit area included. Approximately 1% of the highest luminosity were removed to eliminate the effect of over blooming of urban light. Whereas the 0 DN value, also removed, since it represents the area with cloud cover and the unlit area with the total luminosity less than 20 were detached, considering the drawback of the DMSP-OLS sensor where the night-light intensity might under-represent small settlements that are infrequently lit due to insufficient detection. By focusing on lit area, the unequal variance caused by zero value, that include as unlit area also can be avoid.

4.3.2. Correlation Coefficient

The statistical model aims to identify the relationship between variables that determine the population density and night-light intensity. The correlations were test using Pearson's correlation. This correlation coefficient simply refers to two random variables, instead of refer to dependent variable and independent variable.

Given the observation in provincial level without gas flare removal (N_1=33) and with gas flare removal (N_2=23), it was reported the correlation coefficient r_1 = 0.927 and r_2 =0.950 respectively with the significance value is less than 0.01 (two-tailed), meaning that there is a correlation between population density and light intensity ($r \neq 0$, $p < 0.01$). The correlation coefficient reported close to positive 1, that indicated that there is positive relationship.

Table 10 Correlation Coefficient of Night-Light Intensity and Population Density

		Without gas flare removal		With gas flare removal	
		Ln Population Density	Ln Light Intensity	Ln Population Density	Ln Light Intensity
Ln Population Density	Pearson Correlation	1	.927**	1	.950**
	Sig. (2-tailed)		.000		.000
	N	33	33	23	23
Ln Light Intensity	Pearson Correlation	.927**	1	.950**	1
	Sig. (2-tailed)		.000		.000
	N	33	33	23	23

** Correlation is significant at the 0.01 level (2-tailed)

Table of critical values Pearson's correlation (given in the appendix 6) could be useful to examine the significance of correlation coefficient. Where N_1 = 33, the degree of freedom (*df*), *n-2,* the critical value associated with *df* =31 is $\pm$ 0.462. Since *r* is more positive than the critical value, *r* is significant (0.927 > 0.462), while when N_2 = 23 and the degree of freedom (*df*), *n-2*, the critical value associated with *df* = 21 is $\pm$ 0.549. Since r is more positive than the critical value, the *r* is significant (0.950 > 0.549). This initial result confirms that there is positive correlation between population density and night-light intensity.

To examine the difference of correlation coefficient between night-light intensity and population density with and without gas flare removal, Fisher's *z*-transformation were performed.

$$z = \frac{r_1 - r_2}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}}$$

The significance of the difference between two correlation coefficients, r_1 and r_2 were tested and resulting in $z = -0.68$, $p = 0.0496$ the negative sign shows that the correlation coefficient with gas flare removal (r_2) is greater than the correlation coefficient without gas flare removal (r_1). The correlation coefficient of population density and night-light intensity can also be detected in further regression analysis.

4.3.3. Regression Analysis

The Regression Analysis consists of more than just fitting a linear line through a cloud of data points. It consists analyzing the correlation and directionality of the data, estimating the model, i.e., fitting the line, and evaluating the validity and usefulness of the model.

In simple linear regression, population density is the dependent variable (*Y*) and night-light intensity is the explanatory or predictor variable (*x*). From the previous finding, we have known that the night-light intensity and population density shown positive correlation. As a result of linear regression, presented in the Table 11, for the observation between population density and night-light intensity, without gas flare removal *R* has value 0.927 and with gas flare removal *R*=0.950. Since night-light intensity is the only predictor, this value also confirms the result of previous correlation coefficient.

Table 11 Regression Analysis of Night-Light Intensity and Population Density

Observation	R	R square	Adjusted R square	Std. Error of the Estimate
Without gas flare removal	0.927	0.859	0.854	0.60758
With gas flare removal	0.950	0.903	0.898	0.41360

Coefficient of determination or square function of *R*, show how close the data are fitted the regression line R^2 value is 0.859 for the observation without gas flare removal and 0.903 for observation with gas flare removal. The R^2 indicate how the night-light intensity explains the variance of population density. The regression for the night-light intensity without gas flare removal accounts for 85% of the of population density variance, while the night-light intensity with gas flare removal accounts for 90% of the population density variance. In other word, more variance of population density is accounted by the night-light intensity data with gas flare.

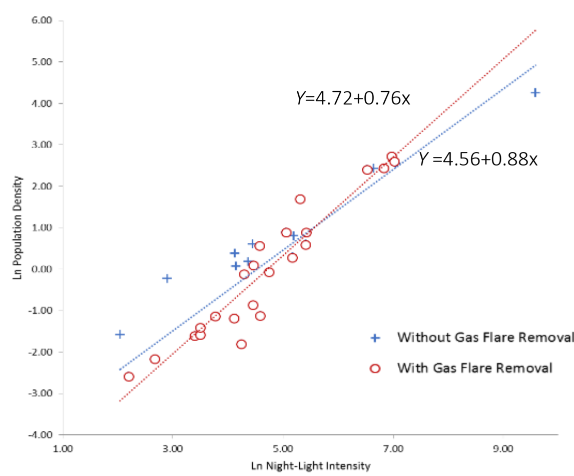


Figure 22 Scatter diagram showing the relationship between night-light intensity and population density with and without gas flare removal

While correlation measured the linear association between the variables of population density and night-light intensity in symmetrically way, the regression set to predict population density from night-light intensity not in symmetrically way. The scatter diagram of the population density as *y*-axis and night-light intensity as *x*-axis shown in Figure 21, where the coefficient estimate the trends and R^2 represent the scatter around the regression line. It illustrates a clear trend of a significant positive correlation of night-light intensity and population density in province level, where these values correlate favourably. The solid line stands for the trend line of regression equation.

Table 12 Regression Coefficient of Night-Light Intensity and Population Density

Observation	Intercept	Slope	SE_{slope}	SSE	SD_X	*n*
Without gas flare removal	4.565	0.883	0.064	81.073	1.671	33
With gas flare removal	4.720	0.761	0.055	36.959	1.618	23

The linear regression estimates the regression coefficient of β_0 and β_1. The value of β_0 (y-intercept) is the mean of population density (y) with night-light intensity (x)= 0, whereas the value of β_1 (the slope) is the increase in the mean of population density (y)for a unit increase in night-light intensity (x). Given 33 observation in provincial level without gas flares removal the equation become y = 4.57+0.88x and with 23 observation with gas flare removal, the equation y =4.72+0.76x.

The variance estimates and the Student's t distribution with $df = N\text{-}2$ were used to examine the difference in the relationship between night-light intensity and population density with and without gas flare removal.

$$t = \frac{b_1 - b_2}{s_{b_1 - b_2}}$$

The t-statistic = 0.676, this is significant on 54 *df (p = 0.05),* so the slope of observation without gas flare removal is significantly higher than observation with gas flare removal. However, as illustrated in scatter diagram of Figure 22, both the night-light intensity with and without gas flare removal can predict correspondingly well, yet the Y variable increases more swiftly using the light intensity (x-variable) with gas flare removal.

Table 13 ANOVA

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	69.629	1	69.629	188.614	.000[b]
	Residual	11.444	31	.369		
	Total	81.073	32			

a. Dependent Variable: Ln Population Density

b. Predictors: (Constant), Ln Light Intensity

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	33.366	1	33.366	195.048	.000[b]
	Residual	3.592	21	.171		
	Total	36.959	22			

a. Dependent Variable: Ln Population Density

b. Predictors: (Constant), Ln Light Density

The Anova table gives the analysis of variance table that shows how the regression equation accounts for variability in the response variable. Given the $p < 0.05$, both regression of night-light intensity with and without gas flare removal statistically significant in predicting the population density.

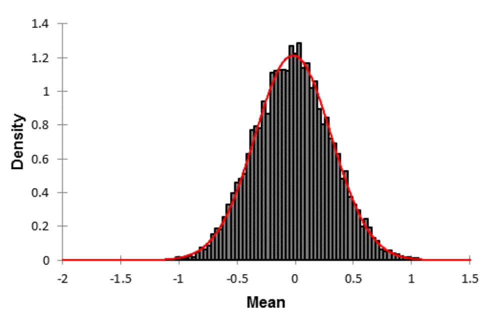


Figure 23 Bootstrap Distribution of Mean of Ln Light Intensity

Bootstrapping were performed to validate the statistical findings. It narrowed the confidence intervals by taking all possible random combinations of samples and then estimating a mean, Bootstrap distribution for the mean of night-light intensity as shown in Figure 23

Table 14 Bootstrap Confidence Interval

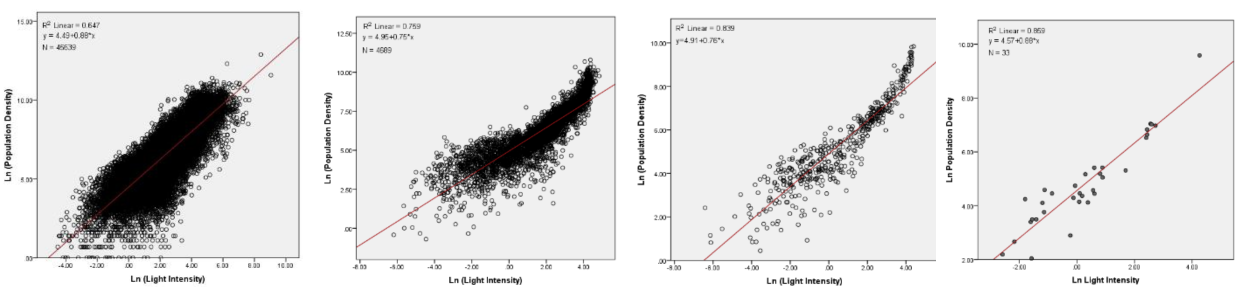
	B	Std. Error	Sig.	Lower Bound	Upper Bound
Coefficient[a]					
(Constant)	4.720	0.86	.000	4.541	4.899
Ln Light Intensity	0.761	0.55	.000	0.648	0.875
Bootstrap for Coefficients[a]					
(Constant)	4.720	0.83	.000	4.560	4.883
Ln Light Intensity	0.761	0.57	.000	0.638	0.845

a. Dependent variable : Ln Population Density

As summarized in the table, the result of bootstrap confidence intervals is relatively narrow; meaning that it can be assumed the statistic findings are valid.

This significance correlation also performed for the other level of administrative unit. Figure 24 summarized the significant correlation between light intensity and population density in lit area from the finer to coarser level of administrative unit.

Village | District | Regency | Province



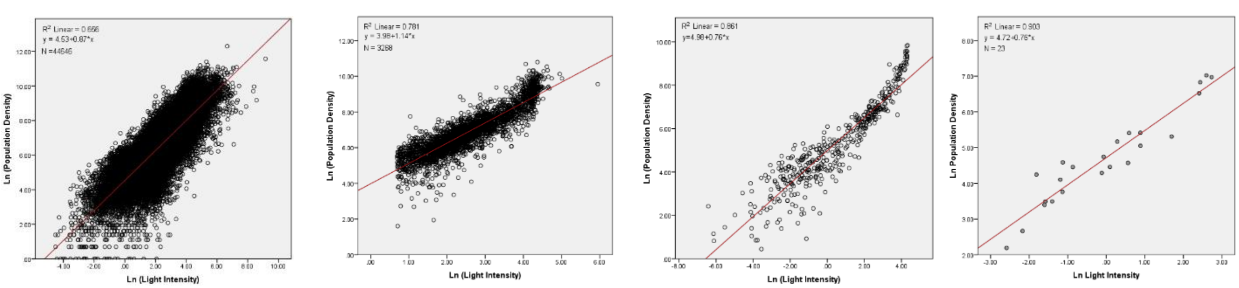
Without Gas Flare Removal

N = 45,639
y = 4.49+0.88x
R^2= 0.647

N= 4689
y = 4.95+0.75x
R^2=0.759

N = 475
y = 4.91+0.76x
R^2=0.839

N = 33
y = 4.57+0.88x
R^2=0.859



With Gas Flare Removal

N = 44,646
y = 4.53+0.87x
R^2 = 0.656

N = 3268
y = 3.98+1.41x
R^2 = 0.781

N = 439
y = 4.98+0.76x
R^2 = 0.861

N = 23
y = 4.72+0.76x
R^2 = 0.903

Figure 24 Significant correlation between night-light intensity and population density in lit area

4.4. Discussion

This subchapter discusses the key findings of the research and the usefulness of gas flare corrected DMSP-OLS night-light data as a proxy of population density.

4.4.1. Gas flare correction of DMSP-OLS night-time light improve the variance of population density

Brightness and spatial extent of night-light are often correlated to population density. However, the gas flare circular shape has saturated pixel of glowing light that can be misled as bright lit urban area. Thus, as a preventive measure to reduce error and improve data quality, an improvement methods were performed by optimizing the available data of gas flare observation to improve the vectors that define the geographic regions containing gas flare.

For the reason that most gas flares are permanently fixed to location and are mostly continuously active for a period of years, The NPP-VIIRS gas flare observation, which were started to record the gas flare combustion in 2012, were included in the analysis. Also, the second measurement of independent observer could narrow down the gas flare position which is useful for to reduce the errors.

Despite setting threshold of lit and unlit area for eliminating background noise and saturation in the night-light image, gas flare removal demonstrated in this study intents to improve the variance of population density from recent census data at various administrative unit (province, regency, district and villages) with light observation.

The statistical model was used to examined the effect of gas flare correction in enhancing the relationship between DMSP-OLS night-light intensity and population density. The relationship between population density and night-light intensity were examined and give positive r value closer to 1, which indicated that there is positive correlation between night-light intensity and population density. This concurs well with the hypothesis that there is a correlation between population density and night-light intensity ($r \neq 0$) with significance level < 0.01. Thus, the night-light intensity can be used as a proxy to predict population density.

Taken as a whole, despite of the slight rise of R2 value, the results show that the gas flares removal has an advantage in enhancing night-light data. As expected, the tests prove that the significant correlation between night-light intensity and population density improved in contrast to the finer level of administrative unit. This fits also indicated in initial summary of light data of administrative unit, where the summary highlighted that any aggregation to coarser level of administrative unit tend to reduce the signal noise. Even though by removing the gas flares, these results only giving slightly increase, it could nevertheless be argued that removing the potential error from the data input, could improve the estimation and reduce additional error component in the model.

In general, apart from this slight improvement of adjusted R2, the result is confirmation of small extent of gas flares presence in the study area that has been reported in previous section of light contamination removal. The results show that the removing gas flare contamination from the stable night-light, improve the correlation between predictor (night-light intensity) and dependent variable (population density). However, since the extend of gas flare is less than 1% of the total area, the correction effects are limited. Furthermore, this result has further strengthened our confidence in hypothesis that adequate removal of gas flare from the night-time light data will give a positive impact for night-light intensity in explaining the variance of population density. In our view, these results represent an initial step toward optimizing the application of night-light data in predicting the population distribution.

Given the slight increase of significance correlation by removing the gas flares contamination in Indonesia, does not mean that this method cannot be implied for another part of the world, particularly for a country that generate gas flares from oil and gas production. However, the development of economic activity in Indonesia and increasing demand of electricity, are likely to continue and influence the increase of gas flare contamination, therefore the method used in this research can be reapplied.

4.4.2. Incorporating Population Distribution and Marine Debris Case

In an archipelagic country where coasts often comprise most of the territory, like Indonesia, populations in coastal areas are growing faster than those in non-coastal areas (Curran, Kumar, Lutz, & Williams, 2002). This is a concern because high population density places on the coasts have meant that higher density is associated with increased risks to coastal and marine ecosystems (Creel, 2003). Coastal cities are surrounded by important but fragile ecosystems that are under pressure from population growth, tourism and large commercial enterprises. These factors contribute to a complex solid waste management situation, which is exacerbated by lack of planning and sanitation infrastructure, common factors in cities in developing countries (de Oliveira & Turra, 2015).

Marine debris might originate from a wide and diverse range of sources. However, in developing countries, most of the plastics waste that enter the marine environment originated from coastal and upriver settlements (Allsopp et al., 2006). High quantities of marine debris may be found on the shoreline close to urban areas (Allsopp, Walters, Santillo, & Johnston, 2006; Uneputty & Evans, 1997). The proximity to urban, industrial and recreational areas also influence the types and amount of litter that are found in the open ocean or along beaches (Galgani, Hanke, & Maes, 2015).

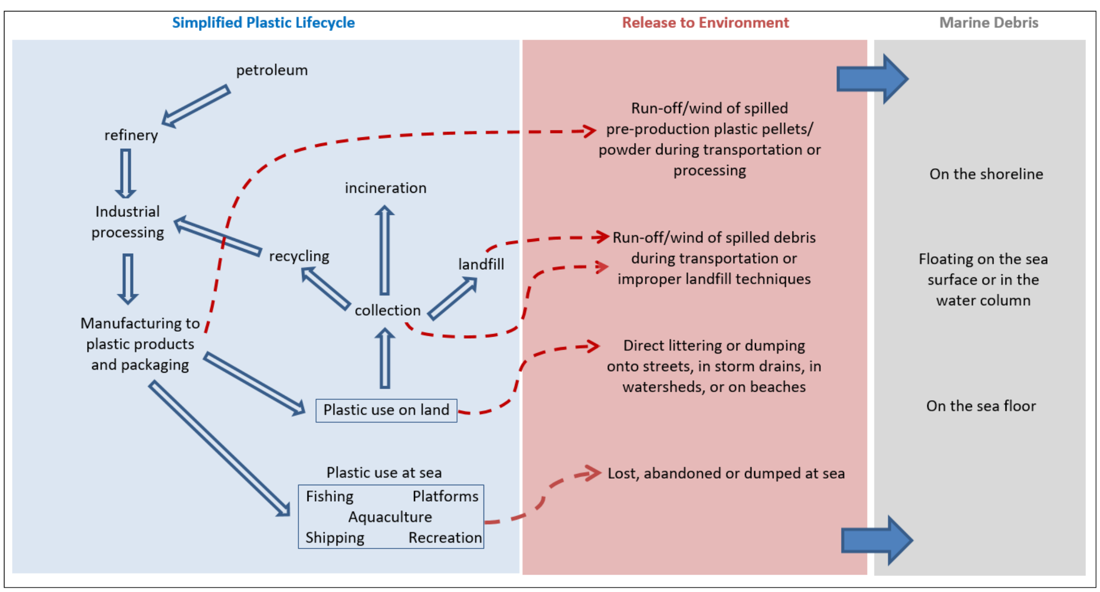


Figure 25 Simplified diagram of lifecycle of plastic and ways it becomes marine debris (Stevenson, 2011)

There are a number of transport pathways by which debris from land-based sources is transported to the shoreline and enters the marine environment, including rivers, drainage or sewerage systems, wind, and direct littering (Barnes et al., 2009; Stevenson, 2011). Moore (2015) has indicated that a significant amount of debris, are transported via watersheds to the shoreline. In estuaries, large rivers are responsible for substantial input of debris to the seabed (Rech et al., 2014). Because of large-scale residual ocean circulation patterns and extensive riverine input (Wei, Rowe, Nunnally, & Wicksten, 2012), marine debris densities are higher in coastal seas (Lee, Cho, & Jeong, 2006). However, in Indonesia, many rivers throughout the country flow into the sea, carrying waste and contaminant from the mainland. Thus, Many problems with water quality and ecosystems are best solved at the watershed level, which encompasses the full area that drains into a particular body of water (Creel, 2003).

Most factors that being considered to estimate the waste generated is the population distribution (Jambeck et al., 2015). In estimating the marine debris generated, commonly accepted life-cycle assessment principle is used, by applying a range of conversion rates from mismanaged waste to marine debris. The marine debris sources were estimated from the mass of waste generated per capita annually; the percentage of plastic waste; and the percentage of plastic waste that is mismanaged and therefore, has the potential to enter the ocean as marine debris and include the ancillary data related to watershed and coastline.

Thus the information of population distribution is crucial in to better understand the origin of marine debris and its distribution in order to assess, determine the solution and evaluate precisely the effectiveness of measures implemented to reduce marine debris pollution (Galgani et al., 2015; Stevenson, 2011). Before an accurate estimate of debris quantities can be made, basic information is needed on sources and inputs (Galgani et al., 2015). In order to carry out a realistic environmental assessment at the regional scale, decision maker have to identify critical areas and to define priorities for policy. Yet, gaining accurate information on how much waste generated is a critical step in targeting management an assessment (Ryan, Moore, Van Franeker, & Moloney, 2009). Population estimation for coastal watersheds could provide useful information for coastal managers, policy and decision maker (Creel, 2003). Thus, input data have to be estimated by adding as much value as possible to existing information, avoiding additional measurements, and integrating what is already available (Rahman, Shi, & Chongfa, 2014).

Although the night-light intensity can be used as a proxy to predict population density in each level of administrative unit, where provincial level give highest significance correlation, the mismanage waste that turn into marine debris cannot be estimated properly.

Aside from population distribution, the presence of neighbouring river that flow to the sea and its proximate location to the dense area are the focal information to get an appropriate and representative estimate of the total amount of mismanage waste that turn into to marine debris. It is important to consider how many people live closed to the river. Since it has indicated that a significant amount of debris, are transported via watersheds to the shoreline. However, in Indonesia, many rivers throughout the country flow into the sea, carrying waste and contaminant from the mainland. Thus, encompasses the good prediction of how many people that live close to the river can be useful for the marine debris estimation. At this point, the information of population density in pixel level become relevant to estimate the waste generated into marine debris.

While investigating the night-light intensity as a proxy of population density in different aggregated administrative level is valuable in an attempt to recognize and reducing the noise in both night-light intensity and population distribution data. Moreover, comparing different aggregated information might help in sidestepping overly optimistic in predicting the coarser level or underrate the outcome of finer level of administrative unit.

Following the same procedure, the night-light intensity as a proxy of population density demonstrated in pixel level, using the night-light intensity with gas flares removal. The sample taken from village unit that has 1 pixel count. In addition, geometry boundary of the village should fall within the pixel size, thus the sample narrowed to the village with closer to 1 km2 extent.

Table 15 Correlation Coefficient of Night-Light Intensity and Population Density in Pixel Level

Correlations

		Ln Population Density	Ln Light Intensity
Ln Population Density	Pearson Correlation	1	.890**
	Sig. (2-tailed)		.000
	N	120	120
Ln Light Intensity	Pearson Correlation	.890**	1
	Sig. (2-tailed)	.000	
	N	120	120

**. Correlation is significant at the 0.01 level (2-tailed).

With 120 observation of village with 1 pixel size (N=120), it was reported the correlation coefficient r= 0.890 with the significance value is less than 0.01 (two-tailed), meaning that there is positive correlation between population density and night-light intensity in pixel level.

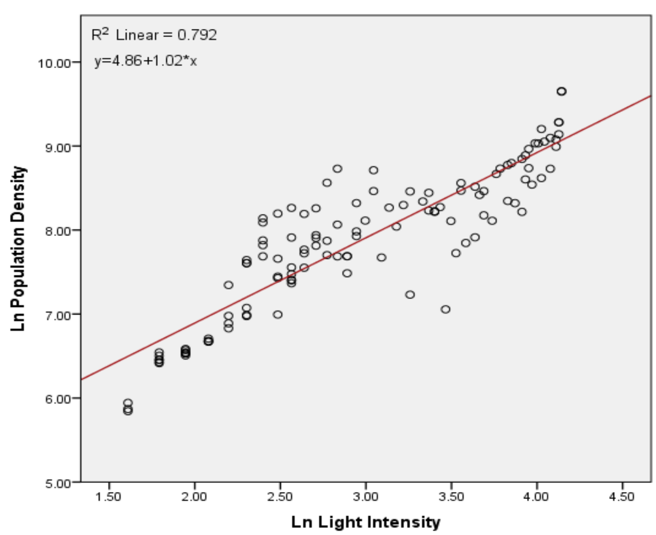


Figure 26 Significant Correlation between Night-Light Intensity and Population Density in Pixel Level

The scatter plot of the population density and night-light intensity shown in Figure 22, where the estimated regression coefficient y = 4.86+1.02x and the coefficient of determination (R^2) = 0.792, it indicates that the night-light intensity explains 79% of the population density variance.

Table 16 Regression Coefficient and Bootstrap for Coefficient of Night-Light Intensity and Population Density in Pixel Level

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	4.863	.146		33.253	.000	4.574	5.153
	Ln Light Intensity	1.015	.048	.890	21.219	.000	.920	1.110

a. Dependent Variable: Ln Population Density

Bootstrap for Coefficients

Model		B	Bootstrap[a]				
			Bias	Std. Error	Sig. (2-tailed)	BCa 95% Confidence Interval	
						Lower	Upper
1	(Constant)	4.863	.007	.142	.001	4.622	5.164
	Ln Light Intensity	1.015	-.003	.045	.001	.918	1.091

a. Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples

Bootstrapping were performed to validate the statistical findings. The bootstrap confidence intervals are relatively narrow; therefore, it can be assumed that the findings are valid.

5. CONCLUSION AND RECOMMENDATION

5.1. Conclusion

This thesis has investigated an account of gas flare as light contamination in DMSP-OLS stable night-light data by optimizing available data of DMSP-OLS gas flare observation and NPP-VIIRS gas flare observation. Our study devised a methodology which combining the technique of data preparation, by setting threshold to delineate lit and unlit area, as well as eliminating light contaminated area with more than one measurement, that will be useful in enhancing the night-light data and narrowed the concerning area. Thus, the certain areas in proximity of gas flare with settlement and urban extent were not cancelled out.

The significance of the difference between two correlation coefficients of night-light intensity and population density, without gas flare removal ($r_{1)}$ and with gas flare removal ($r_{2)}$ were tested and resulting in z = -0.68, p = 0.0496 the negative sign shows that the correlation coefficient with gas flare removal (r_2) is greater than the correlation coefficient without gas flare removal (r_1). This result has further strengthened our hypothesis that gas flare contamination removal in a night-light dataset will give positive impact in correlation coefficient of night-light intensity and population density ($r1 < r2$, $z < 0$).

This work has revealed that an adequate procedure in correcting light contamination would significantly improve the coefficient of determination of light intensity in explaining variance of population density at various administrative unit (province, regency, district and villages). This study also reported the impact of gas flare removal in improving application of night-light data to assess population distribution, particularly in Indonesia.

The regression for the night-light intensity without gas flare removal accounts for 85% of the of population density variance, while the night-light intensity with gas flare removal accounts for 90% of the population density variance. In other word, more variance of population density is a counted by the night-light intensity data with gas flare. This evidence support the hypothesis that gas flare contamination removal in a night-light intensity dataset will give a positive impact in explaining variance of population density ($\bar{R}^2$ with gas flare removal > $\bar{R}^2$ without gas flare removal). Bootstrapping were also performed to validate the statistical findings. The result of bootstrap confidence intervals is relatively narrow; meaning that we can assume the statistic findings are valid.

5.2. Limitations and Improvements for Future Assessment

It is plausible that a number of limitations could have influenced the results obtained. The restricted night-intensity data that account only for year 2010, to reduce uncertainties, this may be improved by inter-calibration with different year of observation.

The degradation of DMSP-OLS from 2013 onward, made it impossible to produce global data for estimating gas flare volumes, however the estimated gas flare recorded up to 2012 can be considered as sufficient measurement, since most gas flares are permanently fixed to location and stable over time. Furthermore, The prospect of being able to do correction for light contamination using different sensor such as NASA- MODIS and NPP-VIIRS serves as a continuous spur for future research. Moreover, the application of NPP-VIIRS for night-light observation and gas flare estimation still open to be explored, as NPP-VIIRS has improvement in spatial resolution, has inflight calibration, no saturation and 14-bit quantization compare to DMSP-OLS.

The upshot of this is the possibility to re-applied the method for another part of the world, particularly for a country that generate gas flares from oil and gas production. Moreover, the development of economic activity in Indonesia and increasing demand of electricity, are likely to continue and influence the increase of gas flare contamination, therefore the method used in this research can be reapplied

REFERENCE

Allsopp, M., Walters, A., Santillo, D., & Johnston, P. (2006). *Plastic Debris in the World's Oceans*. Amsterdam, Netherlands. Retrieved from http://www.unep.org/regionalseas/marinelitter/publications/docs/plastic_ocean_report.pdf

Anderson, S. J., Tuttle, B. T., Powell, R. L., & Sutton, P. C. (2010). Characterizing relationships between population density and nighttime imagery for Denver, Colorado: issues of scale and representation. *International Journal of Remote Sensing*, *31*(21), 5733–5746. http://doi.org/10.1080/01431161.2010.496798

Anejionu, O. C. D., Blackburn, G. A., & Whyatt, J. D. (2014). Satellite survey of gas flares: development and application of a Landsat-based technique in the Niger Delta. *International Journal of Remote Sensing*, *35*(5), 1900–1925. http://doi.org/10.1080/01431161.2013.879351

Anejionu, O. C. D., Blackburn, G. A., & Whyatt, J. D. (2015). Detecting gas flares and estimating flaring volumes at individual flow stations using MODIS data. *Remote Sensing of Environment*, *158*, 81–94. http://doi.org/10.1016/j.rse.2014.11.018

Balk, D. L., Deichmann, U., Yetman, G., Pozzi, F., Hay, S. I., & Nelson, A. (2006). Determining global population distribution: methods, applications and data. *Advances in Parasitology*, *62*, 119–56. http://doi.org/10.1016/S0065-308X(05)62004-0

Barnes, D. K. A., Galgani, F., Thompson, R. C., Barlaz, M., Acha, E., Hermes, W., … Bank, W. (2009). Accumulation and fragmentation of plastic debris in global environments. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *364*(1526), 1985–98. http://doi.org/10.1098/rstb.2008.0205

Bergs, R. (2007). Exploring the Spatial Economy by Night. *Economic Policy*, (2116), 36. http://doi.org/10.1227/01.NEU.0000349921.14519.2A

Bhandari, L., & Roychowdhury, K. (2011). Night Lights and Economic Activity in India: A study using DMSP-OLS night time images. *Proceedings of the Asia-Pacific Advanced Network*, *32*, 218–236. http://doi.org/10.7125/APAN.32.24

Bielecka, E. (2015). A Dasymetric Population Density Map of Poland. *Institute of Geodesy and Cartography*. Retrieved from http://cartesianos.com/geodoc/icc2005/pdf/oral/TEMA5/Session 9/ELZBIETA BIELECKA.pdf

Creel, L. (2003). *Ripple Effect: Population and Coastal Regions*. Washington DC. Retrieved from http://www.prb.org/pdf/RippleEffects_Eng.pdf

Curran, S., Kumar, A., Lutz, W., & Williams, M. (2002). Interactions between Coastal and Marine Ecosystems and Human Population Systems: Perspectives on How Consumption Mediates this Interaction. *Royal Swedish Academy of Sciences Ambio*, *31*(4), 264–268. Retrieved from http://www.ambio.kva.se

de Oliveira, A. L., & Turra, A. (2015). Solid waste management in coastal cities: where are the gaps? Case study of the North Coast of. *Journal of Integrated Coastal Zone Management*, *15*(4), 453–465. http://doi.org/10.5894/rgci544

Deichmann, U. (1996). *A Review of Spatial Population Database Design and Modeling*. Santa Barbara. Retrieved from http://escholarship.org/uc/item/6g190671

Doll, C. N. H. (2008). *CIESIN Thematic Guide to Night-time Light Remote Sensing and its Applications*. New York. Retrieved from http://sedac.ciesin.columbia.edu/tg/

Doll, C. N. H. (2010). Population detection profiles of DMSP-OLS night-time imagery by regions of the world. *Proceedings of the Asia-Pacific Advanced Network*, *30*, 190–206. http://doi.org/10.7125/APAN.30.22

Doxsey-Whitfield, E., MacManus, K., Adamo, S. B., Pistolesi, L., Squires, J., Borkovska, O., & Baptista, S. R. (2015). Taking Advantage of the Improved Availability of Census Data: A First Look at the Gridded Population of the World, Version 4. *Papers in Applied Geography*, *1*(3), 226–234. http://doi.org/10.1080/23754931.2015.1014272

Elvidge, C., Baugh, K., Chi Hsu, F., Zhizhin, M., & Ghosh, T. (2015). *VIIRS Nighttime Lights Algorithm Development*. colorado. Retrieved from https://modis.gsfc.nasa.gov/sci_team/meetings/201505/presentations/plenary/baugh.pdf

Elvidge, C. D., Baugh, K. E., Kihn, E. A., Kroehl, H. W., Davis, E. R., & Davis, C. W. (1997). Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption. *International Journal of Remote Sensing*. http://doi.org/10.1080/014311697218485

Elvidge, C. D., Baugh, K. E., Ziskin, D., Anderson, S., & Ghosh, T. (2010). *Estimation of Gas Flaring Volumes Using NASA MODIS Fire Detection Products*. Colorado. Retrieved from https://ngdc.noaa.gov/eog/interest/flare_docs/NGDC_annual_report_20110209.pdf

Elvidge, C. D., Baugh, K., Tuttle, B., Ziskin, D., Ghosh, T., & Zhizhin, M. (2009). *Improving Satellite Data Estimation of Gas Flaring Volumes Year Two Final Report to the GGFR*. Colorado. Retrieved from http://s3.amazonaws.com/zanran_storage/www.ngdc.noaa.gov/ContentPages/15507168.pdf

Elvidge, C. D., Baugh, K., Zhizhin, M., & Chi Hsu, F. (2012). *Satellite Data Estimation of Gas Flaring Volumes*. Colorado. Retrieved from https://www.arb.ca.gov/fuels/lcfs/lcfs_meetings/ngdc_prstn_07122012.pdf

Elvidge, C. D., Baugh, K., Zhizhin, M., & Hsu, F. C. (2013). Why VIIRS data are superior to DMSP for mapping nighttime lights. *Proceedings of the Asia-Pacific Advanced Network*, 62–69. http://doi.org/10.7125/APAN.35.7

Elvidge, C. D., Baught, K. E., Kihn, E. A., Kroehl, H. W., & Davis, E. R. (1997). Mapping City Lights With Nighttime Data from the DMSP Operational Linescan System. *American Society for Phoyogrammetry andRemote Sensing*, 727–730. Retrieved from http://www.csiss.org/SPACE/workshops/2006/OU/docs/Elvidge_et_al_1997a_001.pdf

Elvidge, C. D., Cinzano{, P., Pettit ", D. R., Arvesen, J., Sutton{{, P., Small{{, C., … Ebener, S. (2007). The Nightsat mission concept. *International Journal of Remote Sensing*, *28*(12), 2645–2670. http://doi.org/10.1080/01431160600981525

Elvidge, C. D., Hsu, F.-C., Baugh, K. E., & Ghosh, T. (2014). National Trends in Satellite-Observed Lighting 1992–2012. In Qihao Weng (Ed.), *Global Urban Monitoring and Assessment Through Earth Observation* (pp. 97–119). Colorado: CRC Press. http://doi.org/10.1201/b17012-9

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. http://doi.org/10.1016/j.cageo.2009.01.009

Elvidge, C. D., Zhizhin, M., Baugh, K., Hsu, F. C., & Ghosh, T. (2016). Methods for global survey of natural gas flaring from visible infrared imaging radiometer suite data. *Energies*, *9*(1). http://doi.org/10.3390/en9010014

Elvidge, C. D., Ziskin, D., Baugh, K. E., Tuttle, B. T., Ghosh, T., Pack, D. W., … Zhizhin, M. (2009). A fifteen year record of global natural gas flaring derived from satellite data. *Energies*, *2*(3), 595–622. http://doi.org/10.3390/en20300595

Emam, E. A. (2015). GAS FLARING IN INDUSTRY: AN OVERVIEW. *Petroleum & Coal*, *57*(5), 532–555.

Esri. (n.d.). Esri Support GIS Dictionary. Retrieved January 23, 2017, from http://support.esri.com/other-resources/gis-dictionary/term/sliver polygon

Galgani, F., Hanke, G., & Maes, T. (2015). Global Distribution, Composition and Abundance of Marine Litter. In *Marine Anthropogenic Litte* (pp. 29–56). http://doi.org/10.1007/978-3-319-16510-3_2

Gallaway, T., Olsen, R. N., & Mitchell, D. M. (2009). The economics of global light pollution. Retrieved from http://www.elsevier.com/copyright

Gao, B., Huang, Q., He, C., Sun, Z., & Zhang, D. (2016). How does sprawl differ across cities in China? A multi-scale investigation using nighttime light and census data. *Landscape and Urban Planning*, *148*, 89–98. http://doi.org/10.1016/j.landurbplan.2015.12.006

Ghosh, T., Anderson, S. J., Elvidge, C. D., & Sutton, P. C. (2013). Using nighttime satellite imagery as a proxy measure of human well-being. *Sustainability (Switzerland)*, *5*(12), 4988–5019. http://doi.org/10.3390/su5124988

Goodchild, M. F; Anselin, L.; Deichmann, U. (1993). A framework for the areal interpolation of socioeconomic data. *Environment and Planning A*, *25*, 383–397. Retrieved from http://www.geog.ucsb.edu/~good/papers/187.pdf

Gustya, I. (2005). *Gas Flaring Reduction in the Indonesian Oil and Gas Sector – Technical and Economic Potential of Clean Development Mechanism (CDM) projects*. Hamburg. Retrieved from Hamburgisches Welt-Wirtschafts-Archiv (HWWA)%0AHamburg Institute of International Economics

Henderson, M., Yeh, E. T., Gong, P., Elvidge, C., & Baugh, K. (2003). Validation of urban boundaries derived from global night-time satellite imagery, *24*(3), 595–609. http://doi.org/10.1080/01431160210144534

Huang, Q., Yang, X., Gao, B., Yang, Y., & Zhao, Y. (2014). Application of DMSP/OLS nighttime light images: A meta-analysis and a systematic literature review. *Remote Sensing*, *6*(8), 6844–6866. http://doi.org/10.3390/rs6086844

Jambeck, J. R., Geyer, R., Wilcox, C., Siegler, T. R., Perryman, M., Andrady, A., … Law, K. L. (2015). Plastic waste inputs from land into the ocean. *Science.org*, *28*(13), 39–138. http://doi.org/10.1126/science.1260879

Jia, P., Qiu, Y., & Gaughan, A. E. (2014). A fine-scale spatial population distribution on the High-resolution Gridded Population Surface and application in Alachua County, Florida. *Applied Geography*, *50*, 99–107. http://doi.org/10.1016/j.apgeog.2014.02.009

Khomarudin, M. R., Strunz, G., Post, J., Zoßeder, K., & Ludwig, R. (2009). Derivation of Population Distribution By Combining Census and Landuse Data : As an Input for Tsunami Risk and Vulnerability Assessment. *International Journal of Remote Sensing*, *6*, 46–56.

Kohiyama, M., Hayashi, H., Maki, N., Higashida, M., Kroehl, H. W., Elvidge, C. D., & Hobson, V. R. (2004). Early damaged area estimation system using DMSP-OLS night-time imagery. *International Journal of Remote Sensing INT. J. REMOTE SENSING*, *10*(11), 2015–2036. http://doi.org/10.1080/01431160310001595033

Lee, D.-I., Cho, H.-S., & Jeong, S.-B. (2006). Distribution characteristics of marine litter on the sea bed of the East China Sea and the South Sea of Korea. *Estuarine Coastal and Shelf Science*, *70*, 187–194. http://doi.org/10.1016/j.ecss.2006.06.003

Linard, C., Alegana, V. A., Noor, A. M., Snow, R. W., & Tatem, A. J. (2010). A high resolution spatial population database of Somalia for disease risk mapping. Retrieved from http://www.afripop.

Liu, X., & Clarke, K. (2003). ESTIMATION OF RESIDENTIAL POPULATION USING HIGH RESOLUTION SATELLITE IMAGERY. Retrieved from http://www.geog.ucsb.edu/~kclarke/Papers/Liu_clarke_2002.pdf

Lowe, M. (2014). Night Lights and ArcGIS: A Brief Guide. Retrieved from http://www.econ.brown.edu/faculty/henderson/lights_hsw_data.html.

Ma, L., Wu, J., Li, W., Peng, J., & Liu, H. (2014). Evaluating saturation correction methods for DMSP/OLS nighttime light data: A case study from China's cities. *Remote Sensing*, *6*(10), 9853–9872. http://doi.org/10.3390/rs6109853

Maantay, J. A., Maroko, A. R., & Herrmann, C. (2007). Mapping Population Distribution in the Urban Environment: The Cadastral-based Expert Dasymetric System (CEDS). *Cartography and Geographic Information Science*, *34*(2), 77–102.

Mellander, C. (2013). *Night-Time Light Data: A Good Proxy Measure for Economic Activity?* (No. 2013-NaN-6). Toronto. Retrieved from http://martinprosperity.org/papers/Night-time LightData-Formatted.pdf

Mennis, J. (2003). Generating Surface Models of Population Using Dasymetric Mapping*. *The Professional Geographer*, *55*(August 2002), 31–42. http://doi.org/10.1111/0033-0124.10042

Michalopoulos, S., & Papaioannou, E. (2012). *Pre-Colonial Ethnics Institutions and Contemporary African Development*. Cambridge. Retrieved from http://www.nber.org/papers/w18224

Moore, C. J. (2015). How much plastic is in the ocean? You tell me! *Marine Pollution Bulletin*. http://doi.org/10.1016/j.marpolbul.2014.11.042

Nelson, A. (2004). African Population Database Documentation. Retrieved November 20, 2016, from https://na.unep.net/siouxfalls/globalpop/africa/part2.html#urban

Oak Ridge National Laboratory. (2014). LandScan - Documentation. Retrieved November 20, 2016, from http://web.ornl.gov/sci/landscan/landscan_documentation.shtml

Pestalozzi, N. (2012). *Nighttime lights as proxy for the spatial growth of dense urbanized areas*. ETH Zürich. Retrieved from http://www.worldatnight.ethz.ch/content/doc/Nicola_Pestalozzi_Master_Thesis.pdf

Pozzi, F., Small, C., & Yetman, G. (2002). Modelling the Distribution of Human Population with Night-time Light Satelilte Imagery and Gridded Population of the world. Retrieved from http://www.isprs.org/proceedings/XXXIV/part1/paper/00061.pdf

Propastin, P., & Kappas, M. (2012). Assessing Satellite-Observed Nighttime Lights for Monitoring Socioeconomic Parameters in the Republic of Kazakhstan. *GIScience & Remote Sensing*, *49*(4), 538–557. http://doi.org/10.2747/1548-1603.49.4.538

Rahman, M. R., Shi, Z. H., & Chongfa, C. (2014). Assessing regional environmental quality by integrated use of remote sensing, GIS, and spatial multi-criteria evaluation for prioritization of environmental restoration. *Environmental Monitoring and Assessment*, *186*(11), 6993–7009. http://doi.org/10.1007/s10661-014-3905-4

Rech, S., Macaya-Caquilpán, V., Pantoja, J. F., Rivadeneira, M. M., Jofre Madariaga, D., & Thiel, M. (2014). Rivers as a source of marine litter – A study from the SE Pacific. *Marine Pollution Bulletin*, *82*(1), 66–75. http://doi.org/10.1016/j.marpolbul.2014.03.019

Roychowdhury, K., Jones, S. D., Arrowsmith, C., & Reinke, K. (2011). A Comparison of High and Low Gain DMSP / OLS Satellite Images for the Study of Socio-Economic Metrics. *Selected Topics in Applied Earth Observations and Remote Sensing IEEE Journal of*, *4*(1), 35–42. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5535229

Ryan, P. G., Moore, C. J., Van Franeker, J. A., & Moloney, C. L. (2009). Monitoring the abundance of plastic debris in the marine environment. *The Royal Society*, 1999–2012. http://doi.org/10.1098/rstb.2008.0207

Schneiderbauer, S., & Ehrlich, D. (2007). EO data supported population density estimation at fine resolution–test case rural Zimbabwe. *Global Monitoring for Security and Stability (GMOSS). Integrated Scientific and Technological Research Supporting Security Aspects of the European Union, JRC Scientific and Technical Reports, EUR*, *23033*(January 2007), 194–209. Retrieved from http://isferea.jrc.it/Activities/ProjectPortfolio/Documents/2004_GMOSS/Schneiderbauer_Ehrlich2007.pdf

SEDAC. (2005). GPW and GRUMP. Retrieved November 13, 2016, from http://sedac.ciesin.columbia.edu/data/collection/gpw-v3/about-us

Small, C., Elvidge, C. D., Balk, D., & Montgomery, M. (2011). Spatial scaling of stable night lights. *Remote of Environment*, *115*(2), 269–280. http://doi.org/10.1016/j.rse.2010.08.021

Stevens, F. R., Gaughan, A. E., Linard, C., & Tatem, A. J. (2015). Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data. *PLoS ONE*, *10*(2). http://doi.org/10.1371/journal.pone.0107042

Stevenson, C. (2011). *Plastic Debris in the California Marine Ecosystem : A Summary of Current Research, Solution Strategies and Data Gaps*. Oakland, CA. Retrieved from http://www.oceansciencetrust.org/wp-content/uploads/2015/08/Plastic-Debris-in-the-CA-Marine-Ecosystem_OST-2011-Report.pdf

Sutton, P. C. (2003). A scale-adjusted measure of "Urban sprawl" using nighttime satellite imagery. *Remote Sensing of Environment*, *86*(3), 353–369. http://doi.org/10.1016/S0034-4257(03)00078-6

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. http://doi.org/10.1080/01431160010007015

Tenerelli, P., Gallego, J. F., & Ehrlich, D. (2015). Population density modelling in support of disaster risk assessment. *International Journal of Disaster Risk Reduction*, *13*, 334–341. http://doi.org/10.1016/j.ijdrr.2015.07.015

Tobler, W., Deichmann, U., Gottsegen, J., & Maloy, K. (1997). World Population in a Grid of Spherical Quadrilaterals. *Int. J. Popul. Geogr*, *3*, 203–225.

Uneputty, P., & Evans, S. M. (1997). The impact of plastic debris on the biota of tidal flats in Ambon Bay (eastern Indonesia). *Marine Environmental Research*, *44*(3), 233–242. http://doi.org/10.1016/S0141-1136(97)00002-0

Wang, L., & Wu, C. (2010). Preface: Population estimation using remote sensing and GIS technologies. *International Journal of Remote Sensing*, *31*(21), 5569–5570. http://doi.org/10.1080/01431161.2010.496809

Wei, C.-L., Rowe, G. T., Nunnally, C. C., & Wicksten, M. K. (2012). Anthropogenic "Litter" and macrophyte detritus in the deep Northern Gulf of Mexico. *Marine Pollution Bulletin*, *64*(5), 966–973. http://doi.org/10.1016/j.marpolbul.2012.02.015

APPENDIX 1 LIST OF DATA

Appendix I List of Data for This Study

No.	Dataset	Data format	Data type	Year of Observation	Spatial Scale	Missing Data	Noise	Sources
1.	Night-time light dataset	Raster	Geotiff	2010	~ 1 km	0 background noise	• Blooming effect that over-represent built-up area • Poorly or infrequently lit that under-represent small settlements	http://www.ngdc.noaa.gov/dmsp/dmsp.html
2.	National Census 2010	dBase File	Table	2010	1. Province 2. Regency 3. District 4. Villages			http://sp2010.bps.go.id
3.	Administration unit	Vector	Polygon	2010	Villages level	Waterbody, Forest and Others labelled as 555, 888 and 999	Around 5 administrative unit with too small polygon	National Statistic Bureau
4.	DMSP-OLS Gas Flare Observation	Vector	Polygon	1997-2012	National			https://ngdc.noaa.gov/eog/interest/gas_ flares_countries_shapefiles. html)
5.	NPP-VIIRS Gas Flare Observation	Vector	Point	2013	National			https://ngdc.noaa.gov/eog/interest/gas_flares_countries_kmz.html
6.	DMSP-OLS Identified Gas Flare	KMZ	Point	2010-2012	National			https://ngdc.noaa.gov/eog/viirs/download_global_flare.html on January 19th 2017

APPENDIX 2 PROVINCE DATA

Province Data - without Gas Flare Removal

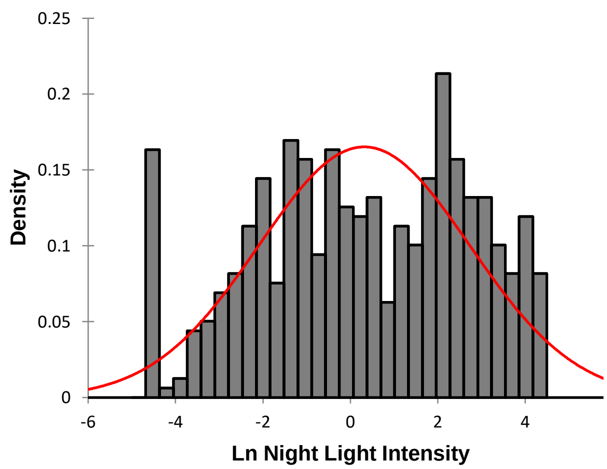
Census ID	Province	Population (people)	Area (Km2)	Sum of Light Luminosity	Population Density (People/km^2)	Light Intensity (DN Value)	Ln (0,01 + Population Density)	Ln (0,01 + Light Intensity)
11	Aceh	4494410	57105.06	68371	79	1.20	4.37	0.18
12	Sumatera Utara	12982204	72419.14	161589	179	2.23	5.19	0.80
13	Sumatera Barat	4846909	42234.14	39638	115	0.94	4.74	-0.06
14	Riau	5538367	89739.05	131355	62	1.46	4.12	0.38
15	Jambi	3092265	48980.96	52999	63	1.08	4.15	0.08
16	Sumatera Selatan	7450394	87267.13	160638	85	1.84	4.45	0.61
17	Bengkulu	1715518	19903.89	8390	86	0.42	4.46	-0.86
18	Lampung	7608405	33850.34	82021	225	2.42	5.42	0.89
19	Kepulauan Bangka Belitung	1223296	16737.59	14863	73	0.89	4.29	-0.12
21	Kepulauan Riau	1679163	8289.86	45005	203	5.43	5.31	1.69
31	DKI Jakarta	9607787	661.21	46765	14531	70.73	9.58	4.26
32	Jawa Barat	43053732	37603.00	484682	1145	12.89	7.04	2.56
33	Jawa Tengah	32382657	34993.05	400176	925	11.44	6.83	2.44
34	DI Yogyakarta	3457491	3237.69	49501	1068	15.29	6.97	2.73
35	Jawa Timur	37476757	48876.04	558940	767	11.44	6.64	2.44
36	Banten	10632166	9471.42	126580	1123	13.36	7.02	2.59
51	Bali	3890757	5710.88	63436	681	11.11	6.52	2.41
52	Nusa Tenggara Barat	4500212	20120.54	36750	224	1.83	5.41	0.60
53	Nusa Tenggara Timur	4683827	47686.41	15533	98	0.33	4.59	-1.12
61	Kalimantan Barat	4395983	146972.98	29544	30	0.20	3.40	-1.60
62	Kalimantan Tengah	2212089	153877.13	17549	14	0.11	2.67	-2.17
63	Kalimantan Selatan	3626616	37496.17	65828	97	1.76	4.57	0.56
64	Kalimantan Timur	3553143	196453.80	156394	18	0.80	2.90	-0.23
71	Sulawesi Utara	2270596	14519.11	35393	156	2.44	5.05	0.89
72	Sulawesi Tengah	2635009	61196.66	19549	43	0.32	3.76	-1.14
73	Sulawesi Selatan	8034776	45686.37	61225	176	1.34	5.17	0.29
74	Sulawesi Tenggara	2232586	36826.30	11113	61	0.30	4.10	-1.20
75	Gorontalo	1040164	12043.58	13301	86	1.10	4.46	0.10
76	Sulawesi Barat	1158651	16603.71	2747	70	0.17	4.25	-1.80
81	Maluku	1533506	46714.44	9870	33	0.21	3.49	-1.55
82	Maluku Utara	1038087	31535.49	7667	33	0.24	3.49	-1.41
91	Papua Barat	760422	98693.96	20396	8	0.21	2.04	-1.58
94	Papua	2833381	316436.87	23909	9	0.08	2.19	-2.58

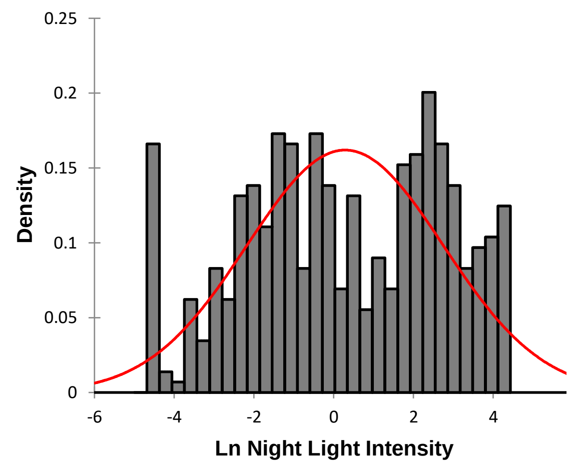
Province Data - with Gas Flare Removal

Census ID	Province	Population (people)	Area (Km2)	Sum of Light Luminosity	Population Density (People/km²)	Light Intensity (DN Value)	Ln (0,01 + Population Density)	Ln (0,01 + Light Intensity)
76	Sulawesi Barat	1158651	16603.71	2712	70	0.16	4.25	-1.81
82	Maluku Utara	1038087	31535.49	7747	33	0.25	3.49	-1.40
17	Bengkulu	1715518	19903.89	8348	86	0.42	4.46	-0.87
81	Maluku	1533506	46714.44	9613	33	0.21	3.49	-1.58
74	Sulawesi Tenggara	2232586	36826.30	11179	61	0.30	4.10	-1.19
75	Gorontalo	1040164	12043.58	13226	86	1.10	4.46	0.09
19	Kepulauan Bangka Belitung	1223296	16737.59	14896	73	0.89	4.29	-0.12
53	Nusa Tenggara Timur	4683827	47686.41	15432	98	0.32	4.59	-1.13
62	Kalimantan Tengah	2212089	153877.13	17537	14	0.11	2.67	-2.17
72	Sulawesi Tengah	2635009	61196.66	19629	43	0.32	3.76	-1.14
94	Papua	2833381	316436.87	23828	9	0.08	2.19	-2.59
61	Kalimantan Barat	4395983	146972.98	29593	30	0.20	3.40	-1.60
71	Sulawesi Utara	2270596	14519.11	35184	156	2.42	5.05	0.89
52	Nusa Tenggara Barat	4500212	20120.54	36309	224	1.80	5.41	0.59
13	Sumatera Barat	4846909	42234.14	39416	115	0.93	4.74	-0.07
21	Kepulauan Riau	1679163	8289.86	45085	203	5.44	5.31	1.69
34	DI Yogyakarta	3457491	3237.69	49702	1068	15.35	6.97	2.73
73	Sulawesi Selatan	8034776	45686.37	60695	176	1.33	5.17	0.28
51	Bali	3890757	5710.88	63448	681	11.11	6.52	2.41
63	Kalimantan Selatan	3626616	37496.17	65811	97	1.76	4.57	0.56
18	Lampung	7608405	33850.34	81836	225	2.42	5.42	0.88
36	Banten	10632166	9471.42	126937	1123	13.40	7.02	2.60
33	Jawa Tengah	32382657	34993.05	399704	925	11.42	6.83	2.44

APPENDIX 3 REGENCY

All Area

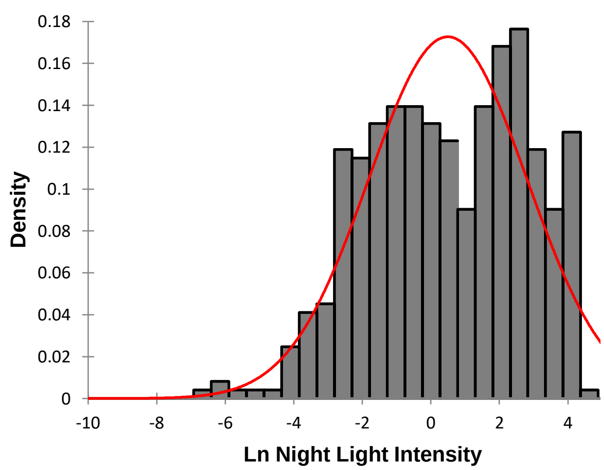


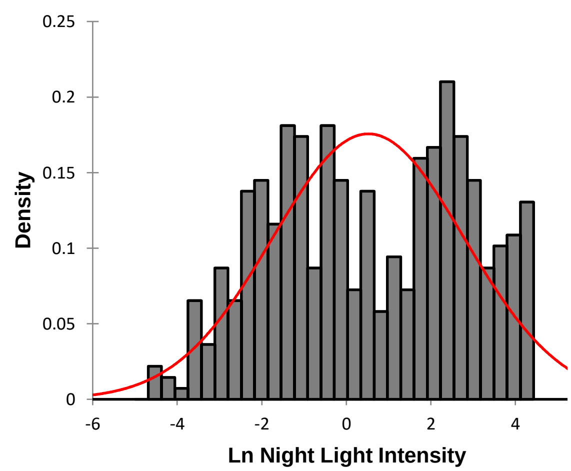


Without Gas Flare Removal

With Gas Flare Removal

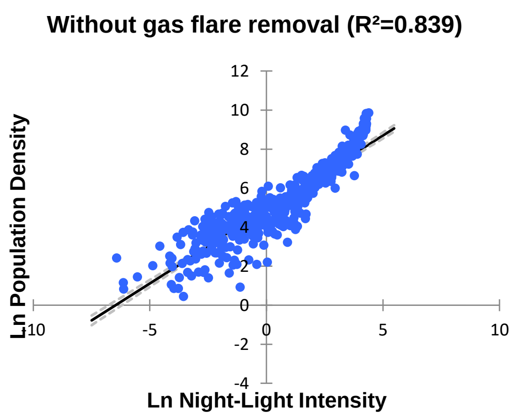
<u>Lit Area</u>

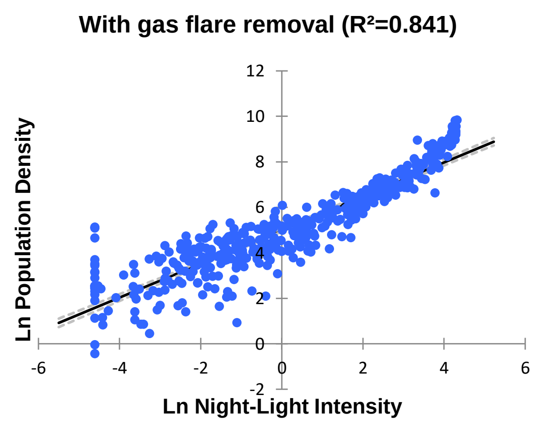


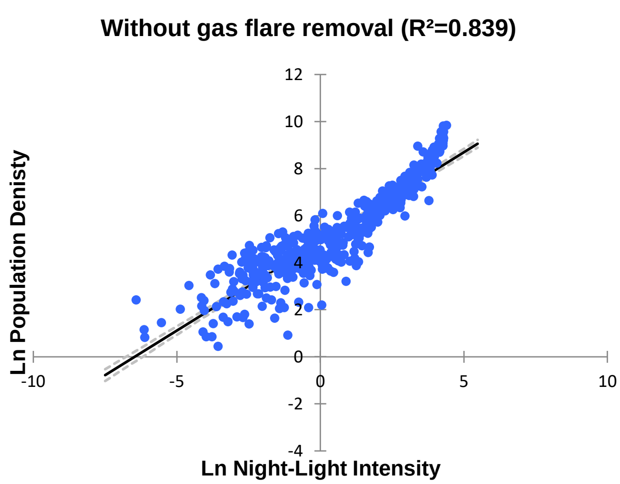


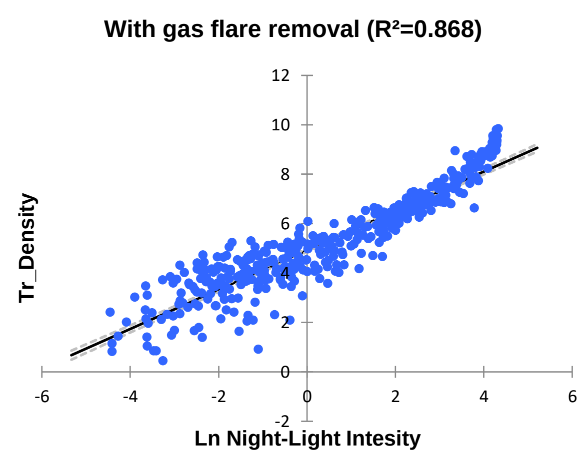
Correlation Matrix

		Without Gas Flare Removal		With Gas Flare Removal	
		Ln Night-Light Intensity	Population Density	Ln Night- Light Intensity	Population Density
All Area	Ln Night-Light Intensity	1	0.792	1	0.932
	Population Density	0.792	1	0.932	0.932
Lit Area	Ln Night-Light Intensity	1	0.916	1	0.932
	Population Density	0.916	1	0.932	0.932



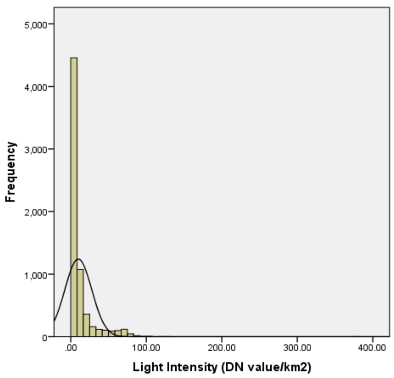




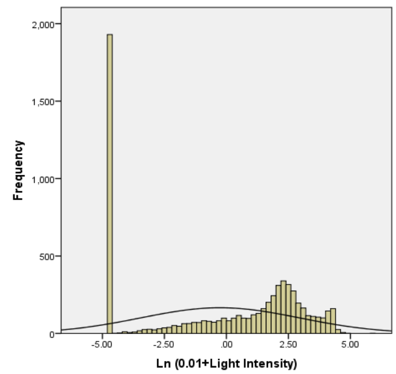


APPENDIX 4 DISTRICT

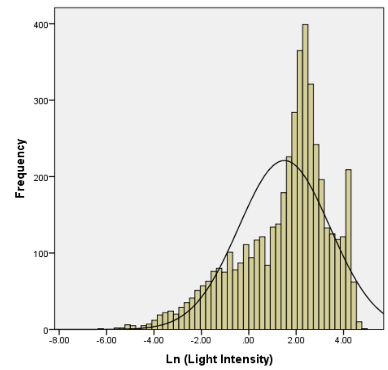
Stable Night-Light



a) Distribution of light intensity

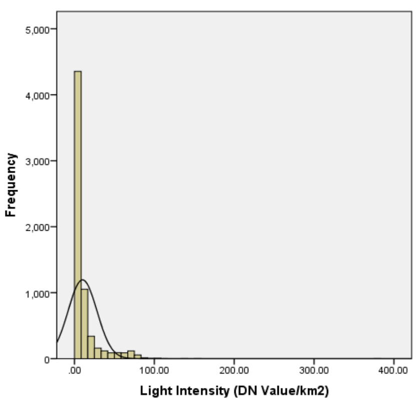


b) Distribution of log of light density by adding small number y=ln (0.01+Light intensity)

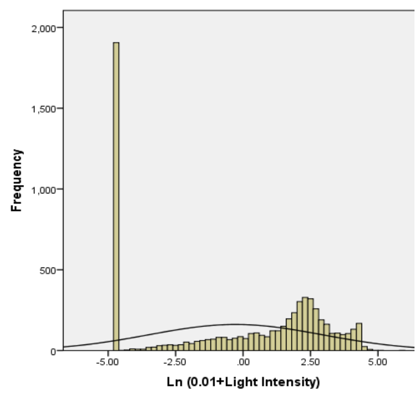


c) Distribution of log light intensity in lit areas y=ln(Light Intensity)

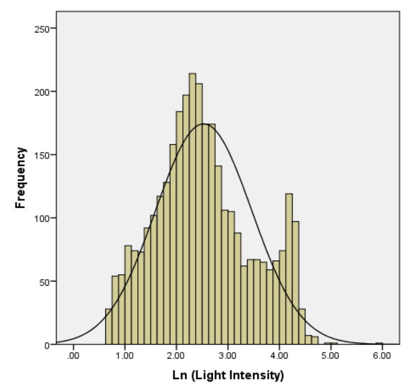
Stable Night-Light without gas flare



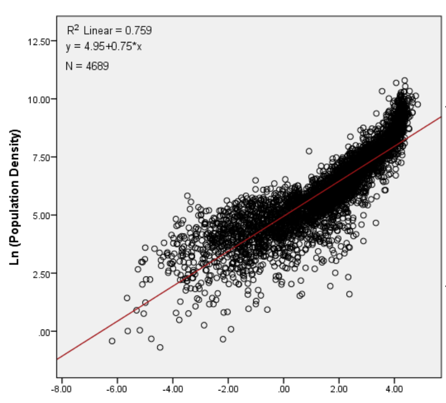
d) Distribution of light intensity



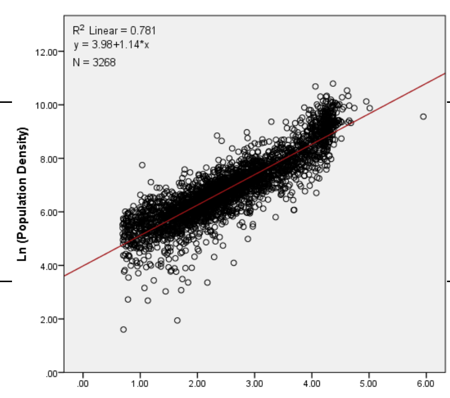
e) Distribution of log of light density by adding small number y=ln (0.01+Light intensity)



f) Distribution of log light intensity in lit areas y=ln(Light Intensity)

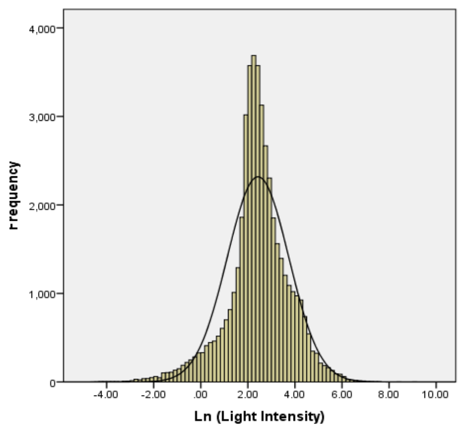


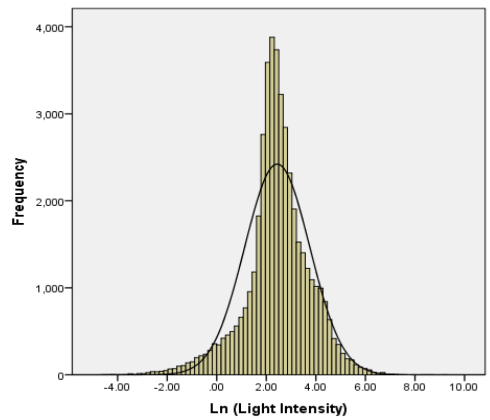
a) Significant correlation between light intensity and population density in lit area without gas flare removal

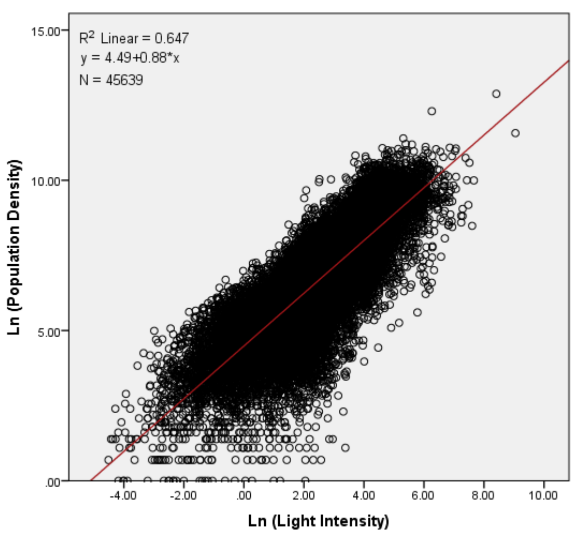


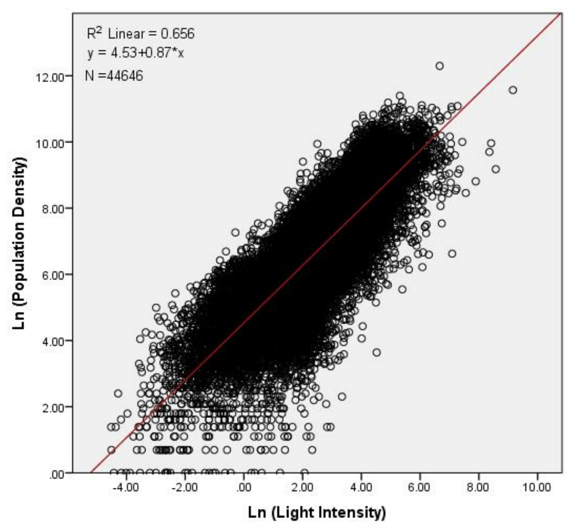
b) Significant correlation between light intensity and population density in lit area with gas flare removal

APPENDIX 5 VILLAGE









APPENDIX 6 PEARSON'S TABLE

Table of Critical Values for Pearson's *r*

	Level of Significance for a One-Tailed Test					
	.10	.05	.025	.01	.005	.0005
	Level of Significance for a Two-Tailed Test					
df	.20	.10	.05	.02	.01	.001
1	0.951	0.988	0.997	0.9995	0.9999	0.99999
2	0.800	0.900	0.950	0.980	0.990	0.999
3	0.687	0.805	0.878	0.934	0.959	0.991
4	0.608	0.729	0.811	0.882	0.917	0.974
5	0.551	0.669	0.755	0.833	0.875	0.951
6	0.507	0.621	0.707	0.789	0.834	0.925
7	0.472	0.582	0.666	0.750	0.798	0.898
8	0.443	0.549	0.632	0.715	0.765	0.872
9	0.419	0.521	0.602	0.685	0.735	0.847
10	0.398	0.497	0.576	0.658	0.708	0.823
11	0.380	0.476	0.553	0.634	0.684	0.801
12	0.365	0.457	0.532	0.612	0.661	0.780
13	0.351	0.441	0.514	0.592	0.641	0.760
14	0.338	0.426	0.497	0.574	0.623	0.742
15	0.327	0.412	0.482	0.558	0.606	0.725
16	0.317	0.400	0.468	0.542	0.590	0.708
17	0.308	0.389	0.456	0.529	0.575	0.693
18	0.299	0.378	0.444	0.515	0.561	0.679
19	0.291	0.369	0.433	0.503	0.549	0.665
20	0.284	0.360	0.423	0.492	0.537	0.652
21	0.277	0.352	0.413	0.482	0.526	0.640
22	0.271	0.344	0.404	0.472	0.515	0.629
23	0.265	0.337	0.396	0.462	0.505	0.618
24	0.260	0.330	0.388	0.453	0.496	0.607
25	0.255	0.323	0.381	0.445	0.487	0.597
26	0.250	0.317	0.374	0.437	0.479	0.588
27	0.245	0.311	0.367	0.430	0.471	0.579
28	0.241	0.306	0.361	0.423	0.463	0.570
29	0.237	0.301	0.355	0.416	0.456	0.562
30	0.233	0.296	0.349	0.409	0.449	0.554
40	0.202	0.257	0.304	0.358	0.393	0.490
60	0.165	0.211	0.250	0.295	0.325	0.408
120	0.117	0.150	0.178	0.210	0.232	0.294
∞	0.057	0.073	0.087	0.103	0.114	0.146

APPENDIX 7 Z TABLE

Two tails of *Z*

Entries in the table represent two-tailed *P* values for *z* statistics

					hundredths					
tenths	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
0.0	1.00000	0.99202	0.98404	0.97607	0.96809	0.96012	0.95216	0.94419	0.93624	0.92829
0.1	0.92034	0.91241	0.90448	0.89657	0.88866	0.88076	0.87288	0.86501	0.85715	0.84931
0.2	0.84148	0.83367	0.82587	0.81809	0.81033	0.80259	0.79486	0.78716	0.77948	0.77182
0.3	0.76418	0.75656	0.74897	0.74140	0.73386	0.72634	0.71885	0.71138	0.70395	0.69654
0.4	0.68916	0.68181	0.67449	0.66720	0.65994	0.65271	0.64552	0.63836	0.63123	0.62413
0.5	0.61708	0.61005	0.60306	0.59611	0.58920	0.58232	0.57548	0.56868	0.56191	0.55519
0.6	0.54851	0.54186	0.53526	0.52869	0.52217	0.51569	0.50925	0.50286	0.49650	0.49019
0.7	0.48393	0.47770	0.47152	0.46539	0.45930	0.45325	0.44725	0.44130	0.43539	0.42953
0.8	0.42371	0.41794	0.41222	0.40654	0.40091	0.39533	0.38979	0.38430	0.37886	0.37347
0.9	0.36812	0.36282	0.35757	0.35237	0.34722	0.34211	0.33706	0.33205	0.32709	0.32217
1.0	0.31731	0.31250	0.30773	0.30301	0.29834	0.29372	0.28914	0.28462	0.28014	0.27571
1.1	0.27133	0.26700	0.26271	0.25848	0.25429	0.25014	0.24605	0.24200	0.23800	0.23405
1.2	0.23014	0.22628	0.22246	0.21870	0.21498	0.21130	0.20767	0.20408	0.20055	0.19705
1.3	0.19360	0.19020	0.18684	0.18352	0.18025	0.17702	0.17383	0.17069	0.16759	0.16453
1.4	0.16151	0.15854	0.15561	0.15272	0.14987	0.14706	0.14429	0.14156	0.13887	0.13622
1.5	0.13361	0.13104	0.12851	0.12602	0.12356	0.12114	0.11876	0.11642	0.11411	0.11183
1.6	0.10960	0.10740	0.10523	0.10310	0.10101	0.09894	0.09691	0.09492	0.09296	0.09103
1.7	0.08913	0.08727	0.08543	0.08363	0.08186	0.08012	0.07841	0.07673	0.07508	0.07345
1.8	0.07186	0.07030	0.06876	0.06725	0.06577	0.06431	0.06289	0.06148	0.06011	0.05876
1.9	0.05743	0.05613	0.05486	0.05361	0.05238	0.05118	0.05000	0.04884	0.04770	0.04659
2.0	0.04550	0.04443	0.04338	0.04236	0.04135	0.04036	0.03940	0.03845	0.03753	0.03662
2.1	0.03573	0.03486	0.03401	0.03317	0.03235	0.03156	0.03077	0.03001	0.02926	0.02852
2.2	0.02781	0.02711	0.02642	0.02575	0.02509	0.02445	0.02382	0.02321	0.02261	0.02202
2.3	0.02145	0.02089	0.02034	0.01981	0.01928	0.01877	0.01827	0.01779	0.01731	0.01685
2.4	0.01640	0.01595	0.01552	0.01510	0.01469	0.01429	0.01389	0.01351	0.01314	0.01277
2.5	0.01242	0.01207	0.01174	0.01141	0.01109	0.01077	0.01047	0.01017	0.00988	0.00960
2.6	0.00932	0.00905	0.00879	0.00854	0.00829	0.00805	0.00781	0.00759	0.00736	0.00715
2.7	0.00693	0.00673	0.00653	0.00633	0.00614	0.00596	0.00578	0.00561	0.00544	0.00527
2.8	0.00511	0.00495	0.00480	0.00465	0.00451	0.00437	0.00424	0.00410	0.00398	0.00385
2.9	0.00373	0.00361	0.00350	0.00339	0.00328	0.00318	0.00308	0.00298	0.00288	0.00279
3.0	0.00270	0.00261	0.00253	0.00245	0.00237	0.00229	0.00221	0.00214	0.00207	0.00200
3.1	0.00194	0.00187	0.00181	0.00175	0.00169	0.00163	0.00158	0.00152	0.00147	0.00142
3.2	0.00137	0.00133	0.00128	0.00124	0.00120	0.00115	0.00111	0.00108	0.00104	0.00100
3.3	0.00097	0.00093	0.00090	0.00087	0.00084	0.00081	0.00078	0.00075	0.00072	0.00070
3.4	0.00067	0.00065	0.00063	0.00060	0.00058	0.00056	0.00054	0.00052	0.00050	0.00048
3.5	0.00047	0.00045	0.00043	0.00042	0.00040	0.00039	0.00037	0.00036	0.00034	0.00033
3.6	0.00032	0.00031	0.00029	0.00028	0.00027	0.00026	0.00025	0.00024	0.00023	0.00022
3.7	0.00022	0.00021	0.00020	0.00019	0.00018	0.00018	0.00017	0.00016	0.00016	0.00015
3.8	0.00014	0.00014	0.00013	0.00013	0.00012	0.00012	0.00011	0.00011	0.00010	0.00010
3.9	0.00010	0.00009	0.00009	0.00008	0.00008	0.00008	0.00007	0.00007	0.00007	0.00007