Student thesis series INES nr 494

Disaggregating economic inequality from space

Neija Maegaard Elvekjær

2019

Department of

Physical Geography and Ecosystem Science

Lund University

Sölvegatan 12

Neija Maegaard Elvekjær (2019).

Disaggregating economic inequality from space

Nedskalning av ekonomisk ojämlikhet sett från rymden

Master degree thesis, 30 credits in *Geomatics*

Department of Physical Geography and Ecosystem Science, Lund University

Level: Master of Science (MSc)

Course duration: *January* 2019 until *June* 2019

Disaggregating Economic Inequality from Space

Neija Maegaard Elvekjær

Master thesis, 30 credits, in *Geomatics*

Supervisors:

Jonathan Seaquist

Department of Physical Geography and Ecosystem Sciences

Ola Hall

Department of Human Geography

Exam committee:

Paul Miller, Department of Physical Geography and Ecosystem Sciences

Emma Li Johansson, Department of Physical Geography and Ecosystem Sciences

Abstract:

Reducing inequality is number 10 of the United Nations sustainability goals, however due to uncertainty and high cost associated with the collection of data, satisfactory inequality data is missing on both temporal and spatial scales. In order to fill the data gap, this thesis attempts to use night-time light (NTL) imagery to model economic inequality on five different spatial scales. The study confirms the use of NTL to model other socio-economic variables such as GDP. However, modelling economic inequality is more complicated, and introduces problems based on the modifiable area unit problem and specific problems related to the definition of the Gini coefficient, a measure of inequality.

Weak associations were found between predicted and observed Gini coefficients. However, more attention to spatial scale is needed. The results of this thesis show that both the correlations, the distribution of error as well as the mean absolute error varies depending on spatial scale of the analysis. Although spatial scale is suggested as a controlling factor, some patterns between the scales are still unexplained. Further investigation into other explanatory variables as well as avoiding errors due to discrepancies between different datasets might clear up the weak trends suggested in this thesis.

Keywords: Visible Infrared Imaging Radiometer Suite (VIIRS), night-time lights, Gross Domestic Product (GDP), economic inequality, regional inequality, spatial inequality, Gini coefficient, spatial scale

Contents

Figure list:

Table list:

Abbreviations used:

DMSP-OLS: The Defence Meteorological Satellite Program Operational Line Scan System

DN: Digital Number

GDP: Gross Domestic Product

LP: Lit Population

MAE: Mean Absolute Error

MAUP: Modifiable Area Unit Problem

NASA: National Aeronautics and Space Administration

NLDI: Night Light Development Index

NOAA: National Oceanic and Atmospheric Administration

NTL: Night-Time Light data

OECD: Organisation for Economic Co-operation and Development

SE: Socio-Economic

SEDAC: Socioeconomic Data and Application Center

SOL: Sum of Light

UN: United Nations

VIIRS: Visible Infrared Imaging Radiometer Suite

1 Introduction

Human activity is one of the major drivers of global change. In order to understand the central relationship between society and the environment, information on the spatial distribution of human activity is needed. Human society can be understood through the spatial and temporal dynamics of population and economics represented by Socio-Economic (SE) variables.

However, SE data such as measures of wealth, development, population and inequality often only exist at a country level, if at all. These metrics are based on household surveys representative at a national level. They are costly and labour intensive to produce and are therefore missing in many developing countries. Furthermore, the spatial scale of the data is insufficient to capture within country variations of SE-variables and to explore relationships between society and environment (Zhou et al. 2015; Ebener et al. 2005).

To solve this problem several studies have applied Night-Time Light (NTL) remotely sensed data in order to model and disaggregate SE-variables from national level data to finer scales. NTL images show the distribution of night light, mostly emitted from public streetlight, thereby representing the spatial distribution of access to electricity and the infrastructure of human society (Bennett and Smith 2017). SE-variables such as population, gross domestic product (GDP), electricity consumption, urbanization, CO_2 consumption, ecological footprints and several development indicators have successfully been modelled with the use of NTL data at various scales (Proville et al. 2017; Ghosh et al. 2013; Ebener et al. 2005). These studies have used the spatially distributed NTL images as a proxy for human activity. By establishing relationships between SE-variables and NTL images, spatially distributed proxies can be produced to cover the missing data gap at fine temporal and spatial scales.

The production of reliable SE-proxies from satellite images increases both the spatial and temporal resolution of the data available for analysis, as well as reducing the problems associated with household survey data, such as human bias and high cost of survey based data collection and problems with inter-comparability between surveys. The development of accurate and precise SE-variables from NTL data could therefore greatly improve the understanding on how SE-variables vary on a smaller scale, and possibly how they are related to environmental variables, such as climate, vegetation etc. Furthermore, global SE-datasets produced with a consistent methodology allows for easy comparison between different countries and regions, thereby enhancing the analysis of the spatial patterns of human society and wellbeing.

Inequality is a growing concern in today's society. Reducing inequality is included as the tenth sustainable development goal, developed by the United Nations. The economic inequality of a society has been shown to affect growth, development and is possibly linked to environmental change. Despite recent trends showing a decrease in global inequality, disparities within countries continue to grow. To further explore these variations on local levels, disaggregated data on economic inequality is needed. However, throughout the large collection of literature concerning SE-variables and NTL, few have attempted to model data of economic inequality data. The differences within society might not be as easily captured as the absolute measures of society, such as GDP and population. Is it possible to model economic inequality from NTL-imagery building upon methods previously used for other SE-variables?

1.1 Aim

This study aims to develop a proxy for economic inequality at five spatial scales. This aim will be accomplished through the following research objectives

The research objectives are to:

1. Identify methods that can be used to construct proxies of SE-variables from NTL data and assess implementation of these methodologies to predict economic inequality.
2. Develop a proxy for economic inequality based on the identified methodologies.
3. Evaluate the model performance at five spatial scales and discuss the implication of area when modelling disparities

This thesis will be divided into three parts, following the three research objectives. Part 1 part will be addressed through a literature review in the background section. This background section will include an introduction to economic inequality, NTL imagery and their application as a proxy for human activity; and the potential of NTL imagery in capturing spatial variation in SE variables, with a focus on inequality. The second part will build upon the theory of previous studies and develop models that can be used for predicting income and economic inequality at various scales. In the third part I will evaluate the developed model and the ability of capturing inequality from NTL, as well as discuss some concerns and the future development of the field.

2 Background

2.1 Economic Inequality

2.1.1 Definition

Inequality is a measure of dispersion within a distribution. The term is used in a wide range of fields and is most commonly used to evaluate the distribution of various SE variables within a population, for example inequality in healthcare, education, opportunity or economy. Inequality measures are generally defined by two components: 1) inequality of what, 2) inequality between whom.

Economic inequality, or income inequality, is used in social and economic sciences to describe the distribution of income among the people within a specified unit such as a region, country or the world. The concept of economic inequality is different from poverty, since it not only reflects the people with the lowest wages but is a measure of the society as a whole.

Economic inequality can generally be divided into two concepts: personal inequality, which is based on the distribution of individual incomes within the society; and regional inequality, also called spatial inequality, which looks at the disparities between the GDP of various regions within a country. Personal inequality is therefore a measure of differences between people; whereas regional inequality is a measure of the differences between regions.

2.1.2 Significance of economic inequality

Economic inequality is a measure of the disparity within human society. The difference between countries and regions within a country depend on resource availability, local policies etc. The distribution of wealth within a society affects development and wealth of the society as a whole. Unequal distribution of wages within a population can affect economic growth and hinder sustainable development. Monitoring this distribution within society is therefore important, in order to strive for a sustainable future.

Recent global assessments show that poverty, hunger and child mortality is declining. However, the distribution of income between global citizens (global inequality) has remained the same despite the reduction in inequality between countries (international inequality) (Alvaredo et al. 2018)**.** This stagnant development comes from an increased inequality within countries, especially those with high economic growth such as China and India.

Inequality trends and their dynamicse is a large field of research. This thesis does not aim to explain this in further details. Economic inequality will be used in this paper as a variable to be modelled in a spatial context. The spatial context of the analysis aims to provide us with a measure of regional inequality, as opposed to individual personal inequality. The politics and economics describing the behaviour of this variable and the spatially modelled inequality will not be addressed further.

2.1.3 Measures of economic inequality

Economic inequality can be expressed through a long range of indices. All measures are slightly different because they express inequality between different parts of the population and because they are based on different economic variables. Each measure has a different focus and is sensitive to a specific part of the population. The choice of metric therefor depends on the proposed research question and the data available.

Ratio metrics, often used by organisations such as the UN, only consider a certain percentage of the top and bottom income; for example, the decile dispersion index compares the richest ten percent with the poorest ten percent in a population. However, the ratio metrics fail to represent the larger part of the population. To include dynamics in the large middle class, metrics such as the Gini index provide a measure of the distribution of the entire population. However, these methods are overly sensitive to variations in the vast middle class rather than the income extremes, both rich and poor, which might be of interest in some cases (Haughton and Khandker 2009).

All economic inequality indices are limited by the economic data that they are computed from. As mentioned by Ghosh et al. (2013), traditional measures of income, such as national statistics fail to account for non-monetary wealth. This includes ownership, ecosystem services, informal economy such as small-scale business on the black market; international aid, for example food stamps; as well as differences in government aid, such as free health care and education. Most inequality indices cannot account for these aspects of wealth.

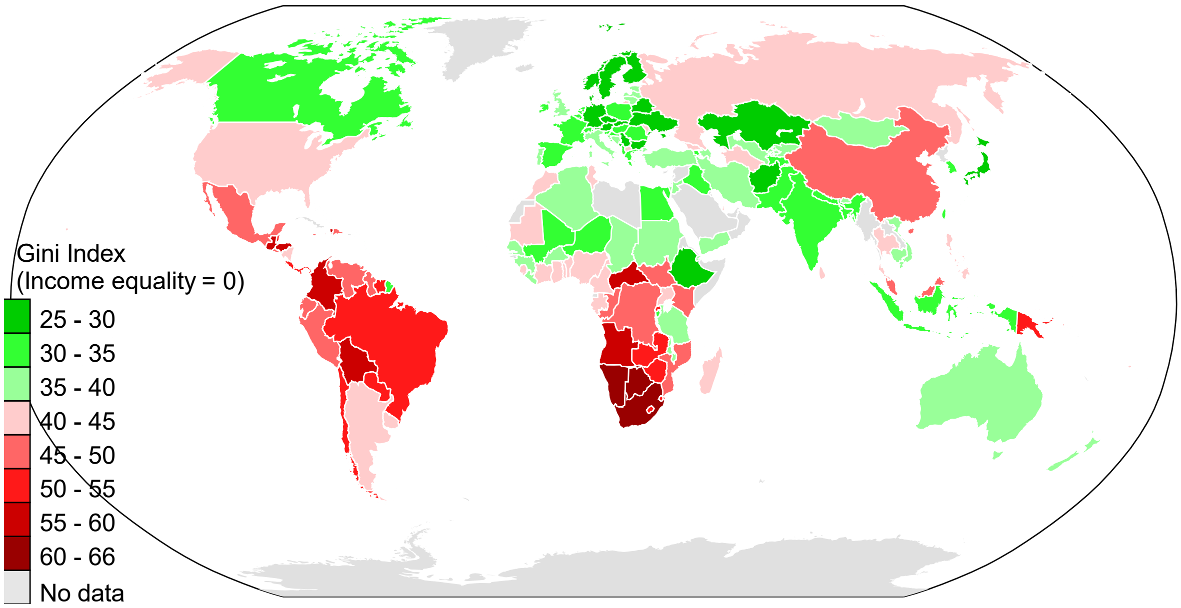
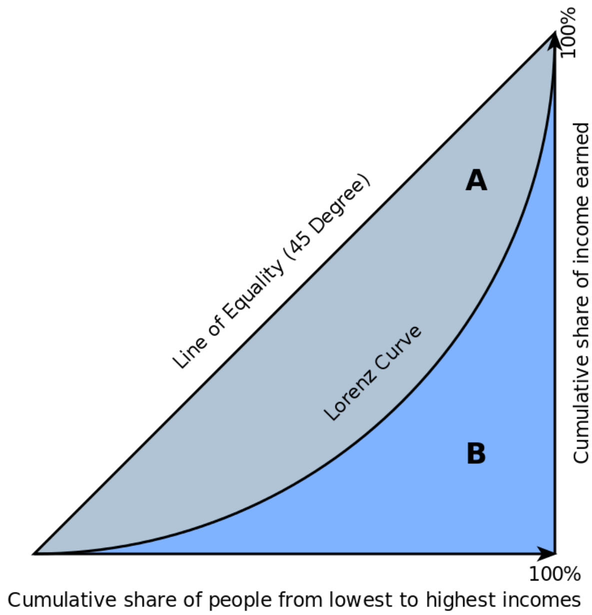


Figure 1 Map of the Gini Index (in percentage),a measure of economic inequality according to latest published data by World Bank in July 2014. Due to missing data, some countries are represented by Gini indices up to 10 years old. Re-produced from Hunter (2014) with permission from author.

This paper will use the Gini coefficient as a measure of economic inequality. This coefficient is the most common metric of inequality, and therefore has the largest amount of data at various spatial scales that is needed for both modelling the relationship with the NTL and for validation of the modelled inequality metric. The Gini index ranges from 0-1 (or from 0-100 percent as in *Figure 1*), where 0 represents perfect equality and 1 represents complete inequality.

The Gini coefficient is computed based on the Lorenz curve (see *Figure 2*). The cumulative share of population is sorted by lowest to highest income on the x-axis and the cumulative share of income is represented by the y-axis. The shape of this distribution is called the Lorenz curve, through this the Gini coefficient is calculated following Equation 1. With perfect equality, cumulative population should match the cumulative income, as every person earns the same, in such a scenario, the Lorenz curve would follow the line of equality, and area A would be zero, resulting in a gini coefficient of zero. The opposite scenario, a distribution of complete inequality, the last person in the cumulative population would earn all the cumulative income. This Lorenz curve would result in area B being zero and the Gini coefficient to be 1.



Equation 1

$$G = \frac{A}{A+B}$$

Figure 2 Gini Coefficient diagram, used to compute the Gini index of inequality. Re-produced from Reidpath (2009) with permission from author.

The Gini coefficient is normally computed from household surveys representative on a national level. However, the lack of a universal system for data collection and the large costs of generating such data has resulted in limited spatial and temporal data coverage. Many developing countries do not have the resources to produce inequality estimates; resulting in no/out-dated Gini coefficients on a national level. The lack of data can be seen from figure 1, where country gini coefficients are missing, and some countries are represented by data as old as 10 years in order to produce this graphic representation of inequality. These problems makes comparison between countries difficult, since the variation in Gini coefficients could be attributed to variations in sampling techniques and temporal variations rather than variations in income distribution (Elvidge et al. 2012).

The strength of the Gini coefficient is that it includes all the members of society into one single measure. However, this is also its weakness, since it summarizes all individuals but does not explain where in the distribution this inequality occur, resulting in very different economies represented by similar Gini coefficients (Alvaredo et al. 2018)**.**

Finer spatial scaled measures of inequality are difficult to come by, as no standard database or standard measurement practice is available to guide users in their analysis. In most countries, both the spatial and temporal resolution of existing data does not allow for detailed analysis and the lack standards makes comparison between areas/datasets difficult. As mentioned previously, current trends in inequality are largely attributed to variation within countries. The generation of data that can show the spatial pattern of income inequality on fine scales is therefore an important step in finding causes and solutions and working towards a more equitable world.

2.2 NTL and its potential for capturing SE variables

Satellites provide globally consistent and repeatable observations of natural phenomenon and human activities (Elvidge et al. 2009) and are the foundation for large-scale and high-frequency studies of humans and our impact on the world (Zhaoxin et al. 2017). This advantage has been used for gathering data throughout many fields such as environmental monitoring of climate, vegetation etc. Satellite imagery provides understanding of environmental factors and human influence from a spatial point of view. However, in order to further understand the effect of human activity, similarly useful spatial data is needed for societal factors such as SE variables. Satellite imagery is the key to combining natural sciences and SE research to explore spatial patterns of human activity and its influence on a global scale. The emerging of NTL imagery provides this opportunity. As seen from *Figure 3,* the nightlight emissions captured by these images originate mostly from public streetlight, and other illuminated human infrastructure. Therefore, the spatial distribution of lights seen from space can serve as a proxy for human activity and has shown to be a useful substitute for SE indicators when the original data is missing or of poor quality, as is the case in many developing countries (Zhaoxin et al. 2017; Proville et al. 2017; Bennett and Smith 2017).

2.2.1 NTL imagery

Two main datasets of global night-time lights are available to the scientific community: the Defence Meteorological Satellite Program Operational Line Scan System (DMSP-OLS) and its successor: the Visible Infrared Imaging Radiometer Suite (VIIRS) aboard the Suomi National Polar-orbiting Partnership (Suomi NPP), a composite of this data is shown in *Figure 3*. A comparison of the two datasets is summarized in Table 1.

The DMSP-OLS is available through annual composites released by the National Oceanic and Atmospheric Administration (NOAA) National Center of Environmental Information (NCEI; formerly the National Geophysical Data Center). The longer time-series (1992-2013) makes this an attractive dataset for temporal analysis. Furthermore, a long range of studies have explored the accuracy of the data and its value as a proxy for SE data (see Table 2). Several limitations have been noted with the DMSP-OLS data. The coarse spatial, radiometric and spectral resolution cannot capture variations in highly saturated areas and fails to capture smaller light sources in areas where light is scarce. This results in up to a third of all urban areas saturated at 63 digital number (DN), the maximum value registered by the sensor, and a large amount of pixels with a DN of 0, the minimum value registered by the sensor (Bennett and Smith 2017). Furthermore, the satellite does not have on-board calibration and records light emissions in digital numbers that need to be converted, and therefore the data needs to be corrected for inter-annual satellite differences through inter-calibration or year fixed effects before any comparison between years can be made (Elvidge et al. 2009; Henderson et al. 2012).

Figure 3 VIIRS image of Europe at night, 2016. Image produced using NASA Worldview snapshot tool (https://wvs.earthdata.nasa.gov/)

The VIIRS data, succeeding the DMSP-OLS provides some solutions to the above-mentioned issues. The VIIRS has a spatial resolution that is 45 times higher at nadir and 88 times at scan edge compared to the DMSP-OLS. Similarly, its radiometric resolution is 256 times finer and it is more sensitive to radiance(Miller et al. 2013). The VIIRs provides on-board calibration and reports radiance in W/cm2/sr. This radiometric resolution and sensitivity to radiance allows for detection of weak sources of light, as well as detection of variations within highly saturated areas such as cities (Bennett and Smith 2017).

However, the use of VIIRS data within the scientific community is still only at a trial stage. Due to the vast amount of literature covering DMSP-OLS some studies still prefer the older data, as calibration methods are more established and comparisons between studies are easier (Bennett and Smith 2017). However, the increased spectral and spatial resolution of the VIIRS data, is a big opportunity and has shown better relationships with SE data (Elvidge et al. 2013; Zhao et al. 2017a; Zhaoxin et al. 2017). Therefore, the use of VIIRS will increase in the future. Furthermore, the temporal resolution of VIIRS, with both monthly and annual composites available, provides opportunities to study the inter-annual variation in light emissions, and thereby societal dynamics (Proville et al. 2017; Bennett and Smith 2017).

Table 1: *Details on the two main night-time light datasets and their satellite sensors. Table adapted from Miller et al. (2013) and Bennett and Smith (2017).*

Attribute	**DMSP-OLS**	**VIIRS**
Time series available	1992-2013	2011-current
Temporal composites available	Annual	Monthly and annual
Spatial resolution	5-7km	0.7-1km
Radiometric resolution	6 bit	14 bit
Radiometric calibration	none	on-board
Areas of saturation	Urban cores	None
Stray light artefacts	Uncorrected	Corrected
Maximum light captured	$3.17 \times 10{-}7$ W·cm−2·sr−1·µm−1	0.02 W·cm−2·sr−1·µm−1
Minimum light captured	$1.54 \times 10{-}9$ W·cm−2·sr−1·µm−1	$3 \times 10{-}9$ W·cm−2·sr−1·µm−1

2.2.2 Previous studies

The relationship between NTL and SE-variables has been established through a large number of scientific studies, and a sample of these studies has been summarized in Table 2. Studies have noted the usefulness of NTL-imagery to capture SE variables. In some cases, estimates based on NTL even outperform outputs from national statistics (Coscieme et al. 2017). Henderson et al. (2012) show that NTL is especially useful in developing countries, where national estimates of SE variables are unavailable or unreliable.

Studies investigating NTL as a proxy for human activity have generally focused on economic factors, population dynamics, consumption patterns and development. Analysis has been made on various scales (see section 2.2.3), most commonly using regression analysis to establish log-linear relationships between light and the SE-variables, and then using these relationships to make predictions on a finer scale.

2.2.2.1 NTL and economy

Relationships between NTL and economic variables have been established through correlating light variables with GDP at various spatial scales (Henderson et al. 2012; Chen and Nordhaus 2011; Ebener et al. 2005; Keola et al. 2015), with wages/income levels (Mellander et al. 2015; Chaturvedi 2011), and informal economy (Tanaka and Keola 2016; Sutton and Costanza 2002). Furthermore, NTL has been used to estimate consumption patterns associated with greenhouse gas emissions and electrification (Doll et al. 2000; Doll and Pachauri 2010; Wang et al. 2018).

The predicted economic outputs generally fit well with validation data. However, Ebener et al. (2005) argues that NTL should only be a qualitative measure of GDP, and that quantitative estimates from such models are not good enough to capture precise numbers of GDP, especially on scales smaller than a national level. Though NTL works well as a proxy for economic estimates in developing countries, when traditional statistics are unavailable or of low quality, NTL has little added value in areas with well-functioning statistical organisations (Chen and Nordhaus 2011).

There is a risk of over/underestimations of GDP from luminosity. In developing countries, the agricultural sector is large part of the GDP. Primarily in rural regions the value of the agricultural sector is underestimated due to its low light emissions (Doll et al. 2006). Similarly, in developed countries, the service sector might generate most of the GDP, however, these areas might not generate as much light as heavy industry. This could also result in an underestimation of GDP, due to the difference in light emission per GDP generated by the different sectors. Overestimation could result from light generating infrastructure that is largely abandoned, such as industrial areas no longer used or the famous "ghost cities" of China (Zhaoxin et al. 2017). Including land cover/land use data to account for these differences have shown to improve GDP estimations from NTL imagery (Keola et al. 2015; Lessmann and Seidel 2017). Furthermore, shadow economies, referring to economic activities not registered in the official statistics, might generate light but not be represented in the validation data. Examples of this could be small roadside shops and exchange of goods in the black market. This unregistered activity is a large part of the economic sector in developing countries. NTL imagery can be an alternative measurement of wealth which also accounts for these missing parts or traditional economic statistics (Ghosh et al. 2013).

2.2.2.2 *NTL, population dynamics and development*

The relationship with population dynamics and urban development has been established through correlating NTL with population and population density. Generally, these relationships are less robust than that of GDP (Mellander et al. 2015; Zhou et al. 2015). Many areas with high densities are saturated in the DMSP-OLS data and therefore do not capture the population properly (Doll and Pachauri 2010; Proville et al. 2017; Sun et al. 2017). The problem of saturation can be solved by using the VIIRS data, and future modelling of population dynamics will benefit from this. Furthermore, most studies suggest that the relationship between population and NTL can be improved by including ancillary information in the analysis, such as land cover (Sun et al. 2017).

A growing number of articles focus on the relationship between various development indicators and NTL imagery. Poverty indices have been developed using population and luminosity (Elvidge et al. 2007; Elvidge et al. 2012; Elvidge et al. 2009; Ghosh et al. 2013; Noor et al. 2008). More recently, Jean et al. (2016) combined night-time images with traditional daytime images to detect poverty through machine learning algorithms. NTL has also been correlated with established measures of sustainable development, such as the Human Development Index (Elvidge et al. 2012). The sum of lights (SOL) can be associated with sustainable development indicators, and used to predict regional sustainable development through time-series based regression (Li et al. 2018). On a local scale, Bruederle and Hodler (2018b) show that NTL is positively correlated with indicators of human development.

Finally, the disparity in luminosity between regions has been studied on various scales in order to distinguish spatial patterns of development. This will be expanded further in section 2.3.

2.2.3 The importance of scale

Relating a SE-variable to NTL-imagery requires consideration of spatial scale. The spatial scale of SE-variables is the extent/area represented by the variable. Several studies have shown how the strength of the relationships as well as the predicted values change with spatial scale. This is an example of the modifiable area unit problem (MAUP). This general geographical concept occurs when different result are observed from the same data depending on the spatial resolution used (scale effect) or depending on the shape of the units used (zoning effect) (Doll et al. 2006). These effects can lead to ecological fallacies, where erroneous inferences are made about fine scale dynamics based on coarser scaled data (Doll et al. 2004).

The importance of spatial scale has been noted by several studies in the field (Xu et al. 2015; Ma et al. 2014; Zhou et al. 2015; Ebener et al. 2005). Zhaoxin et al. (2017) found that the estimation of GDP depends on scale of study area and land cover/industry types. They showed that the fit between GDP and NTL decreases when going from province to city scale. However, Zhaoxin et al. note that the decrease in fit is smaller when using VIIRS data, compared to DMSP-OLS. Analysing fit on a local/micro-scale should be enhanced by using the VIIRS data with a high spatial resolution. However, finer scaled analysis is more likely to be affected by underlying factors that may be smoothed out in coarser resolution analysis. For example, when modelling GDP, factors such as transportation network and existing economic sectors will affect the results. Areas with one dominating sector often appear as outliers in the analysis (Doll et al. 2004) and features relevant to economic outputs might only detectable in fine scaled imagery (Jean et al. 2016).

2.3 How does income inequality fit in this framework?

Economic inequality within society can be said to tie together many of the SE variables previously studied through NTL. Because it is a measure of wage distribution within a population, it is closely related to the economic output of an area, population density and distribution of economic opportunities represented by development and urbanization. However, as seen from section 2.1, the measurement of inequality can be elusive and difficult to capture. Furthermore, the necessary data for computing measures of inequality are not always available at the desired spatial scale. Thus, alternative data sources are needed to compensate national statistical outputs.

NTL provides an opportunity to examine spatial patterns of SE-variables and has thereby enhanced research into differences between regions and inequality of various SE factors. Light gradients between neighbouring areas can be used as an estimate of disparities in the economy (Coscieme et al. 2017).

As in previous sections, it is important to distinguish between two types of economic inequality: the disparity between regions and the disparity between individuals.

Several studies have examined regional inequality (Lessmann and Seidel 2017; Rongwei et al. 2018; Shasha et al. 2015; Xu et al. 2015; Zhou et al. 2015). These studies focus on differences in economic output between regions in a country, mostly focusing on China, which experienced quick and skewed economic development in recent years. These studies show a very distinct pattern, with larger economic outputs from coastal regions with important commercial centers (Rongwei et al. 2018; Zhou et al. 2015). Apart from economic output, Noor et al. (2008) suggests that the distribution of light can be a good proxy for other aspects of inequality such as access to health because the inequality of lights reflects the road distribution and access to electricity. Furthermore, Bruederle and Hodler (2018a), shows that NTL can be related to inequality in education and health on a micro-scale.

Globally, Elvidge et al. (2012) attempted to create a gridded dataset of economic inequality. In this study, the Gini coefficient was calculated based on the distribution of light in the world population resulting in the night light development index (NLDI). It measures how light, represented by the DMSP-OLS data, is spatially distributed relative to the distribution of the global population. However as discussed in the paper, the measure correlated better with other development indexes rather than the traditional Gini coefficient. The weak relationship between NLDI and Gini shows that brightly lit areas in DMSP-OLS images correspond to areas of high population density having a varied mix of individual income levels and not only to lower population densities of few wealthy people. Furthermore, due to the global scale of the study, country specific factors accounting for differences in cultural use of light as well as economic sectors were not included.

Modelling inequality over different countries or regions is problematic due to differences in governance and cultural use of light. The emissions of NTL originate mostly from public streetlight. Different countries/regions have varying policies prioritizing or regulating public streetlight. Thus, the light consumption patterns do not only reflect economic differences between regions, but also includes differences due to local conditions.

A more recent study by Lessmann and Seidel (2017) attempts to correct some of the previously mentioned issues and create a global dataset of regional inequality. In this study, GDP was modelled through subnational GDP per capita and sum DMSP-OLS light. The model also included other factors trying to account for problems with saturation and differences between countries and regions. Inequality within each country was then calculated from the modelled income of all the regions within. The methodology used succeeded in capturing inequality on a national level, based on the differences between the regions in within a country. However, inequality measures on a smaller scale was not produced.

On a micro-scale, Mellander et al. (2015) investigated the relationship between light and SE-variables in Sweden at a scale of 0.5x1 km resolution. They found that the light correlated well with population, establishment density and wages. However, as with GDP, the correlation with wages was often overestimated in urban areas and underestimated in rural areas. This detailed analysis was only possible due to the available fine scaled data on SE-variables in Sweden. Expanding this fine scaled relationship is not possible for most areas of the world, specially the developing countries. And as argued by Chen and Nordhaus (2011), these are exactly the areas were NTL-developed data would be superior to national statistics.

Finally, the NTL emissions mostly represent the consumption of the wealthy in society. Chaturvedi (2011) investigated the relationship between households in different income brackets and light emissions in India. They conclude that the distribution of income plays a larger role than the actual wealth of a region in determining it luminosity level. From their analysis they saw that luminosity was a better predictor of number of households in the higher income bracket than in the lower income bracket. The NTL is therefore better suited as a proxy for wealth in the richer parts of society. Furthermore, because some sources of wealth such as agriculture are not captured sufficiently by the NTL imagery, there is a risk of overestimating poor households.

2.3.1 Concluding remarks

From this literature review we can conclude that NTL has been established as a good proxy for several SE variables through previous research. This is especially valuable in areas with poor or non-existent data, such as developing countries. However recent studies note that NTL data by itself is not enough. Including other variables accounting for land-use, economic sectors and country specific differences can improve this relationship. Furthermore, using finer scaled data such as VIIRS should enhance the estimates of inequality on both large and small scales (Bruederle and Hodler 2018a; Coscieme et al. 2017).

In the following parts, my analysis will build upon the previous frameworks for modelling SE-variables with NTL images. Taking shortcomings of previous studies into account, I will use the VIIRS dataset, avoiding problems with saturation and increasing the sensitivity to low lit areas. Furthermore, the modelled output will be evaluated at several spatial scales, assessing the importance of scale within the framework of modelling economic inequality. Thus, this study will add to the literature suggesting NTL as a proxy for SE variables, more specifically economic inequality. The fine scale of the analysis will help understand inequality patterns within countries, making this a relevant study in relation to current trends.

Table 2*: A summary of previous studies on the relationship between night-time light data and SE variables.*

Reference	**Downscaled variables**	**Spatial scale**	**Data used**	**NTL variable**
Bruederle and Hodler (2018a)	Development	Neighbourhood	DMSP-OLS	Mean
Chen and Nordhaus (2011)	GDP	National Gridded	DMSP-OLS	Sum
Coscieme et al. (2014)	Emergy	National Subnational	DMSP-OLS	SOL
Coscieme et al. (2017)	Disparity and conflict	National Subnational	DMSP-OLS	Light per capita

Reference	**Downscaled variables**	**Spatial scale**	**Data used**	**NTL variable**
Doll et al. (2000)	Electrification rates Greenhouse gas emissions GDP	Subnational	DMSP-OLS	Lit area
Doll et al. (2006)	GDP	Grid	DMSP-OLS	Sum
Ebener et al. (2005)	GDP	National Subnational	DMSP-OLS	Mean
Elvidge et al. (2012)	Development	Subnational Gridded	DMSP-OLS	Radiance
Ghosh et al. (2013)	Well being GDP	National, Subnational and gridded	DMSP-OLS	Sum
Henderson et al. (2012)	GDP Income growth	National Subnational	DMSP-OLS	Weighted average
Jean et al. (2016)	Poverty	Neighbourhood	DMSP-OLS; Daylight images	Mean
Keola et al. (2015)	GDP	National Subnational	DMSP-OLS; landcover	Lit area
Lessmann and Seidel (2017)	GDP Regional inequality	Subnational	DMSP-OLS	Mean
Li et al. (2017)	Regional inequality	National Subnational	DMSP-OLS	SOL
Li et al. (2018)	Development	Subnational	DMSP-OLS	Sum
Ma et al. (2014)	Population density and built-up area	County, district and sub-district	VIIRS	Mean
Mellander et al. (2015)	Population Establishment density Wages	Microscale	DMSP-OLS	radiance
Noor et al. (2008)	Poverty	Subnational	DMSP-OLS	Mean
Rongwei et al. (2018)	Regional inequality	Subnational	VIIRS	Sum
Sun et al. (2017)	Population	Subnational	DMSP-OLS; landcover	Sum
Wang et al. (2018)	Consumption	Gridded	VIIRS	Radiance
Xu et al. (2015)	Regional inequality	National, Subnational metropolitan	DMSP-OLS	Radiance
Zhao et al. (2017a)	GDP	State, county, metropolitan	VIIRS	Sum
Zhou et al. (2015)	Inequality Population GDP	Subnational, Metropolitan	VIIRS	Sum

3 Materials and Method

For the analysis, NTL data was regressed against regional income. The relationship defined was then used to model income and total GDP. The modelled economic output was then used to compute economic inequality, represented through the Gini index. Finally, the modelled income and Gini coefficients were compared to three validation datasets with five different spatial scales. The following subsections describe the datasets and methods in detail. An overview of datasets used is provided in Table 3 and the overview of the work flow in Figure 4. The analysis was preformed using R.

3.1 Data

3.1.1 Night-time light data

Based on previous studies, the VIIRS NTL data was chosen for this analysis. VIIRS was chosen instead of the frequently used DMSP-OLS-OLS, due to the advantages VIIRS has over its predecessor. Comparing the two datasets, the VIIRS data boasts both higher spatial, temporal and spectral resolution and captures SE-variables better than the older DMSP-OLS dataset. Furthermore, no further calibration is needed when using the VIIRS annual composites due to their on-board sensors as well as their extended spectral range resulting in fewer zero and saturated areas (see section 2.2.1).

The night time light images were provided by the Earth Observation Group, NOAA National Center for Environmental Information (NCEI) (Elvidge et al. 2017). Two versions of the data are available, the “vcmcfg”, which exclude the data contaminated by stray lights; and the “vcmslcfg”, which corrects the contaminated data instead of excluding it. The later version was chosen because it includes more data in the high latitudes, and give a more complete spatial and temporal dataset (Rongwei et al. 2018). The annual composite produced by NOAA for the year 2015 will be used in this analysis. Compared to the monthly composites regularly released, annual composites have only been produced by NOAA for the years 2015 and 2016, at the time of writing this thesis. The annual composites have been further processed to remove outliers and background noise by NOAA, thereby attempting to filter out ephemeral lights such as fires and setting non-light pixels to zero (Elvidge et al. 2017).

The economic output will be modelled from the NTL images in order to calculate inequality. Two measures of economy will be tested: total GDP and per capita income. Because a strong relationship has been seen between GDP and light and because inequality measures are calculated based on income, therefor modelling income instead of GDP might result in higher accuracy. To model total GDP, the sum of light (SOL) will be used. A relationship between sum of light and sum of GDP, will gives us a measure of output per light which enables us to distribute the total GDP of the area to smaller spatial units based on their light emissions(Zhao et al. 2017b). The strong relationship between GDP and SOL was shown by Chen and Nordhaus (2011)

When estimating income, the average light emission is used. This variable is a logical match since the income data for each spatial polygon is also an average of income data available in the region. The average light was successfully used to model income by Lessmann and Seidel (2017)

However, modelling SE-variables with only light variables, often results in overestimation in rural areas and underestimations in urban areas. As suggested by Zhao et al. (2017b), accounting for population within the light variable might help reduce such errors. Lit population (LP), computed by multiplying the VIIRS images with a population dataset, can help differentiate between a low-lit pixel with low population density and a low-lit pixel with high population density, but limited access to electricity. This is a differentiation that could be important when modelling GDP per capita and economic inequality on a finer scale.

3.1.2 Socio-economic data

Gross Domestic Product (GDP) was used to model wealth from NTL. The dataset produced by Kummu et al. (2018) is recently compiled includes subnational economic annual statistics for 1990-2015. From the dataset, the regional income for year 2015 was used, and a total GDP grid was computed by multiplying the regional income with the population data, following the suggested methodology in Kummu et al. (2018) but multiplied with a newer population dataset then in the on Kummu used.

Population data were used for calculating GDP from the regional incomes, as well as developing the lit population variable. The Gridded Population of the World, Version 4 (GPWv4) was used for this analysis. This dataset is provided by NASA Socioeconomic Data and Application Center (SEDAC). It provides gridded estimated of population count and population density in 1 km resolution.

SE data, including Gini coefficients and income, for validation purposes was obtained for five levels: national, subnational and three local levels with successively smaller regions. The national economic data was obtained through the World Bank. National level Gini coefficients were retrieved from the World Income Inequality Database (WIID4) which is a compilation of Gini coefficients reported from various institutions on a national level. The most recent entry for each country was selected, and if several existed then the entry with highest quality (a variable provided by the database) was chosen.

Subnational Gini coefficients where obtained from the Regional Wellbeing dataset (OECD 2014). This dataset includes income and Gini coefficients before/after taxes for 402 regions in the OECD countries. This analysis used the Gini coefficient before taxes as the inequality measure. Although not covering the entire world, this dataset has a good temporal coverage. Data for 2015 matching the NTL images was used.

Local level validation data was obtained from the Poverty Mapping Project: Small Area Estimates of Poverty and Inequality (Center for International Earth Science Information Network - CIESIN - Columbia University 2005). This dataset contains poverty, inequality and related measures for subnational administrative units throughout twenty countries in Africa, Asia, Europe, North America, and South America. Gini coefficients are divided into total, rural and urban (the total will be used in this analysis). No income data was included in this dataset. For each country, three levels administrative units are available. Since this dataset represents the smallest area, for which validation data could be found, this represents the local level analysis.

***Table 3**: Overview of datasets used for analysis*

Dataset	***Provider***	***Resolution***	***Description***	***Accessed through***
VIIRS	NASA *Elvidge et al. (2017)*	2015 0.7 km	2015 annual composite "vcm-orm-ntl" (VIIRS Cloud Mask - Outlier Removed - Nighttime Lights)	https://ngdc.noaa.gov/eog/viirs/download_dnb_composites.html
Population	*Center for International Earth Science Information Network - CIESIN - Columbia University (2018)*	2015 1 km	Population count	http://sedac.ciesin.columbia.edu/data/collection/gpw-v4
GDP	Kummu et al. (2018)	2015 10 km	National and subnational data compiled global dataset	https://www.nature.com/articles/sdata20184
National SE-variables	World bank	2015 Countries: 247	Income per capita	https://data.worldbank.org/
National Gini	UNU-WIDER (2017)	1991-2015 Countries 190	Compilation of national Gini coefficients. Latest Gini coefficient was used for each country	https://www.wider.unu.edu/project/wiid-world-income-inequality-database

Dataset	*Provider*	*Resolution*	*Description*	*Accessed through*
Regional SE-data: regional Wellbeing dataset	OECD (2014)	2015 Countries: 36	GDP per capita and Gini coefficients for all sub-regions in the OECD database. Covers most of Europe, USA, Canada, Mexico, Chile, Japan, Australia and Japan	https://stats.oecd.org/Index.aspx?DataSetCode=RWB
Regional SE-data: **Small area estimates**	Center for International Earth Science Information Network - CIESIN - Columbia University (2005)	1991-2002 Countries: 20	Subnational inequality data. Gini coefficients based on household surveys	http://sedac.ciesin.columbia.edu/data/set/povmap-small-area-estimates-poverty-inequality/data-download

3.2 Methods

To accomplish the aim, a model was developed based on previous literature. Lessmann and Seidel (2017) argue that estimate of regional inequality is better when computed with modelled GDP rather than light. This is due to the log linear relationship between light and economy, implying that the relationship between light and income/GDP depends on the light level and that the absolute differences in luminosity between two regions does not indicate the same difference between the income of two other regions that are richer/poorer (Lessmann and Seidel 2017). Furthermore, previous literature suggests that light alone is not enough to predict income (Chen and Nordhaus 2011). Computing inequality indices from modelled GDP rather than pure light data gives us the opportunity to include more explanatory variables.

Based on this reasoning, the first step of the analysis is to develop a model to predict income. Following this the Gini coefficient was calculated from the modelled income. Finally, to evaluate the performance of the developed model, total GDP and Gini coefficients were compared to validation data. A more detailed description of the various steps follows in the sections below; an overview of the entire workflow can be seen from Figure 4.

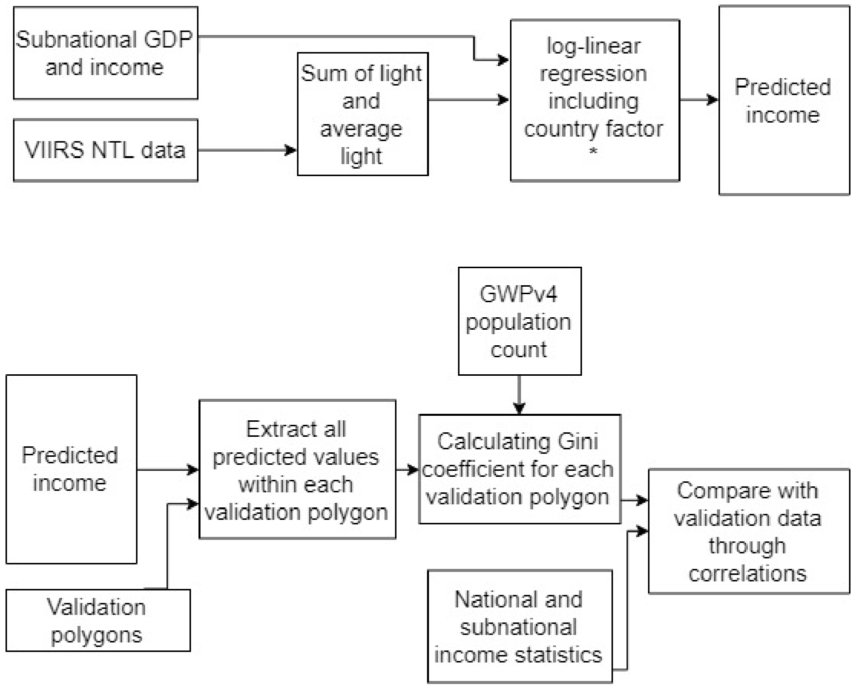


Figure 4 *Flowchart showing an overview of the methodology. Created with flowchart tools found at* [*https://www.draw.io/*](https://www.draw.io/)

3.2.1 Modelling GDP from NTL

The relationship between NTL and SE-variables developed through previous studies can be generalized into Equation 2.

Equation 2

$$\text{Wealth}_{\text{pixel}} = \frac{\text{Wealth}}{\text{Light}} * \text{light emissions}_{\text{pixel}}$$

Where the wealth attributed to an area/pixel is calculated by dividing the total wealth with the total light emission and multiplying this with the light emission from the area/pixel. Throughout this paper we are using two measures for wealth: the GDP e.g. total wealth of the area/pixel; and the income e.g. average wealth of the inhabitants in the area/pixel. Similarly, the Light variable used to model the GDP is the SOL, and the light variable used to model income is the average light emission. The light emissions come directly from the pixel values of the VIIRS images (Zhao et al. 2017b). An overview of the models created can be found in *Table 4*.

Both the GDP and the NTL data are strongly positively skewed. To correct for this heteroscedacity, a logarithmic transformation of both datasets was performed before modelling. The log-transformed data give higher weights to the variation in the data in the lower values and less weight to outlying peaks in both light and GDP (Bruederle and Hodler 2018a). Following previous literature, 0.01 was added to the NTL dataset before taking the logarithm, in order to not miss all the pixels with zero light (Lessmann and Seidel 2017; Bennett and Smith 2017; Ebener et al. 2005). This can also be justified by the logical assumption that even though the NTL of a cell is zero, if the gridded GDP data shows some economic output from the cell due to the presence of population, the cell should be included in the analysis.

To account for the differences between countries due to culture and regulation, country fixed effects were included in the model. A fixed parameter for each country is included in the model in order to account for the differences between countries. Accounting for the variance between and thereby removing it from the picture makes it easier to distinguish the variation within countries, which is our main focus when estimating inequality within countries.

Previous literature suggests that accounting for population density within the model would allow for better capturing of SE-variables. Lit population (LP), computed by multiplying NTL with population count has been used in some cases instead of the NTL values. (Zhao et al. 2017b). This variable will also be tested in the models, since the difference between a low-lit pixel with high population and a low-lit pixel with no inhabitants could be important for modelling the distribution of wealth.

Finally some studies suggest that NTL is insufficient for modelling SE-data (Jean et al. 2016; Sun et al. 2017). Other ancillary data could be included to enhance the relationship between NTL and inequality. However, including variables such as percentage of land covered by agriculture and percentage area lit, did not improve the model very much. These variables were therefore omitted from the analysis after the initial exploratory stages.

Eight different models were tested using parameters suggested above: combining the independent variables (Sum of light/SOL and lit population/LP) and the dependent variables (GDP and income) and the addition of country fixed effects. An overview of final models created is shown in *Table 4*.

__Table 4__ An overview of the final models produced for disaggregating wealth with night-time light imagery. C (country fixed effects); LP (lit population); SOL (sum of light)

Model 1	$\text{Income}_{\text{pixel}} = \dfrac{\textbf{Income of area}}{\text{Mean light of area}} * \text{light}_{\text{pixel}}$
Model 2	$\text{Income}_{\text{pixel}} = \dfrac{\text{Income of area}}{\text{Mean light of area}} * \text{light}_{\text{pixel}} + \text{C}$
Model 3	$\text{GDP}_{\text{pixel}} = \dfrac{\text{GDP of area}}{\text{SOL of area}} * \text{light}_{\text{pixel}}$
Model 4	$\text{GDP}_{\text{pixel}} = \dfrac{\text{GDP of area}}{\text{SOL of area}} * \text{light}_{\text{pixel}} + \text{C}$
Model 5	$\text{Income}_{\text{pixel}} = \dfrac{\text{Income of area}}{\text{Mean LP of area}} * \text{light}_{\text{pixel}}$
Model 6	$\text{Income}_{\text{pixel}} = \dfrac{\text{Income of area}}{\text{Mean LP of area}} * \text{light}_{\text{pixel}} + \text{C}$
Model 7	$GDP_{pixel} = \dfrac{GDP\ of\ area}{Sum\ LP\ of\ area} * light_{pixel}$
Model 8	$\text{GDP}_{\text{pixel}} = \dfrac{\text{GDP of area}}{\text{Sum LP of area}} * \text{light}_{\text{pixel}} + \text{C}$

3.2.2 Economic inequality estimates

Gini coefficients were computed from the predicted GDP and income, following the methods outlined by Zhou (2015). This was done using R, following theory described in section 2.1.3.

3.2.3 Validation of modelled data

Validation of the model performance was done through correlations between predicted and observed validation data. Correlations were performed by groups, representing the spatial scale of the validation polygons. The total light, GDP estimates and the estimated Gini coefficients were validated in this manner.

3.2.4 Spatial scale of the analysis

In order to investigate the influence of spatial scale on predicting inequality, model outputs were correlated with three validation datasets featuring five different spatial scales. These scales are represented by five levels: Level 1 represents national estimate and levels 2-5 represent subnational regions. Data and spatial polygons for level 2 come from the OECD regional statistics; the spatial scale can be said to be large regions within each OECD country. Finally, level 3-5 are smaller regions within selected countries, represented in the small area estimates. Where level 3 are the largest and level 5 are the smallest. The difference in spatial scale can be noted by the average polygon size, see Table 5.

Due to the differences of the datasets, the countries included on various scales area not the same. This is a limitation to the study as the comparison of the results at different levels cannot be done country by country.

Table 5: *Overview of spatial levels of the analysis. Datasets all include Gini coefficients as well as other SE-variables*

Level **(represented by dataset)**	**Number of polygons in data**	**Average area of polygon**
1 (WIID4)	190	596,758 km^2
2 (OECD regional wellbeing)	402	88,724 km^2
3 (Small area estimates level 1)	178	13,988 km^2
4 (Small area estimates level 2)	2315	1,361 km^2
5 (Small area estimates level 3)	3813	783 km^2

4 Results

4.1 Modelling GDP

Based on previous literature eight models were developed for testing (*Table 4*). Out of these, model 4 produced the best fit with the steepest slope, shown by the highest R^2 (see

Table 6 and *Figure 5*). Suggesting that modelling GDP is more accurate using SOL as the light variable and including country fixed effects. Looking at the rest of the models, it is clear from the differences in the R^2 that adding country fixed effects to the linear model improves the model's accuracy. Furthermore, total GDP generates a stronger relationship than income. All linear regressions preformed were statistically significant, with very small P-values.

Comparing the R^2 we determined that lit population did not model GDP or income better than pure light variables, therefore model 5-8 were not pursued further.

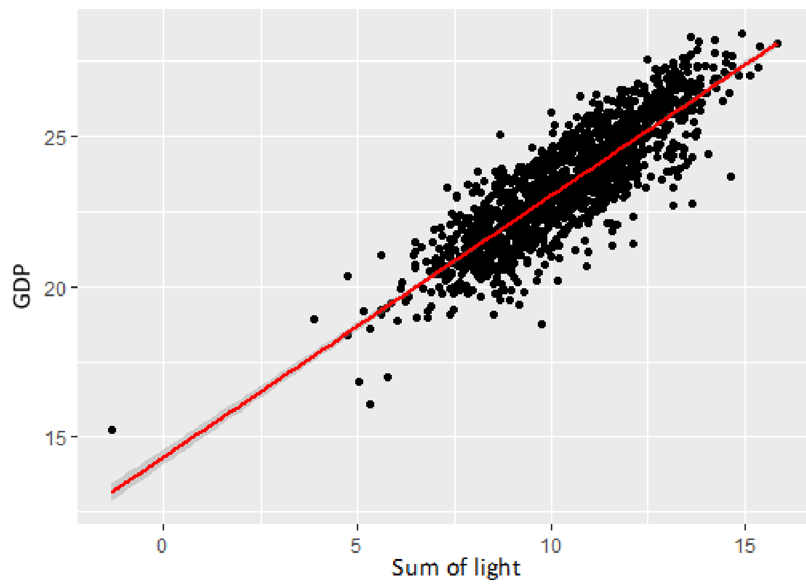


***Figure 5**: Scatterplot of training data. Red line represents linear model 4. Note both axes have been logged transformed*

Table 6 *overview of model parameters. Linear models produced in R based Income data and VIIRS NTL. Stars represent significance levels. SOL = sum of light, GDP = gross domestic product, LP = lit population*

	Model 1	**Model 2**	**Model 3**	**Model 4**
Variables:	log (Income) log(avgLight)	log (Income) log(avgLight)	log (total GDP) log (SOL)	log (total GDP) log (SOL)
Slope	0.31 **	0.15**	0.87 **	0.92 **
Intercept	9.79 **	9.91**	14.34**	13.91 **
R^2	0.27	0.83	0.73	0.92
Adjusted R^2	0.27	0.80	0.73	0.90
Country fixed effects	no	yes	no	Yes

	Model 5	**Model 6**	**Model 7**	**Model 8**
Dependent variable:	log (Income) log (LP)	log (Income) log (LP)	log (total GDP) log (LP)	log (total GDP) log (LP)
Slope	0.09 **	0.08 **	0.55 **	0.46 **
Insterscept	8.08 **	8.96 **	14.53 **	14.72 **
R^2	0.06	0.82	0.73	0.89
Adjusted R^2	0.06	0.79	0.73	0.87
Country effects	no	yes	no	Yes
*Signif. codes: 0.001 '***' 0.01 '**' 0.05 '*'*				

The predicted GDP is strongly correlated to the validation data for all spatial levels (see Figure 6). The national level (1) showed the strongest correlations for all models. Level 2, the subnational OECD regions showed the weakest correlations with the validation GDP.

Model 4 showed the strongest correlations throughout all levels. This confirms that model 4 is the best of all our models to predict GDP, similarly concluded from

Table 6.

4.2 Predicted inequality

Predicted Gini coefficients did not correlate with validation data as well as GDP. As seen from Figure 6, the strongest correlations were found on the smallest spatial scale, level 5. Weak negative correlations were found for levels 1 and 4, and weak positive correlations were found for levels 2 and 5.

Models 1 and 2 resulted in more significant correlations, as they modelled income directly. In contrast to this, models 3 and 4 only produced significant correlations at level 5.

Finally, significant positive correlations between Gini coefficients based on NTL were found on level 3-5. These gini coefficients based on light are more strongly associated with the validation data, than the Gini coefficients based on modelled wealth are.

The following section will explore the error of the predicted Gini coefficients at all the spatial levels through scatterplots and maps showing spatial patterns of the errors. The following results are based on Gini coefficient estimations from model 2, as most significant correlations were seen for this model (see Figure 6).

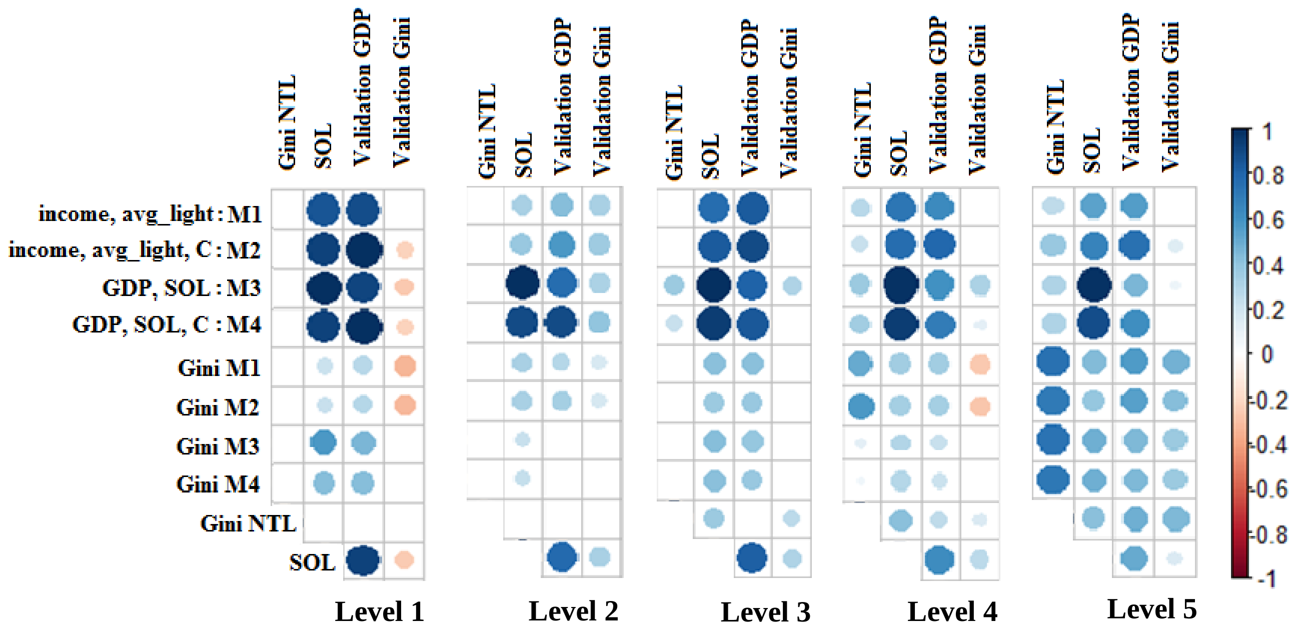


Figure 6: *Correlation matrices for each spatial level of analysis. Color of circle represents strength of correlation and size of circles represents significance level. All insignificant correlations have been left blank. NTL = Nighttime light, SOL = Sum of light, M = model; GDP = gross domestic product*

4.3 Spatial scale

Gini coefficients on all spatial levels were underestimated. This can be seen by plotting the predicted and the validation Gini coefficients in a scatterplot (Figure 7 and Figure 8). For all plots, the point distribution was mainly beneath the 1:1 line, implying that the predicted values were smaller than the observed values.

At the national level, the distribution is slightly negatively skewed (Figure 7). This coincides with the weak negative correlations between predicted and validation Ginis found at this level (Figure 6.)

The distribution at level 2 is different from the rest. The spread of validation Ginis is much smaller than for the other spatial scales, resulting in the absence of the 1:1 line from the plotted area. Furthermore, this distribution shows an artefact from the validation dataset, with a large amount of regions having an observed Gini coefficient of 0.43, while varying in predicted Gini.

At the smaller subnational levels (3-5), the accuracy of the predicted Ginis vary depending on the country, as seen by the clustering of points based on which country the regions are in (Figure 8). Spatial levels 3 and 4 seem to follow the same distribution, although a larger dataset is presented at spatial level 4. Spatial level 5, representing the smallest polygons has a different distribution from the other two. In this distribution the clustering of countries is not as pronounced as with previous levels.

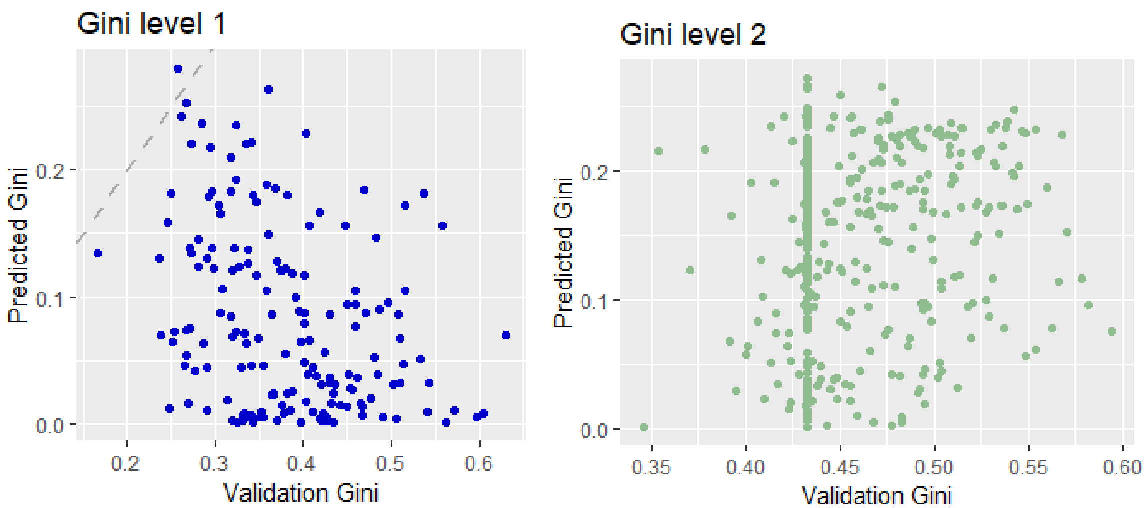


Figure 7: *Scatter plots between predicted and actual Gini coefficients. Spatial level 1 reresent national Gini coefficients with actual values from UNU-WIDER (2017). Spatial level 2 represents subnational Gini coefficients with actual values from OECD (2014). Grey line shows the 1:1 line*

Figure 8: *Scatter plots between predicted and actual Gini coefficients at spatial level 3-5 representing larger (3) to smaller (5) administrative boundaries. Validation Gini coefficients from Center for*

International Earth Science Information Network - CIESIN - Columbia University (2005). Colours represent different countries in the dataset. Grey line shows the 1:1 line

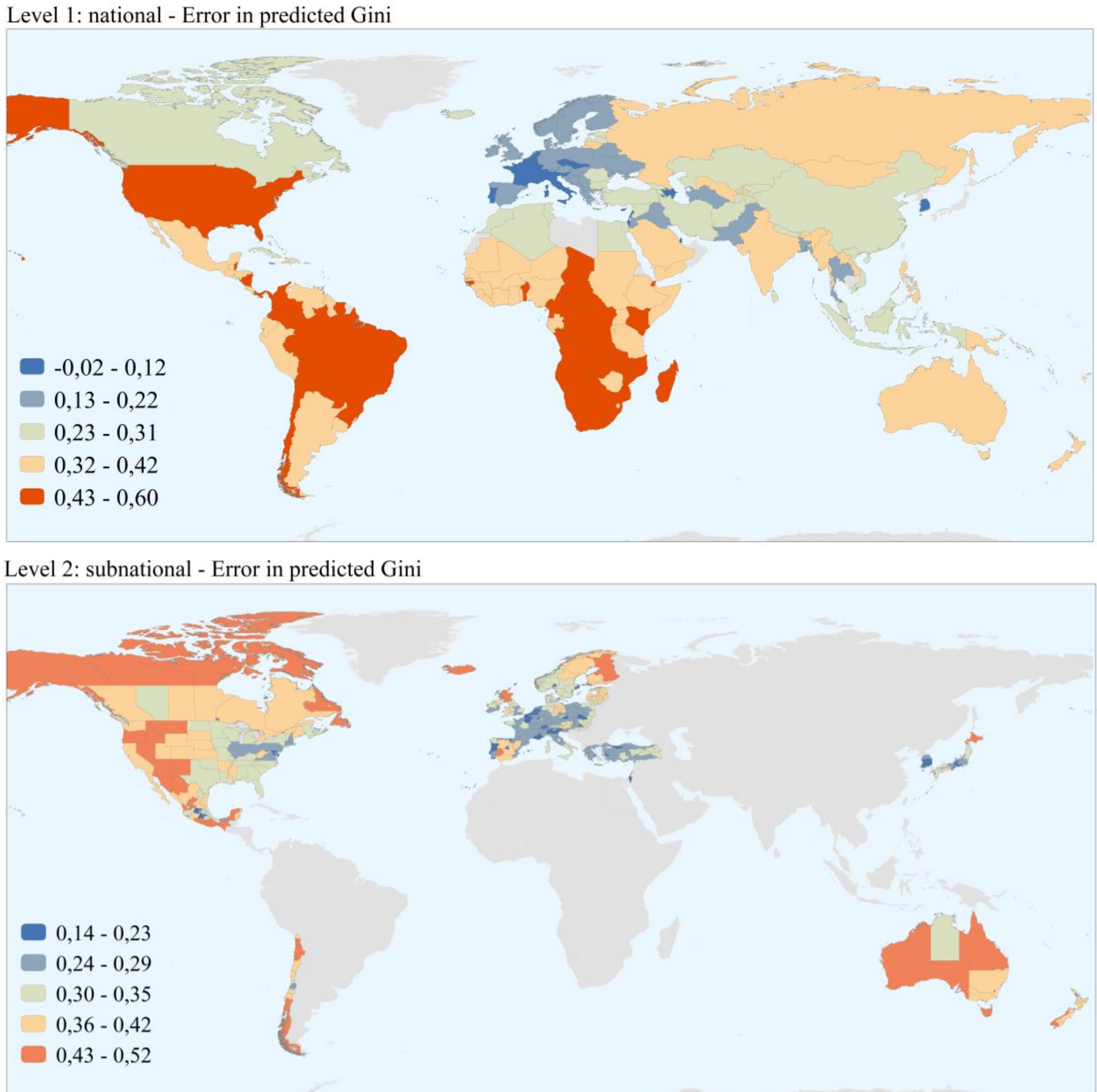


Figure 9: *The spatial distribution of error in predicted Gini coefficients for level 1 (national scale gini coefficients) and level 2 (subnational scale gini coefficients in the OECD countries). The error refers to the difference between predicted and validation gini coefficients, thereby how much our model underestimated inequality. Grey represents area with no data.*

Level 4: subnational - Error in predicted Gini

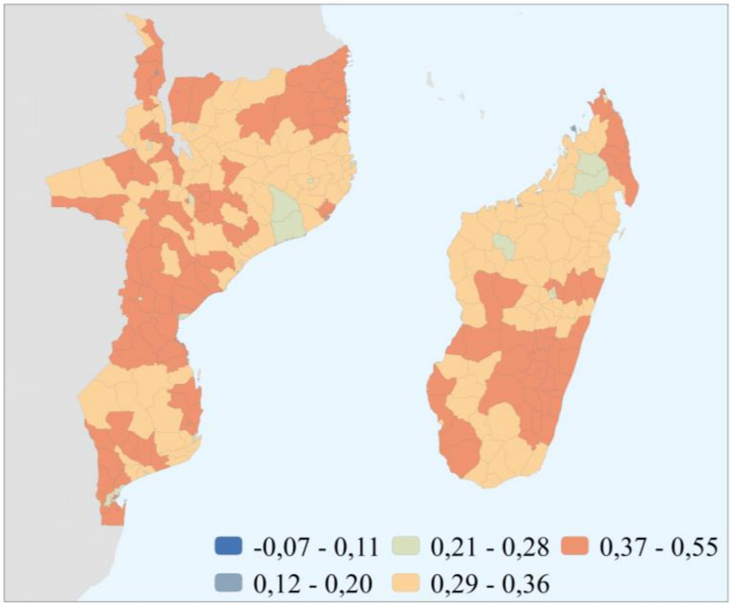


Table 7: *The mean absolute error of predicted Gini coefficients for each spatial level.*

Level	Mean Absolute Error
1	0.30
2	0.32
3	0.19
4	0.21
5	0.33

Figure 10: *The spatial distribution of error in predicted Gini coefficients for level 4 (medium sized administrative boundaries from the dataset by Center for International Earth Science Information Network - CIESIN - Columbia University (2005). An excerpt of the data, covering small area estimates of inequality for Mozambique, Malawi and Madagascar. The error refers to the difference between predicted and validation gini coefficients, thereby how much our model underestimated inequality. Grey represents area with no data.*

To further investigate the patterns in underestimation of the Gini coefficients, the errors were mapped for level 1, 2 and 4, and the mean absolute error (MAE) was calculated for each level.

The National level (1) had medium MAE (see *Table 7*) and contained largest deviation between predicted and observed Ginis. This spatial level shows clear patterns of error. The larger errors seem to be clustered around specific regions, such as sub-Saharan Africa, Latin America etc. (see *Figure 9*). Smaller errors are clustered mostly in Europe

Subnational level 2 had the second largest MAE (see *Table 7*). Smaller errors are mainly located in Europe, South Korea and the eastern coast of the United States (see *Figure 9*). Largest errors are mainly located at higher/lower latitudes.

The smaller subnational levels 3-5 show an increasing MAE with decreasing polygon size, with level 5 having the largest MAE of all the spatial levels. The spatial distribution of the errors for three countries (Mozambique, Malawi and Madagascar) at level 4 is an example of the distribution throughout the dataset (see *Figure 10*). However this dataset included different countries at the various levels, so a country wise comparison between the levels is difficult.

5 Discussion

Through the background section, it is clear that there is a potential for disaggregating SE-variables using NTL images. The literature review gave an overview of the field and evaluated the potential of modelling economic inequality with the established methodology, thereby answering research objective 1. This study attempted to implement previously mentioned concepts to model Gini coefficients for five different spatial scales. Firstly, modelling GDP from NTL data and then calculating the economic inequality from the predicted income. Results from this part of the analysis are discussed in following sections.

5.1 Modelling GDP

This study has shown that NTL is useful in predicting total GDP as well as income on both a national and subnational levels. This agrees with previous studies within the field (Chen and Nordhaus 2011; Henderson et al. 2012). Almost all the models generate performed well, seen from the R^2 values, implying that the models are able to capture the variations in the data. Model 4 showed the best fit of all, and correlations between predicted GDP and validation GDP also confirms that model 4 is the best model out of the eight models tested in this study for predicting economic output.

Models 1 and 5 did not produce a good fit, this can be seen from the low R^2 (see

Table 6). Models 5-8 used LP instead of a pure light variable. As suggested by Zhao et al. (2017b), using the population weighted LP variable produced well-fitting models. However, the models did not perform better than the initial NTL models, and were therefore omitted from further analysis.

Similar to Ebener et al. (2005), this study found that the total GDP has a stronger relationship with light emissions than per capita income, as indicated by the larger R^2 for models predicting the total GDP as well as the stronger correlation between predicted and validation data (see *Figure 6*). However, when including country fixed effects this difference between the model fits becomes much smaller. The addition of country fixed effects has a larger impact on modelling income than modelling GDP. This could be because the differences between countries total GDP is largely affected by the size of the population and this overshadows other country specific effects. Income per capita is not affected by the population size, and therefore effects of other country depending factors are dominating. Furthermore, since the country effects are computed from all the regions within a country, all with different population sizes, standardizing the economic output per capita results in a much better estimate of the country effects. The standardized economic output is not the case for GDP, this is also implied by the p-values of the country fixed effects. The significance of the effects are much larger for model 2 (income) than for model 4 (GDP).

5.2 Predicted inequality

Since the spatially distributed income was modelled successfully, the predicted income grid was used to calculate Gini coefficients representing the disparity between all the cell values within a validation polygon. Following this, predicted inequality was correlated with validation data. The correlations were grouped based on the validation datasets reflecting five different spatial scales.

The resulting correlation coefficients show a mixed picture (see *Figure 6*). Predictions from model 1 and 2 correlated better with the validation dataset, seen by the strength and the significance level of the correlations. This means that modelling income directly from NTL results in better estimates of inequality, in comparison to the two step procedure of modelling GDP from NTL and then converting the modelled GDP to income. Note that in contrast to this, NTL has a stronger association with GDP than with income.

Lessmann and Seidel (2017) claimed that Gini based on modelled income was a better estimate than Gini based on the distribution of light within the population. This study did not find convincing evidence of this. For level 1 and 2 the Gini based on NTL did not result in significant correlations with the observed Gini, however the Gini based on predicted wealth from model 1 and 2 was weakly associated with the validation data (see *Figure 6*). For level 3 and 4, the Gini based on NTL had a weak positive association with the observed Gini meanwhile no modelled Gini showed significant association with the observed coefficients. At level 5 all predicted Ginis were positively associated with the observed values and the strongest association was the Gini based on NTL.

The reason for the mixed signals in correlations between predicted and actual Gini coefficient is unclear. One possibility is that other factors not accounted for in this model explain spatial variation better than the NTL images. As discussed by previous literature, light generating wealth does not give the full picture of the economy of an area. Some income generating activities, such as the agricultural industry might produce luminosity levels that are too low or close to noise, and therefore will not be represented accurately in the modelled economic output. Ebener et al. (2005) found that grouping countries by the value of the agricultural sector improves the relationship between night-time light and economy. Furthermore, Keola et al. (2015) found that including land use in their predictive model produces more accurate predictions of GDP in rural areas. However, at the initial stages of my analysis, the inclusion of percentage area occupied by agriculture as a model parameter did not increase the R^2. Striving for simplicity in the final models generated, this factor was therefore omitted from the rest of the analysis. Including other variables accounting for the GDP generating activities not represented by light could have improved the estimated GDP and in turn the estimated economic inequality.

Another gap in the analysis that could explain the negative and weak correlations in inequality is the difference between the definition of the modelled and the validation variable. Gini calculated from differences in income between regions is an example of spatial/regional inequality. Whereas all validation Gini coefficients represent personal/household income inequality, e.g. the difference in income between selected individuals within a nation. The two measures show slightly different aspects of inequality in a society. Furthermore, the two variables are biased towards different parts of the population. The individual income inequality is based on household surveys, which have been known to be biased towards the lower income bracket, because wealthy people do not see the importance in participating in the surveys. As a contrast to this, VIIRS images largely represent the wealthy part of society, especially in developing countries, since they tend to be the greatest electricity consumers (Chaturvedi 2011). This difference in inequality perspectives is also a matter of scale. From the satellite imagery household lights are not visible in the same way as streetlights. This implies that NTL can be used to indicate Wealth/development for society and not individuals.

The modelled Gini coefficients do not seem to have as well established a relationship with NTL as GDP has. However, Gini coefficients are also more complicated to model. Lessmann and Seidel (2017) highlights that measuring regional income inequality is more challenging than personal income inequality due to the heterogeneity of regions. The number of regions, average size, population etc.

varies between the countries in the dataset used in this analysis. When comparing regions by computing inequality indices, it is important to bear this in mind. The weak relationship between predicted and observed Gini could indicate that brightly lit areas in VIIRS images correspond to areas of high population density having a varied mix of individual income levels and not only to lower population densities of few wealthy people. This example highlights the problems occurring when comparing observed personal income inequality to predicted regional inequality.

The method used in this study varies slightly from that of previous studies. Studies by Rongwei et al. (2018); Xu et al. (2015); Zhou et al. (2015) and Elvidge et al. (2012) all calculate Gini coefficients directly from the light pixel values within the area of interest. However as Elvidge noted, these results correlate better with development indices compared to traditional Gini coefficients. Lessmann and Seidel (2017) developed a model with higher complexity compared to my study, involving many more variables as well as time series data. Gini coefficients are computed from the average predicted income in each region, resulting in predicted Gini coefficients on a national scale. Although the Gini estimations from Lessmann and Seidel (2017) correlated better with validation data than our results do. The inequality measures are only computed for national level. The goal of filling out the data gaps at finer scale is therefore not reached. It may be possible to predict fine scaled economic inequality with better accuracy through further integrating the methodologies.

5.2.1 Spatial patterns in modelling Gini

The accuracy of the predictions is not only dependent on the association between the predicted and observed values. The error of the predictions also needs to be considered. The spatial level with lowest mean absolute error (MAE) was level 3 followed by 4, 1, 2, 5 (see *Table 7*). Only focusing on level 3-5, this shows that larger polygons have lower mean errors than smaller polygons. Level 1 and 2 do not fit within this pattern, which might be a result of differences between the dataset, as discussed in coming sections.

Clear spatial patterns of prediction error can be seen in *Figure 9* and *Figure 10*. On the national level largest errors occur in sub Saharan Africa and in Latin America and the smallest errors occur in Europe. Similar patterns are seen from level 2, smallest errors occur in Europe and the country from Latin America included in these dataset, Chile, also shows large errors at this level. The large error for the United States is still present in level 2 although variation within the country suggests the large errors are clustered along the western. Further analysis is needed in order to explain these patterns in the data. However, inferring from the previous sections, investigating other explanatory factors such as agricultural sector or disparities in the validation data might help explain why my models failed to capture inequality in these areas specifically. Furthermore, clustering of the errors also occur at the small spatial scales (see *Figure 10*). These clear patterns of errors observed at all spatial levels, suggests that an underlying factor important for economic inequality is missing in the model.

Income level could also help explain the varying accuracy in modelled Gini. This explanatory variable was suggested by Chaturvedi (2011) who associated NTL to number of households in the different income brackets; and also by Lessmann and Seidel (2017) who included country average income as an explanatory variable in their model. This study did not account for varying income levels of the countries, apart from the inclusion of country fixed effects, in which some part of the variation might be captured. However slight patterns suggesting its affects can be discussed. Firstly, stronger correlations were found for level 3-5, a dataset only consisting of low-income countries, compared to level 2, a dataset consisting of mostly high-income countries. This suggests that the difference in accuracy between the scales might also be explained by the difference in economic level. Contrasting this, the large errors found in the national level Gini predictions mostly occur in low/lower middle-income countries and smaller errors coincide with high income countries e.g. most of the European countries. Further investigation into this mechanism could improve the estimation of inequality. If the accuracy of the predicted inequality varies based on income level, conclusions made by Chen and Nordhaus (2011) about GDP might not be valid for economic inequality. They suggest that NTL as a proxy for wealth is only valuable in developing countries, where data on economic output is either non existent or of poor quality. However, our results suggest that disparities in wealth within countries modelled with NTL imagery produce better results in high income countries.

Including the different steps of the analysis, a mixed picture has been drawn regarding the modelling of inequality and scale. From the correlations between predicted and observed values, our results indicated the best association at spatial level 5. However, the mean absolute errors suggested that the level with the largest error was level 5. Initially this might be confusing; however, this is an indication of the predicted Gini following the overall variation in the observation data but also being largely underestimated.

5.3 Spatial scale and MAUP

The modifiable area unit problem (MAUP) refers to the errors occurring when inferring that the relationship valid for the training data can also explain the variable in question at other spatial scales (scale effect) or in within other zones (zonal effect). This implies that both the model predicting economic output, as well as the inequality computed from the modelled income, could produce erroneous results at different spatial scales or within different polygon boundaries.

As seen from previous literature, the strength of the relationship between SE-variables and NTL varies depending on scale (Zhaoxin et al. 2017). This is also confirmed by the results of this paper. The accuracy of the predicted GDP is higher for national estimates than for regional estimates, as seen from the strength of the correlations in *Figure 6*. This is an example of the MAUP. When changing resolution/extent of the area of interest, the SE-variables used to represent the area also change. Larger areas, such as national estimates, include a heterogeneous landscape, with high-lit urban areas and low-lit rural areas. The estimated income for these large areas is then based on the averages of all the variation included. The impact of outlying values/areas is therefore smaller for larger areas. However, with smaller area, outliers such as industrial areas that generate a lot of light will have a larger impact on the estimated income. Such effects could lead to erroneous estimations of SE-variables, depending on the scale of analysis.

Furthermore, the MAUP scaling problem extends to most variables within this analysis. For example, the training dataset have a varying number of regions within each country. This difference in the size of the average region in each country has probably affected the analysis.

The second part of the MAUP is the zoning effect. This is not as relevant for this analysis as scale. Nevertheless, slight errors might occur due to this problem. The polygons of the training dataset and the validation dataset do no match completely. When developing the model, the polygon boundaries used by Kummu et al. (2018) to define the regional income dataset were used. When comparing model outputs with validation data, the polygon boundaries of the validation datasets where used. This should not make a large difference, since new polygons is a way to sample and test the predicted data. However, variations in polygon boundaries, also changes the area included in the model and this could create a bias.

Regarding the Gini coefficient two aspects need to be highlighted. Compared to other SE-variables, the Gini-coefficient is a measure of a distribution. The area included within the regional boundary and its disparities is therefore very important and pronounced. Compared to the income estimations, Gini-coefficients are not necessarily “smoothed” with a large area. Increasing area and heterogeneity will increase the disparity within the area and thereby increasing the Gini-coefficient, since there will be a higher chance of including min/max light peaks

When comparing estimated Gini coefficients with official validation data, another scale issue should be mentioned. The sample size for the estimated Gini, reflected by the number of pixels within the polygon, could be much larger than the sample size of the household-based Gini-coefficients, which might affect the distribution. The smallest areas for which Gini coefficients were estimated have the highest correlation with validation data. This could be due to more similar sample sizes, as the number of pixels included in the coefficient is very limited.

5.4 Limitations of this study

5.4.1 Sources of errors in the analysis

The main errors in this analysis stem from the inconsistencies in the various datasets used. The number of regions and size of regions vary for each country, resulting in varying accuracy of country effects as well as the problems in relation to MAUP discussed in the previous section. Furthermore, the official income and Gini coefficients are not based on the same year for all countries and level of analysis. The NTL data was from 2015, so SE-data matching that year would be optimal. However, both the dataset from WIID4 including national Gini coefficients (level 1) as well as the dataset of small area estimates (level 3-5) do not match this time stamp. This might result in some error when comparing predicted and validation income and Gini coefficients.

Finally, the Gini coefficient is just a measure of the dispersion within a distribution. The definition of this distribution varied slightly between the validation datasets. For example, the observed Gini for level 1 was a mix between Ginis based on household income and personal income; the OECD regions divide the observed Gini into before and after taxes and level 3-5 have divided the observed Gini between rural and urban areas – sometimes including the total inequality as well. From these examples it is easy to see that the validation data varies slightly between the datasets. Therefore, the differing patterns and distributions discussed previously might not only be explained by scale, but also due to the discrepancies in the underlying data.

This also underlines the problem of acquiring validation data for economic inequality. As mentioned in the background sections, it is costly to produce Gini estimates through house hold surveys. Furthermore, the methodology for such estimations is not standardized, as seen from the variations between the validation datasets. The gap of consistent validation datasets throughout various spatial scales is probably the most limiting factor of this study and is most likely the reason that economic inequality has not been modelled at fine spatial scales, previous to this study.

5.4.2 Future studies

This study has added to the literature by investigating if NTL satellite images can be used to predict inequality within society. Previous studies have suggested the possibility of predicting inequality (Lessmann and Seidel 2017; Xu et al. 2015; Zhou et al. 2015). This study expands upon the idea by investigating the scale dependency of such predictions. However as implied by the discussed results, predicting inequality is complicated and the models used were not able to predict accurate measures of economic inequality. This suggests that more advanced models are needed in order to capture inequality at various spatial scales. Future improvement of the methodology could be including more explanatory variables such as land-use, population, income levels etc. and differences between countries/region. Accounting for these differences within the model might help to improve the predictions and produce a useful proxy for Gini coefficients within as well as between countries (Lessmann and Seidel 2017).

6 Conclusion

Concluding the research objectives:

1. Through a literature review, evidence from several studies was presented, associating night-time light (NTL) with SE-variables. Light is a good predictor of GDP, and through this relationship economic outputs can be predicted to fill missing data gaps. Similar attempts are suggested for modelling economic inequality. However only a few studies have found associations on coarse spatial scales. Furthermore the lack of observed data limits the spatial and temporal scope of such analysis.
2. Building upon previous knowledge, economic output was modelled successfully with NTL, SOL and country fixed effects. Following this, Gini-coefficients computed based on predicted income were compared to the observed Gini coefficients.
3. The comparisons were inconclusive. The associations, error distribution and the mean absolute error varied by the spatial scale of the analysis. But due to discrepancies in observed datasets as well as other explanatory factors not included. However, clear spatial patterns of the error were seen on all spatial levels, which suggests that important underlying drivers of economic inequality were not captured. Exploring other predictor variables might lead to better estimations of Gini coefficients.

The results of this thesis suggest that modelling economic inequality is more complicated than other parameters previously modelled with NTL. Further investigation of the spatial patterns and dependency on spatial scale is needed before establishing the possible relationship between the Gini coefficient and NTL.

7 References

Alvaredo, F., L. Chancel, T. Piketty, E. Saez, and G. Zucman. 2018. The World Inequality Report

Bennett, M. M., and L. C. Smith. 2017. Review: Advances in using multitemporal night-time lights satellite imagery to detect, estimate, and monitor socioeconomic dynamics. *Remote Sensing of Environment,* 192**:** 176-197. DOI: 10.1016/j.rse.2017.01.005

Bruederle, A., and R. Hodler. 2018a. Nighttime lights as a proxy for human development at the local level. *PLOS ONE,* 13**:** e0202231. DOI: 10.1371/journal.pone.0202231

Bruederle, A., and R. Hodler. 2018b. Nighttime lights as a proxy for human development at the local level. *Plos One,* 13. DOI: 10.1371/journal.pone.0202231

Center for International Earth Science Information Network - CIESIN - Columbia University. 2005. Poverty Mapping Project: Small Area Estimates of Poverty and Inequality. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC).

Center for International Earth Science Information Network - CIESIN - Columbia University. 2018. Gridded Population of the World, Version 4 (GPWv4): Population Count, Revision 11. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC).

Chaturvedi, M. 2011. *Assessing Income Distribution at the District Level for India Using Nighttime Satellite Imagery*.

Chen, X., and W. D. Nordhaus. 2011. Using luminosity data as a proxy for economic statistics. *Proceedings of the National Academy of Sciences of the United States of America,* 108**:** 8589-8594. DOI: 10.1073/pnas.1017031108

Coscieme, L., F. M. Pulselli, S. Bastianoni, C. D. Elvidge, S. Anderson, and P. C. Sutton. 2014. A Thermodynamic Geography: Night-Time Satellite Imagery as a Proxy Measure of Emergy. 969-979.

Coscieme, L., P. C. Sutton, S. Anderson, Q. Liu, and C. D. Elvidge. 2017. Dark Times: nighttime satellite imagery as a detector of regional disparity and the geography of conflict. 118-139.

Doll, C., J. Morley, and J. P. Muller. 2004. *Geographic Information Issues associated with socio-economic modelling from night-time light remote sensing data*.

Doll, C. N. H., J.-P. Muller, and E. C. D. 2000. Night-Time Imagery as a Tool for Global Mapping of Socioeconomic Parameters and Greenhouse Gas Emissions. *Ambio***:** 157.

Doll, C. N. H., J.-P. Muller, and J. G. Morley. 2006. Mapping Regional Economic Activity from Night-Time Light Satellite Imagery. *Ecological Economics,* 57**:** 75-92. DOI: http://www.sciencedirect.com/science/journal/09218009

Doll, C. N. H., and S. Pachauri. 2010. Estimating rural populations without access to electricity in developing countries through night-time light satellite imagery. *Energy Policy,* 38**:** 5661-5670. DOI: https://doi.org/10.1016/j.enpol.2010.05.014

Ebener, S., C. Murray, A. Tandon, and C. Elvidge Christopher. 2005. From wealth to health: modelling the distribution of income per capita at the sub-national level using night-time light imagery. *International Journal of Health Geographics, Vol 4, Iss 1, p 5 (2005)***:** 5. DOI: 10.1186/1476-072X-4-5

Elvidge, C., D. , J. Safran, B. Tuttle, P. Sutton, P. Cinzano, D. Pettit, J. Arvesan, and C. Small. 2007. Potential for global mapping of development via a nightsat mission. *GeoJournal***:** 45.

Elvidge, C., K. E Baugh, M. Zhizhin, and F.-C. Hsu. 2013. *Why VIIRS data are superior to DMSP for mapping nighttime lights*.

Elvidge, C. D., K. Baugh, M. Zhizhin, F. C. Hsu, and T. Ghosh. 2017. VIIRS night-time lights. *International Journal of Remote Sensing,* 38**:** 5860-5879. DOI: 10.1080/01431161.2017.1342050

Elvidge, C. D., K. E. Baugh, S. J. Anderson, P. C. Sutton, and T. Ghosh. 2012. The Night Light Development Index (NLDI): a spatially explicit measure of human development from satellite data. *Social Geography,* 7**:** 23-35. DOI: 10.5194/sg-7-23-2012

Elvidge, C. D., P. C. Sutton, T. Ghosh, B. T. Tuttle, K. E. Baugh, B. Bhaduri, and E. Bright. 2009. A global poverty map derived from satellite data. *Computers & Geosciences,* 35**:** 1652-1660. DOI: 10.1016/j.cageo.2009.01.009

Ghosh, T., S. J. Anderson, C. D. Elvidge, and P. C. Sutton. 2013. Using Nighttime Satellite Imagery as a Proxy Measure of Human Well-Being. *Sustainability, Vol 5, Iss 12, Pp 4988-5019 (2013)***:** 4988. DOI: 10.3390/su5124988

Haughton, J., and S. Khandker. 2009. Handbook on Poverty and Inequality Chapter 6: inequality measures. *World Bank*.

Henderson, J. V., A. Storeygard, and D. Weil, N. . 2012. Measuring Economic Growth from Outer Space. *The American Economic Review***:** 994.

Hunter, T. M. 2014. Map of Gini index 2014. Wiki Commons.

Jean, N., M. Burke, M. Xie, W. M. Davis, D. B. Lobell, and S. Ermon. 2016. Combining satellite imagery and machine learning to predict poverty. *Science (New York, N.Y.),* 353**:** 790-794. DOI: 10.1126/science.aaf7894

Keola, S., M. Andersson, and O. Hall. 2015. Monitoring Economic Development from Space: Using Nighttime Light and Land Cover Data to Measure Economic Growth. 66**:** 322–334. DOI: 10.1016/j.worlddev.2014.08.017

Kummu, M., M. Taka, and J. H. A. Guillaume. 2018. Gridded global datasets for Gross Domestic Product and Human Development Index over 1990–2015. *Scientific Data,* 5**:** 180004. DOI: 10.1038/sdata.2018.4

https://www.nature.com/articles/sdata20184#supplementary-information

Lessmann, C., and A. Seidel. 2017. Regional inequality, convergence, and its determinants – A view from outer space. *European Economic Review,* 92**:** 110-132. DOI: 10.1016/j.euroecorev.2016.11.009

Li, C., G. Li, G. Tao, Y. Zhu, Y. Wu, J. Liu, and X. Li. 2018. DMSP/OLS night-time light intensity as an innovative indicator of regional sustainable development. *International Journal of Remote Sensing*. DOI: 10.1080/01431161.2018.1528022

Li, S., T. Zhang, Z. Yang, X. Li, and H. Xu. 2017. Night time light satellite data for evaluating the socioeconomics in Central Asia. 1237-1243. International Society for Photogrammetry and Remote Sensing.

Ma, T., Y. Zhou, Y. Wang, C. Zhou, S. Haynie, and T. Xu. 2014. Diverse relationships between Suomi-NPP VIIRS night-time light and multi-scale socioeconomic activity. 652-661.

Mellander, C., J. Lobo, K. Stolarick, and Z. Matheson. 2015. Night-Time Light Data: A Good Proxy Measure for Economic Activity? *PLoS ONE,* 10**:** 1-18. DOI: 10.1371/journal.pone.0139779

Miller, S. D., W. Straka, S. P. Mills, C. D. Elvidge, T. F. Lee, J. Solbrig, A. Walther, A. K. Heidinger, et al. 2013. Illuminating the Capabilities of the Suomi National Polar-Orbiting Partnership (NPP) Visible Infrared Imaging Radiometer Suite (VIIRS) Day/Night Band. 5**:** 6717.

Noor, A. M., V. A. Alegana, P. W. Gething, A. J. Tatem, and R. W. Snow. 2008. Using remotely sensed night-time light as a proxy for poverty in Africa. *Population Health Metrics,* 6**:** 5. DOI: 10.1186/1478-7954-6-5

OECD. 2014. Regional well-being.

Proville, J., D. Zavala-Araiza, and G. Wagner. 2017. Night-time lights: A global, long term look at links to socio-economic trends. *PLoS ONE,* 12. DOI: 10.1371/journal.pone.0174610

Reidpath. 2009. Gini coefficient diagram. ed. E. G. coefficient2.svg. Wikimedia Commons.

Rongwei, W., Y. Degang, D. Jiefang, Z. Lu, and X. Fuqiang. 2018. Regional Inequality in China Based on NPP-VIIRS Night-Time Light Imagery. *Remote Sensing,* 10**:** 240. DOI: DOI: 10.3390/rs10020240.

Shasha, L., G. Xingliang, Y. Danlin, D. Yu, and Z. Linjie. 2015. Multi-Scale Analysis of Regional Inequality based on Spatial Field Model: A Case Study of China from 2000 to 2012. *ISPRS International Journal of Geo-Information, Vol 4, Iss 4, Pp 1982-2003 (2015)***:** 1982. DOI: 10.3390/ijgi4041982

Sun, W. C., X. Zhang, N. Wang, and Y. Cen. 2017. Estimating Population Density Using DMSP-OLS Night-Time Imagery and Land Cover Data. *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* 10**:** 2674-2684. DOI: 10.1109/jstars.2017.2703878

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

Tanaka, K., and S. Keola. 2016. *Shedding Light on the Shadow Economy: A Nighttime Light Approach*.

UNU-WIDER. 2017. World income inequality database (WIID 3.4).

Wang, L., H. Fan, and Y. Wang. 2018. Estimation of consumption potentiality using VIIRS night-time light data. *PLoS ONE,* 13**:** 1-19. DOI: 10.1371/journal.pone.0206230

Xu, H., H. Yang, X. Li, H. Jin, and D. Li. 2015. Multi-Scale Measurement of Regional Inequality in Mainland China during 2005–2010 Using DMSP/OLS Night Light Imagery and Population Density Grid Data. *Sustainability, Vol 7, Iss 10, Pp 13469-13499 (2015)***:** 13469. DOI: 10.3390/su71013469

Zhao, N., F. C. Hsu, G. Cao, and E. L. Samson. 2017a. Improving accuracy of economic estimations with VIIRS DNB image products. *International Journal of Remote Sensing,* 38**:** 5899-5918. DOI: 10.1080/01431161.2017.1331060

Zhao, N., Y. Liu, G. Cao, E. L. Samson, and J. Zhang. 2017b. Forecasting China's GDP at the pixel level using nighttime lights time series and population images. *GIScience & Remote Sensing,* 54**:** 407-425. DOI: 10.1080/15481603.2016.1276705

Zhaoxin, D., H. Yunfeng, and Z. Guanhua. 2017. The Suitability of Different Nighttime Light Data for GDP Estimation at Different Spatial Scales and Regional Levels. *Sustainability, Vol 9, Iss 2, p 305 (2017)***:** 305. DOI: 10.3390/su9020305

Zhou, Y., T. Ma, C. Zhou, and T. Xu. 2015. Nighttime Light Derived Assessment of Regional Inequality of Socioeconomic Development in China. 1242-1262.