USING LIGHT AT NIGHT DATA AS A PROXY FOR ECONOMIC ACTIVITY: A CASE STUDY FOR PAPUA NEW GUINEA.

Vrije Universiteit, Amsterdam

Abstract:

Prior research has suggested that Light at Night (LAN) data can be used as a valid proxy for the spatial distribution of economic activity. Most of the research was done for relatively developed countries with high population densities, high urbanization and lots of artificial areas. In this study an attempt was made to establish a correlation between LAN data and economic activity for Papua New Guinea, a little developed country with extremely low urbanization and 65 percent of the land area covered with forests. For the regression analysis LAN data for 2000 and 2010 was used. To estimate local economic activity, a model, based on labour force, land cover data and national statistics, was developed. The model calculates GDP per grid cell at the same resolution as the LAN data. Furthermore the difference for the 2000 and 2010 data was calculated, so a correlation could be established between the LAN data and economic activity for 2000, 2010 and the change of the variables at multiple levels of spatial resolution. The regression analysis resulted in a strong correlation between the LAN data and economic activity, especially at the 5 by 5 km spatial resolution. The correlation was stronger for the 2000 and 2010 values than it was for the change values, particularly at the lower spatial resolutions. So LAN data can serve as a valid proxy for economic activity at high spatial resolution for PNG. But better correlates with yearly economic values than it does with actual change in economic activity.

Daan Petri
2502328
Thesis BSC Earth Science & Economics
Dr. Eric Koomen

1. Introduction

1.1 introduction

In the late 1970s a game changing discovery was made in the world of remote sensing, the Operational Linescan System (OLS) sensor on board of a meteorological satellite was able to record, to some degree of detail, the light that originates from the surface of the earth (Croft, 1978). Since records of the data, collected by the OLS sensors, became publicly available in 1992, it has been used in a number of different settings and purposes. For example: to estimate populations around the globe (Elvidge et al, 1997a), the population without access to electricity (Doll and Pachauri, 2010). Probably more important was the usage of Light At Night (LAN) data to estimate the size of the economy on a national scale (Gosh, Sutton, Powell, Anderson and Elvdige, 2009), the mapping of urban and suburban extent around cities (Roychowdhury, Jones and Arrowsmith, 2009) and for modelling socio-economic parameters and urban population (Doll, Muller and Morley, 2007).

In recent years the focus from global or national scale usage of the LAN data has shifted to a more detailed regional and sub-national level. A study carried out by Roychowdhury Jones, Arrowsmith and Reinke in 2011, proved the usefulness of the DMSP-OLS data on a sub-national scale to estimate population of smaller regions. A year later Levin and Duke showed that LAN data may also be used to inform us on economic activity at the local scale. They both arrived at the conclusion that the DMSP-OLS data was useful as an indicator of certain activities at a local scale but left a lot to be desired. Mainly due to the coarse special resolution of the data (2.7 km^2) that is extracted from the DMSP-OLS, its single spectral band and oversaturation of the urban cores (Levin and Duke, 2012).

Most of the research using LAN data have been done for relatively developed countries where data about social and economic parameters is readily available. Recent examples being Mellander, Stolarick, Matheson and Lobo from 2013, whom investigated how good a proxy LAN data actually was in Sweden. A country with accurate economic statistics available at the highest spatial resolutions. They found that the relation between LAN data and economic data was better correlated with density figures in general and it underestimates the values of economic activity in rural areas and overestimates in urban areas.

So LAN data has proven to be a useful tool that can be used to estimate, to some extent, spatial distribution of economic activity in developed countries. But can it also be used to map such patterns for developing countries? In this study I will use LAN data for Papua New Guinea (PNG) in combination with economic markers, mostly from global datasets, to see if it can be used as a valid proxy for this developing country.

1.2 Aims and hypotheses

The aim of this research is to see whether LAN data is useful, even in little developed countries, to serve as a proxy for the spatial distribution of economic activity. This study will be conducted for Papua New Guinea. It is hypothesized that just as it has proven a useful tool in developed countries and on a global scale, it can be used to approximate the spatial distribution of economic activity within countries. But with increasing spatial resolutions the correlation will decrease, just as correlation strength decreased when spatial resolution was increased from the national to sub-national level (Mellander et al., 2013; Gosh et al., 2009).

2. Methods

2.1 Study area Papua New Guinea

Papua New Guinea is situated in Oceania between Indonesia and Australia, but only shares a border on land with the former. PNG has a population density of only 14.8 person per square mile. Only 12.5 % of the population, which totals around 7.4 million, resides in the city (UN, 2015). Making Papua one of the least urbanized countries in the world. With 40% of the population living below the poverty line (Worldbank, 2015), PNG latest GDP figures total around 16.1 billion dollars (IMF, 2015). Most of the land area, 65 percent to be precise, is covered by primary and secondary forest, it covers a total of 28.4 million ha, to a total land area of 45.29 million ha. Crop and arable land make up around 1 million ha (FAO, 2015). Around 67% of the total labour force is employed in the agricultural sector and is responsible for roughly 40 percent of the total GDP of Papua New Guinea. Average income was a little over 1200 USD per capita for 2010 (FAO, 2015).

With low population density and development level, high forest coverage and extremely low urbanization rate, the question is whether LAN data can still be used as a valid proxy, since earlier research showed that density figures for roads and population are better suited for correlation with LAN data than their corresponding absolute data is (Mellander et al., 2013).

Furthermore, the statistical capacity of PNG is poor, scoring only 46.7 out of a 100, while the average for developing countries is 70 (Worldbank, 2015). Thus PNG is a country where development of accurate statistics is low, therefore their availability will be limited. Which means PNG is a country where LAN data could prove to be a valuable source for accurate, additional statistics on regional development, provided that the initial difficulty concerning the process of establishing an economic model with limited statistics for regional activity can be overcome.

2.2 Method

To establish whether LAN data can be used as a valid proxy for economic activity in PNG, a correlation has to be established between light at night and economic activity. Earlier research into the link between the two variables used GDP and LAN from the DMSP-OLS satellite to establish a correlation has made use of a linear model (Doll, Muller and Morley, 2006; Sutton, Elvidge and Gosh, 2007). Or alternatively LAN and GDP have been recalculated to their respective logarithmic values, for a log-log correlation between the two variables (Doll, Muller and Elvidge, 2000) or a log-linear correlation, where lights is an increasing function of corresponding GDP (Wu, Wang, Li and Peng, 2013).

This study is carried out at a regional scale, which is the same scale as the studies done by, Doll and Sutton from 2006 and 2007. The same method will be used for the regression analysis so the results could eventually be compared. Thus regression analysis is run using a linear model. Moreover there is a mathematical reason why logarithmic data is unsuitable for the LAN data for PNG, a large part of the country emits no light, resulting in radiance values of 0. It is mathematically impossible to calculate a logarithmic for 0; this would result in No Data and render the observation useless.

For light at night data, the OLS-DMSP dataset is used, which is available for multiple years and has been used extensively in different research projects which aimed to correlate LAN data with economic activity. (Doll et al., 2000; Doll et al., 2006; Sutton et al., 2007;Wu et al.,2013).

Gross Domestic Product (GDP) in American dollars is used to estimate the economic activity, where more economic activity is represented by a greater GDP. GDP is the most important variable in analyses of economic growth (Henderson, Storeygard and Weil, 2012) and has been used in earlier research into the correlation between LAN and economic activity (Henderson et al., 2012; Mellander et al., 2013; Gosh et al., 2009).

The obtainment and editing of the datasets used in the regression analysis are discussed in detail in the next section.

2.3 Datasets

2.3.1 Imagery and LAN data

Light at night datasets are made available by the U.S. National Oceanographic and Atmospheric Administration information (NOAA). These dataset are composed following observation from the Operational Linescan System (OLS), a sensor aboard the Defence Meteorological Satellite Program (DSMSP) satellites. The DSMP is used to obtain real-time weather information for the military, during the night clouds, which are one of the indicators of weather systems, are difficult to spot and identify. For that purpose sensors are used to spot the reflection of star and moonlight coming back from clouds. When no clouds are present, the sensor picks up light emanating from the surface of the earth.

LAN data is provide in 30 arc-second grids and covers most of the world. Since earth is not a square, the size of the grid cells decreases in size, the closer you get to the poles. This phenomenon increases as you get closer to the poles, therefore no data is available for the poles, as projection issues would make the data unusable. The OLS datasets are made available on a yearly basis, averaging the amount of captured light over the year for nearly every place on earth.

The radiance values vary from 0 - 63, with 0 being no light and 63 corresponding with the brightest light. For the year 2000, two datasets were available from two different satellites, the F14 and the F15. The datasets were averaged to get the bests approximation of radiance values. For 2010 only one dataset was available from the F16 satellite (NOAA, 2015). The creation of the LAN dataset is discussed more thoroughly in Elvidge, Baugh, Kihn, Kroehl and Davis paper from 1997b.

The LAN data was cleaned to exclude any contamination from flares. The flaring of gas obviously results in light, which is bright enough to be picked up by the OLS sensor on the satellite. Light resulting from flaring would be an additional source of light adding to the non-intentional light emissions from the oil and gas industry. In contrast to other industries, which do not have additional light sources, this would result in higher radiance values for oil and gas industry in comparison to other industries. Thus these higher radiance values cannot be correlated with higher GDP following the same method. Including flaring would result in an overestimation of the GDP for these industries. To maintain consistency for the LAN data in correspondence to the economic data, gas flares are excluded from the LAN data. Thus preventing any contamination and resulting outliers in the data (Lowe, 2014).

2.3.2 Economic activity

For Papua New Guinea little to no official statistics are available and during the course of the study the website for Papua New Guinea's bureau of statistics was unreachable. Using data from different sources, an attempt was made to make a spatial distribution of economic activity for 2000 and 2010 following the same methodology as Nordhaus and others from 2006.

In their methodology Nordhaus et al. used formula 1 to calculate Gross Domestic Product for each gridcell:

(1) $GDP_{gridcell} = GDP_{region} / capita_{region} * population_{gridcell}$.

Regional GDP values are divided by the population of the corresponding region resulting in an estimation of GDP per capita for each region. The GDP per capita is then multiplied by the population estimates for each gridcell. GDP per capita for each region is calculated using formula 2:

(2) $GDP/Capita = ((Percentage\ of\ labour\ force\ active\ in\ agriculture_{region}/100) * total\ labour\ force_{region} * (Total\ GDP\ from\ agricultural\ sector_{country} / total\ labour\ force\ active\ in\ agriculture_{country})) + ((Percentage\ of\ labour\ force\ active\ in\ non\text{-}agricultural\ sector_{region}/100) * total\ labour\ force_{region} * (Total\ GDP\ non\text{-}agriculture_{country}/Total\ labour\ force\ active\ in\ non\text{-}agricultural\ sector_{country})) / total\ population_{region}$.

Labour force for each region was determined using formula 3:

(3) $Labour\ force_{region} = population_{region} * (percentage\ of\ population\ between\ 15\ and\ 65_{region} / 100) * (participation\ rate_{region} / 100)$.

Workforce for each region is split in two groups; an agricultural and non-agricultural workforce. Using aggregated GDP values for the non-agricultural sector and the agricultural sector, GDP per capita was determined for the agricultural and non-agricultural labour force. Because the workforces engaged in agricultural and non-agriculture varies per region, the GDP per capita will also vary. So a different composition of labour force results in different averages. After, grid light density and population density figures were used to further refine the GDP for each grid cell and then results in economic activity values for 1 by 1 degree cells.

For this study a similar methodology was adopted. However, for this economic model, population density and grid light density were left out as variables. Since the goal of the study is to independently verify the use of light data as a proxy for economic activity, it would be very biased to use LAN data in the creation of an economic dataset. Furthermore three sectors instead of two were used for GDP calculations. Which are: mining, agricultural and other (secondary and tertiary) sector. These sectors are used, as labour force active in these sectors account for respectively 35, 28 and 37 percent in 2000 and 31, 25 and 43 percent in 2010 of total labour force (KNOEMA, 2015).

The most import change made to the methodology was the increase in spatial resolution, in contrast to the methodology used by Nordhaus and others, data per Local Level Government (LLG) was collected instead of data on a province level. This increased the resolution from 20 to 297 administrative regions. The output spatial scale was also increased from 1 by 1 degree to the same spatial resolution as the LAN data (500x500 meter), so correlation could be made at the highest possible spatial resolution. The next section discusses the obtainment of variables used in the formulas previously mentioned.

2.3.4 Population

Population data was acquired from the worldpop project. This project was initiated in 2013 and combines mapping projects for three different continents (South-America, Africa and Asia). Population data for PNG was created through a land cover based approach (Gaughan, Stevens, Linard, Jia and Tatem, 2013). With most of the population living in settlement, settlement extent data is used to refine the land cover based approach. Census data is used to identify typical regional population density per land cover type. The same census data was then used to redistribute population over the land areas (Linard, Gilbert and Tatem, 2012).

The data is made at an extremely high resolution of approximately 50 by 50 meters at the equator. Since the LAN data is only available at 500 by 500 meter resolution, the resolution had to be decreased to the same grid size as the LAN data. This meant resampling the data and summing it to a lower spatial resolution of 500 by 500 meter. The world pop data is only available for 2010 and 2015 for PNG. The adjusted version for 2010 was used, this version has its population totals adjusted to fit the UN estimation of total population per country.

Population for 2000 was obtained by comparing the 2000 and 2010 census data, to determine the change of population in percentages for each LLG. These percentages were then used to change the population map that was used for 2010 to fit population totals for 2000. This method was used to ensure similar obtainment in population data and, more importantly, because no population dataset with the same spatial resolution and accuracy exists for PNG for the year 2000.

2.3.5 Labour Force

Census data is available for 2000 (Humanitarian response, 2015) and 2010 (NSO, 2011) for all LLG's. The participation rate and population aged between 15 and 65 was part of the 2000 census, while 2010 census only consisted of population numbers. The 2010 participation rate and population between 15 and 65 were instead obtained from a household survey from the National Statistical Office for 2009.

The 2010 census data was constructed with more LLG's than its 2000 equivalent. No data is available containing information about LLG division for 2010, so the 2010 census data had to be modified to fit the LLG distribution from 2000. Some documentation was available for redistribution of LLG borders, but not all of the change could be explained. Therefore some LLG's could not be redistributed directly and these were incorporated in the totals and then divided over all the LLG's in the same province.

Statistics on the GDP per sector and employment in the agricultural sector are available on a national level (KNOEMA, 2015; FAO, 2015). However, for the purposes of this study, these statistics are required for each LLG. As these statistics are unavailable an approximation had to be made using the national totals as a baseline and refined by various dataset as described below.

2.3.6 Agricultural sector

Foremost the labour force active in agricultural was determined, the agricultural sector is comprised of three elements: agriculture, fishing and forestry. To determine workforce for each of the three sectors proxy datasets had to be used since official statistics are not available. For agriculture, cropland area was used as a proxy, derived from MODIS 2001 and 2011. For the fishing sector, coast length was used for marine fishing, while river and lake area were utilized for inland fishing. Which were proven to be a valid proxy for fishing intensity (Stewart et al., 2010). Deforestation maps for 2000 and 2010, which compare forest cover for consecutive years to conclude where forest area has disappeared, derived from the Modis land cover datasets are used as a proxy for forestry sector. Using formula 4 labour force active in the agricultural sector was estimated and allocated to each LLG:

(4) Workforce active in the agricultural sector $_{LLG}$ = ((Agricultural land cover $_{LLG}$ / Agricultural land cover $_{Country}$ * GDP from agricultural sector $_{Country}$) +(Coastline length $_{LLG}$ / Coastline length $_{country}$ * GDP from marine fishing) + (Lake and river area $_{LLG}$ / Lake and river area $_{country}$ * GDP from inland fishing) + (Deforestation $_{LLG}$ / Deforestation $_{country}$ * GDP from Forestry) / (GDP from Agriculture $_{Country}$ / Labour force Agricultural sector $_{Country}$)

Since this is a very crude way to estimate agricultural employment, the resulting values for some LLGs are implausible. To adjust the values, reclassing was done using statistics on countrywide employment in the agricultural sector as a baseline. Labour Force was adjusted to fit the countrywide averages, but varied per LLG.

2.3.7 Mining sector

About 25% of PNG results from Mining. The economic activity from mining results in LAN data that could not be explained in the original economic model following the methodology from Nordhaus. To correct this, polygons following the mines concessions area were created. These polygons were assigned GDP using formula 5:

(5) $GDP_{Mines} = (Production\ gold_{Mine} * Gold\ Price_{Global\ Average} + Production\ Silver_{Mine} * Silver\ Price_{Global\ Average} + Copper\ production_{mine} * Copper\ price_{Global\ Average}) / (Production\ gold_{total} * Gold\ Price_{Global\ Average} + Production\ Silver_{total} * Silver\ Price_{Global\ Average} + Copper\ production_{total} * Copper\ price_{Global\ Average}) * GDP\ From\ Mining_{country}$

Formula 5 was done for both 2000 and 2010 using corresponding values.

The polygons were then converted into raster data, with the GDP value divided by the number of cells the polygon was converted into. Using the Mosaic to new raster tool in ArcGIS the polygons were added to the map using sum.

2.3.8 Other sectors

Labour force which was not employment in either the mining or the agricultural sector was presumed to be a part of this combined rest category.

2.3.9 GDP

Utilizing countrywide GDP per sector data, GDP for the agricultural, mining and the rest sector were determined on a countrywide basis. Based on total employment per sector GDP per capita for each sector was determined for the whole country. These GDP/capita values were then used to estimate GDP for each LLG based on the size of the labour force in each sector times the GDP/capita from the same sector. This resulted in total GDP per LLG, after which this total was divided by the total labour force of the LLG to have an estimated GDP/Capita for each LLG.

These GDP/Capita values were rasterized so they could be used in the raster calculator. Where population data times GDP per capita plus mining resulted in the estimated total for economic activity in GDP. The manipulation explained here were done following the formulas explained more precise in earlier sections.

2.4 Regression Analysis

To compare the LAN data with the data on economic activity, the same method as Lowe in 2014 was used. A fishnet was created at 5 different spatial resolutions: 0,5x0,5 2,5x2,5, 5x5, 12,5x12,5 and 25x25 km resolution. These resolutions have the same template extent as the LAN data are all multiplications of the grid size used for the LAN data, so borders of zones would overlap with the borders of the LAN grid. These different spatial resolutions were made to average out over saturation of urban cores and to rectify for measurement errors in the data. To try and prevent problems arising from the modifiable areal unit problem (MAUP), an optimal zonal division was sought after (Roychowdhury, Jones, Rienke, Arrowsmith, 2012).

Using the zonal statistics as table, the spatial sum-statistics for the LAN and economic data was calculated for all spatial resolutions. Sum is preferred over mean, as GDP measures economic activity by adding all added values together to estimate economic activity. Averaging GDP would result in unfair comparison between regions, as regions are not similar in size. Larger regions with more economic activity could have a lower mean, as a result of more low values, in comparison to smaller region with a higher mean even though it has less economic activity. The larger region would have a lower mean, but a higher sum in comparison to the smaller region. For this reason all similar research into the relation between GDP and LAN data has preferred sum of mean(Doll et al., 2000; Doll et al., 2006; Sutton et al., 2007; Wu et al.,2013).

Zonal statistics were calculated for the economic and LAN from the 2000, 2010 and the change of these datasets. The resulting tables were joined with the fishnets for their corresponding spatial resolution. This was done so all the observations match for their location and could be correlated. The resulting list of variables is shown in table 1:

Table 1: List of variables.

Variable	N	Mean	Std. Dev.	Min	Max
25x25 km					
Light at Night 2000	429	79.3	514.2	0	8833
Light at Night 2010	429	109.9	627.7	0	10014
GDP 2000	429	1.92e+07	5.45e+07	318.6	5.87e+08
GDP 2010	429	5.85e+07	2.31e+08	946.4	3.99e+09
Difference in GDP	429	3.93e+07	1.86e+08	-3.90e+07	3.42e+09
Difference in LAN	429	30.6	163.3	-651	1759
12,5x12,5 km					
Light at Night 2000	1307	24.3	245	0	7422
Light at Night 2010	1307	33.7	309.6	0	8970
GDP 2000	1307	5885560	2.49e+07	0	5.71e+08
GDP 2010	1307	1.80e+07	9.85e+07	0	2.70e+09
Difference in GDP	1307	1.21e+07	7.74e+07	-4.21e+07	2.25e+09
Difference in LAN	1307	9.4	85.4	-651	1548
5x5 km					
Light at Night 2000	6567	4.8	65.6	0	3376
Light at Night 2010	6567	6.7	83.4	0	3997
GDP 2000	6567	1171376	8144458	0	3.92e+08
GDP 2010	6567	3572852	3.87e+07	0	2.65e+09
Difference in GDP	6567	2401476	3.17e+07	-2.91e+07	2.25e+09
Difference in LAN	6567	1.9	25.8	-503	753
2,5 x 2,5 km					
Light at Night 2000	23305	1.4	20.5	0	1251
Light at Night 2010	23305	1.9	25.8	0	1428
GDP 2000	23305	330076.3	2889614	178.5017	2.13e+08
GDP 2010	23305	1006776	1.26e+07	532.4382	1.23e+09
Difference in GDP	23305	676699.9	1.02e+07	-2.99e+07	1.01e+09
Difference in LAN	23305	0.5	8.9	-283	552
0,5 x 0,5 km					
Light at Night 2000	541645	0.06	0.95	0	60
Light at Night 2010	541645	0.08	1.2	0	63
GDP 2000	541645	14202	172159.6	4.6	20300000

GDP 2010	541645	43304.2	692182.3	13.7	54200000
Difference in GDP	541645	29102.2	557988.4	-8850398	49000000
Difference in LAN	541645	.02	0.49	-55	30

Regression analysis was done for each of the spatial resolutions separately. So Light at night 2000 data at a 5x5 km scale is correlated with the GDP 2000 data from the 5x5 km scale data. The OLS regression will be run using STATA, where Light at Night will serve as the dependent value and GDP will be the independent value.

3 Results

3.1 Light at night

Figure 1, 2 and 3 are the results for the LAN dataset for respectively 2000, 2010 and the change datasets. These figures are a representation of the information captured by the OLS sensors for Papua New Guinea after they were modified using the methodology described.

Most of the country is dark, which was expected, given that most of the land area is covered with forest. A couple of hotspots do stand out though, Port Moresby the capital in the southern part of PNG, the OK tedi mine near the border with Indonesia and Lae, the second city of Papua New Guinea, in the North-East of the main island. The big no data hole in the middle of Papua New Guinea is the result of the correction that was made to exclude the gas-flaring from the analysis.

The distribution of the LAN data is a very diffuse one, with a few dozen spot spread around Papua New Guinea. Almost all of these hotspots seem to correspond with settlements and the large mining operations, ongoing in PNG. In LAN data for developed countries, the spots of the LAN data are usually connected by lines of light, which corresponds with roads. For PNG no such phenomenon exist, given that only 3.5 percent of all roads was paved in 2000 (Trading economics, 2015), it is safe to assume that an even smaller percentage of roads is illuminated. So for PNG road illumination is not a major contributor to the total LAN values of PNG and does not have to be considered for correlation.

Figure 2 clearly shows higher radiation values in comparison to Figure 1, with most of the purple being replaced by blue. Some hotspots have disappeared and a couple new ones appeared. This partly coincides with the opening of new mines and the closing of an older one. The 2010 data also seems to be a little more compact than the 2000 data. With the hotspots being more compact and less radiance values around the cores.

The change values shown in figure 3 seem to confirm this. The spot itself have higher values and experienced an increase in light, while the outskirts of these spots have negative values. The biggest changes corresponds with the closing of the Misima mine on one of the islands in the South-Eastern part of PNG and the opening of the Hidden Valley mine south of Lao. The opening of the hidden valley resulted in a large increase, while the closing of Misima resulted in a large decrease.

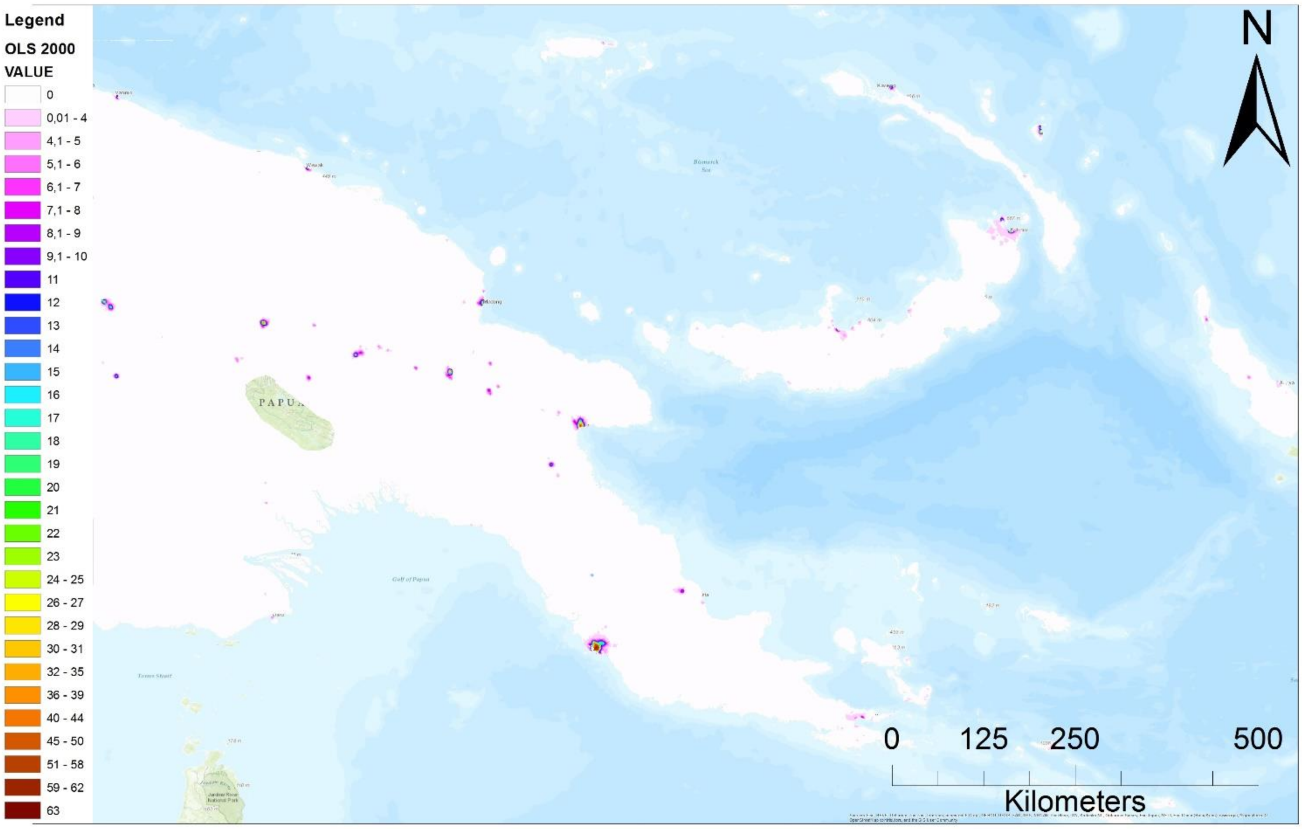

Figure 2: LAN data for 2000

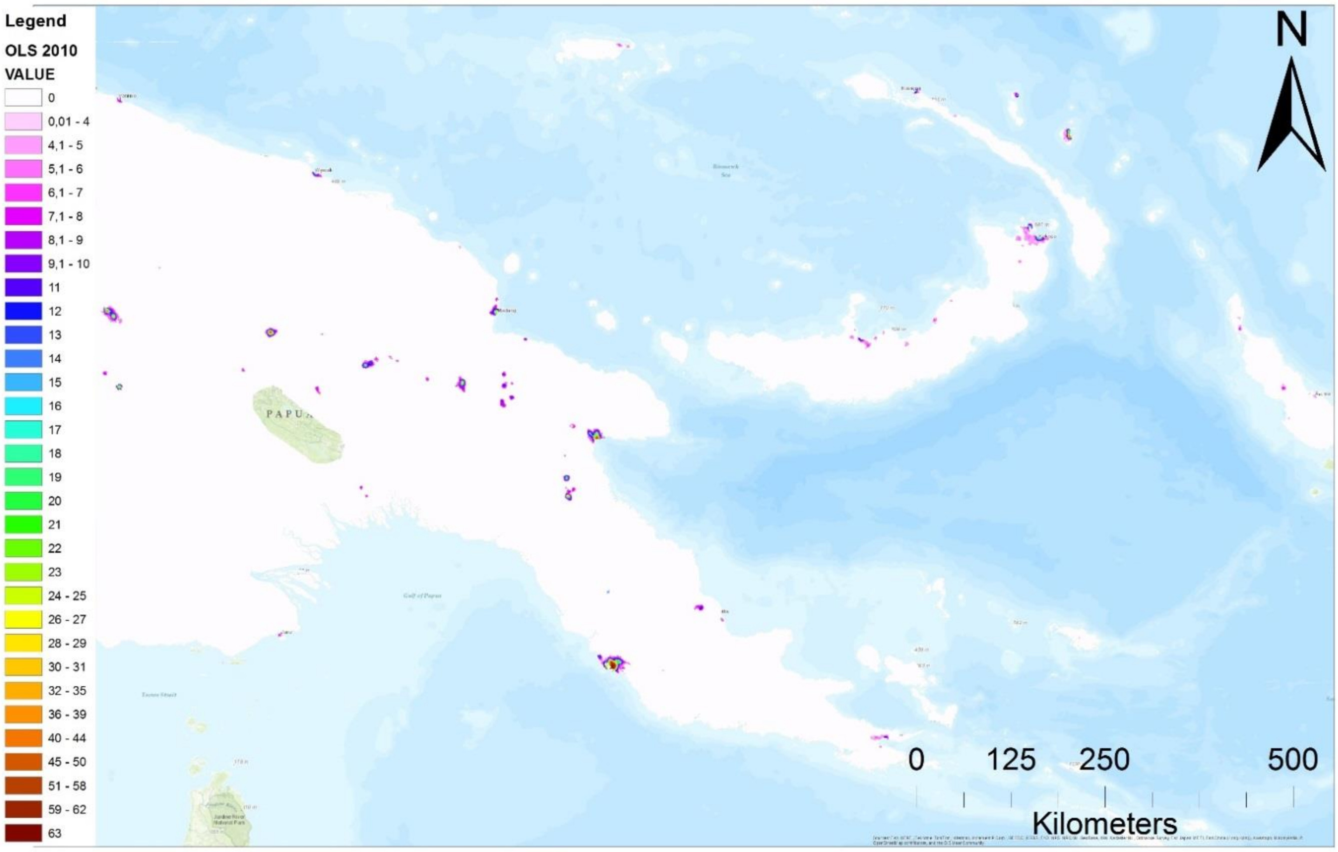

Figure 1: LAN data for 2010

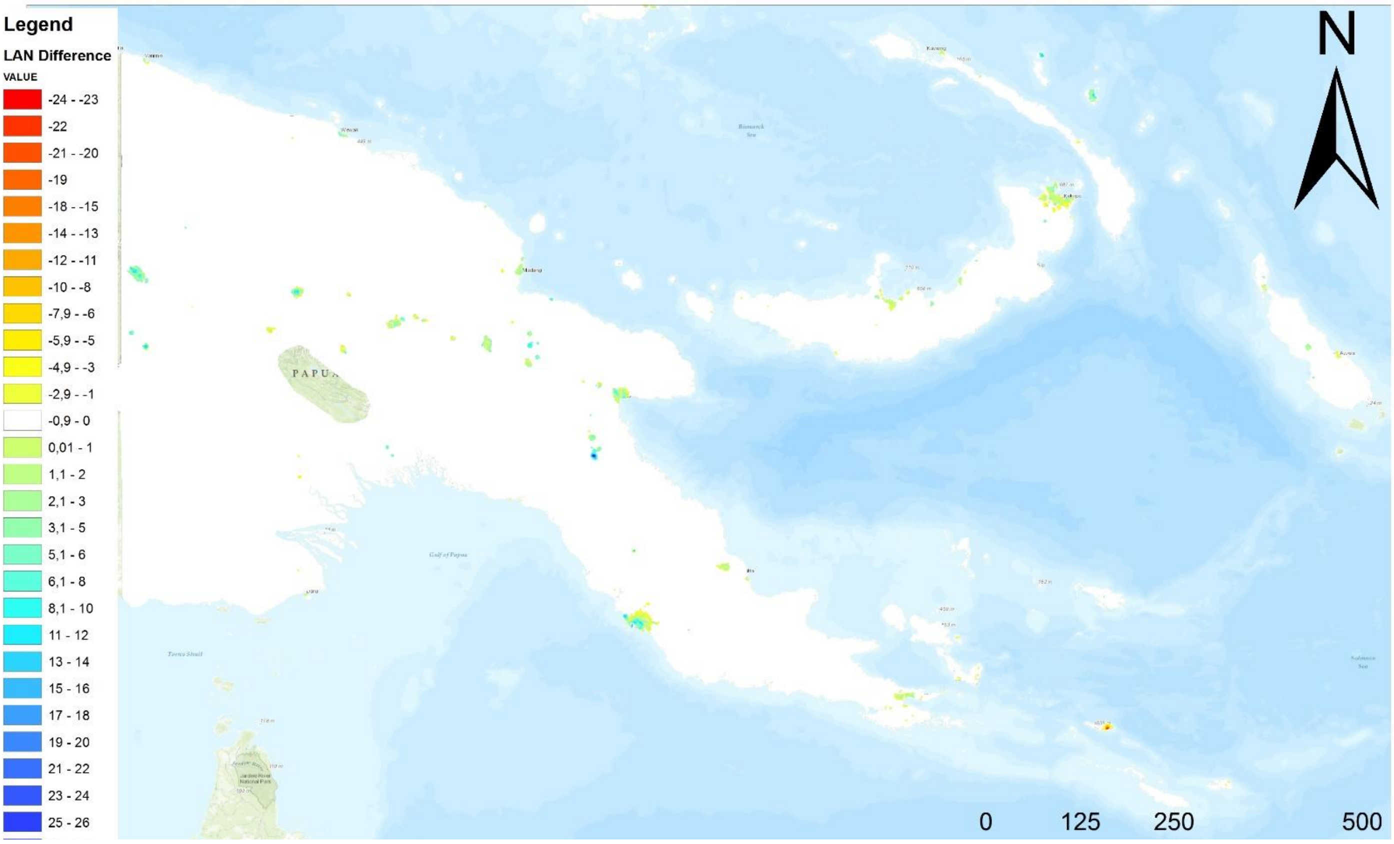

Figure 3: Change in radiance values between 2000 and 2010

3.2 Economic activity data

Figures 5, 6 and 7 are the results of the economic model, all grid cells have a GDP value. The economic activity for 2000 is represented in Figure 4 and shows, the bulk of the economic activity being concentrated in the major mines, the city areas and the highlands. Most of the country has little to no economic activity, which again corresponds with the high forest cover. Exception being the highland area in the centre of PNG and the coastal regions, where economic activity is higher than it is in other rural areas.

On Account of the method used to estimate economic activity, a somewhat unrealistic spread of GDP occurs over areas of PNG, which are largely covered with forest and most likely are uninhabited. This is a result of using GDP averages per capita on a regional level and a population map, which consists of grid cells with very small values (e.g. 0.2), which more or less reflect the probability that people are present rather than realistic population numbers.

Figure 5 shows the economic activity for 2010, the patterns of economic activity have not changed in comparison to 2000. The majority of the economic activity still takes place in the city and mining areas and higher economic activity is still measured in the rural areas of the highlands and near the coast in comparison to other rural areas. Economic activity clearly has increased, as the highest value has increased from 20 to 50 million, raising the entire colour spectrum to higher threshold values in comparison to the 2000 values.

The change map, which is shown in Figure 6, seems to confirm this. Economic activity has increased in the highland areas, while activity in the rural areas in the south-western part of PNG and the rural area north of the highland has decreased. The biggest increase in economic activity took place in the city areas mines. The biggest decrease in activity is the result of the closing of the Misima mine and is presented as green on one of the island in the south-east of PNG

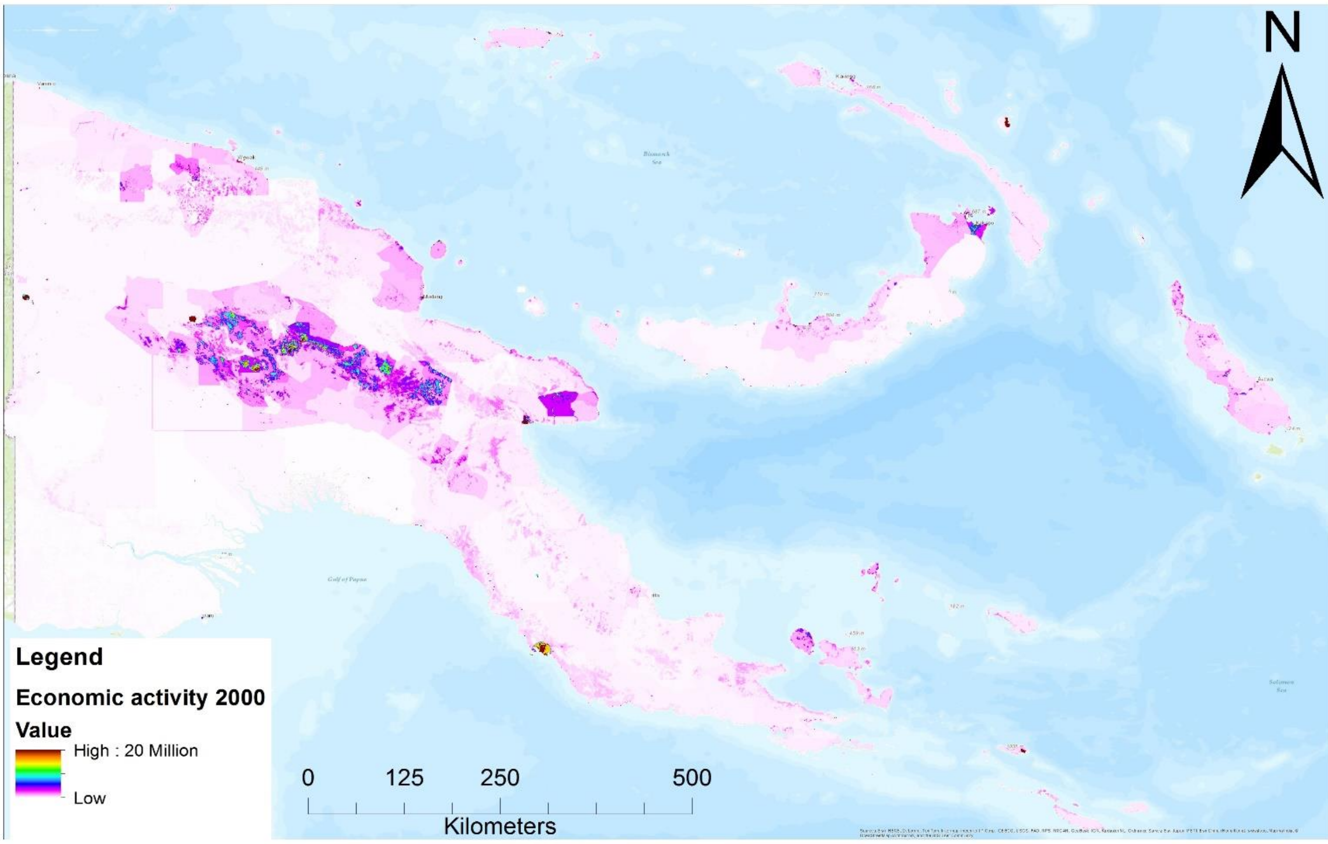

Figure 4: Economic activity for 2000

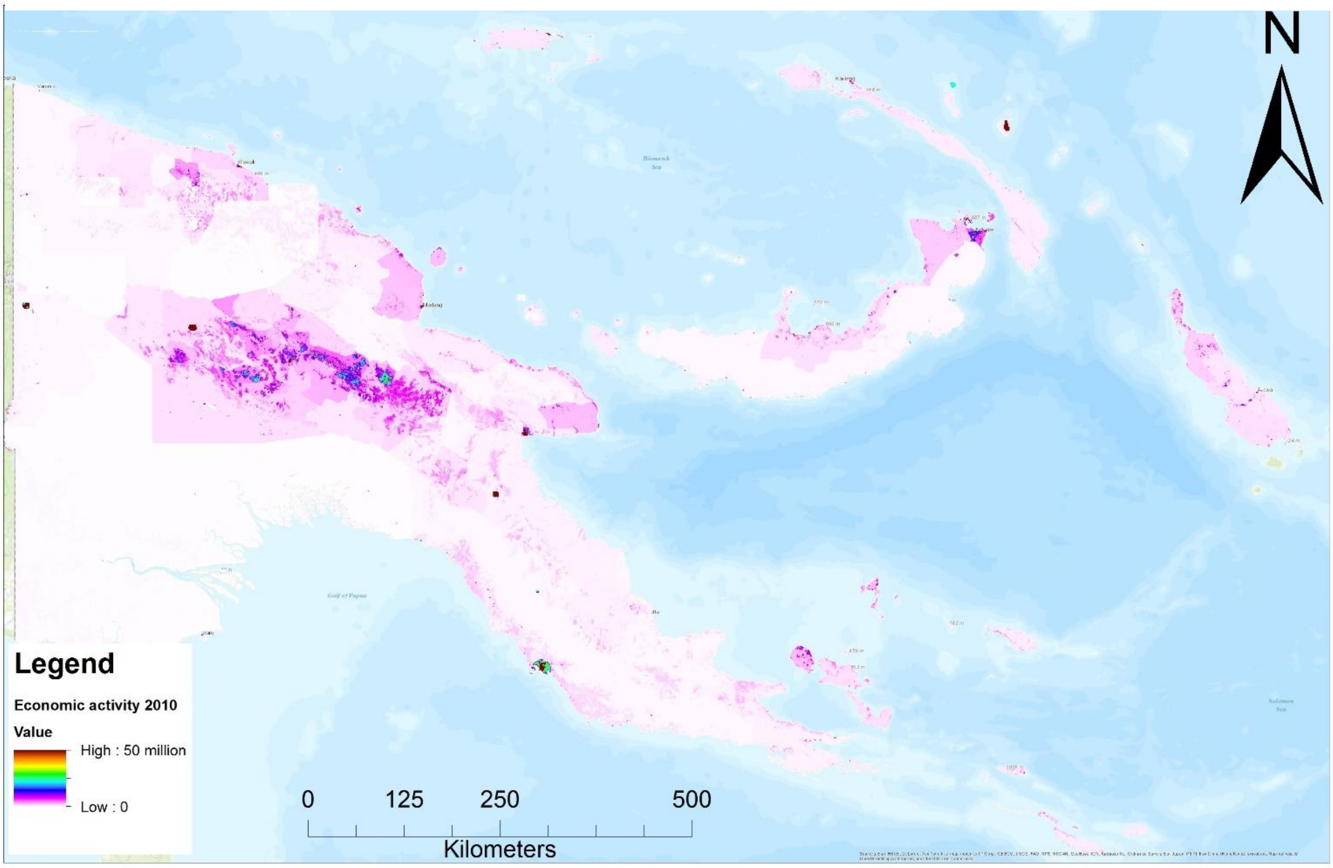

Figure 5: Economic activity for 2010

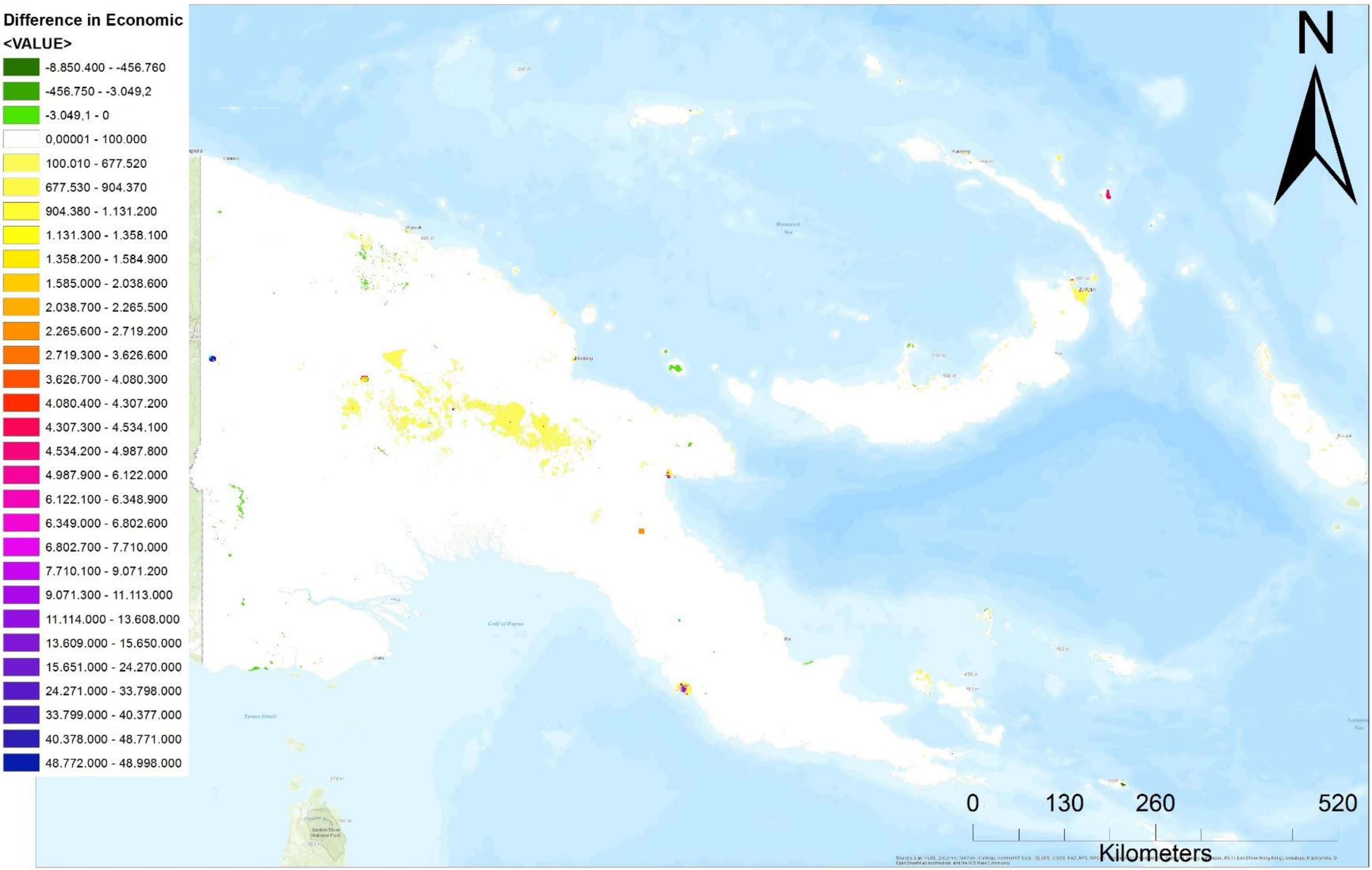

Figure 6: Change in GDP between 2000 and 2010

3.3 Empirical results

Regression analysis, based on the Ordinary Least Squares principles, was done to identify the relation between economic activity and LAN data, it was done for all spatial resolutions. The results are presented in table 2:

Table 2: Regression results

Variable	Prob > F	R-Squared	Coef.
25x25 km			
2000	0.0000	0.60	7.31e-06
2010	0.0000	0.68	4.10e-06
Difference	0.0000	0.49	6.14e-07
12,5x12,5 km			
2000	0.0000	0.72	8.38e-06
2010	0.0000	0.63	3.68e-06
Difference	0.0000	0.46	7.50e-07
5x5 km			
2000	0.0000	0.74	8.51e-06

2010	0.0000	0.72	4.20e-06
Difference	0.0000	0.3	8.92e-07
2,5 x 2,5 km			
2000	0.0000	0.60	6.26e-06
2010	0.0000	0.48	2.91e-06
Difference	0.0000	0.17	3.35e-07
0,5 x 0,5 km			
2000	0.0000	0.34	3.23e-06
2010	0.0000	0.35	2.11e-06
Difference	0.0000	0.07	2.29e-07

The correlation between Light at Night 2000 and Light at Night 2010 was 0.87 and correlation among Economic 2000 and Economic 2010 was 0.69. This shows that, although the two data sets are similar and look alike for both time periods, they are not the same. Moreover, the economic activity seems to have changed more than the LAN did.

All regressions run between the LAN and the economic data are significant at 99% confidence level. Overall, correlation between the 2000 datasets is slightly better than it is for the 2010 data, exception being the 25x25 km data where 2010 correlation is a lot higher and the 0.5x0.5 km data where 2010 data is slightly better. Looking at the correlation results for the different spatial resolutions, the 5x5 km performs best, with both the 2000 and 2010 correlation over 0.7. Even at the highest spatial resolution of 500 by 500 meter correlation values for both 2000 and 2010 reach 0.35.

Overall the correlation values for 2000 and 2010 signify a strong correlation between the economic and LAN data. With 5x5 km spatial resolution performing best and a strong decrease in correlation if spatial resolution is increased further. Correlation values drop from a solid 0.7 to 0.35 values for both 2000 and 2010. The 25 km and 12.5 km spatial resolutions do show strong correlation values, but have spatial resolution that is too coarse to get any higher correlation values since spatial connection is lost at these lower resolutions.

However, the pattern is different for the change values. With an increase in spatial resolution, the correlation of the difference values always decreases. The correlation values for the 25x25 km resolution yields a correlation value of 0,4870, while for the 500x500 meter resolution only results in 0,0724. That is an almost non-existent correlation.

It is clear that at the 500 by 500 meter resolution LAN data is but weakly correlated with the economic data. The high spatial resolution is more sensitive to inaccuracy in data, caused by; for example, different OLS sensors, mistakes in the land use data or locational errors. Change values are even more sensitive, because they are dependent on multiple datasets and the faulty measurements accumulate.

4. Conclusion

Regression results show, that LAN data can be used as a valid proxy for economic activity. Especially to estimate where economic activity is present and where it is likely to be higher in contrast to areas with no or lower radiance values. Correlation is stronger for the 2000 and 2010 datasets than it is for the change dataset.

Decreasing spatial resolution does not benefit the regression beyond the 5 by 5 km point. The correlation still is strong after this, but does not exceed correlation strength for the 5 by 5 km spatial resolution. This lack of increases is the result of loss in spatial relation between the actual origin of the data and the gridcell it is aggregated to.

If the spatial resolution is increased beyond the 5 by 5 km point, the correlation decreases dramatically, dropping almost 0.4 correlation point. I do not think correlation actually decreases, but more likely it is the result of the crude way, I estimated economic activity and the resulting errors in measurement. Increasing spatial resolution beyond the 5 x 5 km point results in more extreme values caused by locational errors, a somewhat unrealistic population distribution, oversaturation of urban cores and the usage of regional averages. Decreasing spatial resolution seems to, in part, correct for these problems and aggregation seems to results in more accurate and realistic results than on the highest spatial resolution.

The regression analysis shows that LAN data can be a useful tool to describe, to some extent, economic activity and the change of this activity, so it is indeed a useful proxy, especially for countries with little accurate statistics. Results show LAN data to be better suited to predict economic activity, than it is to explain change in economic activity. LAN can be used as a stand-alone proxy for economic activity as correlation is strong. However, this should only be done with strong reservation, since correlation is not perfect. For better and more accurate measurement of economic activity, it is therefore recommended to use LAN data in combination with other datasets, to increase robustness and validity.

5. Discussion

In this research, the relation between Light at Night emissions and the spatial distribution of economic activity has been examined. Limitations in the usage of LAN data derived from OLS sensors, are mainly caused by oversaturation of urban cores (Doll, 2008). Although Papua New Guinea has only relatively small and few cities, it would still result in some overestimation. But with only a few cities, little of the LAN data was affected by oversaturation of, so it is less of a problem for the data concerning PNG than it is for developed countries.

Another problem with LAN data is there origin. Over long time periods OLS data is derived from sensors on board different satellites and there is no internal or cross calibration between sensors. So there is no way to know whether changes in radiance are the result of the different sensor or changing ground conditions (Doll, 2008). The 2000 and 2010 data was collected using a different sensor in a different satellite. The difference of the two sensors is noticeable, even when one just glances at the two different datasets. The 2000 data depicts urban spots with larger areas, with a ring of relatively low values around higher central values. The 2010 data shows them in more concentred, smaller areas. For the regression for the models itself this is less of an issue. But when calculating change in these areas, negative values are the result in place where this wouldn't be the case when the same sensor was used.

Change values are extra sensitive to faulty measurement, since they are dependent on multiple datasets. The more variables a dataset uses, the more sensitive it becomes to faults, since the dataset basically accumulates the mistakes of all underlying datasets. This is very noticeable for the LAN data were some of the change values are probably the result of a different sensor instead of changing ground conditions.

The method used to estimate economic activity is a very crude one. Although the census data is as accurate as data can get for PNG, little to no statistics are available beyond those for the national level. This was overcome by making an approximation for statistics on a local level based on other datasets. These approximations are inevitably flawed and will result in unrealistic outcome. These inevitably flaws, will result in faulty observation which, in turn, will result in outliers. Because the resolution of the economic data is so poor, “these estimates are judged to be relatively unreliable” (Nordhaus et al. 2006).

Moreover, the unrealistic spread of GDP as was discussed in section 3.2, could partly be averted with the introduction of a threshold for the population data. This could for example be 1, so large areas most likely uninhabited, will actually have 0 GDP output from the economic model instead of little values as in the economic model for this study.

Better and higher resolution economic data in combination with more accurate LAN data could result in better correlation between the two datasets (Chen and Nordhaus, 2015; Shi et al., 2014) VIIRS data from the SUOMI satellite is a promising new LAN dataset and could prove to be a valuable alternative to the DMSP-OLS data, but still suffers from many childhood diseases and is only available for a limited amount of years.

6. References

Chen, X. Nordhaus, W.(2015). A test of the new VIIRS light data set: Population and economic output in Africa. Remote sensing, 7(1), 4937-4947.

Croft, TA .(1978). Night time images of the earth from space. Scientific American 239(1), 86-96.

Doll, CNH. (2008). CIENSIN thematic guide to night-time light remote sensing and its applications. Technical report, Columbia University.

Doll, CNH. and Pachauri, S. (2010). Estimating rural populations without access to electricity in developing countries through night-time light satellite imagery, Energy Policy, 38(10), 5661-5670

Doll, CNH. Muller, J. and Morley, J. (2006). Mapping regional economic activity from night-time light satellite imagery. Ecological Economics, 57(1), 75-92.

Doll, CNH. Muller, JP. Elvidge CD. (2000). Night time imagery as a tool for global mapping of socioeconomic parameters and greenhouse gas emissions. AMBIO: A Journal of the Human Environment, 29(3), 157-162.

Elvidge, CD. Baugh, KE. Hobson, VR. Kihn, EA. Kroehl, HW. Davis, ER and Cocero, D. (1997b). Satellite inventory of human settlements using nocturnal radiation emissions: a contribution for the global tool chest. Global Change Biology 3(5) 387-395.

Elvidge, CD. Baugh, KE. Kihn, EA. Kroehl, HW. Davis, ER. (1997a). Mapping city lights with the night time data from the DMSP operational linescan system. Photogrammetric engineering and remote sensing, 63(6), 727-734.

Food and Agriculture Organization of the United Nations (2015). Country profile Papua New Guinea. Consulted on May 9, 2015 through <http://www.fao.org/countryprofiles/index/en/?iso3=PNG>.

Gaughan, AE. Stevens, FR. Linard, C. Jia, P. Tatem, AJ. (2013). High resolution population distribution maps for Southeast Asia in 2010 and 2015. PLoS ONE, 8(2), 1-11.

Ghosh, T. Sutton, P. Powell, R. Anderson, S. and Elvidge, CD. (2009). Estimation of Mexico's informal economy using DMSP night time lights data. Remote sensing, 2009(1), 418-444.

Henderson, JV. Storeygard, A. Weil, DN. (2012). Measuring economic growth from outer space. American Economic review, 102(2), 994-1028.

Humanitarian Response (2015). Admin level 3 boundaries attributed with key population statistics from the 2000 senses, obtained from the National statistical office of Papua New Guinea. Consulted on May 15, 2015 through <http://www.humanitarianresponse.info/en/operations/papua-new-guinea/datasets>.

International Monetary Fund (2015). World Economic Outlook Database, April 2015. Consulted on May 6, 2015 through <http://www.imf.org/external/pubs/ft/weo/2015/01/weodata/weoselgr.aspx>.

Knoema (2015). National Accounts Main Aggregates Database, 1970-2013, consulted on May 25, 2015 through <http://knoema.com/UNNAMAD2015/national-accounts-main-aggregates-database-1970-2013>.

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

Linard, C. Gilbert, M. Tatem, AJ. (2011). Assessing the use of global land cover data for guiding large area population distribution modelling. GeoJournal, 76(5), 525-538.

Lowe, M. (2014). Night lights and ArcGIS: A brief guide. Technical report, MIT.

Mellander, C. Stolarick, K. Matheson, Z. and Lobo, J. (2013, June). Night-Time Light Data: A Good Proxy Measure for Economic Activity?. Paper presented to CESIS Electronic Working Paper No.315

National oceanic and atmospheric administration (2015). National geophysical Datacentre version 4 DMSP-OLS night time lights time series. Consulted on May 13, 2015 through <http://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html>.

National statistical Office of Papua New Guinea (2010). Household income and expenditure survey for 2009-2010: NSO.

National statistical Office of Papua New Guinea (2011). Preliminary figures for national population & housing census: NSO.

Nordhaus, W. Azam, Q. Corderi, D. Hood, K. Victor, NM. Mohamed, M. Miltner, A. Weiss, J (2006). The G-econ Database on gridded output: Methods and Data. Technical report, Yale University.

Roychowdhury, K. Jones, SD. and Arrowsmith, C. (2009). Mapping Urban Areas of India from DMSP/OLS Night-time Images. Paper presented to Surveying and Spatial Sciences Institute Biennial International Conference, Adelaide.

Roychowdhury, K. Jones, SD. Arrowsmith, C. and Reinke, K. (2011). A Comparison of High and Low Gain DMSP/OLS Satellite Images for the Study of Socio-Economic Metrics. IEEE Journal of selected topics in applied earth observations and remote sensing, 4(1), 35-42.

Roychowdhury, K. Jones, SD. Arrowsmith, C. and Reinke, K. (2012). Night-time lights and levels of development: A study using DMSP-OLS night time images at the sub-national level. International archives of the photogrammetry, remote sensing and spatial information sciences, 39(b8), 93-98.

Shi, K. Yu, B. Huang, Y. Hu, Y. Yin, B. Chen, Z. Chen, L. Wu, J. (2014). Evaluating the ability of NPP-VIIRS night time light data to estimate the gross domestic product and the electric power consumption of China at multiple scales: A comparison with the DMSP-OLS data. Remote sensing, 6(1), 1705-1724.

Stewart, KR. Lewison, RL. Dunn, DC. Bjorkland, RH. Kelez, S. Halpin, PN. Crowder, LB. (2010). Characterizing Fishing effort and spatial extent of coastal fisheries. Plos One 5(12), 1-8.

Sutton, PC. Elvdige, CD. Ghosh, T. (2007). Estimation of Gross Domestic Product at the Sub-National Scales using nigh ttime Satellite Imagery. International Journal of ecological economics & Statistics 8(7), 5-21.

Trading economics (2015). Roads- paved in Papua New Guinea. Consulted on May 29, 2015 through <http://www.tradingeconomics.com/papua-new-guinea/roads-paved-percent-of-total-roads-wb-data.html>.

United Nations Statistics Division (2015). Country profile for Papua New Guinea. Consulted on May 4, 2015 through <http://data.un.org/CountryProfile.aspx?crName=papua%20new%20guinea>.

World pop (2015). Data Availability. Consulted on May 18, 2015 through <http://www.worldpop.org.uk/data/data_sources/>.

Worldbank (2015). World development Indicators for Papua New Guinea. Consulted on May 8, 2015 through <http://databank.worldbank.org/data/views/reports/tableview.aspx>.

Wu, J. Wang, Z. Li, W. Peng, J. (2013). Exploring factors affecting the relationship between light consumption and GDP based on DMSP/OLS night time satellite imagery. Remote sensing of environment 134(1), 111-119.