Estimation of the Regional GDP using VIIRS nighttime lights

Internship at GRID-Geneva as part of the Geomatics Certificate (UNIGE)

Ena Suarez

1/29/2016

Supervised by: Bruno Chatenoux & Andrea De Bono

From October to January 2016, the UNEP/GRID-Geneva hosted an internship in which the task was to automate the methodology proposed by De Bono et Chatenoux on the A Global Exposure Database presented for the Global Assessment Report on Disaster Risk reduction (GAR) 2015.

Table of Contents

PREFACE

The present work is presented as the result of a full time internship at UNEP[1]/ GRID[2] Geneva, from October to January 2016. The main objective was to improve and automate the methodology proposed by De Bono and Chatenoux (2014) on the conception of a model to estimate the Gross Domestic production at a sub-national level (GRP) using the most recent nighttime lights layers released by the NOAA[3]: The Visible Imaging Radiometer Suite (VIIRS) Day/Night Band Composites. De Bono et Chatenoux (2015) developed a first "semi-automatic" version using the open-source software GRASS and R. They attempted to minimize the manual processes and thus gain in processing speed, all by improving the modeling results.

In this work a new automatic version with some modifications on the original methodology is presented. Commented scripts for GRASS and R are included. The main purpose is to use the results that came out of this study, to improve the Global Exposure Database in the next release of the Global Assessment Report on Disaster Risk reduction (GAR) 2015.

During this internship I was supervised directly by Bruno Chatenoux and Andrea De Bono, both working at GRID-Geneva. This organization (GRID) was launched in 1985 and makes part of UNEP's global group of environmental information centers. Its principal objective[4] is to provide environmental data and related information for decision-making and policy setting.

[1] United Nations Environmental Program
[2] Global Resource Information Database
[3] National Oceanic and Atmospheric Administration of the United States of America
[4] http://www.grid.unep.ch/index.php?option=com_content&view=article&id=1&Itemid=32&lang=en

Introduction

So far the real size of world economies has been measured only at a country level. The Gross Domestic Product (GDP), has been the standard economic parameter to measure economies sizes and to compare wealth, development and poverty between countries. This indicator is estimated by adding all the final goods and services produced within a country in one year (IMF, 2012). The GDP can be expressed in U.S dollars (using the market rates) or in PPP (purchasing power parity)[5]. Market rates are volatile and only relevant to compare internationally traded goods, while PPP rates reflect better the purchasing power of consumers and the relative size of economies (IMF, 2012).

However, the economic activities are not uniformly distributed in space (Krugman, 2011). Thus, national GDP cannot be considered as a reliable measure of the economic value of a region within a country. In this order, it is difficult to analyze environmental data together with economic variables such as the GDP because of their difference in spatial resolution.

Introducing Nighttime lights imagery as a tool to estimate the GDP distribution

A global Exposure Database was developed as part of the Global Assessment Report on Disaster Risk reduction GAR (2015). This database includes a reference grid of 1km spatial resolution in coastal areas and 5km elsewhere, containing information on the capital stock produced per socio economic sector and building typology. To build this global exposure dataset, it is necessary to have an indicator that measures the variation of the economic activity within each country. De Bono and Chatenoux (2015) created a raster of the distribution of GDP (at 1km resolution). This indicator is the result of the integration of tabular information of GDP at regional level or Gross Regional Product (GRP), with the intensities of nighttime lights VIIRS. This approach allows the weighting of the capital stock geographically, limiting the influence of the population and taking into account the regional variations within each country (De Bono and Chatenoux, 2015).

DMSP-OLS Nighttime lights

There are different methodologies to estimate the GRP using Nighttime lights at a country or global level. Initially, studies were based on DMSP-OLS[6] nighttime lights; a digital archive[7] of composite images of stable lights at night (Figure 1). They are collected by NOAA-NGDC[8] since 1992 from the VNIR[9] band of the DMSP sensor (Huang et al. 2014). Elvidge et al. (1997) extracted the area lit from the DMSP-OLS detectable lights to estimate the GDP of 21 countries applying a linear regression. This model showed a coefficient of determination (R^2) of 0.97 (Figure 2), and therefore supports the hypothesis that Nighttime lights can be successfully used to distribute spatially the economic activity.

[5] International Comparison Program (ICP) defines PPP rates as: "…price relatives that show the ratio of the prices in national currencies of the same good or service in different countries…. They show -with a reference to a base economy or region- the price of a given basket of goods and services in each of the economies being compared" (World Bank, 2015). PPP rates are estimated each six years by the ICP (last publication was on 2011). There is data for 199 countries which account for 90% of the world economies. There is no data for 15 economies like Argentina, Libya and Afghanistan but there are some recommended estimations (World Bank, 2015).

[6] Defense Meteorological Satellite program- Operational Line Scanner (DMSP-OLS).

[7] The digital archive of Nighttime lights time series from 1992-2013 is available at http://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html

[8] National Oceanic and Atmospheric Administration and the National Geophysical Data Center

[9] Visible and Near Infrared

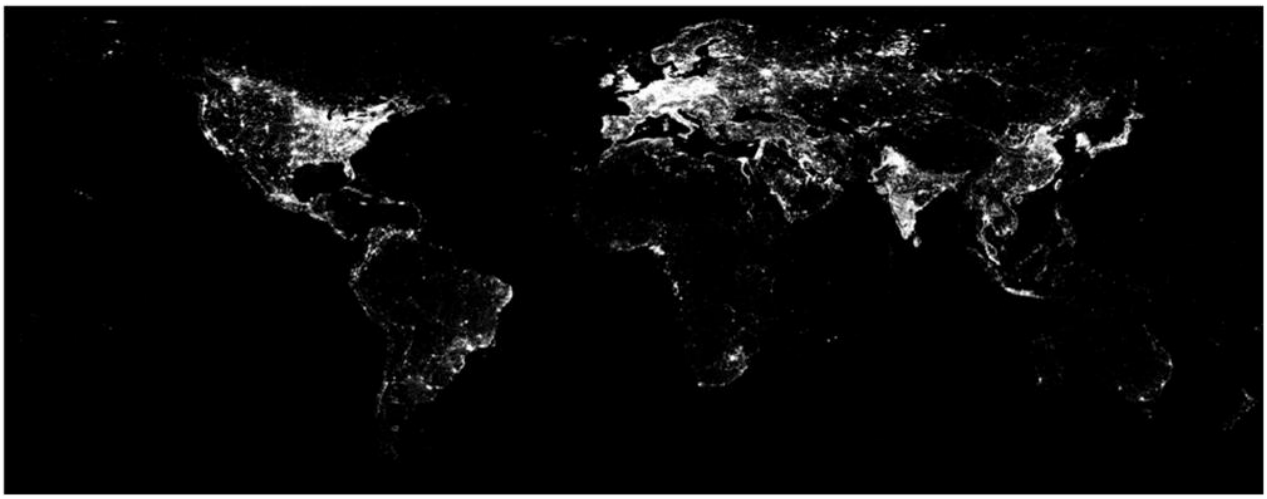

Figure 1 DMSP Nighttime lights composite of 2010 stable lights. Figure extracted from Li et al. (2015)

Later, Doll et al. (2006) correlated nighttime lights with the GRP of the United States and 11 European Union countries and created map for each showing the spatial distribution of the economic activity at a 5 km resolution. For each country a spatial disaggregation process was made to create GRP grids which were merged afterwards to create one map for the U.S.A and another for Europe (Figure 3). One of the main problems they detected was the outliers. Cities with high economic level and high radiance, as Paris and Brussels were major outliers for their countries; therefore, they were threaded separately to increase the accuracy of the model. They suggest incorporating information on land use, which would give an idea of the economic activity of a particular area, outliers would decrease and probably a single model would be applicable to all areas.

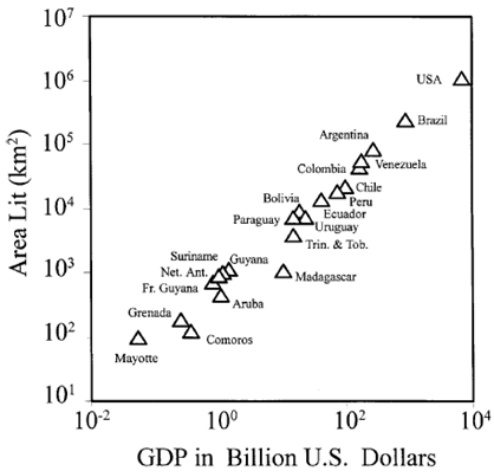

Figure 2 Estimation of 1994 GDP for 21 countries using the area lit (Nighttime lights DMSP). Axes are log transformed. Figure extracted from Elvidge et al. 1997

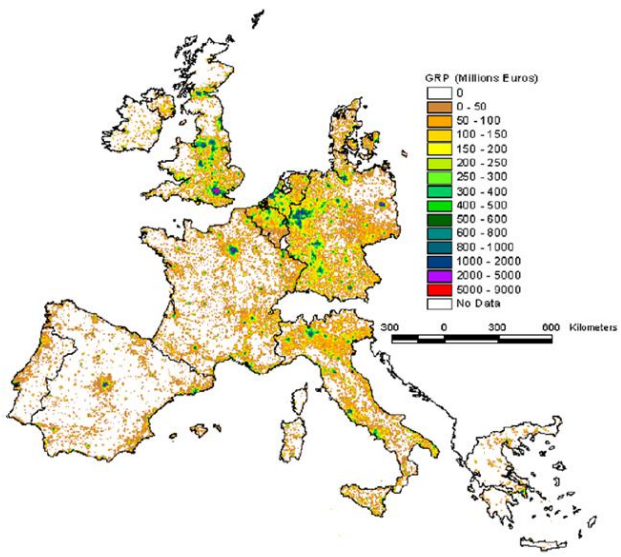

Figure 3 Spatial distribution of the GRP for 11 European Countries, based on DMSP-OLS Nighttime lights. Figure extracted from Doll et al. 2006.

Ghosh et al. (2010) published another methodology on the global distribution of economic activity using shedding light. It included the GRP for the United States, China, India and Mexico and the GDP for the rest of the countries. Additionally, this model included the Informal economy, which is excluded from the official GDP data, and a correction factor for agriculture. Ghosh et al. (2010) argued that in a country where agriculture counts for a high percentage of the national economy, nighttime lights would not be an accurate estimation of the GRP. Nighttime lights would allocate agricultural production in lit areas and not in the fields. Their solution was to add two grids (one of agricultural activity and

another of the non-agricultural activity) to create a map showing the aggregated GDP per countries and regions (Figure 4). This methodology allows the representation of the economic growth at different spatial scales, which is very useful when environmental data needs to be linked with the GDP.

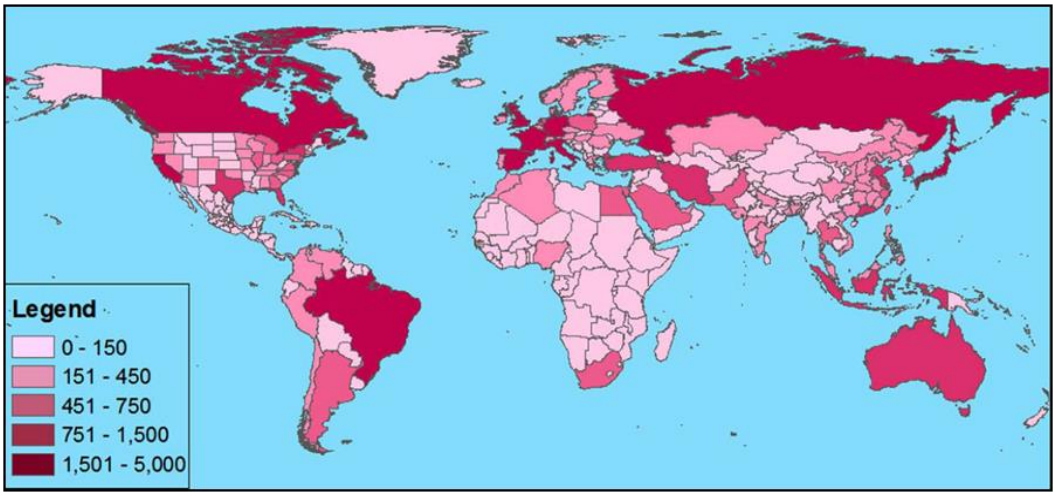

Figure 4 Estimation of the GDP at national or regional level in USD dollars. Figure extracted from Ghosh et al. 2010.

Regarding the data availability, the last version of DMSP Nighttime lights (Version 4) include 30 arc second grids with a nominal spatial resolution of 2.7 km (Baugh et al. 2013). Composites are mostly cloud-free. They include persistent lighting from cities and towns but also gas flares. Background noise and ephemeral effects have been removed, which is the principal advantage of using the DMSP-OLS data (Ghosh et al. 2010; Baugh et al. 2013; Shi et al. 2014). By contrast, there are three main disadvantages of using this data; first, there is light over-saturation of urban centers due to its 6-bit dynamic range (Baugh et al. 2013; Elvidge et al. 2013). In fact, DMSP-OLS data include radiance values up to 10^{-8} $W{\cdot}cm^{-2}{\cdot}sr^{-1}{\cdot}um^{-1}$; higher values receive a maximum pixel value of 63. This creates a distortion of large city centers where there is strong artificial lighting (Shi et al. 2014). Further disadvantages are the low spatial resolution; the lack of onboard calibration and the temporal resolution, e.g. annual composites only (Elvidge et al. 2013).

VIIRS Nighttime lights

NASA[10] and NOAA launched in 2012 the Suomi National Polar Partnership (SNPP) satellite carrying the first visible infrared imaging radiometer (VIIRS), this instrument is designed to collect high quality radiometric data for digital analysis and input into numerical model (Elvidge et al. 2013). VIIRS is the most recent nighttime lights dataset (Figure 5), replacing the DMSP-OLS dataset.

It has three main advantages. First, it has higher spatial resolution, 742 m versus 5km of the DMSP (Elvidge et al. 2013). Second VIIRS data has a higher radiometric detection rate (0.02 to 3 x $10^{\wedge -9}$ $W{\cdot}cm^{-2}{\cdot}sr^{-1}{\cdot}um^{-1}$) and a quantization of 14 bit versus 6 bit of the DMSP-OLS (Elvidge et al. 2013). This would solve the problem of pixel saturation present in the DMSP-OLS data. Third, VIIRS counts with onboard calibration. Therefore, different composites can be joined and Nighttime Lights Time Series can be analyzed without an inter-satellite calibration (Elvidge et al. 2013; Li et al. 2014). Up to 2015, NOOA has released a version 1.0 of nighttime VIIRS day/night Band composites[11] (DNB). These composites are cloud-free and the background noise coming from stray light, lightning, lunar illumination have been eliminated in order to improve the

[10] National Aeronautics and Space administration.
[11] http://ngdc.noaa.gov/eog/viirs/download_monthly.html

image quality. However, composites still contain light detections from fires, aurora, boats and other temporal lights (NOOA). There are monthly nighttime VIIRS DNB composites from January 2014 to November 2015. For 2012 and 2013 there are only beta versions. Hence, one of the first disadvantages of the VIIRS data is that it cannot be used to analyze the economic growth before 2014.

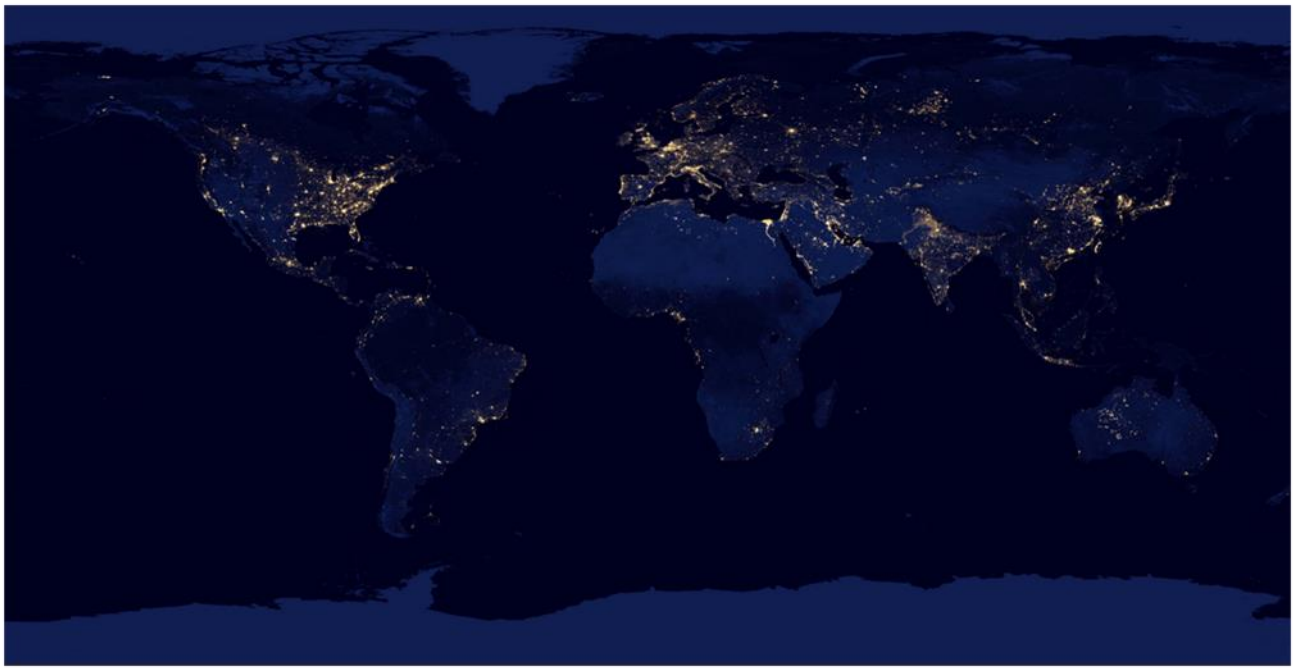

Figure 5 VIIRS Nighttime lights in 2012 (April and October). Source: NASA Earth Observatory/NOAA NGDC

This is a very recent database; therefore, there are few studies exploring the potential of VIIRS to estimate socio-economic parameters, up-to-date we found 9 related articles. These studies are focused on China (8) or Africa (1). Articles about China are addressed on the estimation of poverty (Yu et al. 2015), inequality (Zhou et al. 2015), GDP (Li et al. 2013; Ma et al. 2014; Shi et al. 2014) or the built up-area (Shi et al. 2015). In Africa, the sole study is based on the evaluation of the GDP. For example, in the publication by Shi et al. (2014) their objective was to estimate China's GDP at the regional level using the VIRRS nighttime light data. To do so, they performed a linear regression (form:$\hat{y} = \beta_0 + \beta_1 X$) where the explanatory variable was the sum of all the pixel values in each region and the slope was considered the "regression coefficient". To decrease the effect of outliers they established that Digital number (DN) values at the biggest/largest cities should not be exceeded by DN values of other areas. If there was a pixel with a higher value, it was redefined with the maximal DN value within the pixel's immediate neighbors. They compared results between VIIRS and DMSP data (Figure 6) and they found that VIIRS images were more reliable for estimating GDP; as they obtained a higher correlation coefficients and higher R^2 values. One of the main drawbacks of using the DMSP data was the distortion of large city centers with strong artificial lighting due to its limited dynamic range (Figure 6).

De bono and Chatenoux (2015) developed a different methodology. Their principal objective was to make a global estimation of the capital stock using a raster of the geographic distribution of the subnational gross domestic product (GRP). To create this raster, they used VIIRS nighttime lights as a proxy for human activity. Before correlating lights with the GRP they suggest to preprocess VIIRS nighttime lights composites. They start by creating a light density layer, lights normalization. The next step was to search for the best light density range, by an iterative process, in which the R^2 resulting from the Linear Regression (LR) between GRP and VIIRS nighttime lights was the highest of all the combinations. Once they found the

best light range for each group, as they classified countries by income, they estimated the GRP using the LR equations and the total sum of nighttime lights. They also calculated a correction factor by comparing the initial GRP with the predicted GRP. The final result was a map showing the variation between the GDP at national scale and its regional variations.

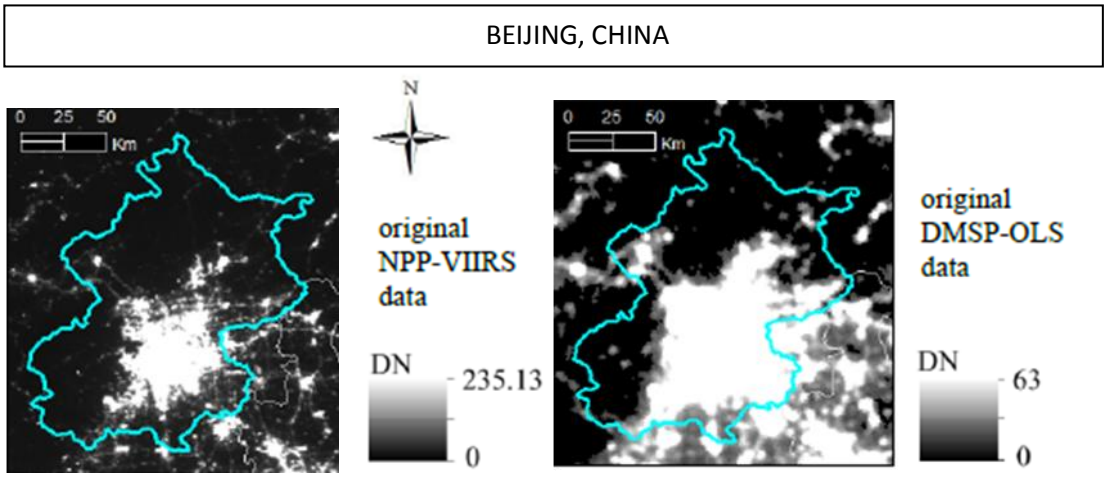

Figure 6 Beijing Nighttime lights in 2012 comparison between the NPP-VIIRS and the DMSP-OLS composites. Figure extracted from Shi et al. 2015

Objectives

There are several investigations demonstrating that nighttime lights are good indicators of global economic activity. Now, in order to improve the global exposure dataset (mentioned above) we will create a linear regression model that predicts the global distribution of the GRP (GDP regional level) using Nighttime lights VIIRS DND-cloud free composites. We will base our methodology principally on De Bono and Chatenoux (2014) but also on previous studies, such as Ghosh et al. (2010) and Shi et al. (2014). Additionally, manual processes will be minimized in order to, accelerate the information processing required to improve the overall model quality. The GRP will be estimated for the year 2014 using 2014 monthly VIIRS composites.

Data collections

Version 1 Nighttime VIIRS Day/Night Band Composites for 2014

Nighttime VIIRS DNB composites (Version 1.0) were obtained from the NGDC-NOAA Earth Observation Group (EOG) official website[12]. These composites have been filtered to remove the cloud cover, stray light, lunar illumination and lighting. However, there is still noise from temporal lights such as boats, fires, volcano or aurora. For each month of the year 2014 there are 6 tiles at a 15 arc-second resolution (GeoTIFF format). Each tile spans 120 degrees up to the equator and the whole composite spans the globe from 75N latitude to 65S. Raster's radiance values (32-bit floating point) units are in 10^9 nanowatts/cm^2/Sr.

[12] http://ngdc.noaa.gov/eog/viirs/download_monthly.html.

Official GDP and GRP worldwide database for 2014

GDP values at market prices (Constant 2005 USD) for the year 2014 were obtained from the World Bank's database[13]. For five countries, [14] for which there was no information for 2014, the GDP was calculated by extrapolating 2012-2014 values. In the specific case of the Syrian Arab Republic, the GDP was extrapolated from 2010 and 2013 values, taken from the United Nations Statistics Division[15].

At the present, there are no worldwide official GRP databases. Merging OECD[16] , EUROSTAT[17] mission and the DECRG[18] we obtained information for 69 countries (Table 1). China, India, USA, Brazil, Russia and Mexico, which are the most populated and largest countries of the world, make part of these databases. Conversely, there is only regional-level information for 11 African countries. Theoretically, data can be obtained from the official countries statistics; however, the quality of the data differs among them. For some countries there are no national standards for regional accounting so there is a lack of consistency in the methodology and low-quality GRP data (Ghosh et al. 2013). This evidence the importance of finding an alternative method to calculate the GDP that doesn't depend on the national accounts or survey means (Pinkovsky et al. 2015).

Table 1 GRP global databases

Database	Timeline	Countries and territories	Characteristics
OECD[19]	2000-2014	41	OECD members including large countries such as China, India, USA
EUROSTAT[20]	2000-2013	28	European Union countries. Detailed accounts
DECRG (World Bank)[21]	2010	50	Collection of GDP and GRP data

The solution was to use the 2014 GDP data that we collected together with the global dataset[22] of the national and subnational GDP for the year 2010 (Unep/GRID-Geneva) and extrapolate values to 2014. In the final database we had information of 191 countries; 69 countries with GRP values (1966 regions) and the rest, 122, only with national GDP data. For every country/region we calculated not only the total GDP/GRP but also the corresponding contributions of the industrial and service sectors.

Global administrative Areas map

All countries have administrative divisions, with few exceptions[23] (Gennaiolli et al. 2013). These divisions can have different levels, for example there are Cantons and Communes in Switzerland or States and counties in the United States. In this study we will focus on the highest level of each country

[13] http://data.worldbank.org/indicator/NY.GDP.MKTP.CD
[14] Trinidad & Tobago, Tunisia, Tuvalu, Vanuatu, Yemen Rep.
[15] http://unstats.un.org/unsd/snaama/dnllist.asp
[16] Organization for economic co-operation and development
[17] The European Union official statistical office: the EUROSTAT mission.
[18] The World Bank's Development Research Group.
[19] https://stats.oecd.org/Index.aspx?DataSetCode=REGACC_TL2
[20] http://www.cgdev.org/blog/global-map-subnational-gdp
[21] http://preview.grid.unep.ch/index.php?preview=data&events=socec&evcat=1
[22] Gross Domestic Product 2010. GIS processing by World Bank DECRG, Washington, DC; extrapolation by UNEP/GRID-Geneva. Available at: http://preview.grid.unep.ch/index.php?preview=data&events=socec&evcat=1
[23] Cook Islands, Hong Kong, Isle of Man, Macau, Malta, Monaco, Niue, Puerto Rico, Singapore, Tuvalu, Vatican City.

administrative or statistics division (administrative level 1). We will use the same administrative subdivisions as the "Gross Domestic Product 2010" dataset, already mentioned in the previous section which includes countries and region boundaries. In this way we will have the national boundaries for countries with GDP data and the regional boundaries for those countries with GRP data.

Countries classification

When correlating nighttime lights with economic activity it is important to classify countries (and its regions) rather than considering them all in a single model. This will give better statistical results and will help to avoid outliers (De Bono & Chatenoux, 2015; Ghosh et al. 2010). The most common option is to classify countries by grouping term. The World Bank divides the world in 7 geographic regions (WB region) as in Table 2 World Bank Countries classification by geographic region.

Table 2 World Bank Countries classification by geographic region

Country Classification by WB Region
1. East Asia & Pacific
2. Europe & Central Asia
3. South Asia
4. North America
5. Latin America & Caribbean
6. Middle East & North Africa.
7. Sub-Saharan Africa

Table 3 World Bank World countries classification by Income

Income group	GNI per capita (USD)
Low	$\leq \$1{,}045$
Lower-middle	$1{,}045 < GNI_{percapita} \leq 4.125$
Upper-middle	$\$4.125 < GNI_{percapita} < 12.746$
High	$12.746 \leq$

Another option is to use the World Bank's country classification by income. It has four groups: low, lower-middle, upper-middle and high-income. This classification is based on the GNI[24] per capita (USD). The group thresholds are updated each year; 2014 classification is described in Table 3. High income economies can also be divided into two groups: High-income OECD[25] members and High-income nonOECD. This additional division distinguishes OECD members as "industrial countries». There is a difference of ≈10,000 USD in their GNI per capita between these two groups (Figure 7).

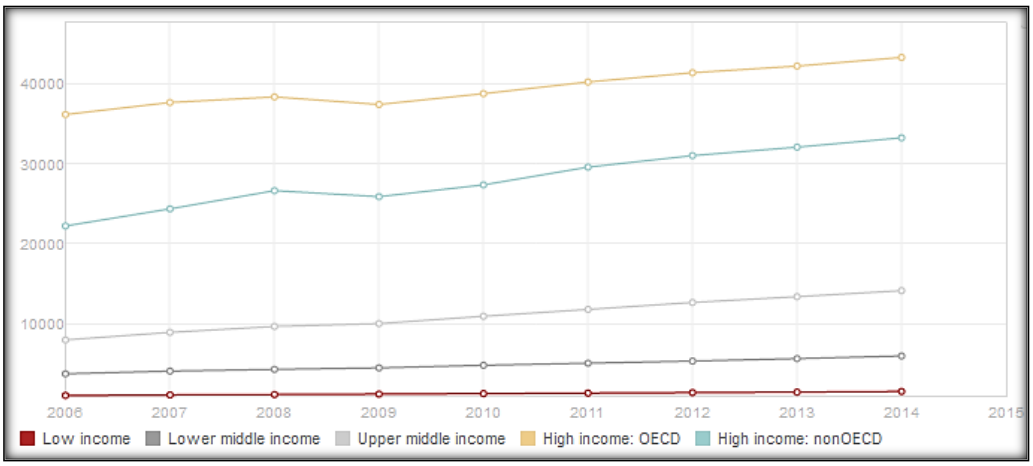

Figure 7 Illustration of world Countries Classification by Income and OECD membership Figure extracted from: World Bank Data.

[24] Gross national product (GNP)
[25] Organization for Economic Co-operation and development (OECD)

The International Monetary Fund (IFM) suggest a more detailed classification of the world countries which is based in two main analytical groups[26]: "advanced economies" and "emerging market and developing countries". The second group can be divided in six sub-groups: 2.1) Commonwealth of Independent States, 2.2) Emerging and Developing Asia, 2.3) Emerging and Developing Europe, 2.4) Latin America and the Caribbean, 2.5) Middle East and 2.6) North Africa, Afghanistan, and Pakistan. The resulting classification for the year 2014 is presented in (Figure 8).

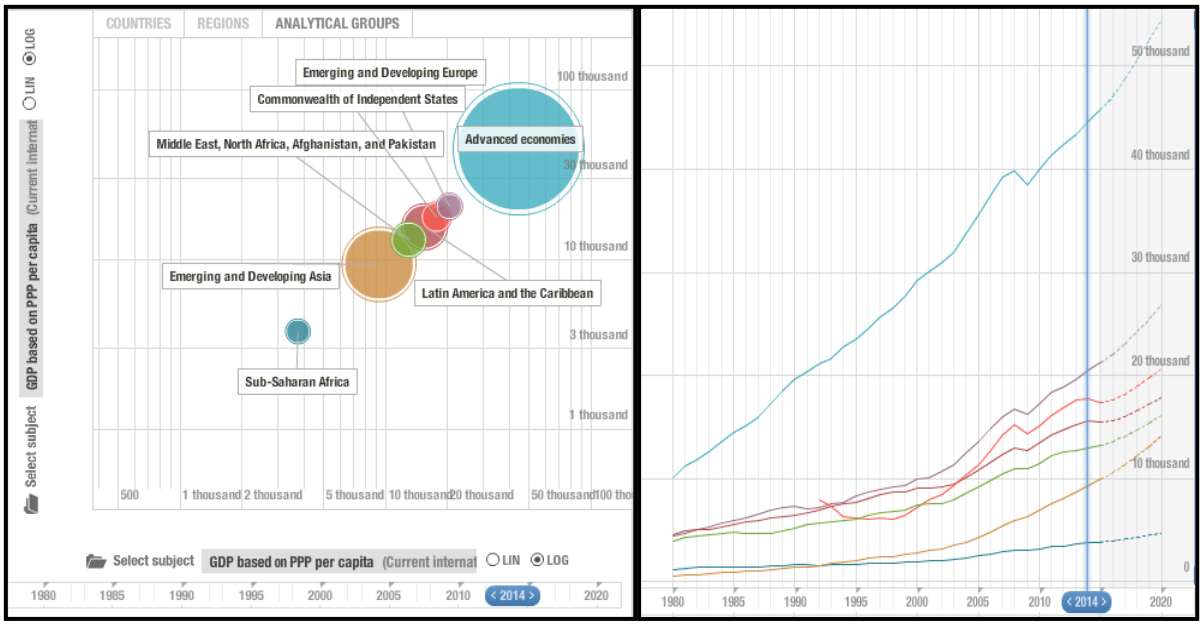

Figure 8 IMF country classification by Analytical Groups. Global distribution of the GDP per capita in PPP for the year 2014. Modified figures from: The IMF data mapper (http://www.imf.org/external/datamapper/index.php)

We will test these three different types of classification; Region (World Bank), Income (World Bank) and analytical groups (IMF), In order to know which classification fits better to predict the GRP.

BUREF –The Global Built-Up Reference Layer 2010

The unpublished built-up Reference (BUREF) data layer is so far considered as the best estimation of built-up areas at a global scale (Peserasi & Carneiro, 2014). It was created by integrating 2010 data of the global urban extent layer, known as MODIS 500[27], and the global population distribution data, the LandScan[28] population grid. Each spatial unit in the BUREF layer represents the percentage of built-up area (0-100) with respect to the total surface of the cell. BUREF's spatial resolution is 30 arc seconds referenced in the WGS84 coordinate system (Peserasi & Carneiro, 2014). We will use this layer to identify the points with the highest percentage of built-up area in each region/country.

[26] IMF methodology on how the classify countries is available at: www.imf.org/external/pubs/ft/weo/2015/02/weodata/
[27] Moderate Resolution Imaging Spectroradiometer (MODIS) 500-m satellite data. More information at: https://nelson.wisc.edu/sage/data-and-models/schneider.php
[28] LandScan official description available at: http://web.ornl.gov/sci/landscan/

Methods

Processing of Nighttime Lights 2014 (VIIRS DNB composites)

1. **Download and processing of monthly VIIRS-DNB composites for 2014:**
   From January to December 2014 there were 6 tiles per month this means **72 raster images** for the whole year. The size of each tile was around 1-2 G thus it was necessary to resize them to accelerate the processing time. Each pixel was multiplied by 100, and the data type was set to UInt 32[29]. Next, 12 mosaics (one per month, Figure9) were produced. Each mosaic spans the earth from 75N latitude to 65N and 180W longitude to 180E. These rasters surpassed the limits of the WSG84 projection by some units (0.00002). In GRASS, this is not accepted therefore, it was necessary to extract a sub-section that expands the globe from 75N to 65S and from 179.999W to 179.999E (Subtract one pixel). These new rasters were still big in size, 11G each, so it was necessary to do a LZW compression to obtain 2G global composites. The original and the compressed files were compared in order to verify that the manipulation of the files didn't modify the data.

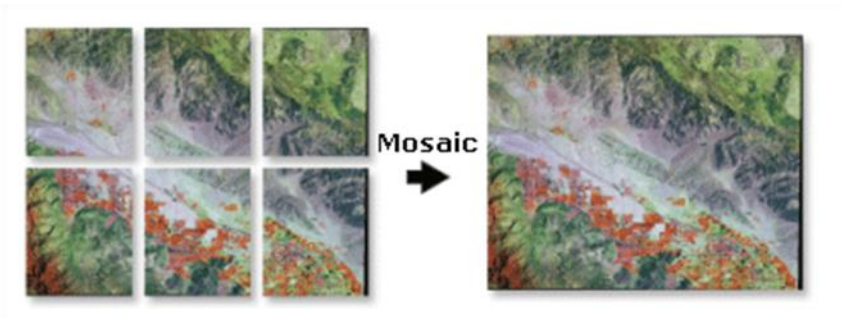

Figure9 Mosaic tool. Create a unique raster from 6 adjacent rasters (tiles)

2. **Creation of a unique VIIRS nighttime lights composite for 2014**:
   Once we had a VIIRS mosaic per month we created a unique composite for 2014. Using the GRASS tool "*Raster map calculator*" we created 4 different maps, the first with the average values of the 12 mosaics, another with the maximum values, the minimum values and the last with the median.

   The "Maximum" and "Minimum" maps didn't reflect the yearly distribution of the nighttime lights, so they were discarded. The "Average" and "Median" maps yielded better results. In order to decide which raster was the best option, we inspected simultaneously both rasters to identify where the difference rely on. We noticed that the "Average" map display higher and more frequeTNLy peak of light than the "Median". These peaks didn't correspond to built-up areas as we verify with the BUREF and Modis layer. In Purovsky District, Russia for example which is rich in lakes and rivers, the "Average" raster showed high and big light peaks in lakes and other water bodies. Contrarily the "Median" rasters didn't show this peaks and it fits well the BUREF layer. Following this, we decided to use the raster of the monthly median VIIRS nighttime lights as the Annual VIIRS nighttime lights composite for 2014.

3. **Lights Normalization:**
   Once we had a unique raster of VIIRS nighttime lights for 2014, we created a light density layer (Figure 10), in order to compensate for projection distortions, as in De Bono and Chatenoux (2014). Each pixel in the annual VIIRS nighttime lights composite was divided by its real area; the resulting units are given in light intensity/km^2.

[29] 32-bit unsigned integer, UInt32. The output range is 0 to 4'294'967'295

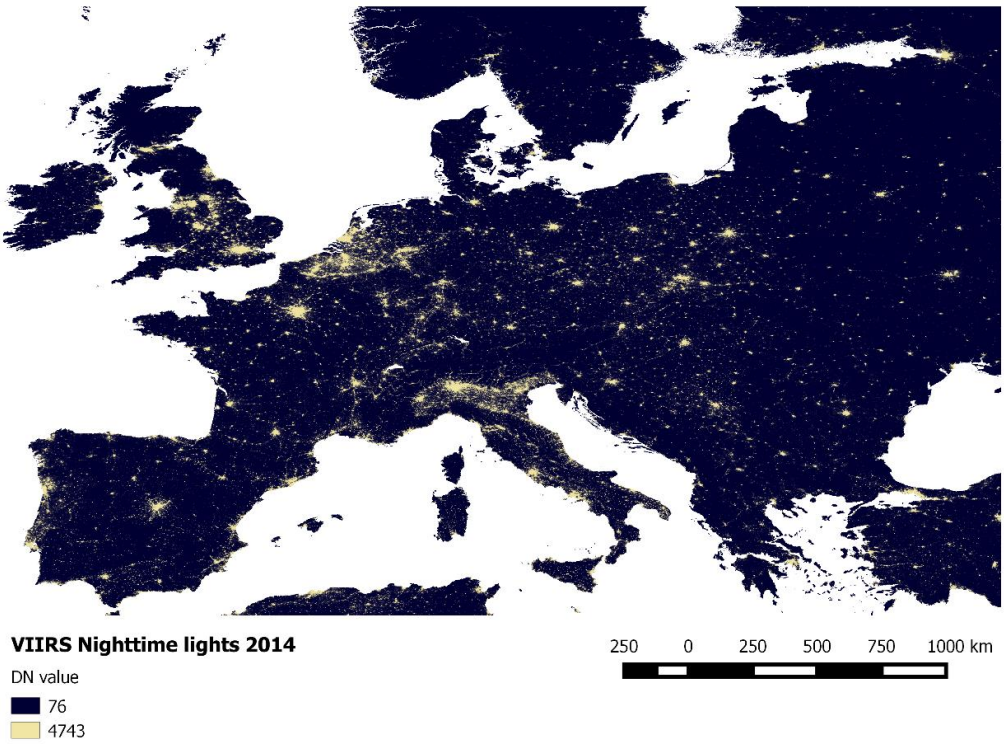

Figure 10 VIIRS nighttime lights composite for 2014, zoomed in Europe. This map represents the median values of Nighttime lights in the year 2014 after lights normalization.

4. VIIRS nighttime lights outlier's correction using BUREF 2010:

As mentioned in Shi et al. (2014) VIIRS composites can contain some outliers as for example stable lights of oil or gas wells and therefore they should be removed from the VIIRS composite. Theoretically, the biggest and most developed city within an administrative unit should have the biggest DN values (Shi et al. 2014; Yu et al. 2015). Then if there are pixels with higher values that this threshold they should be corrected. In our case we decided to correct the 2014 VIIRS composite using the BUREF 2010 layer. For each of the 2088 administrative units we identified the points with the highest percentage of built-up area. Once we had located these points we extracted the corresponding value of Nighttime lights at this point. This DN value was set at the maximum threshold. For each region we had a maximum Light radiance value related to the areas with the highest built-up area. If within an area, there were pixels with greater values they were reassigned with the maximum DN value established for the corresponding region.

In order to have a better understating of the methodology there is an explanatory diagram on Figure 11.

5. Calculate the Total nighttime lights (TNL)

Once we had the corrected composite, we were ready to perform a *Zonal Statistic*[30] analysis of the VIIRS nighttime lights 2014 composite; using the GRP boundaries map as the determinant of the zones of analysis. As a result, we obtained a CVS file with all the univariate statistics: maximum, minimum, average, standard deviation and the sum of VIIRS nighttime lights per region or country. From this table we abstracted the sum and we established the Total

[30] To execute this analysis we used the GRASS function "r.univar" A complete description is available at: https://grass.osgeo.org/grass70/manuals/r.univar.html

Nighttime lights (TNL) as the sum of all the pixels within an administrative region (Li et al. 2013; Shi et al. 2014; De Bono and Chatenoux, 2014)

6. **Iteration: increasing BUREF minimum percentage of urban-built up area.**

Steps 4 and 5 were repeated in sequence, each time increasing BUREF minimum threshold. The first time we did step 4 and 5, we started by setting the minimum BUREF value to 5% (all the data above this number was masked). The Next time, we increased BUREF minimum value to 7% and so on. In resume we masked BUREF layer starting in 5% until we reach 100%, each time with an increasing step of 2%.

The purpose of repeating this process was to collect information from different BUREF and light density ranges and test them all in order to find later, the best range of; light density and minimum built-up area, in which the R^2 is the highest of all the Linear regression between the GDP/GRP and the Total Nighttime Lights.

Steps from 1 to 6 are automated. There is a first script called “VIIRS_data_preparation” which includes the necessary commands to perform 1 and 2, using MSYS and GRASS 7.0. Next, a bash script called “input_data”” resumes step 3, and its intended to run with GRASS 7.0 or later. Steps 4 and 5 can be performed once using the script “VIIRS_step_one”. The user will need to give the desired BUREF minimum level. If the user wants to repeat the process several times as in step 6, call the script “loop_VIIRS_grass” and include the start, end and increment values to perform the loop sequence.

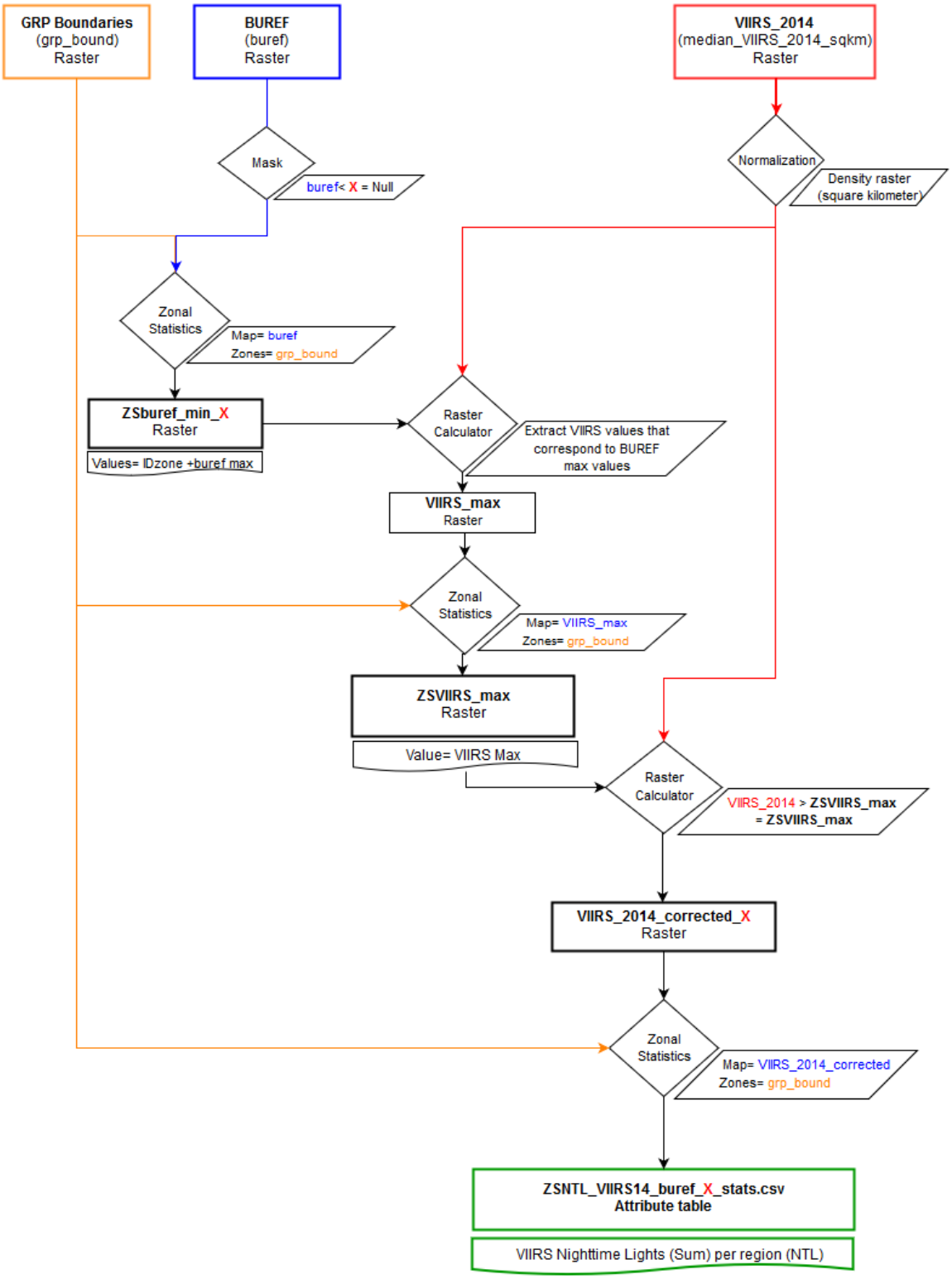

Figure 11 Methodology explaining how to correct VIIRS nighttime lights composites using BUREF 2010, followed by the extraction of the Total nighttime lights per region (NTL). The functions described in this diagram are principally directed to be executed on GRASS 7.0 or later.

Linear regression model $GRP_x \sim \beta_0 + \beta_1 NTL_x$

The next step would be to test the different TNL thresholds to find the best regression parameters to correlate GRP and TNL**.** As one of the main objectives was also to reduce the manual processes, we decided to create an R package[31] called "GRPlights" that included all the functions, data and coding necessary to calculate the GRP using VIIRS nighttime lights. It is also possible to perform a linear regression for each TNL iteration. The data of each regression is stored on an excel file so that data can be easily analyzed by the user of the package.

The structure of the GRPlights package (Figure 12) includes 9 different functions[32] and 3 oracle Storage Area Files (. RDA):

I. GRP_2014: Contains the GDP, GRP, the Industrial, Service and Industrial and service GRP for the year 2014. This database includes a total of 191 countries and 1966 regions.
II. WB_class: Contains World Bank Classification by: Income, Region and Lending.
III. IMF_class: Contains International monetary foundation classification by: Region and development.

7. **Import of the Zonal Statistics Nighttime lights data** (from step 5) to R, using the function "load_VIIRS". A new table with the TNL, the GDP's and region/country ID, will be created.

8. **Countries classification according to the World Bank or the IMF** using the function "countries_class". The table created in step number 7 will now include a column with each region corresponding group name (e.g "**Upper-middle-income", "Low_Income etc.)**

9. **Perform a Linear regression**, using the function "GRP_VIIRS_LR ()" between GRP and the VIIRS nighttime lights for each group (e.g. low income, high income) using the following equation$GRP_x \sim \beta_0 + \beta_1 NTL_x$. Where: x, determines the group type (from the country classification), GRP can be the GDP, GRP, industrial GRP, Services GRP or Industrial and services GRP of a determined group. TNL is the Total Nighttime lights, calculated as the sum of the pixels within a specific region or country that make parts of the determined group x, β_0 is the intercept and β_1 is the regression coefficient.

   A new table with the summary of each linear regression is created. It includes the coefficients, the R^2, R-adjusted, P-value of TNL, Estimates etc.

10. **Export the model data** (from step 9) using the function "update_GRPmodel", as an excel file or update it, if the document already exists and there is new information to add. This function allows the user to save unlimited model summaries. Each excel file contains one sheet called "model_info" with the model summary as in step 9. There is an additional sheet called "comparison" that includes only the Group Code, the type of GRP, the minimum level of BUREF (used to calculate the TNL) and the R^2 adjusted of the corresponding model (in wide format). The information in this sheet can be useful If the user wants to select the best model (best GRP type and BUREF minimum level) by choosing the highest R^2 adjusted.

    - The function "GRP_VIIRS_fit" performs all the steps from 7 to 9 for a unique Zonal Statistics file
    - The function "multifits_VIIRS" performs all the steps from 7 to 9 for multiple zonal statistics files, as in step 6 the user will have to enter the start, end and increment values (BUREF minimum value) to perform the loop sequence.

[31] R package is defined as a «collection of *R* functions, data, and compiled code in a well-defined format» (Quick-R: R Packages, 2016)

[32] A detailed description of each function is included inside of each script.

11. **Perform a linear regression for all the possible combination between TNL, GRP and countries classes** and export an excel file for each type of countries classification using the function "allpossible_LR":

    Using the 2014 GRP database (GRP_2014.RDA) and all the zonal statistics files that we produced on step 6, we executed the function "allpossible_LR". As a result, we obtained 5 excel files:

    I. Grp_model2014_**WB_income**: 920 Linear regressions for 5 different income groups and each GRP_type: GRP, industrial GRP, Service GRP and Industrial and service GRP and 46 different BUREF minimum levels.
    II. Grp_model2014_**WB_region**: 1288 linear regressions for 7 different World Bank regions and each GRP type and 46 different BUREF minimum levels.
    III. Grp_model2014_**IMF_developement**: 1287 linear regressions for 7 different IMF analytic groups (level of development) and each GRP types and 46 different BUREF minimum levels.

    Two additional excel files with some others countries classification

    I. Grp_model2014_**IMF_region**: 1472 linear regressions for 8 different economic regions and each GRP types and 46 different BUREF minimum levels.
    II. Grp_model2014_**WB_lending**: 552 linear regressions for 3 different lending groups (level of development) and each GRP types and 46 different BUREF minimum levels.

12. **Graph the results**, in order to compare the results from all the linear regressions performed on step 11 and to choose the best type of countries classification, the best GRP type to correlate with VIIRS nighttime lights and the BUREF minimum level that yield the model with the highest R^2 adjusted. The script is under the name "allgraphsR.R" and it includes all the necessary commands to create and export graphs from the data collected on step 11.

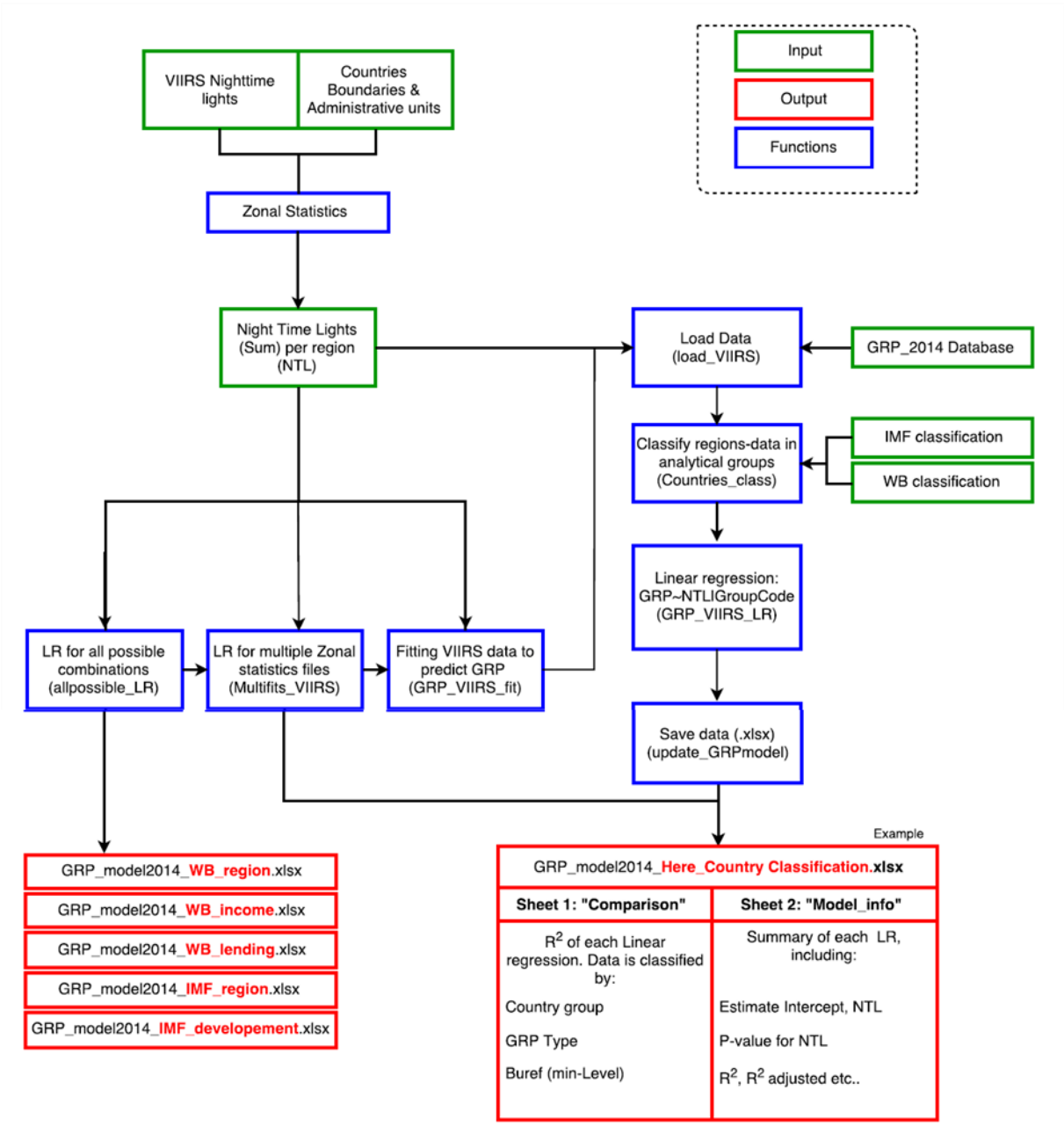

Figure 12 Structure of the R-package GRPlights. This diagrams explains how to use Zonal statistics VIIRS files to correlate Total nighttime lights per region (NTL) and 2014 GRP, following some countries classification. The functions described in this diagram are only for R. 3.2 or later.

Results

We met our objective of performing different linear regression models to predicts the global distribution of the GRP/GRP using Nighttime lights VIIRS DND-cloud free composites for the year 2014. We collected and important amount of information that will need to be processed in order to select which are the best parameters, which is the "best" model.

Additionally, all manual processes were minimized. There are different scripts so that all process described in the methodology can be recreated by other users or used again to process new information. The script "STEPS_to_run" collects all the necessary steps to recreate completely our methodology starting from the lights normalization.

Linear regression results

For each BUREF minimum level we calculated the respective TNL that resulted from using this threshold as a mask. We tried 46 different BUREF thresholds starting with 5% until we got 100% which was the maximum value that BUREF can have. This means that we had 46 different Zonal statistics tables with the sum of the nighttime lights per region/country, which were the result of variating BUREF minimum threshold.

Regression results for countries classified by INCOME (World Bank classification):

Countries and its respective regions were classified by Income in 5 different groups, from low to High Income (Table 3). For each group we performed 4 different linear regressions, each time variating the GRP type (1.GRP, 2. Industrial GRP, 3. Service sector GRP and 4. Industrial and services GRP). We repeated this process for every TNL-BUREF minimum threshold combination. We performed a total of 920 linear regressions (Figure 13).

$$5\ income\ groups\ X\ 4\ GRP\ types\ X\ 46\ BUREF\ thresholds\ =\ 920\ possible\ combinations.$$

From the results collected in Figure 13, we obtained that:

High Income (NON OECD) countries: GRP type doesn't change the relationship with TNL and the R^2 adjusted remains higher than 0.75, except when we considered only the Service sector GRP then the R^2 adjusted is not higher than 0.5. In all cases, regardless of the BUREF threshold, the TNL are statistically significant to model the GRP. The highest R^2 adjusted (0.89) was obtained when we related the Industrial GRP and TNL using a BUREF minimum value of 75%.

High Income (OECD) countries: TNL seems to have the same relationship with the different GRP types. Higher R^2 adjusted values are obtained using the total GRP. When BUREF threshold is higher than 15% ,the R^2 adjusted decrease steeply. As for the non-OECD member, TNL in this group are always statistically significant to model the GRP. The highest R^2 adjusted (0.83) was obtained when we related the total GRP and TNL using a BUREF minimum value of 10%.

Upper middle income countries: Regardless of the GRP type, the R^2 adjusted was never higher than 0.5. When BUREF threshold is higher than 30%, the R^2 adjusted decrease steeply. The TNL in this group are always statistically significant to model the GRP. The highest R^2 adjusted (0.53) was obtained when we related the total GRP or the Service GRP with TNL using a BUREF minimum value of 6%.

Lower middle income countries: It seems to be a better correlation between TNL and the service sector GRP, than with the other GRP types. Regardless of the GRP type, we found a peak in the R^2 adjusted when the BUREF minimum values was 30%, after this point it decreases again. The TNL in this group are not always statistically significant; when BUREF threshold is higher than 85%, TNL are no longer statistically significant to predict the services GRP. The highest R^2 adjusted (0.67) was obtained when we related the Service sector GRP with TNL using a BUREF minimum value of 30%.

Low income countries: This group is not statistically significant in any of the cases. It is not recommended to predict any type of GRP using this Income group. The highest R^2 adjusted (0.58) was obtained when we related the total GRP with TNL using a BUREF minimum value of 87%. For Low income countries the most important contribution on the GRP probably comes from agriculture which certainly not reflected in nighttime lights.

Figure 13 Comparison of the R^2 adjusted values resulting from linear regressions between GRP and TNL at each corresponding BUREF threshold used to calculate the TNL. Each color represents an income group, the shape indicates if the TNL is significant to predict the GRP, and each box regroups and GRP_type. A single point indicates the BUREF threshold value used to calculate the TNL, and the R^2 of the resulting model using this TNL, a determined income group and GRP type.

Regression results for countries classified by geographic Region (World Bank classification):

Countries and its respective regions were classified by geographic region in 7 different groups (Table 2). For each group we performed 4 different linear regressions, each time varying the GRP type. We repeated this process for every TNL-BUREF minimum threshold combination. We performed a total of 1288 linear regressions (Figure14).

$$7\ geographic\ regions\ X\ 4\ GRP\ types\ X\ 46\ BUREF\ thresholds\ =\ 1288\ possible\ models.$$

From the results collected inFigure14, we obtained that:

East Asia & pacific: The highest values for the R^2 adjusted were obtained with the Industrial GRP. When BUREF threshold is higher than 25%, the R^2 adjusted decrease smoothly. The TNL in this group are always statistically significant to model the GRP. The highest R^2 adjusted (0.84) was obtained when we related the industrial GRP with TNL using a BUREF minimum value of 10%.

Europe & Central Asia: The R^2 adjusted values don't show major changes as we change GRP type. However, it decreases rapidly, as the BUREF threshold increases. The TNL in this group are always statistically significant to model the GRP. The highest R^2 adjusted (0.56) was obtained when we related the total GRP with TNL using a BUREF minimum value of 6%, however for the other GRP groups the results were almost the same.

Latin America and the Caribbean: The highest values for the R^2 adjusted were obtained with the service sector GRP. Models using the Service Sector GRP have R^2 adjusted values greater than 0.70, regardless of the BUREF threshold. The TNL in this group are always statistically significant to model the GRP. However, the highest R^2 adjusted (0.76) was obtained when we related the total GRP with TNL using a BUREF minimum value of 20%.

Middle East & North Africa: The R^2 adjusted values don't show major changes as we change GRP type. Almost all models have R^2 adjusted values greater than 0.70, regardless of the BUREF threshold. The TNL in this group are always statistically significant to model the GRP. The highest R^2 adjusted (0.81) was obtained when we related the total GRP with TNL using a BUREF minimum value of 6%.

North America: The R^2 adjusted values don't show major changes as we change GRP type. However, it decreases rapidly when BUREF threshold is greater than 10%, regardless of the GRP type. The TNL in this group are always statistically significant. The highest R^2 adjusted (0.89) was obtained when we related the total GRP with TNL using a BUREF minimum value of 10%.

South Asia: R^2 adjusted values don't show major changes as we change GRP type. There is a peak in the R^2 adjusted when BUREF threshold is 30%, regardless of the GRP type. The TNL in this group are always statistically significant. The highest R^2 adjusted (0.85) was obtained when we related the total GRP with TNL using a BUREF minimum value of 30%; this is also true for the other GRP type as they have close results.

Sub Saharan Africa: The Service GRP is not related with TNL, as they are not statistically significant. The same occurs with the industrial GRP when BUREF threshold is higher than 30%. However, for total GRP, TNL are statistically significant. The R^2 adjusted increases as the BUREF threshold becomes higher. This means that TNL and Total GRP are more correlated as we limit the areas with low built up area. The highest R^2 adjusted (0.78) was obtained when we related the total GRP with TNL using a BUREF minimum value of 57%.

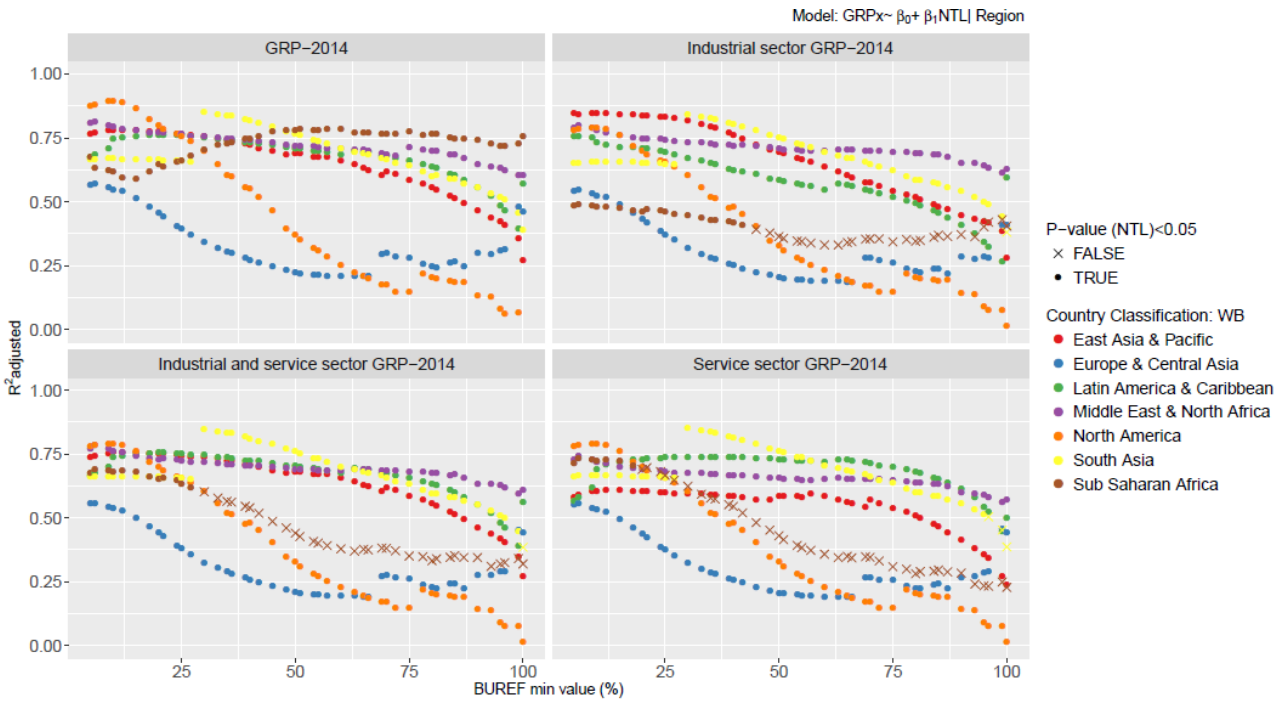

Figure14 Comparison of the R2 adjusted values resulting from linear regressions between GRP and TNL at each corresponding BUREF threshold used to calculate the TNL. Each color represents a geographic region, the shape indicates if the TNL is significant to predict the GRP, and each box regroups and GRP_type. A single point indicates the BUREF threshold value used to calculate the TNL, and the R^2 of the resulting model using this TNL, a determined geographic region and GRP type.

Regression results for countries classified by analytical group or economic development (IMF classification):

Countries and its respective regions were classified in 7 different analytical groups. For each group we performed 4 different linear regressions, each time variating the GRP type. We repeated this process for every TNL-BUREF minimum threshold combination. We performed a total of 1288 linear regressions (Figure15).

$$7\ analytic\ groups\ \ X\ 4\ GRP\ types\ X\ 46\ BUREF\ thresholds\ =\ 1288\ possible\ models.$$

From the results collected in Figure15, we obtained that:

Commonwealth IS: The GRP type doesn't change the relationship with TNL. The R^2 adjusted decreases rapidly after a BUREF minimum value becomes higher than 10%. Before this threshold, R^2 adjusted varies between 0.6 and 0.8. The TNL of this group are statistically significant to model every type of GRP. The highest R^2 adjusted (0.84) was obtained when we related the total GRP with TNL using a BUREF minimum value of 12%.

Emerging and developing Asia: The R^2 adjusted values don't show major changes as we change GRP type. R^2 adjusted values decreases as the BUREF threshold gets closer to 40%. Before this range R^2 adjusted is approximately 0.75 for most of the models. The TNL in this group are always statistically significant to model the GRP. The highest R^2 adjusted (0.79) was obtained when we related the Service sector GRP with TNL using a BUREF minimum value of 6%.

Emerging and developing Europe: TNL in this group are not always statistically significant. As the BUREF minimum value get higher than 30%, R^2 adjusted values decrease and TNL are no longer significant. This occurs for all the GRP types. However, when BUREF minimum value is lower than 30%, R^2 adjusted values are close to 0.75. This means that areas with small percentage of built-up area are important. The highest R^2 adjusted (0.86) was obtained when we related the total GRP with TNL using a BUREF minimum value of 9%.

Latin America & Caribbean: The results are much the same as in the World Bank region classification. They have the same trend. However, the highest R^2 adjusted (0.77) was obtained when we related the total GRP with TNL using a BUREF minimum value of 21%., which is only 1% difference with the world banks group.

Middle East and North Africa: The R^2 adjusted values don't show major changes R^2 adjusted values greater than 0.70, regardless of the BUREF threshold or the GRP Type. The TNL in this group are always statistically significant to model the GRP. The highest R^2 adjusted (0.83) was obtained when we related the total GRP with TNL using a BUREF minimum value of 6%. The results of the World Bank group with this name are almost the same.

Sub-Saharan Africa: TNL are statistically significant only with the total GRP and the Industrial sector GRP models. We obtained higher R^2 adjusted values with the total GRP. However, they decrease as the BUREF threshold becomes higher. From the significant models, the highest R^2 adjusted (0.81) was obtained when we related the total GRP with TNL using a BUREF minimum value of 6%.

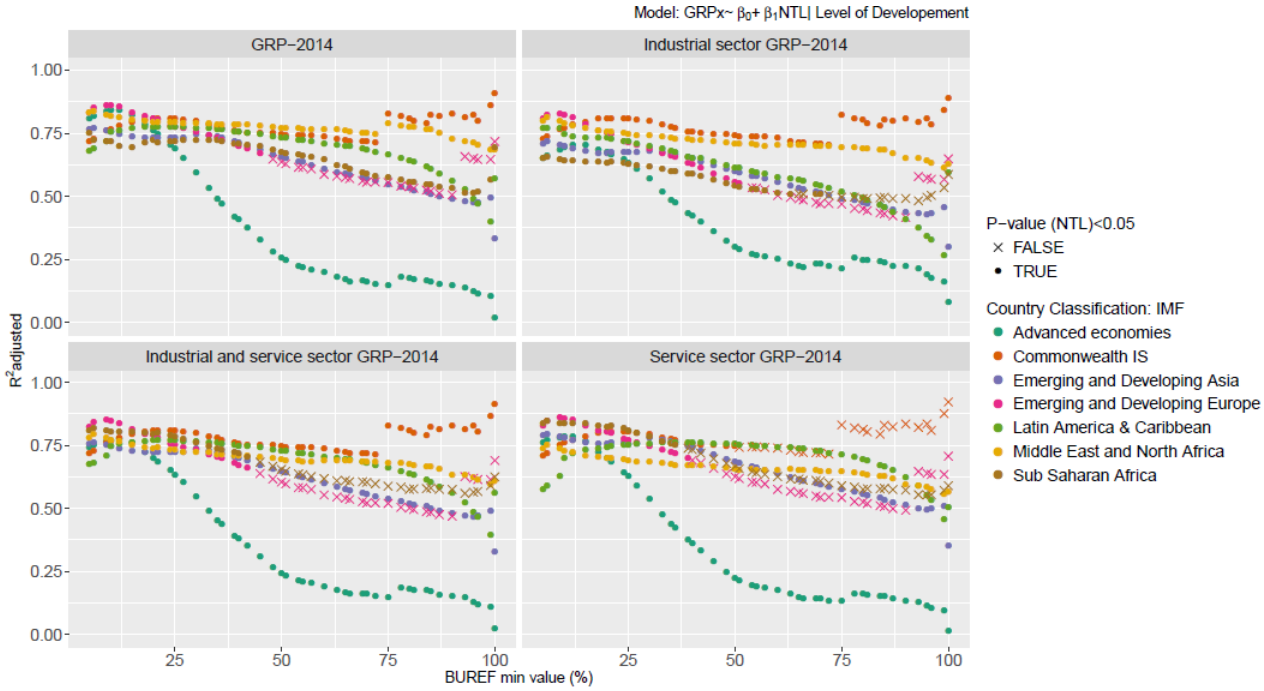

Figure15 Comparison of the R2 adjusted values resulting from linear regressions between GRP and TNL at each corresponding BUREF threshold used to calculate the TNL Each color represents an analytical group, the shape indicates if the TNL is significant to predict the GRP, and each box regroups and GRP_type. A single point indicates BUREF threshold value used to calculate the TNL, and the R^2 of the resulting model using this TNL, a determined analytical group and GRP type.

A resume with the selection of the best models for each GRP_type and country classification is available in the excel file "LR comparison" with can be automatically generated using the script "all_graphs. R"

Limits & perspectives

There is still a huge amount of information to process. The next step would be to select which is the best countries classification system that can fit our requirements, and check if its valid to have different BUREF threshold between groups.

Once a model is defined, we would continue to follow De Bono and Chatenoux (2014) methodology. We would need to integrate the resultant model equation to the light density to create a new layer with the predicted GRP. Afterwards a correction factor will be calculated by comparing the initial GRP with the predicted GRP. We will add this value to each region's TNL. At the end we would create map showing the variations between the GDP at national scale and its regional variations, a map with the spatial distribution (variations) of the economic activity within each country.

Conclusion

The results show that VIIRS nighttime lights are a good parameter to estimate the GRP. However, it is necessary to perform some corrections in the original composites, in order to improve the overall model results. This is where the BUREF layer becomes essential. We not only restricted the maximum light radiance to places with the highest built-up area, but we also tested which were the effects of varying BUREF (minimum value) on the linear regression between GRP and TNL. Our results show that effectively, there are variations on the coefficient of determination (R^2) as we vary the BUREF threshold (or minimum value). The direction of this effect depends on the classification group. As we didn't perform a global regression, but rather, divide the countries in different groups, there are different results for each one of

them. Three classification systems were chosen: Income, geographic region and analytical group (based on the economic development).

For the first classification system by Income, suggested by the World Bank (Table 3), we obtained that the model parameters that yielded the highest R^2 and that were at the same time statistically significant, differ between each group. This means that there is no a unique GRP type that yields the best results in all the groups. In this case, for three groups (Low income, Lower middle income and Upper middle income) TNL could predict better the total GRP than the others GRP type. The same occurs for the BUREF threshold, it varies among the groups. For example, the Upper middle income group obtained the highest R^2 when the BUREF threshold was 6%, regardless from the GRP type. Oppositely, for the High Income (non OECD member) group the best results were obtained when the BUREF threshold was 75%. This mean that if we select this classification system the model parameter will be different for each group. Different BUREF threshold would not pose a problem, if we are able to justify why for some groups a low percentage built - up areas don't reflect the changes in lights and GRP and for others all levels of Built-up area are important. For the GRP type would be different as we cannot compare for example the Industrial GRP for High income countries with the Service sector GRP of the lower middle income group. One solution would be to simply select the best model for each GRP type but it would be necessary to select the same BUREF threshold for all the GRP type within the same countries group. Another limiting of this classification system was that a whole group, "Low income", was not statistically significant to any GRP. Our hypothesis was that the countries in this group have a more important contribution of agriculture on their GRP which is not reflected by nighttime lights.

From the geographic region classification, suggested by the World Bank (Table 2), we obtained more homogeneous results. First the difference between the R^2 of the best GRP models were not high, no bigger than 5%. Moreover, for all the best models using the total GRP have a R^2 h higher than 0.75, except for Europe & Central Asia group, which highest R^2 was 0.5. The difference in PIB between Europe and Central Asia must explain this result. Now concerning the BUREF thresholds they also varied within the groups, however this time the difference was shorter than in the Income classification. Europe and Central Asia group obtained the highest R^2 when the BUREF threshold was 6%, regardless from the GRP type; and the Sub-Saharan group, when the BUREF threshold was 57%, which was the highest from all the other groups. If we would want to select this classification system, it would be enough by estimating only the total GRP with the TNL of each geographic region and establishing a "maximum" BUREF threshold of 30%. In this way we would obtain 7 models, one for each geographic region, with a R^2 adjusted higher than 0.7.

From the last countries classification system by analytical groups, suggested by the IMF, we obtained also good results. If for example, we restrict BUREF threshold to be maximum 25%, we found would obtain 7 models, one per each group, with coefficients of determination ranging from 0.77 to 0.85 (Table 4). This type of parameterization allows us to detect combinations if we want to predict the total GRP or the Industrial plus Services GRP. For both GRP types results are satisfactory, then it would depend on the type of information that you would want to extract from the GDP to make a decision.

It is also important to emphasis that we succeed to automate all the manual processes as we planned. This means that the methodology can now be applied by other users or consulted for further amelioration. Everything, except for the original VIIRS composite, will be stocked in a single folder so that it can be easily transfer to other users or readers.

Table 4 Collection of the best results for the Country classification by analytical group.

Group Name	*GRP Type (2014)*	*R^2 adjusted*	*BUREF (min-value in %)*	*GRP Type (2014)*	*R^2 adjusted*	*BUREF (min-value in %)*
Advanced economies	GRP	0,84	12	Industrial +Services GRP	0,77	12
Commonwealth IS	GRP	0,77	25	Industrial +Services GRP	0,77	25
Emerging and Developing Asia	GRP	0,77	6	Industrial +Services GRP	0,76	6
Emerging and Developing Europe	GRP	0,86	9	Industrial +Services GRP	0,85	9
Latin America Caribbean	GRP	0,78	21	Industrial +Services GRP	0,77	21
Middle East and North Africa	GRP	0,84	6	Industrial +Services GRP	0,79	6
Sub-Saharan Africa	GRP	0,75	5	Industrial +Services GRP	0,82	6

Personal conclusion

The development of this internship allowed me to learn multiple new things in a short time. From using the open source software GRASS, and creating bash scripts to developing even more my knowledge in R, which allowed me to create, for the first time, my own R package. I also acquire work techniques that we don't usually learn in the university. I developed my capability to resolve problems, learn from scratch, decrease the time of procrastination etc.

I would like to thank GRID-Geneva for the opportunity of making this internship; it certainly developed on me a certain passion for Geomatics. I hope that this information that we developed can be useful in future researches.

Bibliography

Baugh, K., Hsu, F., Elvidge, C., & Zhizhin, M. (2013). Nighttime Lights Compositing Using the VIIRS Day-Night Band: Preliminary Results. *Proceedings Of The Asia-Pacific Advanced Network*, *35*(0), 70. http://dx.doi.org/10.7125/apan.35.8

De Bono, A., Chatenoux, B., 2014. A global exposure model for GAR 2015. GAR, UNISDR, 20 p. www.preventionweb.net/english/hyogo/gar/2015/en/bgdocs/risk-section/De%20Bono,%20Andrea,%20Bruno%20Chatenoux.%202015.%20A%20Global%20Exposure%20Model%20for%20GAR%202015,%20%20UNEP-GRID.pdf

Doll, C., Muller, J., & Morley, J. (2006). Mapping regional economic activity from night-time light satellite imagery. *Ecological Economics*, *57*(1), 75-92. http://dx.doi.org/10.1016/j.ecolecon.2005.03.007

Elvidge, C., Baugh, K., Kihn, E., Kroehl, H., Davis, E., & Davis, C. (1997). Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption. *International Journal Of Remote Sensing*, *18*(6), 1373-1379. http://dx.doi.org/10.1080/014311697218485

Gennaioli, N., La Porta, R., Lopez-De-Silanes, F., & Shleifer, A. (2013). *Human capital and regional development*. Cambridge, Mass.: National Bureau of Economic Research.

Ghosh, T., Anderson, S., Powell, R., Sutton, P., & Elvidge, C. (2009). Estimation of Mexico's Informal Economy and Remittances Using Nighttime Imagery. *Remote Sensing*, *1*(3), 418-444. http://dx.doi.org/10.3390/rs1030418

Ghosh, T., Powell, R., Elvidge, C., Baugh, K., Sutton, P., & Anderson, S. (2010). Shedding Light on the Global Distribution of Economic Activity. *The Open Geography Journal*, *3*(1), 147-160. http://dx.doi.org/10.2174/1874923201003010147

Huang, Q., Yang, X., Gao, B., Yang, Y., & Zhao, Y. (2014). Application of DMSP/OLS Nighttime Light Images: A Meta-Analysis and a Systematic Literature Review. *Remote Sensing*, *6*(8), 6844-6866. http://dx.doi.org/10.3390/rs6086844

Li, X., Xu, H., Chen, X., & Li, C. (2013). Potential of NPP-VIIRS Nighttime Light Imagery for Modeling the Regional Economy of China. *Remote Sensing*, *5*(6), 3057-3081. http://dx.doi.org/10.3390/rs5063057

Peserasi, M., & Carneiro, S. (2014). BUREF – PRODUCING A GLOBAL REFERENCE LAYER OF BUILT-UP BY INTEGRATING POPULATION AND REMOTE SENSING DATA. *Institute For The Protection And Security Of The Citizen. UNISDR*.

Pinkovskiy, M., & Sala-i-Martin, X. Lights, Camera,...Income! Estimating Poverty Using National Accounts, Survey Means, and Lights. *SSRN Electronic Journal*. http://dx.doi.org/10.2139/ssrn.2423393

Shi, K., Yu, B., Huang, Y., Hu, Y., Yin, B., & Chen, Z. et al. (2014). Evaluating the Ability of NPP-VIIRS Nighttime Light Data to Estimate the Gross Domestic Product and the Electric Power Consumption of China at Multiple Scales: A Comparison with DMSP-OLS Data. *Remote Sensing*, *6*(2), 1705-1724. http://dx.doi.org/10.3390/rs6021705

World Bank. 2015. Purchasing Power Parities and the Real Size of World Economies: A Comprehensive Report of the 2011 International Comparison Program. Washington, DC: World Bank. doi:10.1596/978-1-4648-0329-1. License: Creative Commons Attribution CC BY 3.0 IGO.

SCRIPTS

STEPS TO RUN

```
## How to ESTIMATE the global GDP using VIIRS nighttime lights
## Authors: Ena Suarez <ena.suarez@unepgrid.ch>- for GRID-UNEP GENEVA
# Download VIIRS data and prepate for further analysis using MSYS:
    Run coomands on script: "VIIRS_data_preparation"

# 1) Import Principal data: VIIRS nighttime lights (NTL) Raster for 2014 (median of monthly tiles), Worldwide regional bounds (administrative areas) and BUREF raster (The  Global  Built-up  Reference  Layer)
    sh -x "DIRECTORY"/VIIRS_project/Scripts/input_data_linux.sh
# 2) Correction of  VIIRS NTL using BUREF layer (limit maximum DN values to the greatest urban area of each region, in order to find the best range, for example  we started by a min buref value of 5% and increasing steps of 5%,
#  For each iteration we will obtain a Zonal statistics talbe that will give us the necesary information to perform a linear regression between the GRP and VIIRS NTL, the results will be compiled on excel files
   ##single run
   sh -x "DIRECTORY"/VIIRS_project/Scripts/VIIRS_step_one_linux.sh $1
   ## Loops parameters: $1 min, $2 max, $3 step
   ## Grass iteration (Change directory on the script)
   sh -x "DIRECTORY"/VIIRS_project/Scripts/loop_VIIRS_grass.sh $1 $2 $3
# 3) R installation
R
# 4) Install necesary packages and functions (Change directory on the script) (Change directory on the script)
   source('DIRECTORY/VIIRS_project/Scripts/install_analysispack.R')


# 5) Perform a linear regression for each of the Zonal statistics tables and save all data on a excel file(there will be an excel file per country classification
# This would be in total of 5 -->># file  "ZSNTL_VIIRS14_buref_$X_stats" sequence for loop $1 min $2max $3 step
# There is a copy of the R_functions that wokr with the Rpackage "cats" ,that I created, on the Folder scripts/R_VIIRS_functions.  (The original files are on the folder Cats/R, but moodifications will change the R_package)
allpossible_LR($1,$2,$3, directory="DIRECTORY""/VIIRS_project")

# 6) Grahp the results, save them as pdf files and create an excel file with the selection of the best models for each classification system.

 source('DIRECTORY/VIIRS_project/Scripts/allgraphsR.R')

# Quit R session and dont save Image work
q()
n

###Examples on How to use GRPlights package:
# R installation
R
# Install GRPlights  R-package
source('DIRECTORY/VIIRS_project/Scripts/install_analysispack.R')

## Load data VIIRS and GRP
```

```
world_ntl_grp<-load_VIIRS(output ="H:/VIIRS_project/model_GRP",
NTL_data="H:/VIIRS_project/NTL_data/ZSNTL_VIIRS14_buref_5_stats.csv",
GRP_2014="H:/VIIRS_project/GRP_data/GRPs_2010_2014.txt",level=5)

## Filter data by analysis groups
classing<-countries_class(data=world_ntl_grp, group="WB",subgroup="income")

## Linear Regression
LR<-GRP_VIIRS_LR(data=classing,"GRP_14",only_R2=FALSE)

## Save in excel
update_GRPmodel(output_folder="H:/VIIRS_project/model_GRP",output_name="grp_model2014xx",
GRPmodel=LR, R_type="R2adj")

## Resume all the steps done above and update files
GRP_VIIRS_fit(directory="H:/VIIRS_project",level=5,group="WB",subgroup="income",R_type="R2adj")

##Performs the GRP_VIIRS_FIT funcion in a for loop from a to b with n as step
multifits_VIIRS(a,b,n,"WB","income",directory="H:/VIIRS_project",R_type="R2adj")

## performs all the possible combinations for the recoleted VIIRS data
allpossible_LR(a,b,n, directory="H:/VIIRS_project")
```

VIIRS_DATA_PREPARATION

```
# Unzip:
 for f in *.tgz; do tar xzvf  $f; done
# EREASE ALL THE fILES DIFFERENT FROM AVG_RADE9
# Multiply *100 and set type: UInt32:
for f in *.tif; do gdal_calc.py -A $f --outfile=$f.tif --calc="100*A" --type="UInt32" ; done

# EREASE ALL THE fILES DIFFERENT FROM the new .tif
# Merge:
##step 1: Create a catalog
gdalbuildvrt  june14.vrt SVDNB*.tif
##step 2: Convert catalog to tif
gdal_translate june14.vrt alljune.tif
#### Clip-Resize (cut one pixel from each side)
for f in *.tif; do gdal_translate $f $f.tif -projwin -179.9979166 74.99791663 179.9979195 -64.99375113;
done
####Compress LZW:
 for f in *.tif; do gdal_translate $f $f.tif -co COMPRESS=LZW ; done
############ MAP CALCULATOR  in GRASS to create one unique image of nighttime lights for 2014
Set resolution to:  0:00:15  with g.region
r.mapcalc "median_VIIRS_2014 = median( alljanuary_cl@PERMANENT + allfebruary_cl@PERMANENT +
allmarch_cl@PERMANENT + allapril_cl@PERMANENT + allmay_cl@PERMANENT + alljune_cl@PERMANENT
+ alljuly_cl@PERMANENT + allaugust_cl@PERMANENT + allseptember_cl@PERMANENT +
alloctober_cl@PERMANENT + allnovember_cl@PERMANENT + alldecember_cl@PERMANENT)"
r.mapcalc "average_VIIRS_2014 = ( alljanuary_cl@PERMANENT + allfebruary_cl@PERMANENT +
allmarch_cl@PERMANENT + allapril_cl@PERMANENT + allmay_cl@PERMANENT + alljune_cl@PERMANENT
+ alljuly_cl@PERMANENT + allaugust_cl@PERMANENT + allseptember_cl@PERMANENT +
alloctober_cl@PERMANENT + allnovember_cl@PERMANENT + alldecember_cl@PERMANENT )/12"

r.mapcalc "maximum_VIIRS_2014= max( alljanuary_cl@PERMANENT, allfebruary_cl@PERMANENT ,
allmarch_cl@PERMANENT , allapril_cl@PERMANENT , allmay_cl@PERMANENT , alljune_cl@PERMANENT ,
alljuly_cl@PERMANENT , allaugust_cl@PERMANENT , allseptember_cl@PERMANENT ,
alloctober_cl@PERMANENT , allnovember_cl@PERMANENT , alldecember_cl@PERMANENT )"

r.mapcalc "mimimum_VIIRS_2014 = min( alljanuary_cl@PERMANENT, allfebruary_cl@PERMANENT ,
allmarch_cl@PERMANENT , allapril_cl@PERMANENT, allmay_cl@PERMANENT, alljune_cl@PERMANENT ,
alljuly_cl@PERMANENT , allaugust_cl@PERMANENT , allseptember_cl@PERMANENT ,
alloctober_cl@PERMANENT , allnovember_cl@PERMANENT , alldecember_cl@PERMANENT  )"
```

INPUT_DATA_LINUX .SH

```
#!/bin/bash
## This script was built to run in GRASS 7.0 or later
## Replace the word -->"DIRECTORY" by the directory where you have the folder VIIRS_intial_Rasters
## Authors: Ena Suarez <ena.suarez@unepgrid.ch>- for GRID-UNEP GENEVA
## Correction of Nighttime lights Part 1  #####----------------------------

####Import raster data: VIIRS median composite for the year 2014
r.in.gdal --overwrite input="DIRECTORY"/VIIRS_intial_Rasters/median_VIIRS_2014.tif
output=median_VIIRS_2014

#Create a real pixel raster (in square meter) and a density raster (in sqare kilometer) from a given raster
g.region raster=median_VIIRS_2014
PI=3.1415926536
R=6371007.2
r.mapcalc --overwrite "sqm = round((sin(y()+nsres()/2) - sin(y()-nsres()/2)) * (ewres() * $PI/180) * $R^2)"
r.mapcalc --overwrite "median_VIIRS_2014_sqkm1 = round (float(median_VIIRS_2014)/ sqm * 1000000)"

r.out.gdal --overwrite input=median_VIIRS_2014_sqkm1
output="DIRECTORY"/VIIRS_intial_Rasters/median_VIIRS_2014_sqkm1.tif format=GTiff

######   IMPORT GRP_Bound and eliminating all data outside the set region#####------------------
#set region
g.region raster=median_VIIRS_2014_sqkm
# GRP import limit the SHAPE file import to the current region (-r and snap=overlapping features)
v.in.ogr -r --overwrite "DIRECTORY"/VIIRS_intial_Rasters/Grp_bound/grp_bound.shp out=grp_bound
snap=1e-4
#Erease columns that are not useful
v.db.dropcolumn map=grp_bound columns=ZS_count,ZS_sum,ZS_mean,MIN_ls_reg,OBJECTID
#Convert from vector to raster using the column ID as attribute
v.to.rast --overwrite input=grp_bound output=grp_bound_map use=attr attribute_column=ID

### Create a raster only with lights within the countries boundaries
r.mapcalc --overwrite "median_VIIRS_2014_sqkm = if( grp_bound_map@PERMANENT>=1 ,
median_VIIRS_2014_sqkm1, null())"

##### Create a thershold for the layer "BUREF" #####-----------------------------
#set region
g.region raster=median_VIIRS_2014_sqkm

# Call Buref layer and set Thershold for calibration
r.in.gdal input="DIRECTORY"VIIRS_intial_Rasters/Buref/buref09_cut.tif output=buref

exit 0
```

VIIRS_step_one_linux.sh

```
#!/bin/bash
## This script was built to run in GRASS 7.0 or later
## Replace the word -->"DIRECTORY" by the directory where you have the folder VIIRS_intial_Rasters
## Parameters: $1, indicate  which would be the minimum BUREF value (values can varie from 0-100)
## Authors: Ena Suarez <ena.suarez@unepgrid.ch>- for GRID-UNEP GENEVA
## Correction of Nighttime lights Part 2  #####----------------------------
## check the number of arguments
if (( $# != 1 )); then
   echo "Illegal number of parameters"
   exit
fi
#set region
g.region raster=median_VIIRS_2014_sqkm
# MASK:Create new raster with the set thershold "$1"
r.mapcalc --overwrite "buref_min_$1 = if(buref< $1, null(),buref)"
r.mapcalc --overwrite "median_VIIRS_2014_$1 = if(buref< $1, null(),median_VIIRS_2014_sqkm)"
#Zonal statistics between GRP and  BUREF mask
r.univar -t map=buref_min_$1 zones=grp_bound_map output=ZSburef_stats_$1.csv separator=comma --
overwrite
#Import CVS table
db.in.ogr input=ZSburef_stats_$1.csv output=ZSburef_stats_$1 --overwrite
#Work on a copy of grp_bound with the current mapset
g.copy --overwrite vector=grp_bound,ZSburef_min_$1
#Join tables using key columns ID and ZONES
v.db.join map=ZSburef_min_$1 column=ID other_table=ZSburef_stats_$1 other_column=zone
##Add column with ID_burefmax and field calculator
v.db.addcolumn ZSburef_min_$1 columns="burefmax integer"
v.db.update map=ZSburef_min_$1 layer=1 column=burefmax query_column=max
v.db.addcolumn ZSburef_min_$1 columns="ID_burefmax integer"
v.db.update map=ZSburef_min_$1 layer=1 column=ID_burefmax query_column=ID+burefmax

#Convert from vector to raster using the column ID as attribute
v.to.rast --overwrite input=ZSburef_min_$1 output=ZSburef_min_$1_map use=attr
attribute_column=ID_burefmax
#create a raster with VIIRS max values following Buref
r.mapcalc --overwrite "VIIRS_max = if (buref_min_$1 ==(ZSburef_min_$1_map - grp_bound_map),
median_VIIRS_2014_$1 , null())"
#### Zonal statistics
r.univar -t map=VIIRS_max zones=grp_bound_map output=ZSVIIRS_stats.csv separator=comma --overwrite
db.in.ogr --overwrite input=ZSVIIRS_stats.csv output=ZSVIIRS_stats
g.copy --overwrite vector=grp_bound,ZSVIIRS_max
v.db.join map=ZSVIIRS_max column=ID other_table=ZSVIIRS_stats other_column=zone
v.db.addcolumn ZSVIIRS_max columns="VIIRSmax integer"
v.db.update map=ZSVIIRS_max layer=1 column=VIIRSmax query_column=max
v.to.rast --overwrite input=ZSVIIRS_max output=ZSVIIRS_max_map use=attr attribute_column=VIIRSmax
r.mapcalc --overwrite "VIIRS_2014_limit_buref_$1 = if( median_VIIRS_2014_$1 >=  ZSVIIRS_max_map &
ZSVIIRS_max_map > 0  , ZSVIIRS_max_map , median_VIIRS_2014_$1)"
#### Zonal statistics to CALCULATE THE TOTAL NIGHTTIME LIGHTS PER REGION
r.univar -t map=VIIRS_2014_limit_buref_$1 zones=grp_bound_map
output="DIRECTORY"/VIIRS_project/NTL_data/ZSNTL_VIIRS14_buref_$1_stats.csv separator=comma --
overwrite

exit 0
```

LOOP_VIIRS_GRASS.SH

```
#!/bin/bash
## This script was built to run in GRASS 7.0 or later
## Replace the word -->"DIRECTORY" by the directory where you have the folder VIIRS_intial_Rasters
## Parameters:  $1, $2, $3  sequence values to run  loop {START..END..INCREMENT}
##Authors: Ena Suarez <ena.suarez@unepgrid.ch>  for GRID-UNEP GENEVA
## Loop sequance to repetar serveral time the "Correction of Nighttime lights Part 2"  #####----------
---------

for i in $(seq $1 $3 $2)
do
  sh -x "DIRECTORY"/VIIRS_project/VIIRS_step_one_linux_checked.sh ${i}
  echo "Welcome $i times"
done
```

INSTALL_ANALYISISPACK.R

```
# Install GRPlights R-package
# Require: Roxygen and devtools packages
# Authors: Ena Suarez <ena.suarez@unepgrid.ch>- for GRID-UNEP GENEVA

  library(roxygen2)
  library(devtools)
  setwd("Directory""/VIIRS_project/GRPlights")
  document ()

  setwd("Directory""/VIIRS_project/")
  install("GRPlights")
```

LOAD_VIIRS.R

```
#' A load VIIRS and grp data function
#'
#' This function allows you to upload your Zonal Statistics Nighttime lights data into R for further analysis
#' @param Zonal statistics for VIIRs and GRP for the same yeard divided in GDP, GRP, GRP from agriculture,
industry and industry and services
#' @keywords VIIRS GRP
#' @export
#' @examples
#' load_VIIRS()
load_VIIRS<-function(output=NULL,NTL_data=NULL,GRP_2014=NULL, level=NULL) {

  ### NTL should be in CVs format, GRP data should be in Txt format.

  ###--- Set directory
  setwd(output)

  ###--- Input data for VIIRS Nighttime lights (NTL)
  NTL_stats <- read.csv(NTL_data)

  ###--- Input data for GRP database
  GRP_2014 <-read.table(GRP_2014, header=T)

  #----Create a table with important variables
  world_NTL=data.frame(ID_NTL=NTL_stats$zone,NTL=NTL_stats$sum)
  world_NTL$ID_NTL<-as.factor(world_NTL$ID_NTL)
  world_GRP=data.frame (ISO3_CODE=GRP_2014$ISO3_CODE,
                ID_GRP=GRP_2014$ID,
                GDP_14=GRP_2014$GDP14,
                GRP_14=GRP_2014$grp14,
                GRP_agri14=GRP_2014$grp_agri14,
                GRP_ind14=GRP_2014$grp_ind14,
                GRP_serv14=GRP_2014$grp_serv14,
                GRP_indserv14=GRP_2014$grp_indserv14)

  world_GRP$ID_GRP<- as.factor(world_GRP$ID_GRP)

  world_ntl_grp <- merge(world_NTL,world_GRP, by.x="ID_NTL", by.y="ID_GRP")
  world_ntl_grp$level<-level


  return(world_ntl_grp)
}
```

COUNTRIES_CLASS.R

```
#' Divide countries data in analysis groups.
#'
#' This function allows you to divide the Global data (VIIRS and GRP) in different groups to have a better
analysis. Data can be divided by region, income or lending
#' @param  World recolted data, group of analysis (world bank or International monetary fund countries
classification) and subgroup(The options are income, region, lending, development and economic region)
#' @keywords VIIRS GRP Income Region lending
#' @export
#' @examples
#' countries_class()
countries_class<-function(data=NULL, group="WB", subgroup="region") {

  # New version of length which can handle NA's: if na.rm==T, don't count them
      if (group=="IMF") {
    # IMF Country Classification
            grouping<-merge(data, IMF_class, by.x="ISO3_CODE", by.y="country_code")
            grouping$group<-"IMF"
            }else {
    # World Bank classification
              grouping<-merge(data, WB_class, by.x="ISO3_CODE", by.y="CountryCode")
              grouping$group<-"WB"

            }
  ##//// divide World bank groups in subgroups Region, Income or Lending
     if (subgroup=="region"& group=="WB"){
       ##### Subgroups WB by region
       grouping<-subset(grouping, GroupCode=="EAS" | GroupCode=="ECS"| GroupCode=="LCN"|
GroupCode=="MEA"| GroupCode=="NAC"| GroupCode=="SAS"| GroupCode=="SSF")
       grouping$subgroup<-"region"
     } else if (subgroup=="income"& group=="WB") {

      ##### Subgroups WB by Income
      grouping<-subset(grouping, GroupCode=="LIC"| GroupCode== "LMC"| GroupCode== "UMC"|
GroupCode== "OEC"| GroupCode=="NOC")
      grouping$subgroup<-"income"

    } else if (subgroup=="lending"& group=="WB") {

     ##### Subgroups WB by lending
      grouping<-subset(grouping, GroupCode=="IDX"| GroupCode== "IDB"| GroupCode== "IBD")
      grouping$subgroup<-"lending"
    }

    ##//// divide IMF in two subgroups

    else if (group=="IMF"& subgroup=="development"){
    ##Advanced economies:
    grouping<-subset(grouping, IMF_group=="Emerging_and_developing"|
IMF_group=="Advanced_economies")
    grouping$subgroup<-"development"

   }else if (group=="IMF"& subgroup=="region") {
```

```
  ###Emerging market and deloping economies:
  grouping<-subset(grouping, IMF_group=="Euro_Area"| IMF_group=="Major_Advanced_economies"|
IMF_group=="Other_adv_Economies" | IMF_group=="Commonwealth_IS"|
IMF_group=="Emerging_and_Developing_Asia"| IMF_group=="Latin_America_Caribbean"|
IMF_group=="Middle_East_and_North_Africa"| IMF_group=="Sub-Saharan_Africa")
  grouping$subgroup<-"region"
 }

 return(grouping)
}
```

GR_VIIRS_LR.R

```
#' Linear regression between GRP and the VIIRS nighttime lights by analysis groups
#'
#' This function allows you to Performs a linear regression between GRP and the VIIRS nighttime lights and
save the R2, coefficient and graphs is desired
#' @param   Grouped data of the GRP and the VIIRS data grouped by analitical groups
#' @keywords VIIRS GRP prediction
#' @export
#' @examples
#' GRP_VIIRS_LR()
GRP_VIIRS_LR<-function(data=NULL,column,only_R2=TRUE) {

  ### Import R packages
  library(lme4)
  library(nlme)

  ##detect type of grouping

  groupx<-data[1,c('group')]
  subgroupx<-data[1,c('subgroup')]
  GRPx<-data[[column]]

  ########## Linear regresion of GRP and VIIRS nighttime lights
  if (groupx=="WB") {
    datac<-data.frame(NTL=data$NTL, GRP=GRPx,GroupCode=data$GroupCode,
GroupName=data$GroupName)

  }else {
    datac<-data.frame(NTL=data$NTL,
GRP=GRPx,GroupCode=data$IMF_group,GroupName=data$IMF_group)

  }

  #linear regression and main results
  tryfit<-lmList(GRP~NTL|GroupCode, data= datac)
  xx<-summary(tryfit)$coefficients
  LR<-as.data.frame(xx)
  LR$R2<-summary(tryfit)$r.squared
  LR$R2adj<-summary(tryfit)$adj.r.squared


  ### Linear regression
  if (only_R2==TRUE) {
    LR[,1]<-NULL
    LR[,1]<-NULL
    LR[,1]<-NULL
    LR[,1]<-NULL
    LR[,1]<-NULL
    LR[,1]<-NULL
    LR[,1]<-NULL
    LR2<-LR
    LR2$level<-unique(data$level)
```

```
  } else if (only_R2==FALSE) {
    LR2<-LR
    LR2$level<-unique(data$level)
  }
  else{ LR2<-NULL}

  model<-LR2
  model$GroupCode<-rownames(model)
  row.names(model)<-NULL
  model<-model[,c(ncol(model),1:(ncol(model)-1))]
  model$GRP<-column

  groups<-factor(datac$GroupCode)
  groups<-levels(groups)
  vectorx<-NULL

  for (i in groups)
  {
    newdata <- subset(datac, GroupCode==i,select=c(NTL, GRP))
    my.lm <- lm(GRP ~ NTL, data = newdata)
    pr <- residuals(my.lm)/(1 - lm.influence(my.lm)$hat)
    PRESS <- sum(pr^2)
    # anova to calculate residual sum of squares
    my.anova <- anova(my.lm)
    tss <- sum(my.anova$"Sum Sq")
    # predictive R^2
    pred.r.squared <- 1 - PRESS/(tss)
    vectorx<-c(vectorx,pred.r.squared)
  }

  model$predicted_R2<-vectorx
  colnames(model)[9] <- "P-value_NTL"
  xx<-data.frame(GroupName=unique(datac$GroupName),y=unique(datac$GroupCode))
  model=merge(model,xx,by.x="GroupCode",by.y="y")

  return(model)

}
```

UPDATE_GRPMODEL.R

```
#' Create or update an excel file with the model data
#'
#' This function allows you to save as much as models as you want
#' @param output(path), data or new data to add- There are three types of R: normal ("R2"), adjusted
("R2adj") and predictive "predicted_R2".
#' @keywords VIIRS GRP
#' @export
#' @examples
#' update_GRPmodel
update_GRPmodel<-function(output_folder=NULL,output_name=NULL, GRPmodel=NULL, R_type) {


  ### Starts Function

  ## Set directory
  setwd(output_folder)
  ## Paste output path
  output_path<-paste(output_folder,"/",output_name,".xlsx",sep="")

  ## Verify if there is an existing project with this name, in this case UPDATE, if not SAVE it as a new project
  library(xlsx)
  library (tidyr)

  #Select the INPUT data to compare and Transform dataset from long to wide format:
  comparison<-subset(GRPmodel,select=c("GroupCode","level",R_type,"GRP"))
  levelx<-unique(GRPmodel$level)

  ### Create workbook
  grp_project<- createWorkbook()
  ##Create 2 sheets one with the whole information of the model and other with the R2 for future
comparison
  sheet_1<-createSheet(grp_project,sheetName ="comparison")
  sheet_2<-createSheet(grp_project,sheetName ="model_info")
  sheet_3<-createSheet(grp_project,sheetName ="long_format")


  if (file.exists(output_path)==TRUE) {

    old_modelp2<- read.xlsx(output_path, sheetName ="model_info")
    old_modelp3<- read.xlsx(output_path, sheetName ="long_format")

    #Join New and old data

    comparing<-unique(rbind(old_modelp3,comparison))

    # Save data in the long format
    addDataFrame(comparing, sheet_3, startRow=1, startColumn=1,row.names=FALSE)

    #Transform to wide format
    colnames(comparing)[3] <- "R_type"
    comparing_wide<-spread(comparing,level,R_type)
```

```
    colnames(comparing)[3] <- R_type
    comparing_wide<-comparing_wide[order(comparing_wide$GRP),]

    ## Add data of the linear regression to the sheet of comparison
    addDataFrame(comparing_wide, sheet_1, startRow=1, startColumn=1,row.names=FALSE)
    # Save statistical data about the LR
    addDataFrame(old_modelp2, sheet_2, startRow=1, startColumn=1,row.names=FALSE)

    ## If there is information about the coefficients save it in a new sheet
    if (ncol(GRPmodel)==15) {
      addDataFrame(GRPmodel, sheet_2, startRow=nrow(old_modelp2)+2,
startColumn=1,row.names=FALSE,col.names=FALSE)
    }
  }

  if (file.exists(output_path)==FALSE){

    ##Add data to long format
    addDataFrame(comparison, sheet_3, startRow=1, startColumn=1,row.names=FALSE)


    ## Add data of the linear regression to the sheet of comparison
    colnames(comparison)[3] <- "R_type"
    comparison_wide<-spread(comparison,level,R_type)
    colnames(comparison)[3] <- R_type
    addDataFrame(comparison_wide, sheet_1, startRow=1, startColumn=1,row.names=FALSE)

    ## If there is information about the coefficients save it in a new sheet
    if (ncol(GRPmodel)==15) {
      addDataFrame(GRPmodel, sheet_2, startRow=1, startColumn=1,row.names=FALSE)

    }

  }
  saveWorkbook(grp_project, output_path) # Save workbook
}
```

GRP_VIIRS_FIT.R

```
#' This functions makes all the necesary steps for fitting VIIRS data to GRP data
#'
#' This function allows you to do all the possible processing of the data
#' @param directory of the project and the thershold level that will be added to the database
#' @keywords VIIRS GRP linear fitting
#' @export
#' @examples
#' GRP_VIIRS_fit
GRP_VIIRS_fit<-function(directory=NULL,level=NULL,group="WB",subgroup="income",R_type) {

  n1= directory
  n2=paste(directory,"/NTL_data/ZSNTL_VIIRS14_buref_",level,"_stats.csv", sep="")
  n3=paste(directory,"/GRP_data/GRPs_2010_2014.txt", sep="")
  n4=level
  outputname=paste(group,"_",subgroup,sep="")

  #Load data VIIRS and GRP
  world_ntl_grp<-load_VIIRS(output =n1, NTL_data=n2, GRP_2014=n3,level=n4)

  # Filter data by analysis groups
  classing<-countries_class(data=world_ntl_grp, group, subgroup)

  # Linear Regression
  LR1<-GRP_VIIRS_LR(data=classing,"GRP_14",only_R2=FALSE)
  # Linear Regression
  LR2<-GRP_VIIRS_LR(data=classing,"GRP_ind14",only_R2=FALSE)
  # Linear Regression
  LR3<-GRP_VIIRS_LR(data=classing,"GRP_serv14",only_R2=FALSE)
  # Linear Regression
  LR4<-GRP_VIIRS_LR(data=classing,"GRP_indserv14",only_R2=FALSE)

  LRALL<-rbind(LR1,LR2,LR3,LR4)

  #Save data in an excel file

update_GRPmodel(output_folder=paste(directory,"/model_GRP",sep=""),output_name=paste("grp_model
2014",outputname,sep="_"), GRPmodel=LRALL,R_type)

}
```

MULTIFITS_VIIRS.R

```
#' This functions collects all the linear regressions about ceratin group-subgroup in an excel file for further comparison
#'
#' This function allows you to do create an excel file with the desired model comparisons
#' @param The min and max therholds with n step between them that you want to analyse divided in a ceratin group-subgroup and the type of GRP and the directory where the excel file will be saved.
#' @keywords VIIRS GRP linear fitting compilation
#' @export
#' @examples
#' multifits_VIIRS
multifits_VIIRS<-function(min,max,step,group,subgroup,directory=NULL,R_type)
  for (i in seq(min,max,by=step))
  {
    verify<-paste("H:/VIIRS_project/NTL_data/ZSNTL_VIIRS14_buref_",i,"_stats.csv",sep="")
    if (file.exists(verify)){
    GRP_VIIRS_fit(directory=directory,level=i,group=group,subgroup=subgroup,R_type)
    }
    else{
      verify="bla"
    }
  }
```

ALLPOSSIBLE_LR.R

```
#' This functions performs all the possible linear regression with the input thershold given by the user
#'
#' This function allows you to do create an excel file with all possible combinations of groups,subgroups and
GRP types that are available
#' @param The min and max therholds with n step between them that you want to analyse divided in a
ceratin group-subgroup and the type of GRP and the directory where the excel file will be saved.
#' @keywords VIIRS GRP linear fitting compilation WB IMF
#' @export
#' @examples
#' allpossible_LR
allpossible_LR<-function(min,max,step,directory=NULL,R_type="R2adj") {

   ###GRP_2014
   multifits_VIIRS(min,max,step,"WB","income",directory=directory , R_type)
   multifits_VIIRS(min,max,step,"WB","region",directory=directory , R_type)
   multifits_VIIRS(min,max,step,"WB","lending",directory=directory , R_type)
   multifits_VIIRS(min,max,step,"IMF","developement",directory=directory , R_type)
   multifits_VIIRS(min,max,step,"IMF","region",directory=directory , R_type)


 }
```

ALL_GRAPHSR.R

```
library(roxygen2)
library(devtools)
setwd("C:/Users/Enna/Documents/VIIRS_project/cats")

document ()

setwd("C:/Users/Enna/Documents/VIIRS_project")
install("cats")
library(cats)

#----------------------------------------

# Directory : This can be R2_predicted, R2, R2adj
output_folder="D:/VIIRS_project/model_GRP"

# Classification: IMF_developement, IMF_Region, WB_income, WB_lending,WB_region
output_name_1=paste("grp_model2014_","WB_region",sep="")
output_name_2=paste("grp_model2014_","WB_income",sep="")
output_name_3=paste("grp_model2014_","WB_lending",sep="")
output_name_4=paste("grp_model2014_","IMF_region",sep="")
output_name_5=paste("grp_model2014_","IMF_developement",sep="")

library("xlsx")
library("ggplot2")
library("plyr")

## Set directory
setwd(output_folder)
## Paste output path
output_path_1<-paste(output_folder,"/",output_name_1,".xlsx",sep="")
output_path_2<-paste(output_folder,"/",output_name_2,".xlsx",sep="")
output_path_3<-paste(output_folder,"/",output_name_3,".xlsx",sep="")
output_path_4<-paste(output_folder,"/",output_name_4,".xlsx",sep="")
output_path_5<-paste(output_folder,"/",output_name_5,".xlsx",sep="")

# Import data

# comparison<- read.xlsx(output_path, sheetName ="comparison")
model_info_WB_region<- read.xlsx(output_path_1, sheetName ="model_info")
model_info_WB_income<- read.xlsx(output_path_2, sheetName ="model_info")
model_info_WB_lending<- read.xlsx(output_path_3, sheetName ="model_info")
model_info_IMF_region<- read.xlsx(output_path_4, sheetName ="model_info")
model_info_IMF_developement<- read.xlsx(output_path_5, sheetName ="model_info")



model_info_WB_region<-unique(model_info_WB_region)
model_info_WB_income<-unique(model_info_WB_income)
```

```
model_info_WB_lending<-unique(model_info_WB_lending)
model_info_IMF_region<-unique(model_info_IMF_region)
model_info_IMF_developement<-unique(model_info_IMF_developement)


# Create new variable that evaluates if p-value is significant & summary tables
model_info_WB_region$p_test<-ifelse(model_info_WB_region$P.value_NTL<0.05,"TRUE","FALSE")
model_info_WB_region$R2adj<-as.numeric(as.character(model_info_WB_region$R2adj))
model_info_WB_region$GRP<-revalue(model_info_WB_region$GRP,c("GRP_14"="GRP-
2014","GRP_ind14"="Industrial sector GRP-2014","GRP_indserv14"="Industrial and service sector GRP-
2014","GRP_serv14"="Service sector GRP-2014"))
Smodel_info_WB_region<-summary_LR_GRP_VIIRS(model_info_WB_region,measurevar="R2adj",
groupvars=c("GroupCode","GRP"))
Smodel_info_WB_region<-
merge(Smodel_info_WB_region,model_info_WB_region,by.x=c("GroupCode","GRP","Max"),by.y=c("Group
Code","GRP","R2adj"),all.y=FALSE)
Smodel_info_WB_region<- Smodel_info_WB_region[, c(1,23, 2,3,21 )]
colnames(Smodel_info_WB_region)[4] <- "Max_R2adj"


model_info_WB_income$p_test<-ifelse(model_info_WB_income$P.value_NTL<0.05,"TRUE","FALSE")
model_info_WB_income$R2adj<-as.numeric(as.character(model_info_WB_income$R2adj))
model_info_WB_income$GRP<-revalue(model_info_WB_income$GRP,c("GRP_14"="GRP-
2014","GRP_ind14"="Industrial sector GRP-2014","GRP_indserv14"="Industrial and service sector GRP-
2014","GRP_serv14"="Service sector GRP-2014"))
Smodel_info_WB_income<-summary_LR_GRP_VIIRS(model_info_WB_income,measurevar="R2adj",
groupvars=c("GroupCode","GRP"))
Smodel_info_WB_income<-
merge(Smodel_info_WB_income,model_info_WB_income,by.x=c("GroupCode","GRP","Max"),by.y=c("Gro
upCode","GRP","R2adj"),all.y=FALSE)
Smodel_info_WB_income<- Smodel_info_WB_income[, c(1,23, 2,3,21 )]
colnames(Smodel_info_WB_income)[4] <- "Max_R2adj"


model_info_WB_lending$p_test<-ifelse(model_info_WB_lending$P.value_NTL<0.05,"TRUE","FALSE")
model_info_WB_lending$R2adj<-as.numeric(as.character(model_info_WB_lending$R2adj))
model_info_WB_lending$GRP<-revalue(model_info_WB_lending$GRP,c("GRP_14"="GRP-
2014","GRP_ind14"="Industrial sector GRP-2014","GRP_indserv14"="Industrial and service sector GRP-
2014","GRP_serv14"="Service sector GRP-2014"))
Smodel_info_WB_lending<-summary_LR_GRP_VIIRS(model_info_WB_lending,measurevar="R2adj",
groupvars=c("GroupCode","GRP"))
Smodel_info_WB_lending<-
merge(Smodel_info_WB_lending,model_info_WB_lending,by.x=c("GroupCode","GRP","Max"),by.y=c("Gro
upCode","GRP","R2adj"),all.y=FALSE)
Smodel_info_WB_lending<- Smodel_info_WB_lending[, c(1,23, 2,3,21 )]
colnames(Smodel_info_WB_lending)[4] <- "Max_R2adj"


model_info_IMF_region$p_test<-ifelse(model_info_IMF_region$P.value_NTL<0.05,"TRUE","FALSE")
model_info_IMF_region$R2adj<-as.numeric(as.character(model_info_IMF_region$R2adj))
model_info_IMF_region$GRP<-revalue(model_info_IMF_region$GRP,c("GRP_14"="GRP-
2014","GRP_ind14"="Industrial sector GRP-2014","GRP_indserv14"="Industrial and service sector GRP-
2014","GRP_serv14"="Service sector GRP-2014"))
```

```
Smodel_info_IMF_region<-summary_LR_GRP_VIIRS(model_info_IMF_region,measurevar="R2adj",
groupvars=c("GroupCode","GRP"))
Smodel_info_IMF_region<-
merge(Smodel_info_IMF_region,model_info_IMF_region,by.x=c("GroupCode","GRP","Max"),by.y=c("Grou
pCode","GRP","R2adj"),all.y=FALSE)
Smodel_info_IMF_region<- Smodel_info_IMF_region[, c(1,23, 2,3,21 )]
colnames(Smodel_info_IMF_region)[4] <- "Max_R2adj"


model_info_IMF_developement$p_test<-
ifelse(model_info_IMF_developement$P.value_NTL<0.05,"TRUE","FALSE")
model_info_IMF_developement$R2adj<-as.numeric(as.character(model_info_IMF_developement$R2adj))
model_info_IMF_developement$GRP<-revalue(model_info_IMF_developement$GRP,c("GRP_14"="GRP-
2014","GRP_ind14"="Industrial sector GRP-2014","GRP_indserv14"="Industrial and service sector GRP-
2014","GRP_serv14"="Service sector GRP-2014"))
Smodel_info_IMF_developement<-
summary_LR_GRP_VIIRS(model_info_IMF_developement,measurevar="R2adj",
groupvars=c("GroupCode","GRP"))
Smodel_info_IMF_developement<-
merge(Smodel_info_IMF_developement,model_info_IMF_developement,by.x=c("GroupCode","GRP","Max
"),by.y=c("GroupCode","GRP","R2adj"),all.y=FALSE)
Smodel_info_IMF_developement<- Smodel_info_IMF_developement[, c(1,23, 2,3,21 )]
colnames(Smodel_info_IMF_developement)[4] <- "Max_R2adj"


#----------------------------------------
# World Bank: Income
#----------------------------------------

## Graph the results (to change, R,and group)
WB_income1<-ggplot(data=model_info_WB_income, aes(x=level, y=R2adj,
group=GroupCode,shape=p_test, colour=GroupCode)) +
  #geomline(aes(colour=GroupCode),size=1)+
   geom_point(aes(shape=p_test,colour=GroupCode),size=2.5)+
  scale_shape_manual(values=c(4,20), name="P-value (NTL) <0.05")+
  scale_colour_brewer(palette="Set1",
                  name="Country Classification: WB",
                  breaks=c("LIC","LMC","UMC","NOC","OEC"),
                  labels=c("Low Income","Lower Middle Income","Upper Middle Income","High Income: Non
OECD","High Income: OECD"))+
  ylim(0, 1) + xlab ("NTL-BUREF Thresholds") + ylab(~R^{2}*"adjusted")+
  ggtitle("Model: GRPx~"~beta[0]*"+"~beta[1]*"NTL| Income ")+
  theme(plot.title = element_text(hjust = 1,size=11),legend.key = element_rect(colour = 'white', fill = 'white',
size = 0.5, linetype='dashed'),
       axis.text = element_text(size = 12), legend.text = element_text(size = 12),legend.title =
element_text(size = 12))

## Labeller for Facet-wrap graph

WB_income2<-WB_income1 + facet_grid( GroupCode~ GRP )
WB_income3<-WB_income1 + facet_wrap( ~ GRP, ncol=2)+theme(strip.text.x=element_text(size =
12))+guides(shape = guide_legend(order = 1),colour = guide_legend(order = 2))
```

```
#----------------------------------------
# World Bank: Region
#----------------------------------------

## Graph the results (to change, R,and group)
WB_region1<-ggplot(data=model_info_WB_region, aes(x=level, y=R2adj, group=GroupCode,shape=p_test))
+
  #geomline(aes(colour=GroupCode),size=1)+
   geom_point(aes(shape=p_test,colour=GroupCode),size=2.5)+
  scale_colour_brewer(palette="Set1",
                  name="Country Classification: WB",
                  breaks=c("EAS","ECS","LCN","MEA","NAC","SAS","SSF"),
                  labels=c("East Asia & Pacific","Europe & Central Asia","Latin America & Caribbean","Middle
East & North Africa", "North America", "South Asia", "Sub Saharan Africa"))+
  scale_shape_manual(values=c(4,20), name="P-value (NTL)<0.05")+
  ylim(0, 1) + xlab ("NTL-BUREF Thresholds") + ylab(~R^{2}*"adjusted")+
  ggtitle("Model: GRPx~"~beta[0]*"+"~beta[1]*"NTL| Region ")+
  theme(plot.title = element_text(hjust = 1,size=11),legend.key = element_rect(colour = 'white', fill = 'white',
size = 0.5, linetype='dashed'),
       axis.text = element_text(size = 12), legend.text = element_text(size = 12),legend.title =
element_text(size = 12))

## Labeller for Facet-wrap graph

WB_region2<-WB_region1 + facet_grid( GroupCode~ GRP )
WB_region3<-WB_region1 + facet_wrap( ~ GRP, ncol=2)+theme(strip.text.x=element_text(size =
12))+guides(shape = guide_legend(order = 1),colour = guide_legend(order = 2))
#----------------------------------------
# World Bank: Lending
#----------------------------------------

## Graph the results (to change, R,and group)
WB_lending1<-ggplot(data=model_info_WB_lending, aes(x=level, y=R2adj,
group=GroupCode,shape=p_test)) +
  #geomline(aes(colour=GroupCode),size=1)+
   geom_point(aes(shape=p_test,colour=GroupCode),size=2.5)+
  scale_colour_brewer(palette="Set1",
                  name="Country Classification: WB",
                  breaks=c("IBD","IDB","IDX"),
                  labels=c("IBRD only","IDA blend","IDA only"))+
  scale_shape_manual(values=c(4,20), name="P-value (NTL)<0.05")+
  ylim(0, 1) + xlab ("NTL-BUREF Thresholds") + ylab(~R^{2}*"adjusted")+
  ggtitle("Model: GRPx~"~beta[0]*"+"~beta[1]*"NTL| Lending Group")+
  theme(plot.title = element_text(hjust = 1,size=11),legend.key = element_rect(colour = 'white', fill = 'white',
size = 0.5, linetype='dashed'),
       axis.text = element_text(size = 12), legend.text = element_text(size = 12),legend.title =
element_text(size = 12))

## Labeller for Facet-wrap graph

WB_lending2<-WB_lending1 + facet_grid( GroupCode~ GRP )
WB_lending3<-WB_lending1 + facet_wrap( ~ GRP, ncol=2)+theme(strip.text.x=element_text(size =
12))+guides(shape = guide_legend(order = 1),colour = guide_legend(order = 2))
```

```
#----------------------------------------
# IMF: Region
#----------------------------------------

## Graph the results (to change, R,and group)
IMF_region1<-ggplot(data=model_info_IMF_region, aes(x=level, y=R2adj,
group=GroupCode,shape=p_test)) +
  #geomline(aes(colour=GroupCode),size=1)+
   geom_point(aes(shape=p_test,colour=GroupCode),size=2.5)+
  scale_colour_brewer(palette="Set1",
                  name="Country Classification: IMF",
                  labels=c("Commonwealth IS","Emerging and Developing Asia","Euro Area","Latin America
Caribbean","Major Advanced economies","Middle East and North Africa","Other adv.Economies","Sub
Saharan Africa"))+
  scale_shape_manual(values=c(4,20), name="P-value (NTL)<0.05")+
  ylim(0, 1) + xlab ("NTL-BUREF Thresholds") + ylab(~R^{2}*"adjusted")+
  ggtitle("Model: GRPx~"~beta[0]*"+"~beta[1]*"NTL| Region ")+
  theme(plot.title = element_text(hjust = 1,size=11),legend.key = element_rect(colour = 'white', fill = 'white',
size = 0.5, linetype='dashed'),
       axis.text = element_text(size = 12), legend.text = element_text(size = 12),legend.title =
element_text(size = 12))

## Labeller for Facet-wrap graph

IMF_region2<-IMF_region1 + facet_grid( GroupCode~ GRP )
IMF_region3<-IMF_region1 + facet_wrap( ~ GRP, ncol=2)+theme(strip.text.x=element_text(size =
12))+guides(shape = guide_legend(order = 1),colour = guide_legend(order = 2))
#----------------------------------------
# IMF: Developement
#----------------------------------------

## Graph the results (to change, R,and group)
IMF_developement1<-ggplot(data=model_info_IMF_developement, aes(x=level, y=R2adj,
group=GroupCode,shape=p_test)) +
  #geomline(aes(colour=GroupCode),size=1)+
  geom_point(aes(shape=p_test,colour=GroupCode),size=2.5)+
  scale_colour_hue(h=c(-20, 300),name="Country Classification: IMF",
                labels=c("Advanced economies","ASEAN 5","Commonwealth IS","Emerging and
developing","Emerging and Developing Asia",
                 "Emerging and Developing Europe", "Euro Area", "European union", "Latin America Caribbean",
"Major Advanced economies", "Middle East and North Africa","Other adv Economies", "Sub Saharan
Africa"))+
  scale_shape_manual(values=c(4,20), name="P-value (NTL)<0.05")+
  ylim(0, 1) + xlab ("NTL-BUREF Thresholds") + ylab(~R^{2}*"adjusted")+
  ggtitle("Model: GRPx~"~beta[0]*"+"~beta[1]*"NTL|  Level of Developement ")+
  theme(plot.title = element_text(hjust = 1,size=11),legend.key = element_rect(colour = 'white', fill = 'white',
size = 0.5, linetype='dashed'),
       axis.text = element_text(size = 12), legend.text = element_text(size = 12),legend.title =
element_text(size = 12))

## Labeller for Facet-wrap graph

IMF_developement2<-IMF_developement1 + facet_grid( GroupCode~ GRP )
```

```
IMF_developement3<-IMF_developement1 + facet_wrap( ~ GRP,
ncol=2)+theme(strip.text.x=element_text(size = 12))+guides(shape = guide_legend(order = 1),colour =
guide_legend(order = 2))

#----------------------------------------
# IMF: Developement:2
#----------------------------------------
model_info_IMF_developement2<-subset(model_info_IMF_developement, GroupCode==
"Advanced_economies" | GroupCode=="Commonwealth_IS"|
GroupCode=="Emerging_and_Developing_Asia"| GroupCode=="Emerging_and_Developing_Europe"|
GroupCode=="Latin_America_Caribbean"| GroupCode=="Middle_East_and_North_Africa"|
GroupCode=="Sub_Saharan_Africa",select =1:16 )
Smodel_info_IMF_developement<-
summary_LR_GRP_VIIRS(model_info_IMF_developement2,measurevar="R2adj",
groupvars=c("GroupCode","GRP"))
Smodel_info_IMF_developement<-
merge(Smodel_info_IMF_developement,model_info_IMF_developement2,by.x=c("GroupCode","GRP","Ma
x"),by.y=c("GroupCode","GRP","R2adj"),all.y=FALSE)
Smodel_info_IMF_developement<- Smodel_info_IMF_developement[, c(1,23, 2,3,21 )]
colnames(Smodel_info_IMF_developement)[4] <- "Max_R2adj"


## Graph the results (to change, R,and group)
IMF_developement4<-ggplot(data=model_info_IMF_developement2, aes(x=level, y=R2adj,
group=GroupCode,shape=p_test)) +
  #geomline(aes(colour=GroupCode),size=1)+
  geom_point(aes(shape=p_test,colour=GroupCode),size=2.5)+
  scale_colour_brewer(palette="Dark2",name="Country Classification: IMF",

breaks=c("Advanced_economies","Commonwealth_IS","Emerging_and_Developing_Asia","Emerging_and_
Developing_Europe","Latin_America_Caribbean","Middle_East_and_North_Africa","Sub_Saharan_Africa"),
              labels=c("Advanced economies","Commonwealth IS","Emerging and Developing Asia","Emerging
and Developing Europe", "Latin America & Caribbean", "Middle East and North Africa", "Sub Saharan
Africa"))+
  scale_shape_manual(values=c(4,20), name="P-value (NTL)<0.05")+
  ylim(0, 1) + xlab ("NTL-BUREF Thresholds") + ylab(~R^{2}*"adjusted")+
  ggtitle("Model: GRPx~"~beta[0]*"+"~beta[1]*"NTL| Level of Developement")+
  theme(plot.title = element_text(hjust = 1,size=11),legend.key = element_rect(colour = 'white', fill = 'white',
size = 0.5, linetype='dashed'),
      axis.text = element_text(size = 12), legend.text = element_text(size = 12),legend.title =
element_text(size = 12))

## Labeller for Facet-wrap graph

IMF_developement5<-IMF_developement4 + facet_grid( GroupCode~ GRP )
IMF_developement6<-IMF_developement4 + facet_wrap( ~ GRP,
ncol=2)+theme(strip.text.x=element_text(size = 12))+guides(shape = guide_legend(order = 1),colour =
guide_legend(order = 2))


#----------------------------------------
# Add horizontal and vertical lines at highest R2adj points
#----------------------------------------
```

```
WB_income2<-WB_income2 + geom_hline(aes(yintercept=Max_R2adj),data=Smodel_info_WB_income,
colour="#990000", linetype="dashed") +
  geom_vline(aes(xintercept=level), data=Smodel_info_WB_income,
          colour="#990000", linetype="dashed")+scale_x_continuous(breaks=seq(5,100,10))

WB_region2<-WB_region2 + geom_hline(aes(yintercept=Max_R2adj),data=Smodel_info_WB_region,
colour="#990000", linetype="dashed") +
  geom_vline(aes(xintercept=level), data=Smodel_info_WB_region,
          colour="#990000", linetype="dashed")+scale_x_continuous(breaks=seq(5,100,10))

WB_lending2<-WB_lending2+ geom_hline(aes(yintercept=Max_R2adj),data=Smodel_info_WB_lending,
colour="#990000", linetype="dashed") +
  geom_vline(aes(xintercept=level), data=Smodel_info_WB_lending,
          colour="#990000", linetype="dashed")+scale_x_continuous(breaks=seq(5,100,10))

IMF_region2<-IMF_region2+ geom_hline(aes(yintercept=Max_R2adj),data=Smodel_info_IMF_region,
colour="#990000", linetype="dashed") +
  geom_vline(aes(xintercept=level), data=Smodel_info_IMF_region,
          colour="#990000", linetype="dashed")+scale_x_continuous(breaks=seq(5,100,10))+
           theme(strip.text.x = element_text(size=8),strip.text.y = element_text(size=0))

IMF_developement5<-IMF_developement5+
geom_hline(aes(yintercept=Max_R2adj),data=Smodel_info_IMF_developement, colour="#990000",
linetype="dashed") +
  geom_vline(aes(xintercept=level), data=Smodel_info_IMF_developement,
          colour="#990000", linetype="dashed")+scale_x_continuous(breaks=seq(5,100,10))+
          theme(strip.text.y = element_text(size=0))
#----------------------------------------
# Add horizontal and vertical lines at highest R2adj points
#----------------------------------------
WB_income2
WB_region2
WB_lending2
IMF_region2
IMF_developement5


WB_income3
WB_region3
WB_lending3
IMF_region3
IMF_developement6



#----------------------------------------
# save graphs as pdf
#----------------------------------------
setwd("C:/Users/Enna/Documents/VIIRS_project/graphs")
ggsave("WB_income3.pdf",WB_income3,width=13,height = 7)
ggsave("WB_region3.pdf",WB_region3,width=13,height = 7)
ggsave("WB_lending3.pdf",WB_lending3,width=13,height = 7)
ggsave("IMF_region3.pdf",IMF_region3,width=13,height = 7)
ggsave("IMF_developement6.pdf",IMF_developement6,width=13,height = 7)
```

```
ggsave("WB_income2.pdf",WB_income2,width=15,height = 8)
ggsave("WB_region2.pdf",WB_region2,width=15,height = 8)
ggsave("WB_lending2.pdf",WB_lending2,width=15,height = 8)
ggsave("IMF_region2.pdf",IMF_region2,width=15,height = 15)
ggsave("IMF_developement5.pdf",IMF_developement5,width=15,height = 8)


#----------------------------------------
# save summary as excel file
#----------------------------------------
LR_comparisons<- createWorkbook()
sheet_1<-createSheet(LR_comparisons,sheetName ="WB_income")
sheet_2<-createSheet(LR_comparisons,sheetName ="WB_region")
sheet_3<-createSheet(LR_comparisons,sheetName ="WB_lending")
sheet_4<-createSheet(LR_comparisons,sheetName ="IMF_developement")
sheet_5<-createSheet(LR_comparisons,sheetName ="IMF_region")

addDataFrame(Smodel_info_WB_income, sheet_1, startRow=1, startColumn=1,row.names=FALSE)
addDataFrame(Smodel_info_WB_region, sheet_2, startRow=1, startColumn=1,row.names=FALSE)
addDataFrame(Smodel_info_WB_lending, sheet_3, startRow=1, startColumn=1,row.names=FALSE)
addDataFrame(Smodel_info_IMF_developement, sheet_4, startRow=1, startColumn=1,row.names=FALSE)
addDataFrame(Smodel_info_IMF_region, sheet_5, startRow=1, startColumn=1,row.names=FALSE)
saveWorkbook(LR_comparisons, "C:/Users/Enna/Documents/VIIRS_project/graphs/LR_comparisons.xlsx")
```