Spatio-temporal analysis of Urban sprawl of North East India using DMSP-OLS and VIIRS-DNB night-light data from 1992 to 2017.

Rachit Agrawal, Pranjal Dave, Kumar Anant Raj

Supervised by: Soham Mukherjee

26/03/19

1 Introduction

Remote sensing is used widely for the study of the features of the surface of the earth. Night-light time imagery forms an essential data set in remote sensing[3]. It is essential for study of human activities on earth and their effect. Such datasets are also found to be of great importance in study of the atmosphere and other natural processes. A few common examples are detection and monitoring of city lights, fires, dust storms, volcanoes, gas flares, population/economic geography, etc. As the name suggests, Night-time light imagery data is collected during the earth during the night. During the night, we can detect two types of features, first are the self-illuminating features and the moon-illuminated features. Some examples of self-illuminating features include gas flares, forest fires, human-caused disasters, volcanic lava, and bioluminescence. Moonlight illuminating features include snow cover, sea ice, volcanic ash and various surface features such as mountains, deserts, rivers, and moon glint. Another source of illumination for the clouds can be airglow. It is the luminosity occurring due to chemical reactions involving ozone, sodium, and oxygen and nitrogen molecules in the upper atmosphere (www.meted.ucar.edu, 2017). Airglow can be exploited for the study of clouds during the darkest nights. Night-time visible imaging was initiated

by the Defense Meteorological Satellite Program (DMSP) Operational Linescan System (OLS) in the 1960s, with a mission which was the only source of visible nighttime images until the launch of Suomi National Polar-orbiting Partnership visible Infrared Imaging Radiometer Suite (SNPP-VIIRS). The successor of DMSP-OLS was launched in October 2011 has continued the acquisition of night time visible images with some quality enhancements over DMSP-OLS. The DMSP satellite of the U.S Department of Defence provides global coverage every 24 hours. The onboard OLS sensor acquires images in visible, near infrared and thermal infrared region of the electromagnetic spectrum. It is the longest running time series of nighttime lights data. DMSP-OLS is an oscillating scan radiometer with two spectral bands: Visible Near-infrared (VNIR) for Night-time Light (NTL) and Thermal Infrared. The night-time overpass is between 20:30 and 21:30. The data is reported as Digital Number (DN) values on a six-bit scale that ranges from 0 (no light) to 63 (maximum light). A total of 9 satellites from F10 to F18 have collected OLS data. The data is mainly of three types: stable lights, radiance calibrated and average digital number. These categories are based on different parameters of detection frequency, radiance, and digital number respectively. These differences make the datasets usable in different applications such as urban extent, population, economic activity, greenhouse gas emissions, light pollution, and disaster management.

2 Problem statement

Nocturnal manmade lighting is considered as a good indicator of various socio-economic studies e.g.; Urban expansion, GDP estimation and many more (Doll, 2008). DMSP-OLS and VIIRS-DNB acquired night-light data are of the mostly used nightlight products and frequently used in many socio-economic studies (Li et al., 2017). However, these two satellite products have been acquired in different time periods; DMSP-OLS provides coverage from 1992 to 2013 whereas, VIIRS-DNB is available from end of 2012 only (Shao et al., 2014). The radiometric quantification is also observed to be different of these two products; DMSP-OLS product is provided in DN values with a 6 bit radiometric resolution whereas VIIRS-DNB consists of radiance values with a 14 bit radiometric quantification (Xie, Weng, Weng, 2014). Hence a relative radiometric normalization (inter-calibration in current context) between DMSP-OLS and VIIRS-DNB acquired data is necessary in order

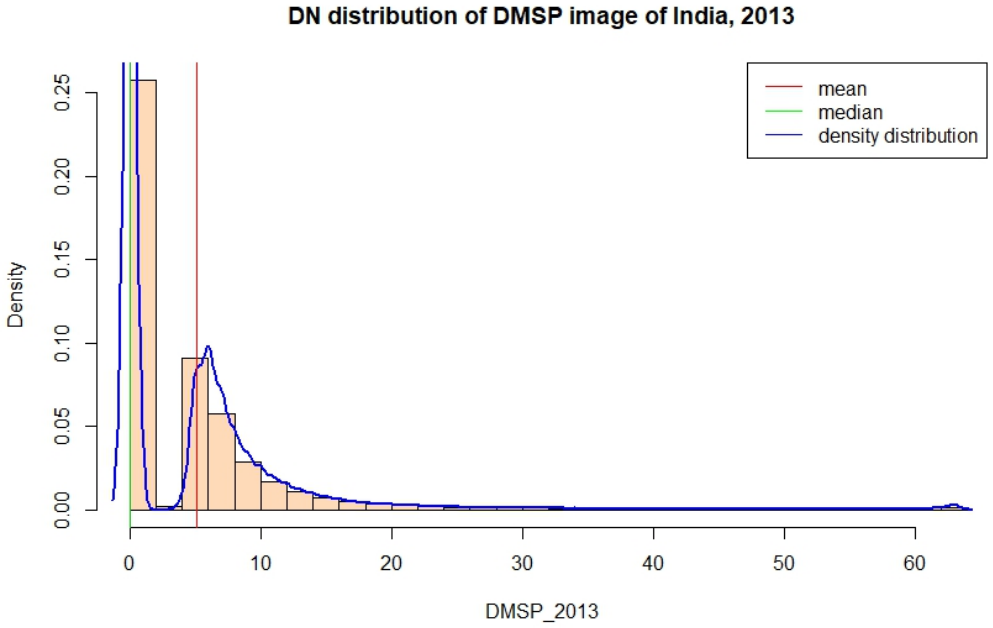


Figure 1: DN distribution of DMSP images of India

to study the urban expansion. The study will develop an inter-calibration model between both the products and also evaluate the quality of calibration through estimating the urban expansion of North East India for a time period of 1992 to 2017 and checking the accuracy and precision of the estimate.

3 Dataset used:

1. DMSP-OLS Version 4 Stable light yearly composite product from 1992 to 2013
2. VIIRS-DNB Version 1 averaged radiance monthly composite from 2012 to 2017.

4 Challenge

VIIRS-DNB data are provided in radiance values whereas DMSP-OLS data are provided in DNs with a varying radiometric resolution between them. Thus, it is necessary and as well as challenging to develop an inter-calibration algorithm which will also be free of selection bias between the datasets mentioned. OGC standard WMS provides a platform to publish geospatial data on the web. In case of time series raster data, transition from one date to another should be smooth. Thus, it is important to develop a framework to handle time series information.

5 Objectives

1. Generate calibrated DMSP-OLS and VIIRS-DNB products.
2. Sum of Light (SOL) of different districts in NE India through 1992 to 2017.

6 Methodology

6.1 For DMSP-OLS:

1. All pixels with Digital Number(DN) value 63 were removed because of the oversaturation of DMSP data for Digital Number values greater than 63.
2. A threshold value of 5 for Digital Number(DN) was estimated to reduce the blooming effect. Hence all pixels with $DN < 5$ were eliminated.

6.2 For VIIRS:

1. Outliers were removed by using median absolute deviation(MAD).
2. Faint light features (including pixels of negative radiance values) were removed by using a process of local averaging.
3. The data was also brought to the same resolution as the DMSP-OLS data.

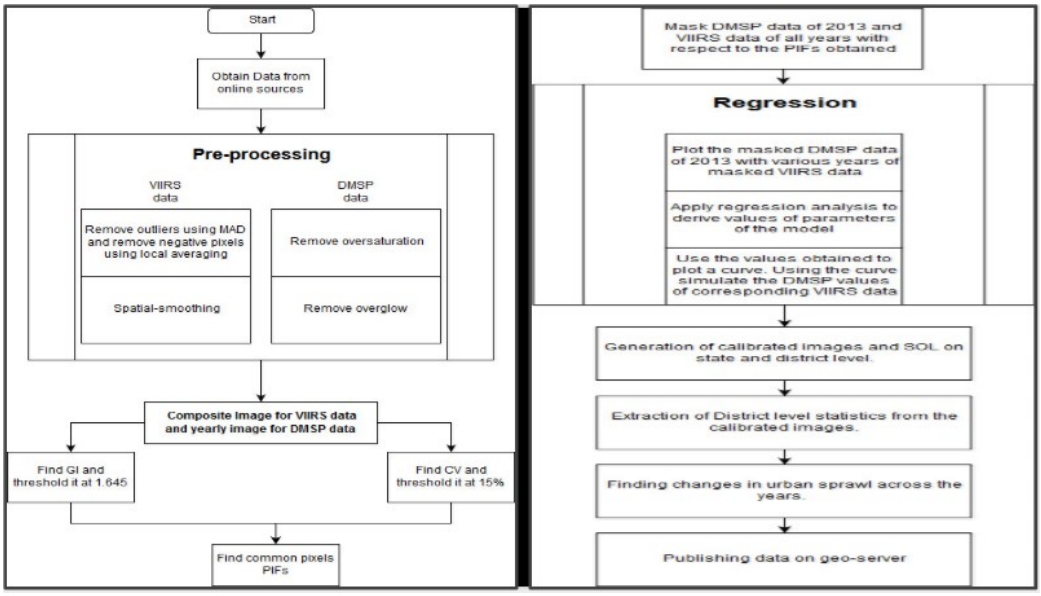


Figure 2: Methodology used

7 Median Absolute Deviation(MAD)

It measures the variability of a univariate sample of Quantitative data. It calculated as Median of absolute deviation from the median of data.

$$MAD = b * M_i(x_i - M_j(x_j))$$

where M_i denotes median of x_i

1. The multiplication by b is crucial, as otherwise, the formula for the MAD would only estimate the scale up to a multiplicative constant the constant b = 1.4826 assuming normality of data.

2. This method is used because it is mostly unaffected by the presence of outliers in data. It has a breakdown point of 0.5 whereas the breakdown point of the classical interquartile range is 0.25.

3. Data points outside the range of (M - 2.5* MAD, M + 2.5*MAD) were removed as outliers.

8 Gas Flare:

1. Gas flaring is widely used to dispose of dissolved natural gas present in petroleum in production and processing facilities where there is no infrastructure to make use of the gas
2. Most gas flaring occurs at remote petroleum production or processing facilities, many of which are offshore
3. Flaring sites are indicated for grid cells with average temperature exceeding 1200 K and per cent detection frequencies of 1% or more.
4. Currently no precise algorithm is there. A gas flaring vector set is available, with more than 12,000 flaring sites identified in 2015
5. The site present in India were identified and were removed from the yearly composites.

9 Spatial Smoothing of VIIRS Data

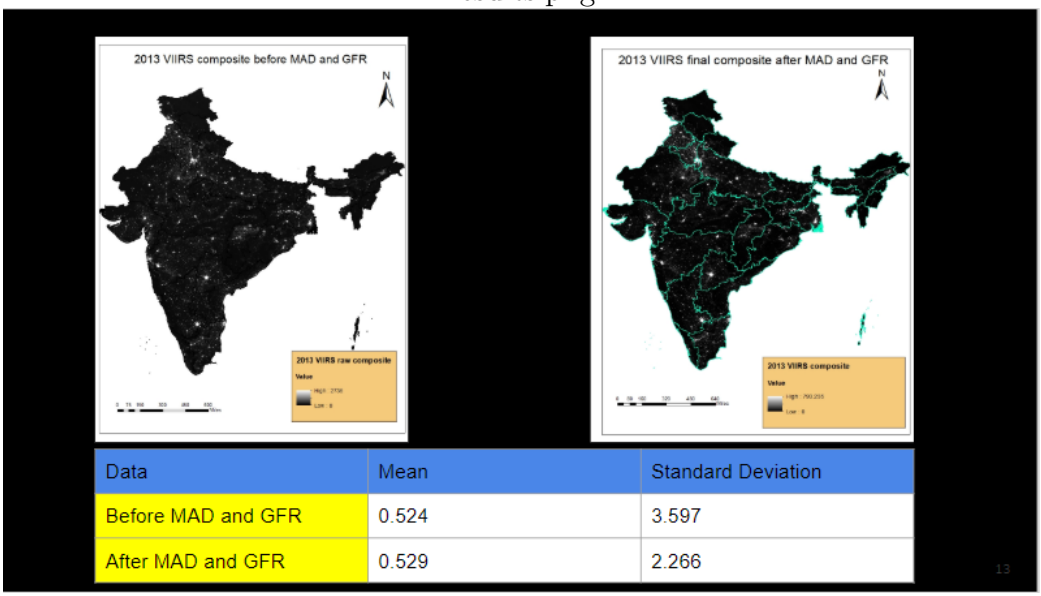
1. The spatial resolution of VIIRS composites is 0.004167Â°, while that of the DMSP/OLS is 0.008333Â°. Therefore, the VIIRS composites were aggregated to a spatial resolution of 0.008333Â°.
2. This was done using the concept of local averaging where a 2X2 submatrix of VIIRS data was aggregated and the result was put into a single pixel of matrix of dimension equal to that of DMSP/OLS.
3. This process was necessary to bring the area corresponding to same geographical location to the same indexed pixel in both DMSP/OLS and VIIRS data.

10 Composite Generation

1. The VIIRS data are available for each month.
2. The pre -processing steps are applied to each months' data.
3. Then the average of the monthly composites is taken to find the VIIRS data for the whole year.

4. The averaging process takes care of the ' No data 's values to ensure they don'st interfere in the process of finding the composite.

results.png



Data	Mean	Standard Deviation
Before MAD and GFR	0.524	3.597
After MAD and GFR	0.529	2.266

Figure 3: 2013 VIIRS composite

11 Statistics used

11.1 Coefficient of Variability(CV)

PIFs consists of pixels with both spatial homogeneity and low spatial variability. Thus, on the extraction of relatively bright clusters with variable window size, regions with low spatial variability are also extracted. The coefficient of variability (CV)[6] is used in the project to select local spatially stable regions in the study area. The areas with a low CV over different years are considered as regions with stable and low spatial variability.

$$CV = \frac{S}{\bar{x}}$$

Where S is the measure of standard deviation of luminosity measurements, $\bar{x}$ is the mean of luminosity measurements in a pre-defined window.

The CV is calculated for the images from 2009 to 2017 on a variable window size of 3X3 , 5X5 and 9X9. A threshold of CV is computed to select the areas with low spatial variability. Such use of threshold can also be seen in works of Bannari et al. (2005), Kneubhuler et al. (2006), Scott et al. (1996) and Odongo et al. (2014) where they have considered an area as temporally stable for vicarious calibration if the CV remains within a threshold of 3%ĊV value of 5% 10% and 15% are used as thresholds. Due to less no. of points available after thresholding for 5% and 10% , the threshold level was pushed to 15% for better and reliable regression.

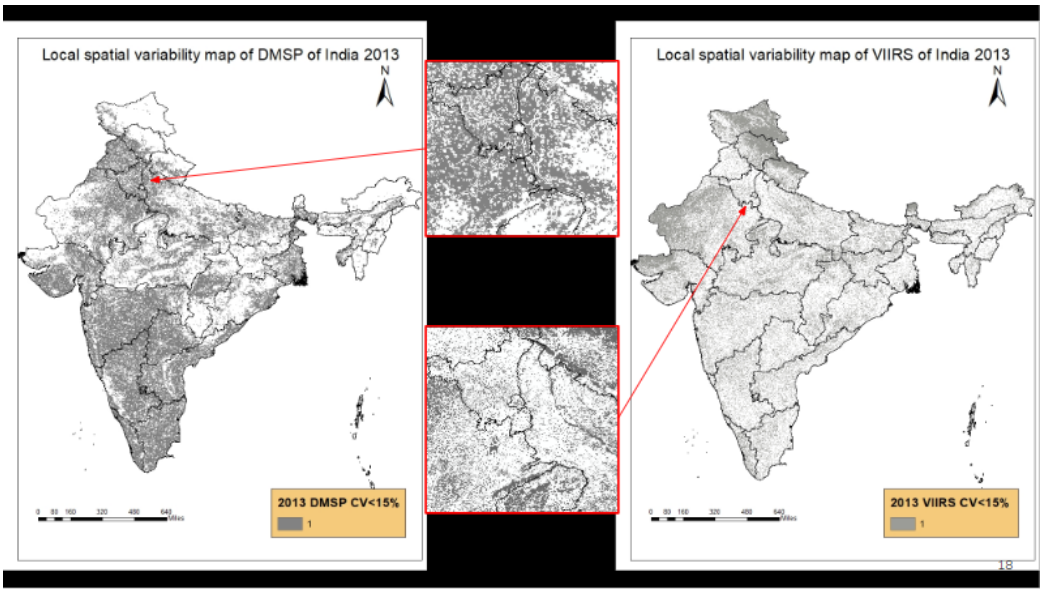


Figure 4: CV statistics result on the regions near Delhi

11.2 Getis-ord(GI*)

Getis Ord statistic [6] is a measure of spatial association generated from the concentration of weighted points in a defined local area where the whole target site is spatially non-stationary (Getis Ord, 1992). The area is subdivided into n regions; i =1, 2,3. . ., n. Each region is identified with a point with known coordinates. Each point in the regions is given weight as per the change in spatial attribute of the variable whose spatial association is being evaluated (DN value here). The variable should be positive and natural. Gi*

provides test scores of the hypothesis about the spatial association of the sum of weights assigned to points in sub areas within the distance of the point in ith region. The Gi* formula is as following in equation.

$$W_i^* = \sum_j W_{ij}(d)$$

$$\bar{x} = \frac{\sum_j x_j}{n}$$

Here, n is the total number of subregions, $\bar{x}$ is the global mean of x, s is the standard deviation of x, $w_{ij}(d)$ is a matrix of spatial weights with binary and symmetric weight equal to one ($w_{ij} = 1$) for all pixels found within distance d of pixel i considered and a weight equal to zero ($w_{ij} = 0$) for all pixels found outside d, $\sum w_{ij}(d)x_j$ is the sum of varying values of light intensity (DN) of the imagery within distance d of pixel i till n and is the DN value. d is the distance obtained from the fixed window size in which Gi* is implemented (d=1; 3X3 window, d=2; 5X5 window). The Getis statistic gives a measure of clustering, and it is used to identify clusters of comparatively higher values ($Gi^* > 0$; hotspot) and clusters of comparatively lower values ($Gi* < 0$; cold spot) than the global mean of the image attribute. A positive output of Gi* indicates hotspots (cluster of high valued pixels) and a negative output indicates cold spot (cluster of low valued pixels) (Getis and Ord, 1992). In the research, the image attribute is light intensity of DN values in the study area. Gi* is used to identify relatively bright pixel clusters ($Gi* > 0$) and dark pixel clusters ($Gi* < 0$). As the most stable light emitting luminous areas are assessed for the selection of PIF, only bright clusters (hotspots) are identified. The similar selection of the values of Gi* to identify bright pixel clusters can also be found in work of Bannari et al. (2005) and Odongo et al. (2014) where they have used $Gi* > 0$ to identify clusters of relatively bright pixels.In the research, only the Gi* output value with more than 1.645 is considered as a potential cluster in the research among all the values in hotspot as Gi* value greater than 1.645 are associated with a significance level of 0.10. It does ensure the maximum probability of random cluster formation is limited by only 10% of all the clusters have formed and a 90% probability of the formed cluster to have high spatial autocorrelation.

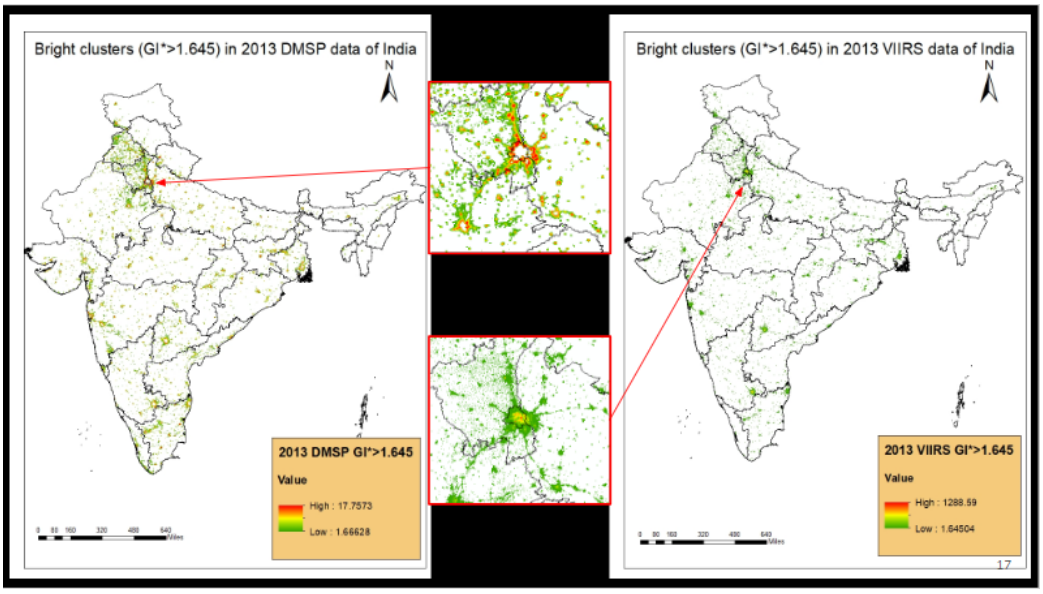


Figure 5: GI* statistics result on the regions near Delhi

12 Pseudo Invariant Features(PIFs)

PIFs are the regions that are both spatially homogeneous and of low local spatial variability. A combination of CV and Gi* statistics are used to extract out PIFs. The PIF for the year 1992, 2001 and 2006 are used to compute the optimal window size to execute the LISA measures. These PIFs with different CV thresholds are taken input in [R-3.3.2]TM as rasters and plotted keeping PIF of 2001 at X-axis. It shows the number of points in the scatterplot which gives the idea of change of DN values among years. OLS is implemented on plots between PIF2006 and PIF2001, PIF1992 and PIF2001 for each CV threshold and a linear regression line is fitted in the scatterplot. Rand number of observations are used as the metric to select the better fit of the trendline among the PIFs generated with different CV thresholds. The CV threshold that generates PIFs with a long range of DN values with a high number of observations in the OLS is selected as optimal CV threshold for generating PIF. Thresholding was done for 5% 10% and 15% and PIFs were extracted from the image. The no. of PIFs extracted for 5% and 10% were insufficient for plotting a reliable regression. Hence 15% was taken as the threshold for

CV. To select the optimal window size PIFs are generated with pixels Gi* ¿1.645 and CV ¡ âoptimal threshold under an increased window size of 5X5. These PIFs for the year 1992, 2001 and 2006 are also plotted in [R-3.3.2]TM similar to the approach of selecting CV threshold above. The window size between 3X3 and 5X5 that generates a higher number of observations along with a big range of DN values and high R2 is selected as the optimal window size. If 5X5 results as better size, the same process is repeated with a window size of 7X7 and compared with the results of 5X5 else the research follows window size of 3X3.

DMSP PIFs = Gi* AND CV

VIIRS PIFs = Gi* AND CV

Total PIFs = DMSP PIFs AND VIIRS PIFs

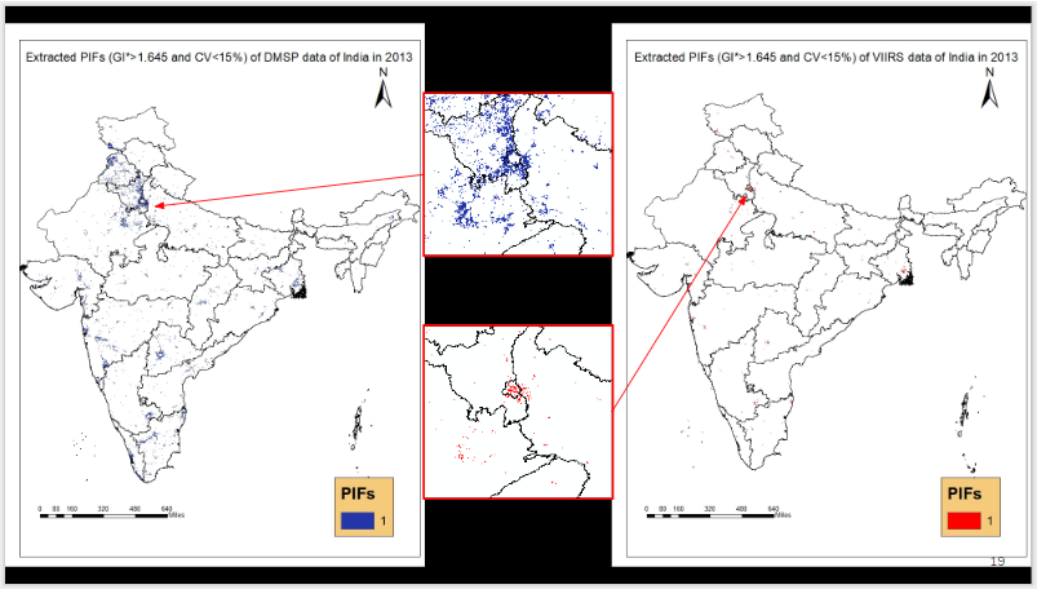


Figure 6: PIFs on the regions near Delhi

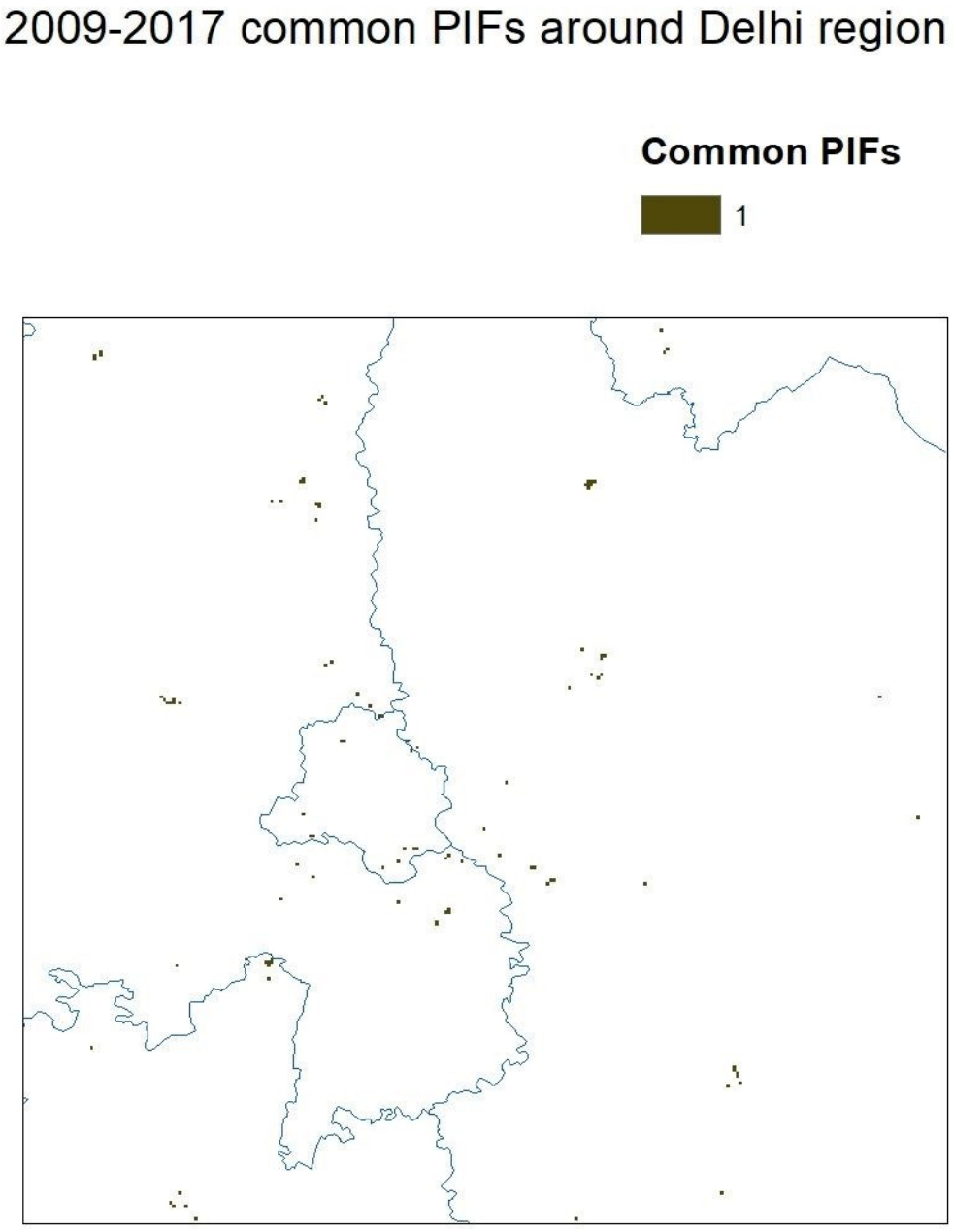


Figure 7: 2009-17 common PIF near Delhi

variance.png

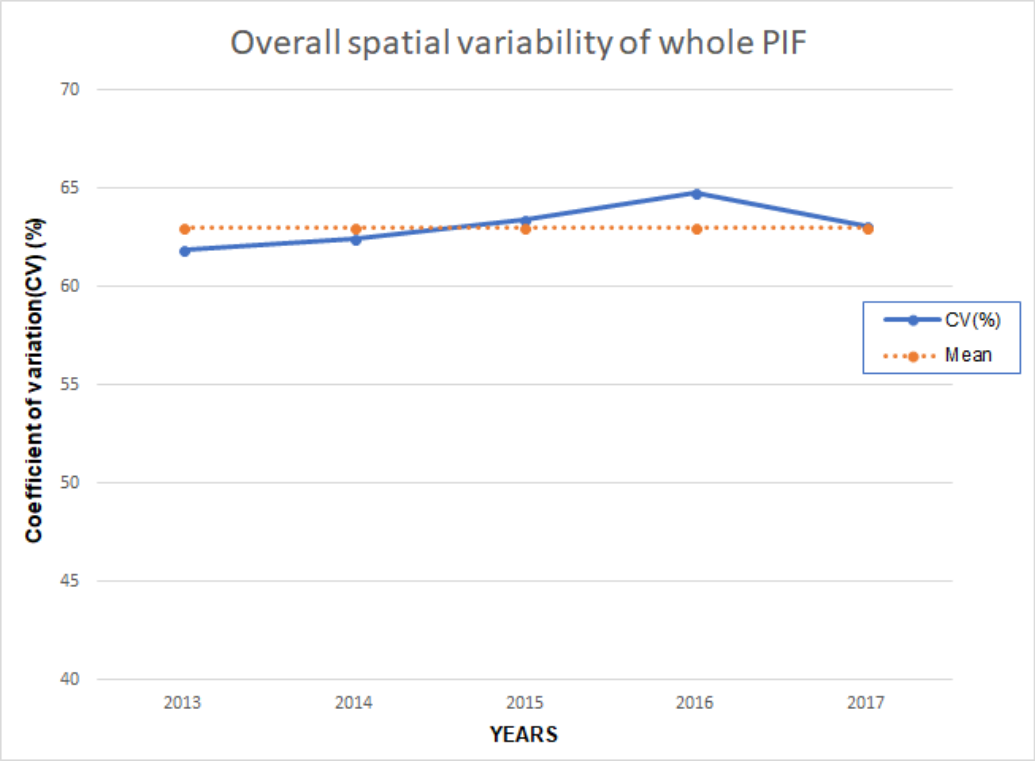


Figure 8: Spatial Variance of the no. of PIFs

13 Inter-calibration between DMSP-OLS data of varying years

Unfortunately, the DMSP satellite has no onboard calibration for the visible band. Due to atmospheric changes in the subsequent years, the satellite images needed to be intercalibrated. This intercaliberation was done by Soham Mukherjee[4]. They selected the satellite year F121999 as a reference year and calibrated other images using regression

A second-order regression model was developed for each satellite year, as shown in Table

Satellite	Year	C_o	C_1	C_2	R^2
F10	1992	−2.0570	1.5903	−0.0090	0.9075
F12	1996	−0.0959	1.2727	−0.0040	0.9319
F12	1999(Ref.)	0.0000	1.0000	0.0000	1.0000
F14	2003	0.7390	1.2416	−0.0040	0.9432
F16	2007	0.6394	0.9114	0.0014	0.9511
F18	2011	1.8956	0.7345	0.0030	0.9095

Figure 9: Regression Coefficients for different images

The form of the calculation is

$$Y = C_0 + C_1 x + C_2 x^2$$

. Calculated values that run over 63 are truncated at 63.

14 Inter-calibration between DMSP-OLS and SNPP-VIIRS

The importance of post-launch calibration is to validate the sensor's data products and to ensure correct data. We use inter-calibration of VIIRS and DMSP-OLS to predict the extent of urbanization. We use spatially, and temporary homogenous region called PIFs and assume that these regions do not change much temporary. The net PIFs were used to generate a regression model to inter-calibrate the two datasets. We tested for logarithmic, exponential, quadratic and Linear[2],[5]. The logarithmic model fitted the dataset most appropriately. The post-launch calibration includes direct calibration method in which test sites with the ground reference are considered and Inter-calibration in which one data product is calibrated using another data product to ensure the consistency and homogeneity of the datasets.

14.1 Using Logarithmic function

The relationship between DMSP and VIIRS fits the logarithmic model:

$$y = a\log(x) + b$$

Where, x denotes VIIRS value, y denotes the DMSP/OLS value, a and b are coefficients. Inter-calibration equations: $y(dmsp) = a\log(viirs) + b$

Where y signifies the calculation of calibrated (simulated VIIRS) by altering the equation with coefficients -

Year	2013	2014	2015	2016	2017
a	11.2582	11.4623	12.2764	11.9637	12.7412
b	24.6421	27.8722	23.7286	25.8012	23.4470
RMSE	9.7624	9.913	9.787	9.832	9.991
R^2	0.72	0.75	0.74	0.83	0.78

Figure 10: Logarithmic regression coefficient

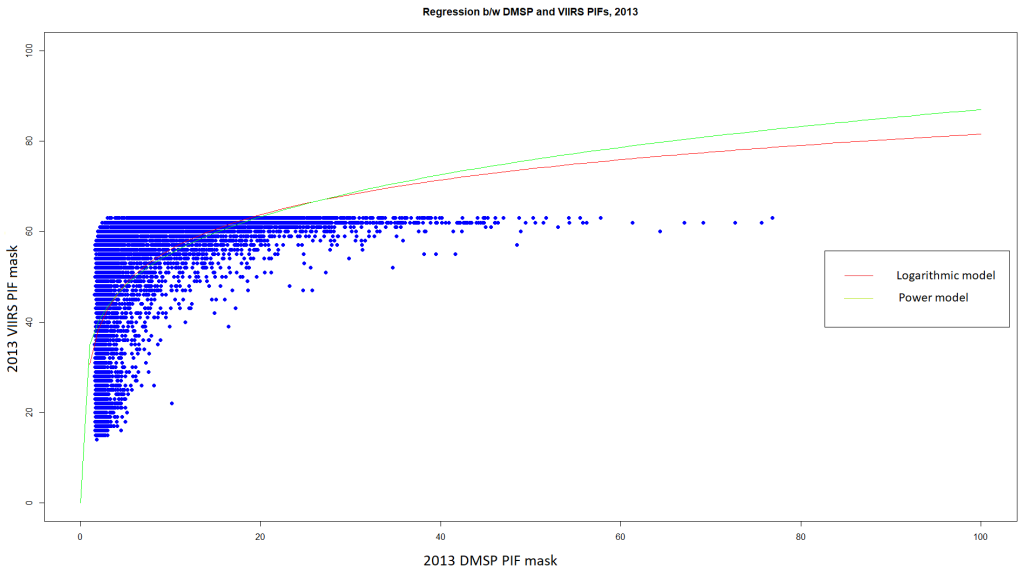


Figure 11: Plot of logarithmic regression

14.2 Using Power function

The power function is found to improve the comparability and to describe the non-linear relationship between the two datasets:

$$y = ax^b$$

Where, x denotes VIIRS value, y denotes the DMSP/OLS value, a and b are coefficients. Inter-calibration equations:

$$y(dmsp) = a * x(viirs)^b$$

Where y signifies the calculation of calibrated (simulated VIIRS) by altering the equation with coefficients.

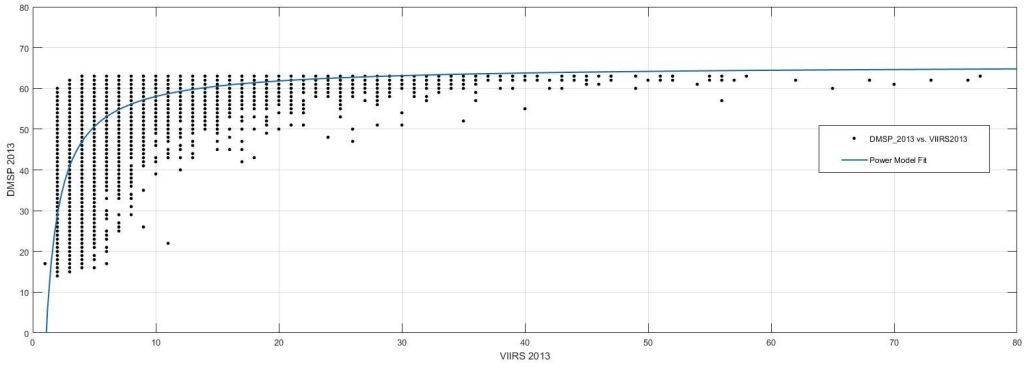


Figure 12: Plot of power regression

15 Optimization of generated simulated images

Due to lesser available data points in the lower region of DMSP data, the regression done was not reliable enough. Hence further optimization was done to remove the bias of different no. of points for different DMSP DN values.

15.1 Selection of Regions

Firstly 4 dark regions were selected each corresponding to either of East, West,North or South and these regions were clipped from India for DMSP-2013 and VIIRS-2013,2014,2015,2016,2017.

15.2 Calculating Mode

Mode of VIIRS-20XX values corresponding to each DMSP-2013 value for all 4 regions were calculated.

15.3 Calculating Mean of Mode

Mean of mode of VIIRS-20XX values corresponding to each DMSP DN value from the 4 regions were calculated.

15.4 Calculating residuals of Mean of Mode

Residuals of mean of mode from their actual DMSP DN value were calculated

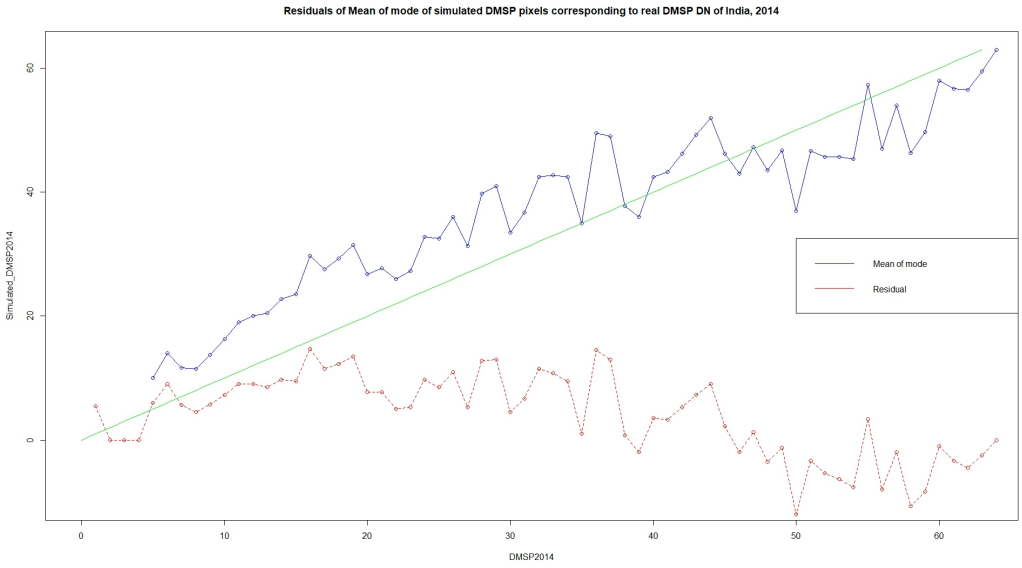


Figure 13: Residuals of mean of modes for year 2014

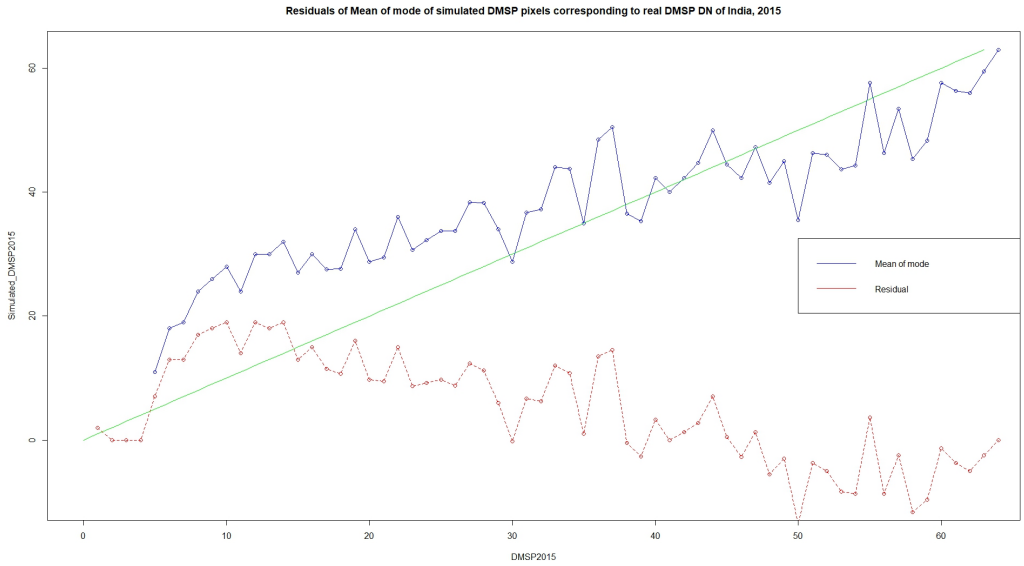


Figure 14: Residuals of mean of modes for year 2016

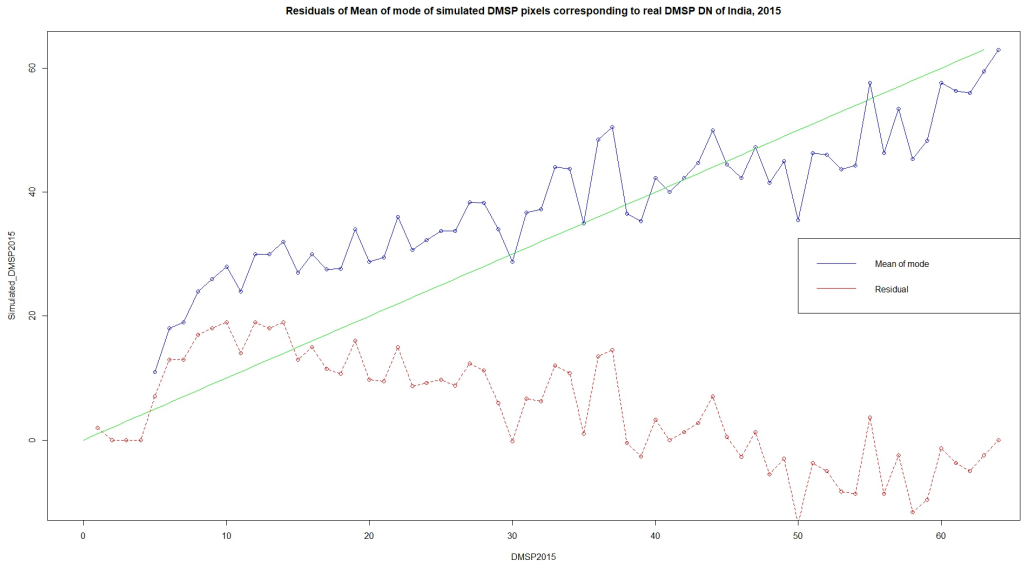


Figure 15: Residuals of mean of modes for year 2016

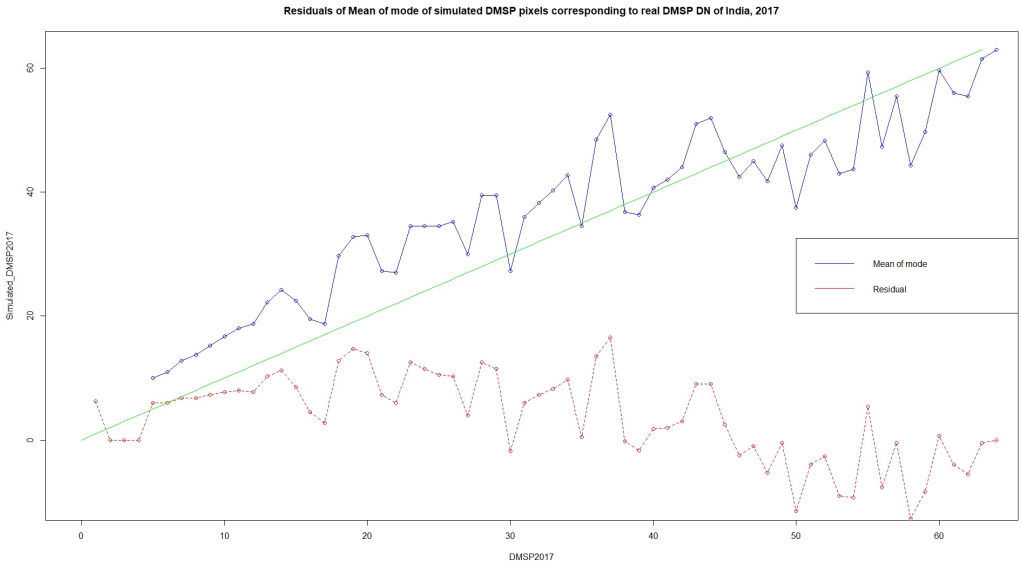


Figure 16: Residuals of mean of modes for year 2017

15.5 Generating optimized DMSP

Each pixel of Regression given simulated DMSP was subtracted from residual of mean of mode of the corresponding pixel's real DMSP-2013 DN value.

For 2013, optimized DMSP was generated with the **Zero Handling** . Pixels of regression give simulated DMSP were set to zero which correspond to DN zero in real DMSP-2013.

For other years,optimized DMSP was generated without the zero handling.

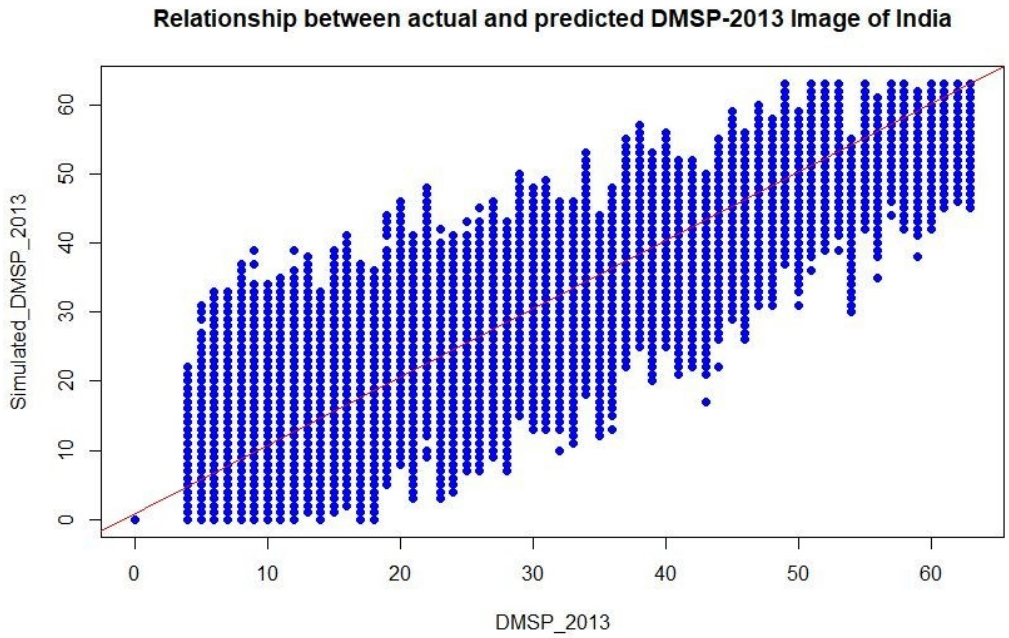


Figure 17: Plotting regression between real and predicted DMSP

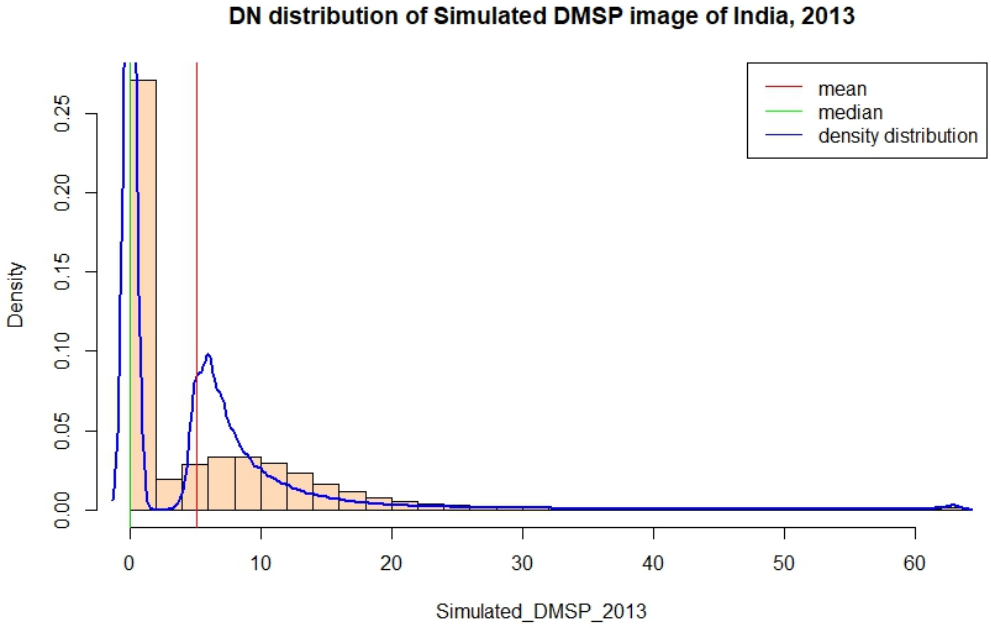


Figure 18: DN distribution in simulated DMSP 2013

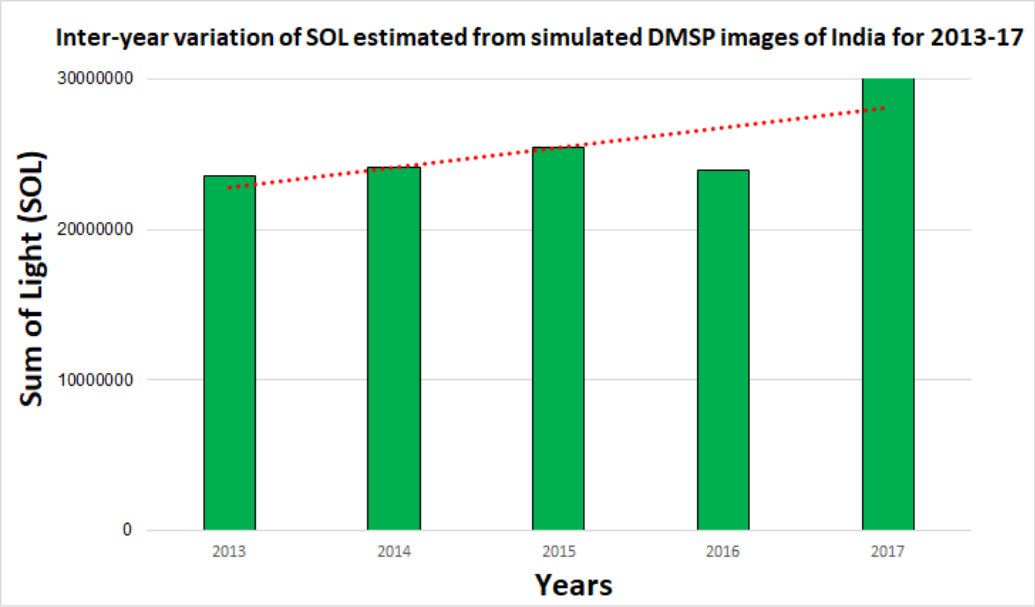


Figure 19: Estimated Sum Of Lights(SOL) for simulated DMSP 2013-17

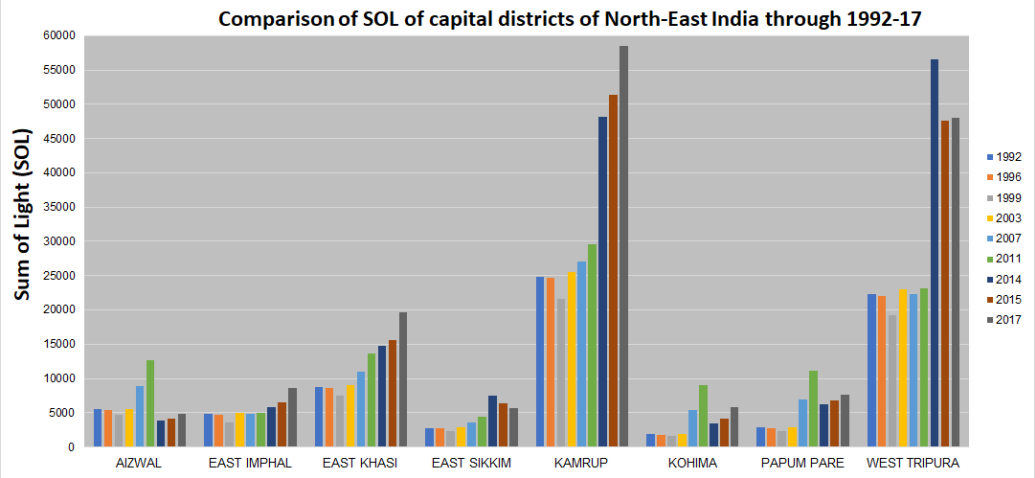


Figure 20: Plotting regression between real and predicted DMSP

References

[1] Michel Goossens, Frank Mittelbach, and Alexander Samarin. *The LaTeX Companion*. Addison-Wesley, Reading, Massachusetts, 1993.

[2] RESHMA JESWANI *Evaluation of the consistency of DMSP-OLS and SNPP-VIIRS Night-time Light Datasets*.

[3] Christopher N.H. Doll *CIESIN Thematic Guide to Night-time Light Remote Sensing and its Applications 2008*.

[4] SOHAM MUKHERJEE *Quality analysis of Inter-calibration of DMSP-OLS night-time images 2017*.

[5] Christopher D Elvidge, Kimberly Baugh, Mikhail Zhizhin, Feng Chi Hsu Tilottama Ghosh 2017 *VIIRS night-time lights*

[6] NAS Hamm, EJ Milton VO Odongo *Indicators of spatial autocorrelation for identification of calibration targets for remote sensing 2014*