Quality analysis of inter-calibration of DMSP-OLS night-time images

Soham Mukherjee

March, 2017

IIRS Supervisor

Mr. Prasun Kumar Gupta

Dr. S.K. Srivastav

ITC Supervisor

Dr. Valentyn Tolpekin

INDIAN INSTITUTE OF REMOTE SENSING
Indian Space Research Organisation
Department of Space, Government of India

UNIVERSITY OF TWENTE.
ITC FACULTY OF GEO-INFORMATION SCIENCE AND EARTH OBSERVATION

Quality analysis of Inter-calibration of DMSP-OLS night-time images

SOHAM MUKHERJEE
Enschede, The Netherlands, 2017

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

SUPERVISORS:

IIRS Supervisors	:	Mr Prasun Kumar Gupta
		Dr. S.K. Srivastav
ITC Supervisor	:	Dr. Valentyn Tolpekin

THESIS ASSESSMENT BOARD:

Chair	:	Prof. Dr. Ir. A. Stein
External	:	Dr. R. D. Garg (IIT ROORKEE)
OBSERVERS:		
ITC Observers	:	drs. J.P.G. (Wan) Bakx
IIRS Observers	:	Dr. Sameer Saran

INDIAN INSTITUTE OF REMOTE SENSING
Indian Space Research Organisation
Department of Space, Government of India

UNIVERSITY OF TWENTE.
ITC FACULTY OF GEO-INFORMATION SCIENCE AND EARTH OBSERVATION

Dedicated to my beloved parents and my teachers

ABSTRACT

Defense Meteorological Satellite Program Operational Linescan System (DMSP-OLS) night-time images are used to monitor human related temporal and spatial changes on the Earth surface like urbanization and socio-economic activities. But due to lack of on-board calibration these datasets cannot be directly used in the studies as the uncalibrated images consist of drastic variation in total brightness throughout years of acquisition. Hence a radiometric calibration (inter-calibration) is necessary to perform before using the images in a study. The inter-calibration approach requires regions with constant luminosity throughout the years. However, it requires predefined knowledge of the study area. This study formulates a method to automatically extract regions with constant luminosity throughout time without need of any prerequisite knowledge of study area. Such regions with stable luminosity are called Pseudo Invariant Feature (PIF). PIFs are considered as locally spatially homogeneous and of low local spatial variability in nature. Getis statistic (Gi*) is used to determine spatial homogeneity and Coefficient of variation (CV) to determine spatial variability. Gi* and CV are computed in 3×3 and 5×5 respectively and compared to select optimal window size. Gi* > 1.645 and CV < 10% are set as the criteria to determine areas that are locally spatially homogeneous and of low local spatial variability respectively. Integration of output of Gi* and CV that satisfy the thresholds, generate PIF. These PIFs are also checked for overall variability to check the consistency of total variability of all pixels in PIF through the years of study. Extracted PIFs show very consistent variability in the study time period. PIFs are used to model regression for radiometrically calibrating the images. The approach of regression model is to remove unstable pixels or outliers from affecting the fit of line in data. Thus, robust regression methods; Least Median of Squares (LMedS) and Least Trimmed Squares (LTS) are performed along with Ordinary Least Squares (OLS). These three estimators are compared on basis of coefficient of determination (R^2), Root mean square error (RMSE) and execution time. LTS is found to be of lower RMSE and execution time than OLS and LMedS and also has high R^2 which selects LTS as optimal parameter estimation technique for regression for inter-calibration. All the images are calibrated with the coefficients obtained from LTS for each year. Sum of light (SOL) index is generated to evaluate the inter-calibration quality. Yearly SOL values generated from uncalibrated images for 1992-2006 shows a R^2 of 0.40 whereas for calibrated images R^2 goes up to 0.89. Yearly SOL of both uncalibrated and calibrated images is also regressed with socio-economic parameters; gross domestic product (GDP) and urban population (UP) of India for 1992-2006 to find relationship between NTI and socio-economic parameters. A second order polynomial relation is seen between SOL and socio-economic parameters. R^2 of 0.31 and 0.38 is observed in relation of uncalibrated SOL with GDP and UP respectively whereas for calibrated images, the R^2 goes up to 0.93 and 0.89 for GDP and UP respectively. The study reflects the variation in relationship between DMSP-OLS NTI and socio-economic parameters with respect to inter-calibration.

Keywords: DMSP-OLS, inter-calibration, pixel saturation, PIF, robust regression, estimation of GDP, UP.

ACKNOWLEDGEMENTS

First and foremost, I thank GOD for giving me the determination and strength all throughout the last eighteen months as well as through my life. I thank my parents, family and my beloved brother *Hindol* for giving me the constant support, encouragement, motivation and most importantly for believing me till today.

I want to express my heartiest thankfulness to all my teachers from IIRS and ITC for helping me to learn the concepts of various subjects. Another special thanks for technically supporting me and teaching new and brilliant technical software and programming language throughout this course

Mr Prasun Kumar Gupta, my first supervisor from IIRS, who is extremely helpful as a guide and also as a teacher. He has provided great support and motivation towards completing this research work. I highly respect his thoughts about life and work, technical knowledge, support, unbiased criticism and most importantly providing guidance and time in difficult situations. Dear Sir, thank you so much for helping me throughout and it has been a privilege for me to work under your guidance. I am grateful to you.

Dr. S. K. Srivastav, my second supervisor from IIRS, is an eminent person with clarity in thoughts and ideas. His vast knowledge and huge experience were a boon for me. Even in his intense busy schedule, he has been a strong guiding force all throughout my research work. I feel blessed to have a supervisor like him. Thank you sir for providing new ideas and teaching me discipline in life.

I also like to thank Dr. Nicholas Hamm, my ex-supervisor from ITC, who has been a constant support in my research. Your innovative ideas and vast knowledge helped me to structure my work scientifically. Your clear identification of mistakes in the mails has guided me throughout to construct and refine my research. It has been a pleasure to work with you.

Dr. Valentyn Tolpekin, my supervisor from ITC has been extremely helpful and critical for the later parts of my research. His crucial comments over mail helped me a lot to rectify my mistakes in understandings and also in writing the thesis properly. Besides, I also want to express my heartiest thanks and regards for his polite and helpful nature, and also for providing necessary advice during our mid-term assessment and especially during the stay at Netherlands.

I also like to thank Dr. Sameer Saran, Head of Department, Geoinformatics Department, for giving me the most important chance to pursue the MSc degree and coming up to this far. Thank you sir for providing constant support during hard times and also for the necessary resources to finish the research.

At last but not the least, big thanks to all my friends and I take this opportunity to appreciate them individually, *Hati, Nandi, Jitu, Avishek, Prerna, Vaibhav ,Panda da , Surajit da, Daman, Sayan da and Atreya* for their constant encouragement and all sorts of help.

I would also like to thank especially my amigos of *V. United*, especially *Souvik* and *Pritthish* for helping me all throughout and providing enormous mental support and happiness throughout the research work phase.

I also express my deep gratitude towards all the staff members of the GID department and CMA, IIRS, especially *Mr. Dharmendar* and *Ms. Sangeeta* for providing all kinds of official support whenever required.

ABBREVIATION USED

Approx.: Approximation

CV: Coefficient of variation

DMSP: Defence Meteorological Satellite Program

DN: Digital Number

GDP: Gross Domestic Product

Gi*: Getis statistic

LMedS: Least median of squares

LTS: Least trimmed squares

NCO: Number of cloud free observation

NTI: Night-time images

NTL: Night-time light

OLS: Operational Linescan System

OLS: Ordinary least squares

PIF: Pseudo Invariant Feature

SOL: Sum of light

UP: Urban Population

VNIR: Visible and near infra-red

TABLE OF CONTENTS

LIST OF FIGURE

LIST OF TABLE

1. INTRODUCTION

1.1. Background

Defense Meteorological Satellite Program's Operational Linescan System (DMSP OLS) sensor generated imagery is one the most used nighttime imagery in different fields of research like estimating the change in economy, change in urbanization, change in population, night-time electricity consumption and more. There are total six satellites (F10, F12, F14, F15, F16 and F18) sent in the DMSP satellite program which give observation over a long time; from 1992 to 2013. DMSP satellite moves in a sun synchronous orbit with a mean altitude of 833 km and it makes night-time pass between 20:30 and 21:30 every night (Elvidge et al., 2001). DMSP satellite moves in orbit around the Earth 14 times per day, thus it provides a global coverage every 24 hours. OLS sensor acquires data in two different bands, near infrared (0.4-1.1 μm) and thermal infrared (10.5-12.6 μm). OLS sensor acquires data at a native spatial resolution of 2.7 km but the stable light categorized yearly composites have been resampled at a spatial resolution of 1 km (Doll, 2008). The 21 year long (1992-2013) coverage enables researchers to analyze changes in a time series of night time images. Though the DMSP satellite program was initially established to study the distribution of clouds and cloud top temperature at a global scale (Huang et al., 2014), later it was found by Croft (1973) that nighttime images acquired in Visible and Near-Infrared (VNIR) band from OLS sensor can also observe faint Very Near Infra-Red emission sources, such as city lights, auroras, gas flares, and fires. This study can also be justified by the ability of the sensor to detect radiation up to 4 times lower than standard detectable magnitude at night (Wu et al., 2013).

1.2. Motivation and Problem Statement

Though DMSP-OLS acquired images are used in context of socio-economic studies, it suffers from some technical disadvantages also. The major disadvantages of DMSP-OLS acquired images are:

- **Absence of inter-calibration**: DMSP-OLS satellite series does have multiple satellites acquiring data for the same year (e.g. Satellite F12 and F14 acquired data for the year 1997) in some years of coverage. It can be assumed that the total brightness measured by multiple DMSP-OLS satellites for the same year should be identical. But Doll (2008) found in his study an inconsistent nature in total sum of DN values of DMSP-OLS images of different years; the F14 1997-2002 data series was found to be consistently dimmer than its co-temporal satellite F12 and F15 (F14 series is co-temporal with F12 at its early range (1997-1999) and F15 with its late range (2001-2002)). It showed a difference in observed brightness in data of same year acquired by multiple satellites. A similar, though less difference between the F10 and F12 acquired total DN value for 1994 was seen. In the late years, a decrease in total DN was seen for the year 2003. Though the area residing high lit area were same or increased, the brightness detected by the F15 for 2003 was substantially lower than of 2001. The total brightness of F14 2001 was also compared with the total brightness for F15 2001 and it was found to be significantly less. Studies show these variations in DMSP-OLS data are caused due to acquisition of data by sensors of varying configuration to record the data over different year (Cauwels, Pestalozzi, and Sornette, 2014). The degradation in the quality of the sensor and different atmospheric conditions at acquisition time also affect the digital number (DN) values of the images (Wei et al., 2014). The variation in atmospheric and surface properties like atmospheric absorption and scattering, terrain illumination, cloud cover also cause variation in data of the surface obtained by satellite sensors. Thus it is seen that the DN values of acquired images of DMSP-OLS over time represent redundant data instead of data consisting true illumination of the surface. DMSP-OLS satellites acquire imagery at variable gain settings and also lack onboard calibration. Hence a relative radiometric normalization (named as inter-calibration in DMSP-OLS context) among the images is necessary to normalize the brightness of different yearly composites relatively to each other. This inter-calibration enables researchers to extract the characteristics and change on surface over time.

- **Light saturation**: DMSP-OLS sensor acquired image has six bit radiometric quantification levels, ranging pixel values from 0 to 63 (Doll, 2008). These images suffer from saturation problem, especially in urban centers where nighttime light intensity exceeds the maximum radiometric quantification level (63) which eventually gets truncated to 63 only as it is the maximum possible value (Elvidge et al., 2009). This pixel saturation problem distorts the ground information and limits the study of variation of luminosity in human settlements. Thus removal of such saturated pixel or correction of saturation is necessary in order to perform any research with DMSP-OLS images.

Several methods are already used for inter-calibration. These methods rely on selection of PIFs for performing inter-calibration between DMSP-OLS night-time images of different years. PIFs are area of stable light over time; the light emission of those areas is stable over a certain long period of time. However, selection of such PIFs is done manually and no traditional automated selection method is available. Manual selection of PIFs introduces variation in inter-calibration result. Selection of different areas may lead to varying results in regression diagnostics as the PIF DN values differ in different PIFs which can generate different trend in calibrated data (Li et al., 2016). Thus research is necessary to find an automated PIF (s) selection method followed by comparative research to determine the calibration model producing the most accurate results.

In this research, attempts are made to create an automated PIF selection method followed by comparison of the different models of inter-calibration of DMSP-OLS night-time images to find the most accurate one. Specific statistical criteria are set to determine the PIFs like, spatial homogeneity and local spatial stability of image pixels. Inter-calibration among images is constructed by modelling regression between

PIF of chosen reference image and the rest images. The inter-calibrated images are free from brightness fluctuations and can be used in socio-economic studies.

1.3. Research Identification

Through this research, attempts are made to rectify the brightness fluctuation observed in the DMSP-OLS night-time image (NTI) and to estimate the actual values of DNs in such images. Many works have showed that the inter-calibration of DMSP-OLS night-time images has been mainly performed with respect to PIFs. However, selection of PIFs requires much prior knowledge of the study area and manual selection cause variability in the inter-calibration results. Thus an automated method to select such PIF followed by inter-calibration is developed in this research to overcome the issues stated in previous line.

1.4. Research Objectives

The main objective of this research is to develop an optimal automatic PIF selection method through a study followed by inter-calibration through comparison of ordinary and robust linear regression models between the selected PIFs of reference DMSP-OLS NTI and uncalibrated DMSP-OLS NTI. The inter-calibrated images are checked for relationship with socio-economic parameters like Gross Domestic Product (GDP) and Urban Population (UP). The main objective can be reached by defining the following sub-objectives:

1. To develop an optimal automatic PIF selection method for inter-calibration of DMSP-OLS night-time images acquired in different years.

2. To perform inter-calibration with respect to selected PIF among DMSP-OLS night-time imagery acquired in different years through a comparative study of ordinary and robust regression model comprising different parameter estimations methods.

3. To study the improvement of quality of inter-calibrated images over uncalibrated images in socio-economic context (GDP, UP).

1.5. Research Questions

The following are the research questions identified from the research objectives:

1. What are the spatio-temporal assessment criteria for selecting PIFs in DMSP-OLS nighttime images?
2. What is the optimal parameter estimation method associated with linear regression for inter-calibration of PIFs?
3. What is the improvement in DMSP-OLS night-time images after inter-calibration?
4. What is the relationship between inter-calibrated images and socio-economic variables like GDP and UP?

1.6. Innovation Aimed At

The innovations intended in this study are:

1. Manual selection of PIF introduces variation in the calibrated result because selection of different potential PIF may produce different results in regression and the inter-calibration cannot be considered as optimal. Thus to solve this issue, an automated method is developed in the research for selecting PIFs with high spatial homogeneity and low spatial variability without any implementation of previous knowledge of the study area, which is the key to perform radiometric inter-calibration.
2. Selection of optimal parameter estimation method associated with linear regression in context to perform inter-calibration of DMSP-OLS nighttime images acquired by different OLS sensors over time.

1.7. Research Approach

To answer the research questions and research objectives of this work, DMSP-OLS version 4 stable light yearly composites of night-time images from 1992 to 2006 are downloaded and India is masked from these images as study area. To select the spatially homogeneous clusters in the images Getis statistic (Gi*) is implemented on the images. Coefficient of Variation (CV) is applied on the images to check the local spatial variance in the images. On generation of Gi* and CV images of all the years a criteria is established to select the regions that are spatially homogeneous and of low spatial variability. The pixels that satisfy both the criteria are selected as PIF. On extraction of PIF, regression is modelled among the yearly composites to inter-calibrate the yearly images. For regression, the image of 2001 (F152001) is used as the reference. Robust and ordinary parameter estimation methods associated with linear regression are compared in context of inter-calibration to select the optimal or best performing one. The used parameter estimation techniques are: Ordinary Least Squares (OLS), Least Median of Squares (LMedS) and Least Trimmed Squares (LTS). On selection of optimal parameter estimation method all the images are calibrated with respect to the reference image. Finally, the inter-calibrated images will be checked for relationship between the images and socio-economic parameters like GDP and UP.

1.8. Thesis structure

The whole thesis is organised into five chapters. ***Chapter one*** includes the background information of the research work along with the important facts of the topic, the motivation and problem statement, research questions and overview of the approach taken for the research. ***Chapter two*** describes the details of the related work that has been done in the past by various researchers. ***Chapter three*** includes the information of the study area chosen and the dataset used along with the details of the methodology adopted. ***Chapter four*** shows the results obtained along with the discussion of the results. Finally, the conclusion of the research work with recommendations leading to future research is mentioned in ***chapter-five.***

2. LITERATURE REVIEW

This chapter has different sections giving an introduction to the previous research works on radiometric calibration of satellite images (Section 2.1); approaches to inter-calibration of DMSP-OLS images (Section 2.2); various regression techniques applied in order to perform inter-calibration (Section 2.3), about solving the saturation problem in DMSP-OLS NTI (Section 2.4) and also the socio-economic studies with DMSP-OLS NTI (Section 2.5).

2.1. Radiometric Calibration of Satellite images

Calibration of satellite sensor acquired imagery is considered to be a fundamental step for validation of satellite data derived products. Moreover, image acquisition of Earth's environment needs radiometric calibration to obtain accurate geophysical and socio-economic information through time (Bannari et al., 1999). But for the inaccessibility of orbiting satellites, calibration methods based on test site with ground data are opted (Bannari et al., 2005). Such sites are usually selected on some specific criteria. The important ones among such criteria are radiometric spatial uniformity and temporal stability (Bannari et al., 2005). Derksen et al. (1998) showed that spatial uniformity information can be obtained from the study of spatial autocorrelation. They also stated in their work that in any remotely sensed image, positive spatial autocorrelation refers to cluster of similar DNs whereas a negative spatial autocorrelation can be found in clusters of dissimilar DNs. However, the study did not state the reliability of spatial autocorrelation if it is observed to vary significantly over the study area. Anselin (1995) developed Local indicators of Spatial Association (LISA) to find the spatial association of study area that are undetectable by using spatial autocorrelation. Getis and Ord (1992) and Bailey (1994) showed in their work that Getis statistic (Gi*) is such a LISA that gives measure of local clustering. It is considered as an indicator of spatial dependence and spatial homogeneity (similar spatial structure in DN values under a fixed area or window) when similar DNs are clustered together. Odongo et al. (2014) showed the use of Gi* to determine clusters of relatively bright pixels than the mean brightness observed in cloud-free LANDSAT TM images of Tuz Golu, Turkey for spatio-temporally assessing it as a potential radiometric calibration site. Spatial stability of a test site can be assessed by using coefficient of variation (CV) along with a measure of variability of reflectance within a defined area in image (de Vries et al., 2007). Rondeaux et al. (1998) showed that spatial stability of the calibration site can be determined by studying the extent to which CV varies for the target site over different years. Odongo et al. (2014) stated if CV remains inside threshold of 3%, the site is considered to be spatially stable.

2.2. Inter-calibration techniques of DMSP-OLS night-time images

In case of specific approaches to perform inter-calibration of DMSP-OLS nighttime images, several methods are already in use. The conventional technique used for inter-calibration was proposed by Elvidge et al. (2009). This method constitutes manual selection of PIF followed by quadratic regression method to estimate DN values of images of different years with respect to a reference year image. This reference image was selected on the basis of selecting the image that contained maximum magnitude of DN values. This method was adopted and updated by Liu et al. (2012) by applying second-order regression in their optimal threshold method. Wu et al. (2013) used five different regression models (exponential, linear, logarithmic, quadratic and power) and compared them to optimize the inter-calibration. A different method was proposed by Wei et al. (2014) where Pseudo-Invariant Features (PIF) are used for inter-calibration. PIF are subsets of the DMSP-OLS nighttime image where luminosity is observed to not change drastically over time. It was also shown in the study that PIFs can also be selected with exclusion of the highest valued DN to avoid the light saturation problem. Wei et al. (2014) created an urban detection threshold to select urban areas where luminosity is observed to be maximum. From this threshold Wei et al. (2014) extracted urban area in the image selected as reference (F142000). Wei et al. (2014) manually selected PIFs in the extracted urban area and performed regression between PIF of the selected reference year image and with same PIFs in other year images to relate light intensity values of DE saturated PIFs between reference year image and subsequent calibration awaited image.

2.3. Regression techniques used in inter-calibration

The selection of PIFs is seen to be performed manually in most of the works. But it introduces variation in the calibrated result because selection of different potential PIF may produce different results and the inter-calibration cannot be considered as optimal (Li et al., 2016). Li et al.(2013) constructed an automatic PIF selection algorithm by discarding changing pixel values by considering them as outliers and also found strong linear relation between socio-economic variables and his inter-calibrated product pixel values. Though Li et al.(2013) only used the 2.5σ empirical rule for detecting the outliers (DN values representing similar location on the surface but changes drastically in magnitude in different yearly image composites). Thus it can be considered as a limitation in his work. Recently Li et al. (2016) included different weighted parameter estimation methods, i.e. Classic M-Estimators, Random Sample Consensus (RANSAC) and Least Median of Square (LMedS) for linear regression model and made a comparative study to select the best performing estimation method on the basis of robustness (better fit of model in data by detecting and reducing effect of outliers in fit of model in data). But the study done by Li et al. (2016) does neither cover the use of the obtained optimal estimation scheme in the automatic inter-calibration nor it finds the degree of relationship of the inter-calibrated image pixels with socio-economic indices. The study did not also resolve the saturation problem in DMSP-OLS images. Besides, Least Trimmed Square (LTS) was also not included as an estimator along with previously mentioned estimation techniques whereas LTS is found to be asymptotically efficient (better efficiency in limit when the sample size tends to infinite) than LMedS (Rousseeuw, 1984). LTS is also efficient in handling Gaussian noise in the data (Massart et al., 1986).

2.4. Rectification of saturation problem in DMSP-OLS night-time images

To rectify the light saturation problem in DMSP-OLS NTI, Hara et al. (2004) proposed a linear correction method for light saturation problem on a regional scale. A method was proposed by Letu et al. (2012) where night-time light (NTL) intensity was assumed to be constant in saturated areas during 1996-1999, a linear regression model was developed between non-saturated part of stable light image of 1999 and the radiance calibrated image in 1996-1997 and this model was used to correct the saturation problem of 1999 image. But Zhang, Schaaf and Seto (2013) showed that this method is not compatible for developing countries like China, because NTL of these countries may change rapidly because of their fast urban growth. Besides, this method also requires a radiance calibrated image as reference. Ziskin et al.(2010) generated a no saturation NTL image of 2006 by combination of NTL data obtained in low gain with operational data captured in high gain. This mixed gain mode acquired images do not suffer from light saturation but number of such images are very less and generation of such radiance calibrated images is also quite labour-intensive. Wei et al. (2014) showed a comparatively easier way to handle saturation by removing the saturated DNs (DN value=63) from the PIFs for studying urban growth over time. Ma et al. (2014) made a comparative study of four different correction methods of light saturation, i.e., linear and cubic regression on the regional scale and two different indexes obtained from combining DMSP-OLS NTI and Normalized Difference Vegetation Index (NDVI) on the pixel scale. They came with the conclusion that there is no such optimal method of saturation correction rather the selection of saturation correction method is needed to be selected as per study area and purpose of research.

2.5. Socio-economic studies with DMSP-OLS night-time images

Several studies have been carried out to estimate the change in socio-economic parameters like, GDP, Total income, poverty, urban population and more in national and regional levels. Doll, Muller, & Morley (2006) performed a study with radiance calibrated DMSP-OLS product of 1996 and regional GDP data at different regional scales for 11 European countries; the study showed a high positive correlation between the DMSP-OLS data and GDP. Sutton, Elvidge, & Ghosh (2007) made a study with the "stable light" product for US, India, China and Turkey along with population density to find relation with GDP. A log linear relationship was found between NTI and GDP. However, the study consisted of only a single year composite instead of a time series. Later studies considered a time series of several year composites and searched for the relationship for developed countries like US as well as also for developing countries like India and concluded that a positive increasing trend of economic growth can be estimated from NTI (Chen & Nordhaus, 2011) (Kulkarni et al., 2011). Besides estimating the GDP, NTI has also been used to estimate population also on both global and regional scale. Sutton (1997) used data of US census of 1990 with NTI to estimate the change in population of urban areas in USA. Sutton et al., (2001) also tried to estimate the global population of 1997 at country specific scale which produced good estimation of the world population of 1997 around 6.3 billion where the actual population was noted to be of 5.9 billion. Some researches were concentrated on developing nations also. Amaral et al. (2006) and Lo (2001) made studies to estimate the change in urban population with NTI for Brazil and China respectively; both the studies showed positive correlation between NTI and urban population.

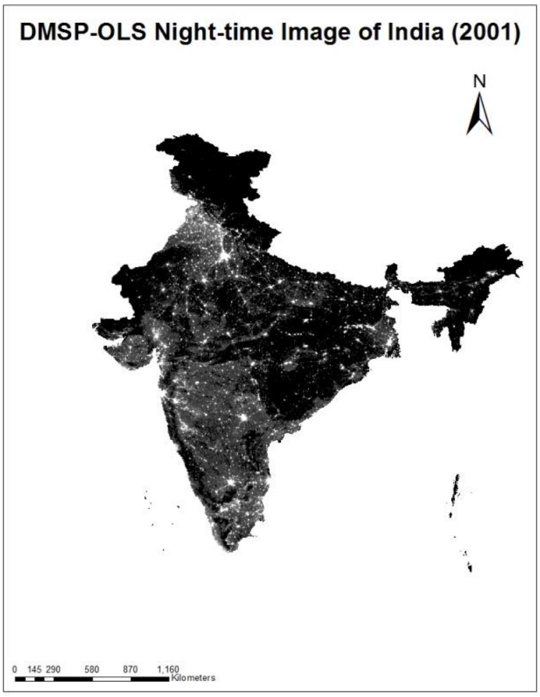
3. STUDY AREA, DATA USED AND METHODOLOGY

This chapter explains the study area with the reasons for choosing the study area and the data used for completing the work along with the methodology.

3.1. Study Area

The study area selected for the work is India. The country shares its political borders with several other countries. India shares boundary with Pakistan and Afghanistan on the west and Bangladesh and Myanmar on the east. The northern boundary is shared with China, Tibet, Nepal and Bhutan. India has a central latitude and longitude of 20.5937° N, 78.9629° E respectively. The reasons for selecting India as the study area are:

1. India is an emerging country and is reported to have a fast urbanization rate over the time. DMSP-OLS nighttime data for the period of 1992-2006 is taken as input dataset in the research and it can be interpreted that there are much change to detect in the urban scape as well as in GDP and UP of India in this period. Thus selection of such study area gives a measure of robustness of the inter-calibration.
2. India does consist of many big cities like, Delhi, Kolkata, Chennai, Mumbai, Bangalore etc. These cities are developed and it can be assumed that these cities do contain regions with high illumination throughout the time period of the study. Such regions can be considered as PIF. Hence, these regions can be used to compare with the automatically selected PIFs as ground validation process.



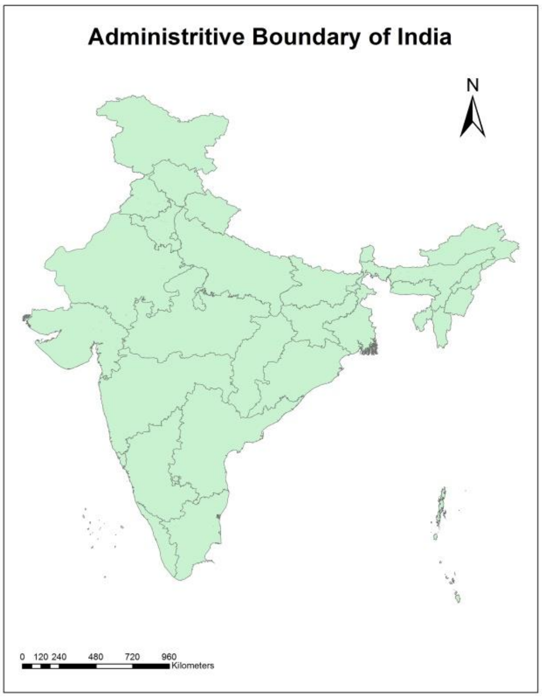


Figure 3.1: a) DMSP-OLS Image of India (F152001) b) Map of India

3.2. Data used

In this research work, DMSP-OLS acquired version 4 stable light products for the time period of 1992 – 2006 of India is used as the motif of the research has been to study inter-calibration methods dedicated to DMSP-OLS acquired images specifically.

Specification	DMSP-OLS
Spatial Resolution (m)	1000 m
Band	VNIR
Wavelength	0.4 – 1.1 micro meters
Radiometric quantization	6 bit
Field of view	3000 km
Return interval	Daily
Orbit	Polar

Table 3.1: DMSP-OLS satellite specifications

The Defense Meteorological Satellite Program, (DMSP) is the meteorological program of the US Department of Defense, which originated in the mid-1960s. It was launched with primary objective of collecting worldwide cloud cover on a daily basis (Kramer, 1994). It was declassified in 1972 and made available to open use of scientific community as archives from 1992 (NASA, 1997). After declassification, DMSP satellites did undergo many changes. The latest series (Block-5D) is incorporated with the OLS sensor. The photo multiplier tube (PMT) in the OLS sensor facilitates low level light amplification in visible and VNIR range. The gain applied to signal varies every 0.4 milliseconds. The variation of gain depends on the predicted illumination of the scene from the payload elevation and lunar phase (Doll, 2008). This high gain mode of acquisition permits detection of visible and VNIR band light sources up to10^{-9} $Watts \bullet cm^{-2} \bullet sr^{-1}$. So, besides acquiring the cloud top cover information DMSP-OLS acquired imagery also captures light emitted from Earth surface at night, i.e., city lights, gas flaring, shipping fleets and forest fires (Croft, 1973). Since digital archiving of DMSP-OLS data has started numerous products have been released. In the current research, version 4 stable lights are used as the dataset from 1992 to 2006. Stable light data is the cleaned up average of visible lights on cloud-free acquisition in a year that contains the lights from cities, towns, and other sites with persistent lighting, including gas flares. Ephemeral events, such as forest fires and oil station fires have been discarded. Here the background noise is also identified and replaced with values of zero. DN values in stable lights range from 1-63 expressing a 6-bit radiometric quantization. Stable light product is resampled at a grid of 1 km and consists of WGS84 co-ordinate system.

There are total six satellites sent to space in the DMSP satellite program which gives observation over a long time; from 1992 to 2013. For the current research timespan (1992-2006), data from the initial four sensors are used. The following table 3.2 shows the different sensors over time and the selected sensors for the corresponding years.

Year\Sat.	F10	F12	F14	F15	F16	F18
1992	F101992					
1993	F101993					
1994	F101994	F121994				
1995		F121995				
1996		F121996				
1997		F121997	F141997			
1998		F121998	F141998			
1999		F121999	F141999			
2000			F142000	F152000		
2001			F142001	F152001		
2002			F142002	F152002		
2003			F142003	F152003		
2004				F152004	F162004	
2005				F152005	F162005	
2006				F152006	F162006	
2007				F152007	F162007	
2008					F162008	
2009					F162009	
2010						F182010
2011						F182011
2012						F182012
2013						F182013

Table 3.2: DMSP-OLS data acquired by different satellites over time; cells in green colour represent the data taken; blue coloured cell represents reference year in inter-calibration

The cells highlighted in green colour in table 3.2 are the products that have been used in the study. The stable light products for the mentioned years are used as stable light products do consist of the stable surface lights only. This product does not consist the illumination from forest fires or oil shores which helps in studying the urbanization pattern and growth over time (Doll, 2008). The cell highlighted in blue (F152001) is used as the reference image in the process of inter-calibration.

To find the relationship with socio-economic parameters GDP and UP are taken as the two different indices. Yearly GDP data of India in US $ is collected for the time of 1992 to 2006. GDP data is observed in 10^{11} order; in research, the data is converted to 10^4 order for better visualization of results in graphs. Urban Population (UP) data of India is also collected as yearly total urban population format. Population is observed to be of 10^3 order in magnitude. This UP data is available from 1996 onwards. But the study time period is from 1992 onwards, hence the data for 1992-1995 is generated from the available data using an exponential equation (Equation 6, described in section 4.3.2). Both GDP and UP data are obtained from **www.indiastat.com**.

3.3. Methodology

The main objective of this work was to develop an automatic inter-calibration method for DMSP-OLS night-time images of different years. This Section of the chapter describes the steps taken to accomplish the objectives of section 1.3.

The flow chart of the methodology adopted and developed has been presented in Figure 3.2 below.

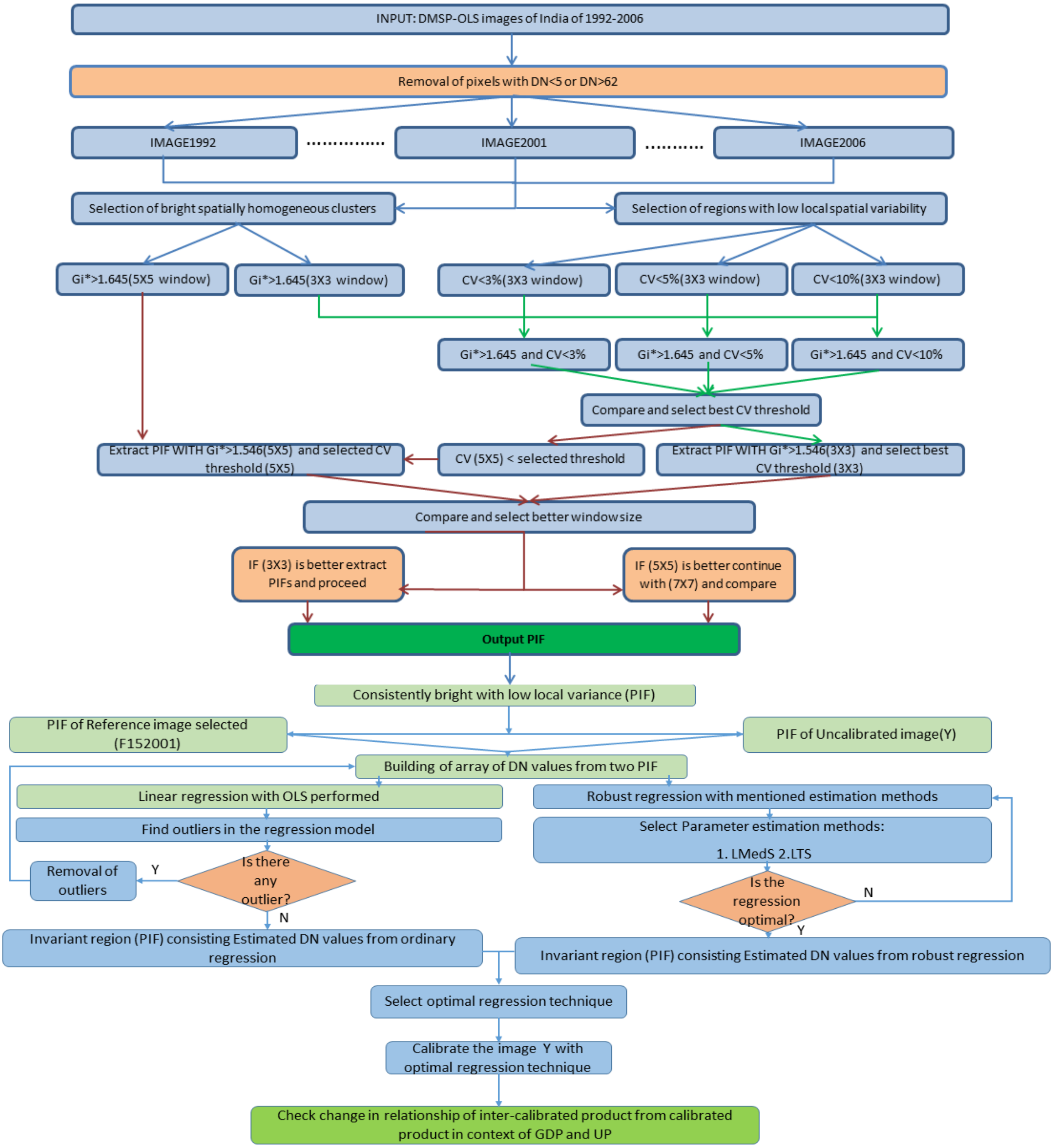


Figure 3.2: Methodology flow diagram of research; Gi*: Getis statistic; CV: coefficient of variation

3.3.1. Data pre-processing

The annual composites of DMSP-OLS version 4 “stable light” for the period of 1992-2006 are obtained from the website of NOAA National Geophysical Data Center (NGDC) website (**http://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html**). These images are of global coverage but the initial goal is to identify PIFs inside India. So the rest surrounding area is masked to make it sure that those areas will not affect the calculation of mean ($\bar{x}$) , standard deviation (s). In most of the years there were two satellites collecting the night light images. Only one image of one satellite with better image quality is selected for each year. The number of cloud-free observations (NCO) is used for the selection metric of images. Satellite images with higher NCO are selected for the research. NCO can be obtained from the archive technical specification information. The inter-calibration process relies on a reference image with which rest of the images will be calibrated. The selection of reference image is done on basis of enough availability of stable light pixels in the image. The span of the data for the work is 15 years and it can be assumed that urbanization will definitely promote the nightlight growth. So image of the middle year is selected as the reference image because it does have enough number of stable light pixels. 2001 image acquired by F15 satellite is used as the reference image.

DMSP-OLS night-time image data consists of many non-lit pixels with a value zero and very low values (DN <5) and also many pixels with saturated values (DN=63). The saturated values do not represent true ground luminosity due to the truncation of DN value to 63 for the limited 6-bit radiometric quantization. In case of the non-lit and faintly lit lights, the non-lit pixels (DN=0) do not represent any vital information about the surface and corresponding low DN values are majorly seen as a result of blooming effect of nearby lit pixels (Doll, 2008). Thus, the zero valued pixels along with pixels with a comparatively lower value (DN < 5) are discarded from the images. Removing the pixels with DN value less than 5 removes the blooming effect from the images. Blooming is overestimation of brightness of pixels due to combined effect to error in geo-location, larger overlap in the pixels (60% overlap), coarser spatial resolution and atmospheric vapour content (Doll, 2008). Due to blooming the subsequent parts of the bright areas also seem to have a low luminosity even though the surface represent non-lit pixel. Besides, blooming is also seen in big highway traffics between large urban areas. Thus to remove the overestimation of brightness pixels with DN <5 are discarded. The saturated pixels (DN=63) are also discarded from the images which handles the saturation problem. It keeps only the comparatively more relevant information of luminosity of the artificial lighting at night.

3.3.2. Selection of PIFs in study area

The uncalibrated images(Y) and reference image(X) (F152001) are kept as two different sets. One set only contains the image X and the other set contains Y images (DMSP-OLS NTI from 1992 to 2006 excluding reference year). For selection of PIFs some criteria are developed on satisfying which, a region can be called as a PIF. The criteria are set up as following:

1. Luminosity of the pixels inside the PIF will not change drastically throughout the years.
2. Pixels inside the PIFs shall not contain blooming and shall be bright (DN > 5).

3. PIFs should not contain saturated pixels (DN = 63).
4. Pixels inside the PIF should be locally spatially homogeneous in nature (Gi* >1.645).
5. Pixels should have low local spatial variability (CV shall be low).
6. The overall spatial variability of the PIFs of different years should not vary drastically with time.

3.3.2.1. Getis Statistic (Gi*)

Gi* provides 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 co-ordinates. Each point in the regions is given a 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 1.

$$\boldsymbol{G_i^*(d) = \frac{\sum w_{ij}(d)x_j - W_i^*\bar{x}}{s[W_i^*(n - W_i^*)/(n-1)]^{1/2}}} \qquad (1)$$

Where,

$$\boldsymbol{W_i^* = \sum_j w_{ij}(d)} \qquad (2)$$

$$\boldsymbol{\bar{x} = \frac{\sum_j x_j}{n}} \qquad (3)$$

n is the total number of sub regions, $\bar{x}$ 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 x is the DN value. d is the distance obtained from the fixed window size in which Gi* is implemented (d=1; 3×3 window, d=2; 5×5 window).

The Getis statistic gives a measure of clustering and it is used to identify clusters of comparatively higher values (Gi*>0; hotspot) and cluster 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 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% proability of the formed cluster to have high spatial autocorrelation.

The Gi* are calculated for each image at a variable window size of 3×3 and 5×5 initially. The selection of 3×3 window can be found in the work of Kneubhuler et al. (2006), de Vries et al. (2007) and Odongo et al. (2014) where they have got satisfactory results for selection of spatial homogeneous vicarious calibration site. However the selection of the optimal window size is performed with evaluation of linear OLS results(decribed below).

3.3.2.2. Coefficient of variation (CV):

PIFs do consist of both spatial homogenity and low spatial variability. Thus, on extraction of relatively bright clusters with variable window size, regions with low spatial variability are also extracted. Coefficient of variability(CV) is used in the research 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.

The CV gives a standardized measure of variability. It is defined as following:

$$\boldsymbol{CV = \frac{S}{\bar{x}}} \qquad (4)$$

Where,

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

CV is calculated for the images from 1992 to 2006 on variable window size of 3×3, 5×5 and 9×9. 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%. CV value of 3%, 5% and 10% are used as the thresholds and combined with result of Gi*> 1.645 both in same window sizes for all the respective window sizes mentioned. This combination of Gi* and CV output produces the regions that are both spatially homogeneous and of low local spatial variability (PIF) at different window sizes. 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]™ 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. R^2 and 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 high number of observations in the OLS is selected as optimal CV threshold for generating PIF. To select the optimal window size PIFs are generated with pixels Gi* >1.645 and CV < "optimal threshold" under an increased window size of 5×5. These PIFs for the year 1992, 2001 and 2006 are also plotted in [R-3.3.2]™ simillar to the approach of selecting CV threshold above. The window size between 3×3 and 5×5 that generates higher number of observations along with a big range of DN values and high R^2 is selected as the optimal window size. If 5×5 results as better size, the same process is repeated with a window size of 7×7 and compared with the results of 5×5 else the research follows window size of 3×3.

Initial preprocessing (Removal of nonlit and saturated DN) is performed in ESRI ArcGIS™-10.2. The RTW tool compaitable in IDLE/ENVI™-5.0 suite provided by Willson (2011) is used to compute Gi*and CV is implemented with a self written non-linear kernel in Python™ with variable window size in the images of study area. "GDAL™" and "NumPy™" packages in Python™ are used to develop the kernel.

3.3.3. Implementation of regression

Regression model is used to relate light intensity values within the extracted spatio-temporal stable area between the previously selected reference year (t_r) image and uncalibrated images of different years (t_i) and to estimate the DN values of the target year t_i with respect to reference year. Linear regression with Ordinary least square (OLS) estimation and robust regression with two different estimation methods are applied and compared as per goodness of linear fit of model in data. In the case of OLS, the method does search and does discard outliers (changing pixel values from the reference) that cannot be fitted by the regression model. Outliers in OLS estimation is selected by checking the interval of standardized residuals (δ^*). Residual (δ) between the predicted DN value and observed DN value is observed and δ^* is computed from the following equation:

$$\boldsymbol{\delta^*} = \frac{\boldsymbol{\delta} - \overline{\boldsymbol{\delta}}}{\boldsymbol{\sigma(\delta)}} \qquad \textbf{(5)}$$

Where, $\overline{\delta}$ is the average of δ and $\sigma(\delta)$ is the standard deviation of all the residuals δ generated for each estimation for the dependent variable (DN of calibration awaited PIF). If the δ^* of sample point falls outside the interval (-2, 2), the sample point is eliminated as outlier at 95% confidence level. Thus only the pixels that have not changed drastically over the years are only left. Such pixels do form the PIFs because these pixels don't change very much in value over time. But as OLS is very sensitive to presence of outliers in data, inter-calibration is also implemented with two different linear parameter estimation techniques of robust regression; Least Trimmed Squares (LTS) and Least Median of Squares (LMedS) (for details of these methods see Appendix-A). These mentioned estimators above are selected to study which one works better with robust regression in context of reducing the effect of outliers in the fitting of trend line in data. Besides, selection of the specified parameter estimators are also in line with study of Li et al. (2016) with addition of LTS method. LTS is added in the research as it executes faster and reduced the effect of outlier in fit of linear trend line to a comprehensive extent (Mount et al., 2014). Besides, all these mentioned estimators have a variable breakdown point, which adds good generalization of methods in the research. These parameter estimation techniques of robust regression model along with ordinary regression (OLS) will be compared with each other to select the most accurate one. Accurate one will be selected on basis of coefficient of determination (R^2), Root mean square error (RMSE) and execution time (for details of R^2 and RMSE, see Appendix-A). R^2 gives the measure of how close the fitted regression line is close to data; precision, RMSE gives the measure of the difference between magnitude between predicted values and observed values; error. The optimal technique will be used to calibrate the DMSP-OLS night-time images with respect to the reference year (F152001).The final calibrated output image pixels will be used to generate Sum of light (SOL) index. SOL index is the sum of all pixels in DMSP-OLD NTI of specific year.

SOL is the index to check the efficiency and consistency in inter-calibration. SOL of a DMSP-OLS NTI of a specific year for a region is obtained from the sum of all pixel values for that certain region. SOL is calculated for both uncalibrated and calibrated images of same year to evaluate the improvement. However perfect increasing or decreasing consistency cannot be obtained in SOL for different years. In calculation of SOL of calibrated images it is important to radiometrically normalize the newly obtained high and low DN values. After calibration some DN values may exceed 63; the maximum value of the sensor, or may go below 0; the minimum value of the sensor. SOL of such calibrated images present redundant total brightness. Thus, any DN value greater than 63 is truncated to 63 only and any value less than 0 is also approximated to 0 and then only SOL is calculated. The flow diagram to generate SOL is shown below in Figure 3.3:

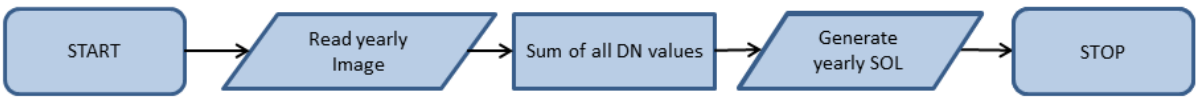


Figure 3.3: Flowchart of generating SOL index

SOL index of images of different years will be plotted to check the degree of convergence. Better the convergence is, better the quality of inter-calibration can be assumed. Besides, trend line is also fitted in plot of SOL of both uncalibrated and calibrated images for the time period of 1992 to 2006. R^2 is used as the metric to determine the better fit of trend line in the SOL data.

The work of implementing regression is done in [R-3.3.2]™ and SOL is generated by a self-written program in Python™-2.7.12. "GDAL™" and "NumPy™" packages in Python™ are used to develop the program to generate SOL.

3.3.4. Quality analysis in Socio-economic context

DMSP-OLS observes artificial manmade light on the Earth surface. Increase in manmade light indicates the growth in urban population also. With increase in urban population, the change in economic growth of a country is also inevitable. Thus, manmade lighting can be a good indicator to observe the change of these socio-economic variables. The proposed research is concentrated on GDP and UP of the study area and attempts are made to find the relationship between these variables and SOL index. The relationship is checked for both calibrated and uncalibrated images and the change in relationship with respect to calibration is studied. In case of change is observed in relationship between socio-economic variables (GDP, UP) the study does quantify the change besides the study on how the inter-calibrated images differ in quality for socio-economic studies. The degree of relationship is measured by coefficient of determination (R^2). The study does search for linear and polynomial second order relationship between inter-calibrated DMSP-OLS NTI pixels and mentioned socio-economic variables and the results are compared to select the optimal relationship. The relationship between SOL and GDP/UP are measured using Microsoft Excel (2010) and [R-3.3.2]™.

4. RESULTS AND DISCUSSION

This chapter describes the outputs achieved by using different methods to extract Pseudo Invariant Features (PIF) (mentioned in Section 3.3) followed by implementation and comparison of ordinary and robust linear regression in order to find optimal regression technique for performing inter-calibration of the DMSP-OLS images of India. This chapter also describes the study to find relationship between the DMSP-OLS images and Socio-economic parameters. GDP and UP of India is used as the socio-economic parameters. The following Section 4.1 presents the results of implementation of Gi* and CV on DMSP-OLS version 4 "stable light" yearly composite for time period of 1992 to 2006 to select PIFs of each year. The next Section of 4.2 shows the results of implementing and comparing normal linear regression (OLS) and robust linear regression (LMedS and LTS) for the extracted PIFs from DMSP-OLS NTI of India from 1992 to 2006 keeping PIF of 2001 image as reference or independent variable in implementing regression to inter-calibrate DMSP-OLS images. Section 4.3 presents the results on relationship of DMSP-OLS images and socio-economic parameters like GDP and UP with respect to calibration. Relationship between GDP UP and DMSP-OLS NTI is checked for both uncalibrated images and calibrated images and the effect of inter-calibration in change in relationship with socio-economic parameters is also discussed.

4.1. Selection of PIF in images

This Section is sub-divided into multiple chapters as per the workflow. Section 4.1.1 states the data analysis and pre-processing of the data. Implementation of Gi* to select spatially homogeneous clusters and corresponding outputs are stated and described in section 4.1.2, Implementation of CV in order to select regions with low spatial variability and corresponding results are described in section 4.1.3 and section 4.1.4 consists of integration of results of section 4.1.2 and 4.1.3 to extract the Pseudo Invariant Features (PIF) in study area with respect to luminosity. DMSP-OLS NTI of India for the time period of 1992-2006 is used in the study.

4.1.1. Analysis of data

As mentioned earlier, DMSP-OLS night time composite for the time period of 1992 to 2006 are used in research. "Stable light" product of DMSP-OLS for all the years is used. These composites do provide global coverage, thus the study area is needed to be extracted from all the yearly composites. An iterative model in ArcGIS is used to clip the area of India from all the yearly composites. Figure 4.1a, 4.1b and 4.1c show the DMSP-OLS NTI of India for 1992, 2001 and 2006.

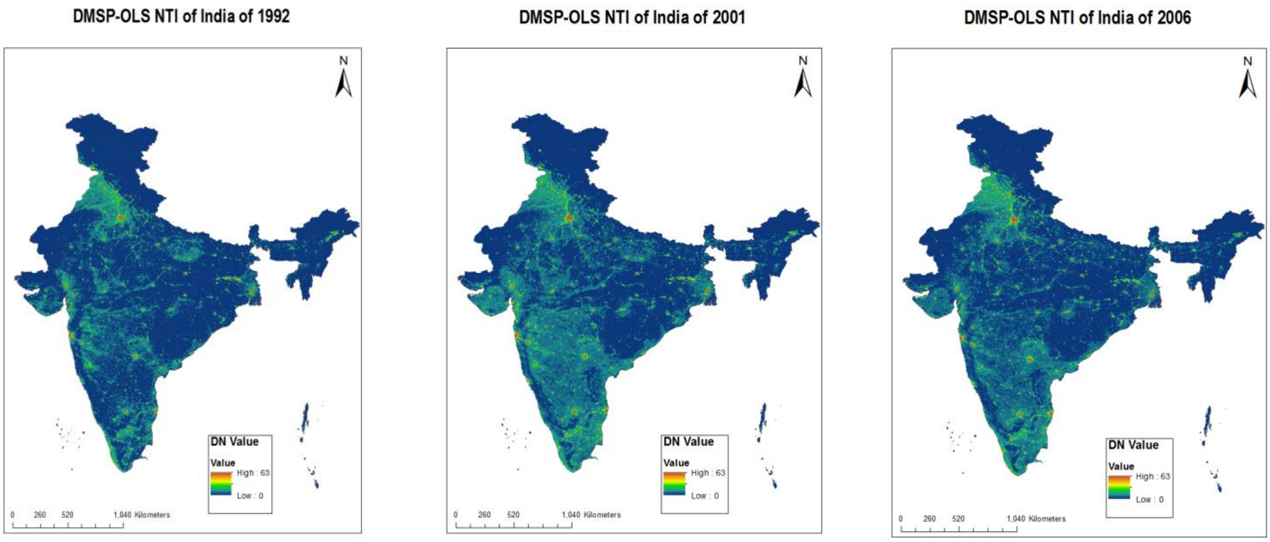


Figure 4.1: DMSP-OLS NTI of India: a) For year 1992 b) For year 2001 c) For year 2006; Blue colour denotes low pixel values and red colour means high pixel values

Figure 4.1a, 4.1b and 4.1c shows the variation of illumination over different regions in India over time. The minimum value is observed to be of 0 and maximum value is 63. Blue colour in the map represents pixels with low values and red colour correspondingly signifies pixels with high values. In all the images, high valued pixels are observed above centre which can be identified as Delhi, India's capital. In the east, a red cluster can be observed around Bay of Bengal. That can also be identified as Kolkata, one of the major cities in India. It can be inferred from such observation of bright points that bright clusters do reside in the big cities as these cities have dense urbanization which boosts the night-time light illumination of these areas. Saturated pixels can also be observed inside the cities which refer to areas with very high luminosity. These saturated regions are needed to be discarded in further study as the saturated pixels do not represent the real luminosity of the surface. A large number of dark regions are also in the study area. These non-lit areas along with low DN values (DN < 5) are also discarded in further study. The iterative model used to clip the study area from global dataset is updated with a mask that will discard the pixels that have DN < 5 or DN > 62. The overview of the model is shown below in Figure 4.2.

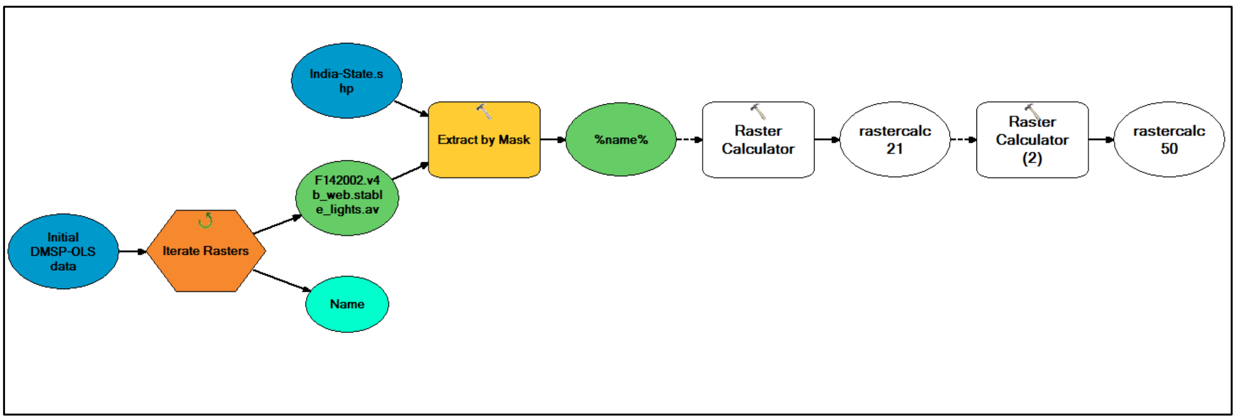


Figure 4.2: Snapshot of the model used in ArcGIS

The model iterates over all the global images and extracts the region of study area followed by extraction of pixels with DN >5 and DN<63. Figure 4.3a, 4.3b and 4.3c show the DMSP-OLS NTI of India with DN>5 and DN<63 of 1992, 2001 and 2006 respectively.

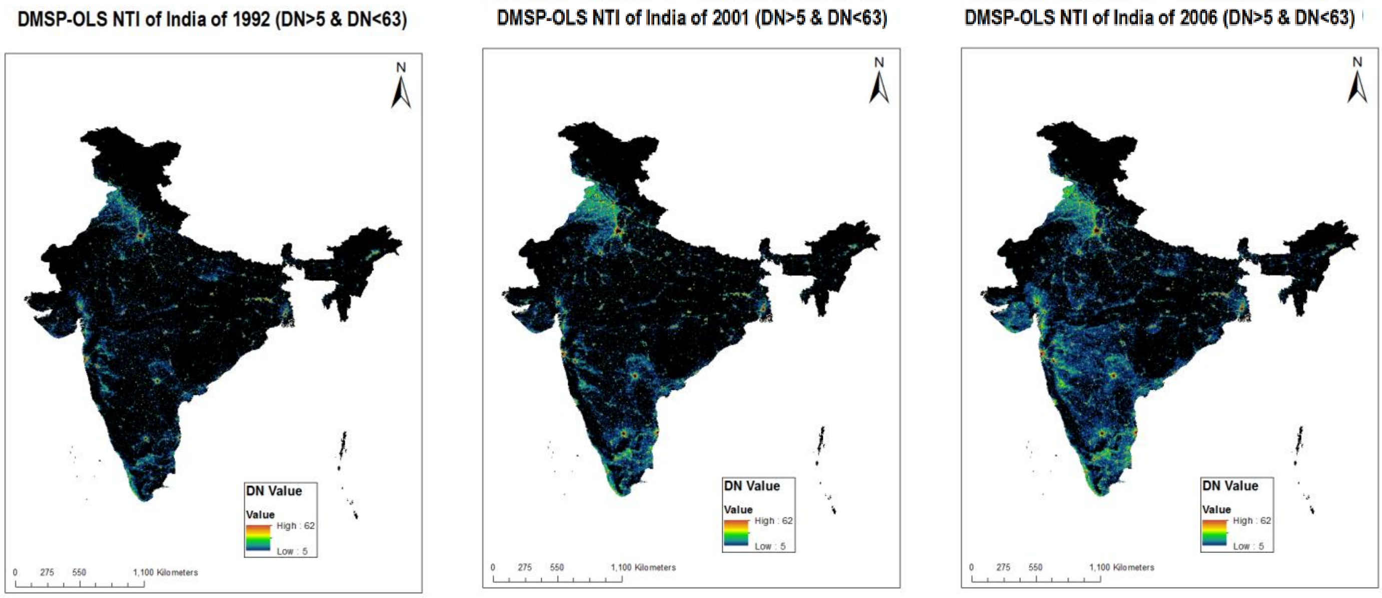


Figure 4.3: DMSP-OLS NTI of India with DN >5 and DN<63; a) for 1992 image, b) for 2001 image, c) for 2006 image; blue colour denotes low valued DN and red colour denotes high valued DN

Figure 4.3a, 4.3b and 4.3c show only the lit and saturation free pixels inside India for the year 1992, 2001 and 2006. The blue colour in the map represents lower valued lit pixels and red colour signifies brightly lit pixels with a maximum value of 62. The black colour in the maps represents NA values. In all the images, big cities like Delhi are observed to be of high luminosity and no sudden change can be observed in the cities. However, taking the whole India into consideration, an increasing trend in luminosity with time is observed from the images in the study time period. This growth indicates increase in urbanization. The trend of urbanization is however discussed in Section 4.3. These yearly composites with filtered DN values are used to in order to extract PIF with respect to luminosity. The following sections 4.1.2, 4.1.3 and 4.1.4 contain the results achieved in extraction of PIF.

4.1.2. Extraction of Spatially homogeneous clusters

Spatially homogeneous clusters are clusters that consist of DN values of similar spatial characteristics. It is discussed in Section 3.3.2 that Getis statistic (Gi*) gives a measure of clustering in images with respect to spatial association and similarities in DN values. Gi* is implemented using The RTW tool provided by Willson (2011) using ENVI on the masked pixel (DN value ranges between 5 and 62) images for the time period of 1992-2006 on initially a window size of 3×3 window and later it was also computed on a bigger window size of 5×5 and compared with each other to select optimal window size (for details see Section 4.1.4). Generated output images contain the clustered patches. The output images of implementing Gi* in a 3×3 window on 1992, 2001 and 2006

DN filtered images of India are shown below in Figure 4.4a, 4.4b and 4.4c. The outputs of these particular years are shown as it covers the whole time period of study and these yearly composites also show change in nature of luminosity with respect to time. In the next subsections under section 4.1, output of these three yearly composites are illustrated and discussed instead of all images.

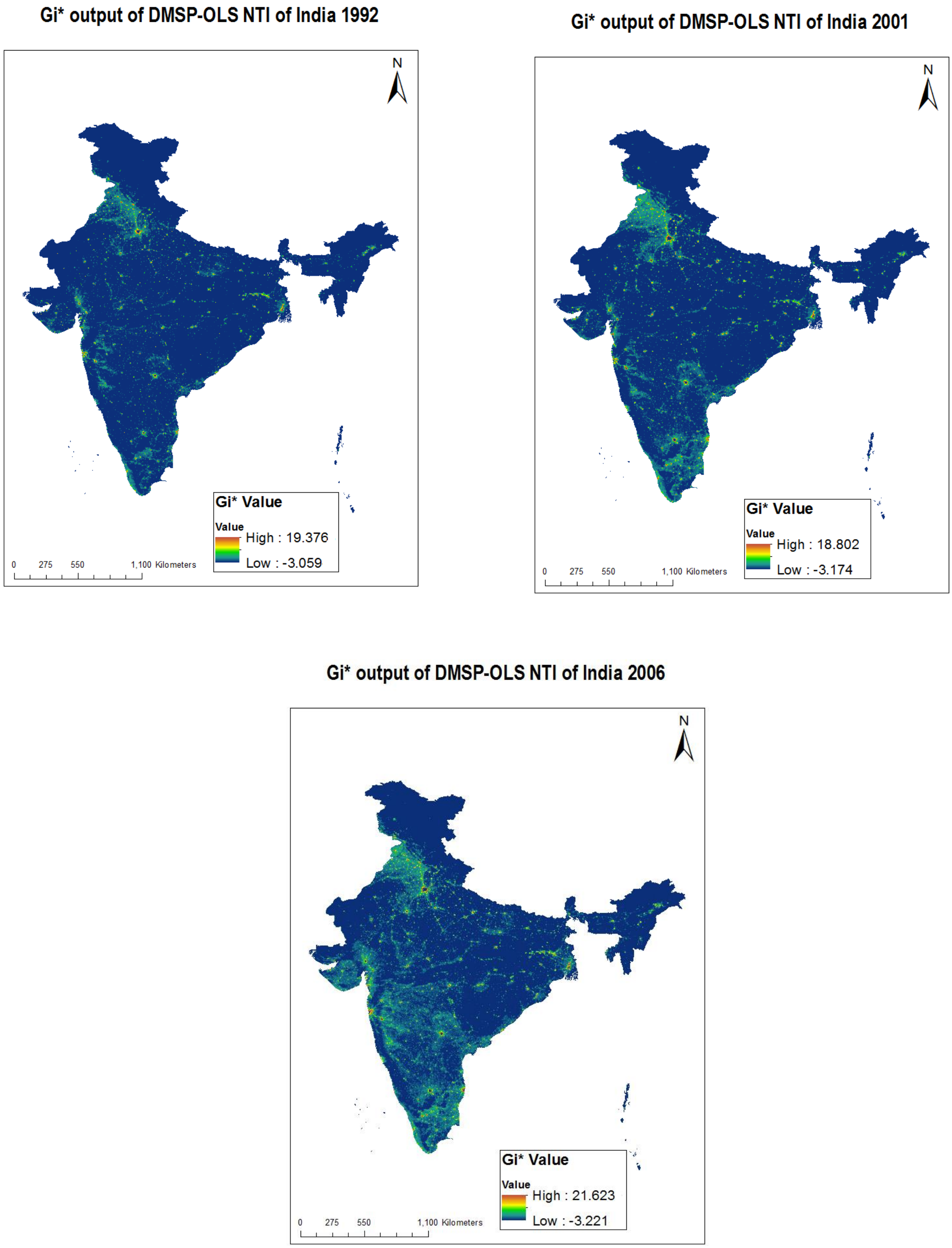


Figure 4.4: Gi* output of DMSP-OLS NTI of India of a) 1992, b) 2001, c) 2006; Gi* gives a measure of spatial association in local domain; blue colour denotes dark clusters and red colour denotes bright clusters

From Figure 4.4a, 4.4b and 4.4c spatially homogeneous clusters inside India with respect to luminosity in different years can be seen. These clusters are observed to have minimum value less than zero (approximately around -3.20) and a high value approximately around 20. As discussed in section 3.3.2, the pixels with Gi*> 0 refer to hotspot or spatially homogeneous bright clusters. In the current study, in order to extract PIF the bright clusters are only used. In Figure 4.4, the brightest clusters are observed around big cities like Delhi, Kolkata, and Mumbai etc. These bright clusters are not varying drastically with respect to location throughout the study time period. But on the other hand, many low lit areas observed in 1992 are changed into brightly lit areas afterwards. Thus, the location of dark clusters does seem to have more variability than of the bright clusters. A threshold of Gi* > 1.645 is used for all the Gi* output images in the research. This threshold selects the clusters at a significance level of 0.10 (Discussed in Section 3.2.2). Output images of Gi* are taken input into [R-3.2.2]™ using "raster" library. The pixels that satisfy Gi* >1.645 are considered as bright clusters and rest of the pixels that have a lower value than 1.645 are discarded. On successful extraction of pixels with a Gi* value more than 1.645, it shows clusters of pixels that are bright and locally spatially homogeneous. The regions of the cluster are also compared with study area information to check whether the extracted clusters are valid or not. Figure 4.5 shows the tri-temporal image containing the extracted pixel clusters (Gi* > 1.645) from 1992 masked image, 2001 masked image and from 2006 masked image with an administrative boundary map of India.

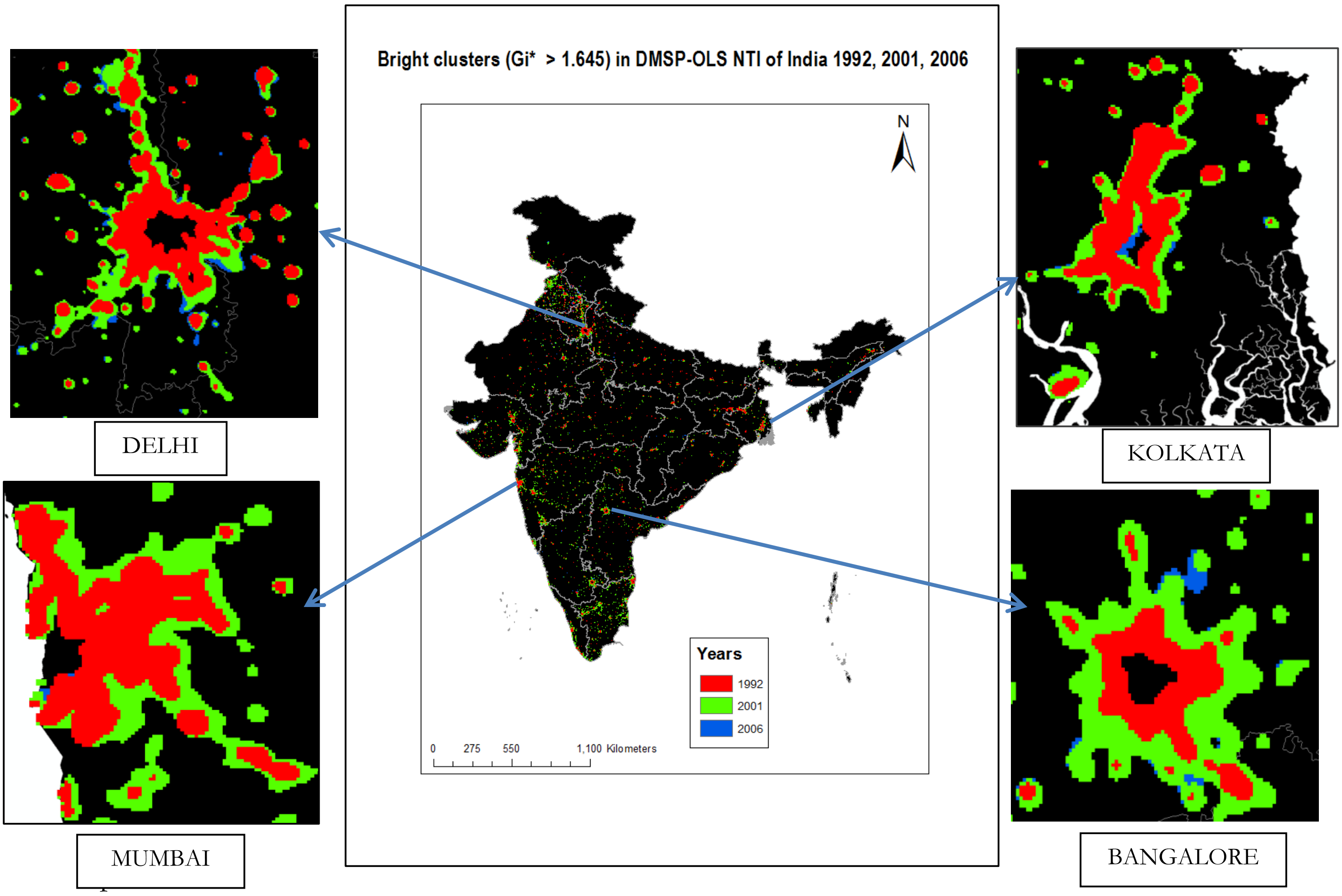


Figure 4.5: Tri-temporal map of pixels with Gi* > 1.645 in 1992, 2001 and 2006. Gi* > 1.645 denotes significant clustering (0.10) of bright DN values; Red colour denotes output for year 1992, green for 2001 and blue for 2006

Figure 4.5 shows the extracted bright clusters calculated in a 3×3 window are distributed all over India. Brightest clusters are observed mainly in cities in different states. These cities are heavily urbanized and densely populated regions. As a result of high urbanized areas, the night-time luminosity of these cities is also expected to be high. The brightest clusters among all the generated ones are represented above in zoomed view. It can be seen from the zoomed image chips in Figure 4.5 that in these big cities that with advance in time the number of bright pixels have also increased. As an example, the patch of bright pixels around Delhi in 2001 is bigger than of the same in 1992. It indicates increasing trend in urbanization explicitly also. More the bright pixels that have less variation in magnitude are situated near; more the pixel clusters are spatially homogeneous. Thus, it can be assumed that the output of the Gi* is valid and spatially homogeneous.

In order to extract PIF, the regions that are spatially homogeneous throughout the research time period are necessary to be extracted. For that all the bright clusters generated (Gi* > 1.645) for each year are converted to binary 1 and rest are converted to 0. On generation of all the binary bright clusters, Boolean AND operation is performed. It labels the regions as 1 which have a value of 1 for all the years; the regions that are bright and spatially homogeneous in all the years. Figure 4.6 shows the bright clusters that are locally spatially homogeneous through 1992-2006.

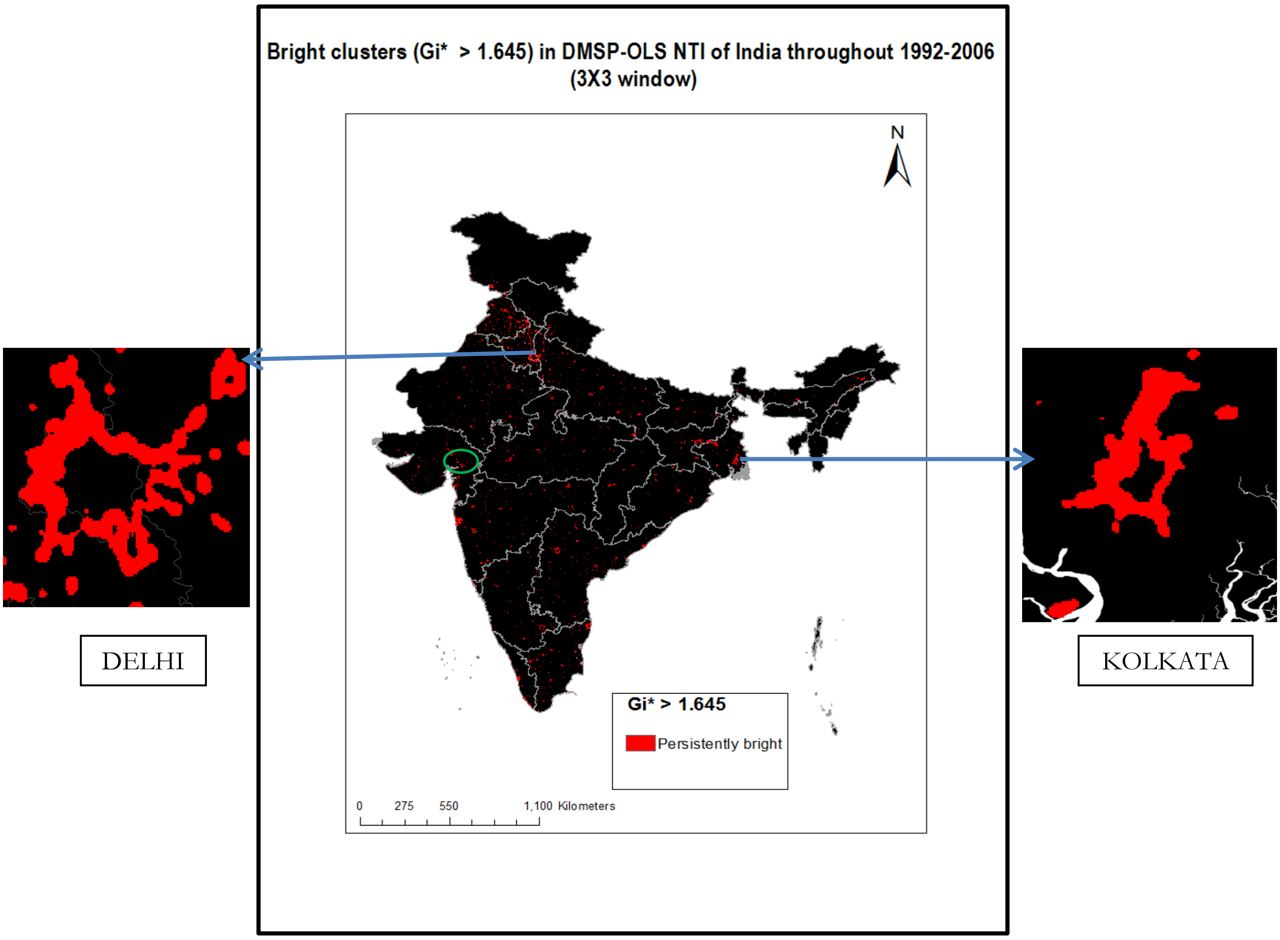


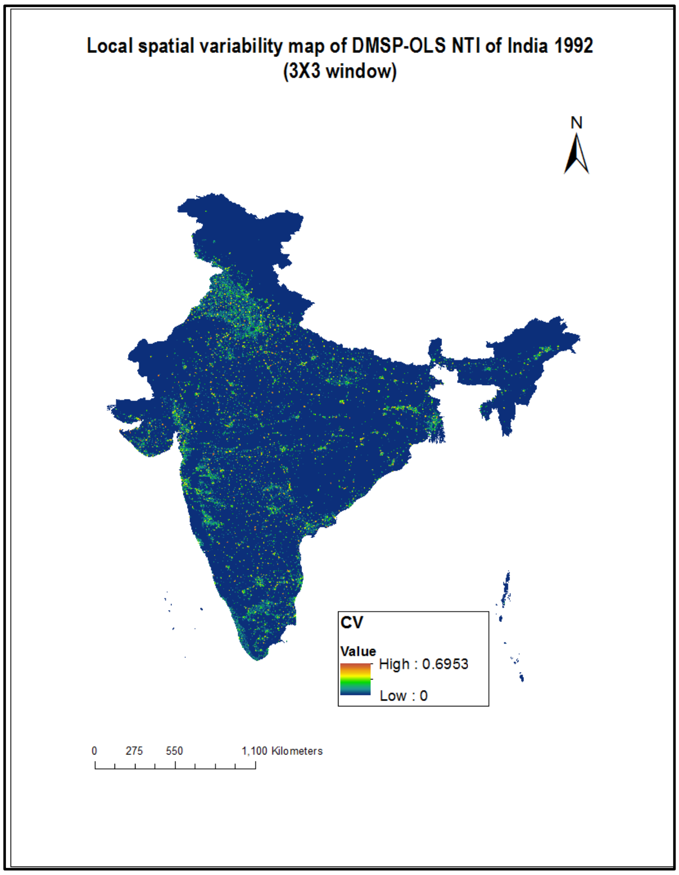
Figure 4.6: Areas that have Gi* > 1.645 output in all the years in 1992-2006; these areas are spatially homogeneous in nature and bright; the green circled region shows spatially homogeneous bright cluster with less DN range than cities

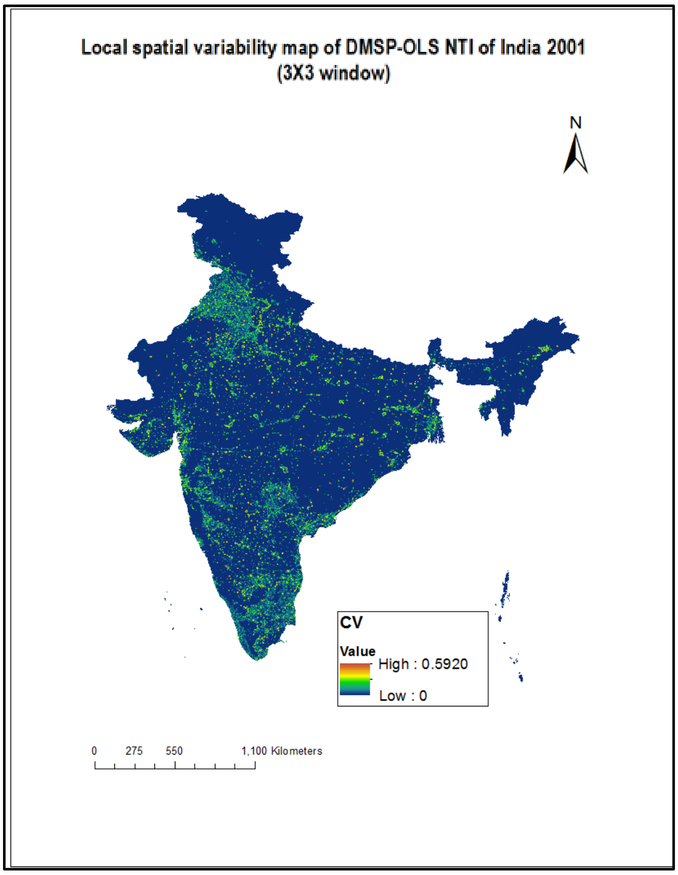
In Figure 4.6, the red coloured patches are bright clusters that are consistently bright and spatially homogeneous throughout the research time period. These clusters are generated with a significance level of 0.10 which limits the chance of random clustering to only 10%. In a zoomed view, Figure 4.6 also consists of the patches around big cities like Delhi, Kolkata and more. DN values inside these patches have less variation in magnitude; e.g. in cluster around Delhi the DN value is observed to vary between 48 and 62 which is approximately 20% of the range of DMSP-OLS NTI. Whereas in a cluster situated in north of Mumbai (shown in green circle in Figure 4.6) which is not a very big city, DN value varies between 34 and 53. Thus it can be assumed from such nature of clusters that the clusters around big cities are usually brightest and are locally spatially homogeneous over time. These patches are integrated along with output of CV in order to extract PIF. The next Section 4.1.3 states the implementation of CV to select regions with low spatial variability.

4.1.3. Extraction of regions with low spatial variability

The CV gives a measure of the spatial variability of the reflectance (here luminosity) for a given band within a defined area of an image. The defined area can be a small fixed window (e.g., 3×3, 5×5, 9×9 window) (Hamm, Milton, & Odongo, 2011). Initially a window size of 3×3 is used to implement CV. A comparative study is performed to select the optimal threshold to select pixels with low local spatial variability. Three different thresholds are compared in the study; CV < 3%, CV < 5% and CV < 10%. The region extracted from each threshold is combined with results obtained from Gi* outputs of corresponding years and as output of the integration, three different PIF are generated (for details of PIF generation see section 4.1.4) based on three different thresholds of CV taken. Initially, PIFs are generated for 1992, 2001 and 2006 image only. OLS is used to plot regression between the obtained PIFs of 1992 and 2006 keeping 2001 one as reference. These results of regression are compared to see which CV threshold is most feasible with respect to regression. As the main motif of extracting PIF is to perform inter-calibration, regression is selected as also an metric to decide the optimal CV threshold. On generation of the optimal CV threshold, the above mentioned procedure of generating PIF and plotting regression is performed in different window sizes (Gi* and CV computed in 5×5 window) and the output is compared with of the previously generated 3×3 regression results. The better resulting window is selected as optimal and PIF of all the years are generated.

CV is calculated using a program written in Python™ programming language. GDAL and NumPy packages are used in order to develop the program. The program calculates local CV under a user input window size. As mentioned in above paragraph, the program is initially used to generated CV map of the DN filtered DMSP-OLS NTI of India in a 3×3 window. Figure 4.7a, 4.7b and 4.7c shows the local spatial variability map of DMSP-OLS DN filtered NTI of 1992, 2001 and 2006 respectively.





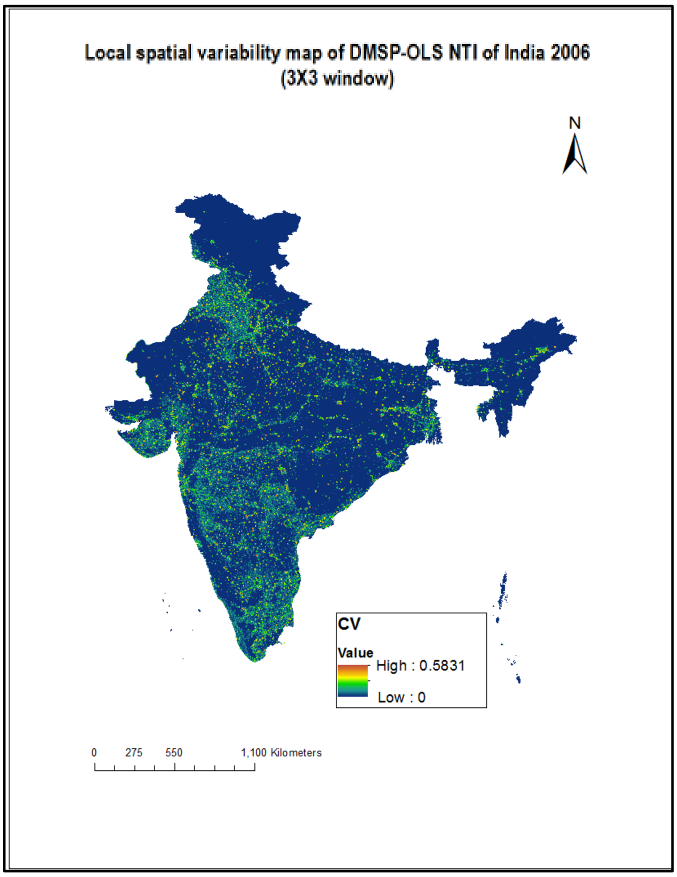


Figure 4.7: Map of local spatial variability of DMSP-OLS NTI of a) 1992, b) 2001 and c) 2006; CV is calculated inside a window; CV high value decreases with time for images of India

In Figure 4.7a, 4.7b and 4.7c, local spatial variability is mapped for DMSP-OLS NTI of India. Blue colour signifies pixels with zero or low variability whereas yellow and red colour signifies moderate and high spatial variability respectively. Image of 1992 is observed with highest local variability of 69.53% and image of 2006 is observed to have lowest local variability of 58.31%. It can be noted that the variability in data has reduced with the advancement of time. Consistent increase in CV values from previous year in same location is also observed. It refers to increase in lit pixels or increase in luminosity in pixels with time. In previous years, as numbers of bright pixels are less there is higher variability in luminosity between low-lit and bright-lit pixels. But with progress in time, many low-lit pixels have increased in luminosity and the variation between pixels has become lower. Thus, a decreasing trend in maximum spatial variability can be noted. However, pixels with such high spatial variability cannot be considered as of stable and low local variability. Thus a threshold is computed to select locally less spatially variable pixels. Three different values are compared in current research for selection of optimal threshold. These values are,

1. CV < 3%
2. CV <5%
3. CV <10%

Figure 4.8, 4.9 and 4.10 below shows the composite image of CV < 3%, CV < 5% and CV < 10% for the year 1992, 2001 and 2006 respectively. These Figures show the points that have a lower spatial variability than the mentioned thresholds.

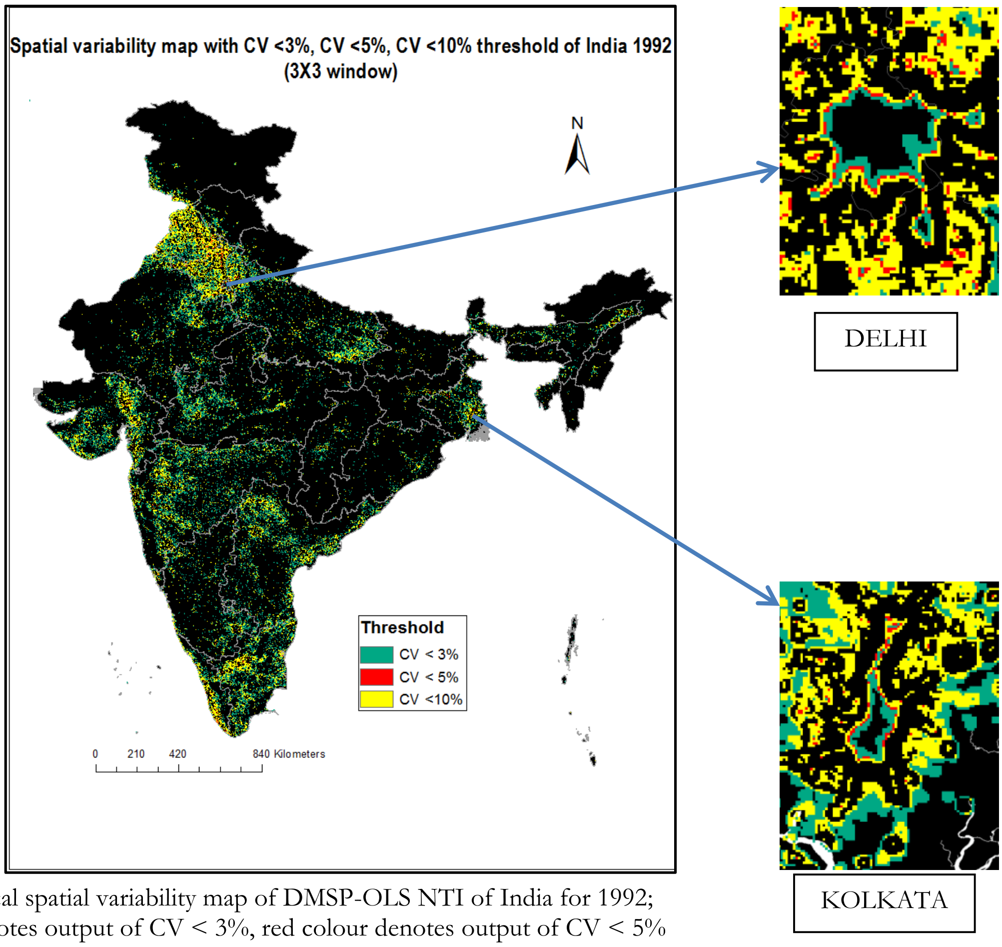


Figure 4.8: Local spatial variability map of DMSP-OLS NTI of India for 1992; cyan colour denotes output of CV < 3%, red colour denotes output of CV < 5% and yellow colour denotes output of CV < 10%

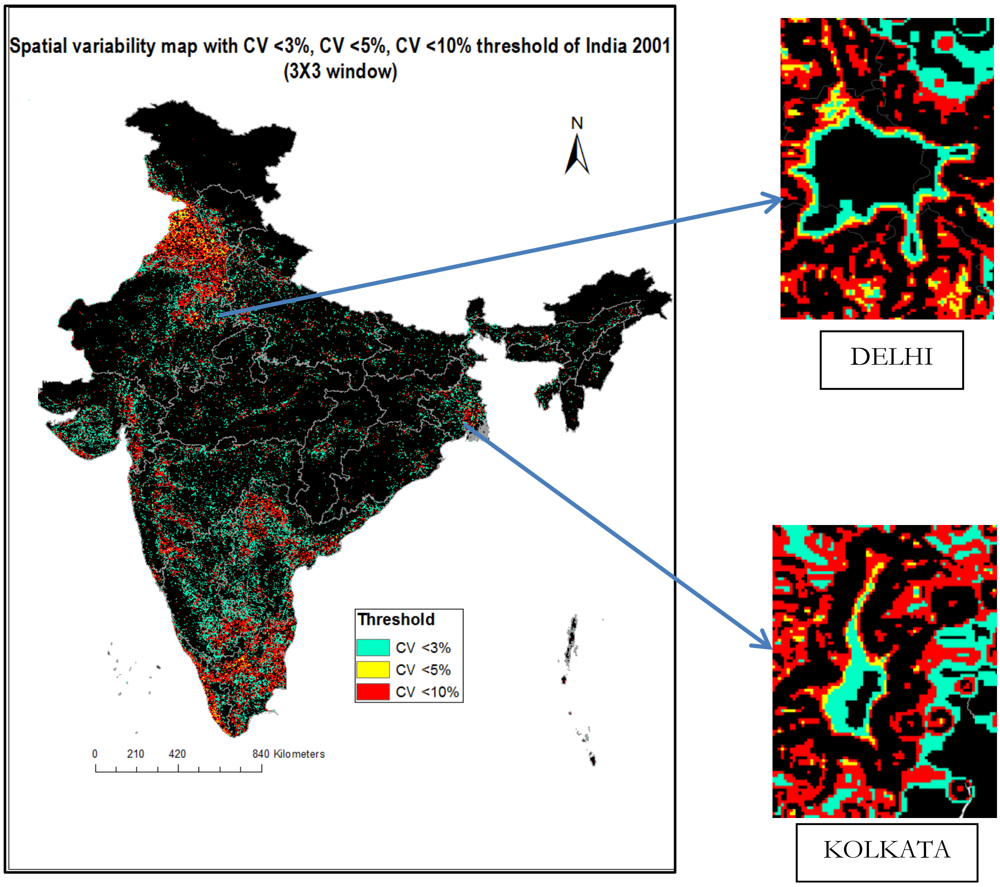


Figure 4.9: Local spatial variability map of DMSP-OLS NTI of India for 2001; cyan colour denotes output of CV < 3%, yellow colour denotes output of CV < 5% and red colour denotes output of CV < 10%

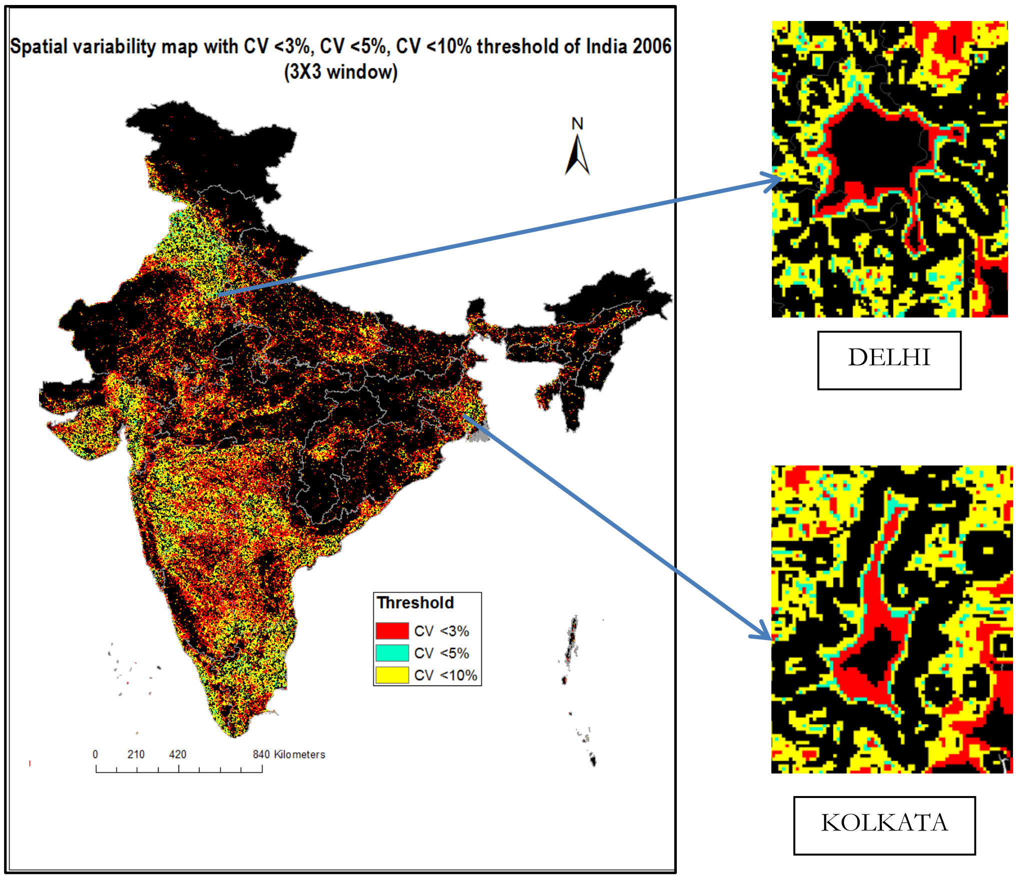


Figure 4.10: Local spatial variability map of DMSP-OLS NTI of India for 2006; red colour denotes output of CV < 3%, cyan colour denotes output of CV < 5% and yellow colour denotes output of CV < 10%

In Figure 4.8, pixels below the mentioned three different CV thresholds are composited in a single image for all the thresholds for the year 1992. It shows that the number of points that have a CV <3% are lower than of CV <5% and correspondingly than CV <10%. With increasing the threshold, numbers of points are observed to increase. From comparing Figure 4.8 with Figure 4.9 and 4.10 respectively, it can be seen that regions around the big cities like, Delhi and Kolkata do have low spatial variability. Inside the core parts and centre of the cities the CV is observed to be NULL. It can be justified from the fact the city centres do consist of saturated pixels (DN=63) that were discarded at beginning of research. Thus, there are only background values are present which are NA values. The low spatial variability around the cities occurs as cities consist of bright pixels and the variability of luminosity is less in the cities. Similarly, spatial variability is also less inside the cities.

In order to extract PIF from the images, regions with consistent low local spatial variability are extracted. For that all the regions with low local spatial variability (CV < 3%, CV < 5% and CV < 10%) are converted to binary 1 and rest are converted to 0. On generation of the binary regions of low local variability of all the years, Boolean AND operation is performed. It labels the regions as 1 which has a value of 1 for all the years; the regions that have low local variability in all the years. Figure 4.11 shows constantly spatially less variable regions through 1992-2006.

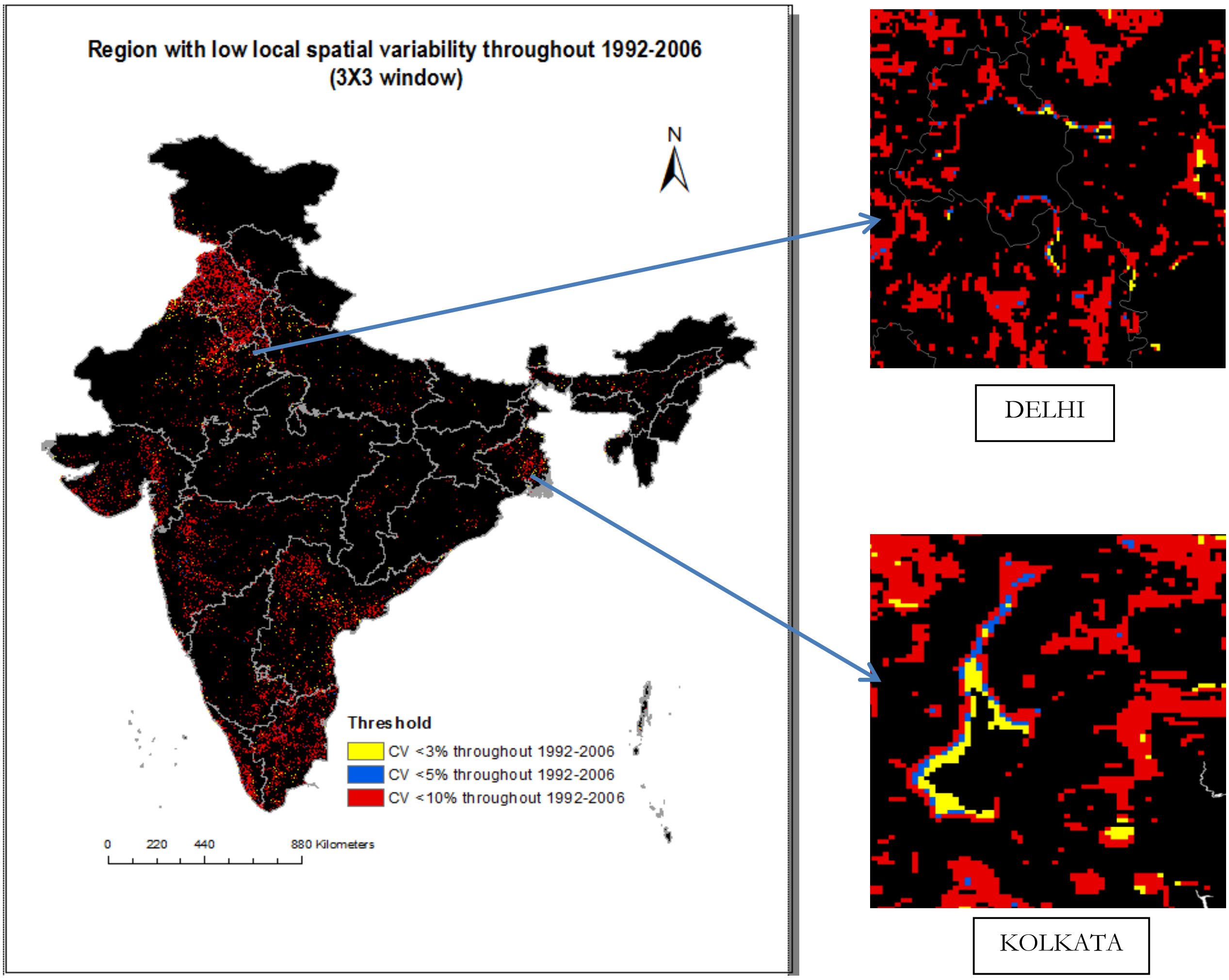


Figure 4.11: Areas with consistent low local spatial variability through 1992-2006. Yellow, blue and red colour shows the areas with CV < 3%, CV <5% and CV <10% respectively in all the years

In Figure 4.11, points that are consistently of low local spatial variability (CV <3%, CV <5% and CV <10%) are shown. Such points are observed around big cities also though the number of points varies with respect to CV threshold. The numbers of points obtained with all three CV thresholds are shown below in table 4.1.

CV threshold	Number of points
CV < 3%	41896
CV <5%	46675
CV <10%	265307

Table 4.1: Number of pixels obtained from various CV thresholds throughout 1992-2006

CV<10% threshold is seen to generate the maximum number of points for all the years. These regions are selected as regions with low local variability which satisfies condition of extraction of PIF. The next Section 4.1.4 states the result of integrating spatially homogeneous and low local spatial variable region in generation of PIF with different CV thresholds. Finally, by comparing these different PIFs optimal CV threshold is selected and optimal window size to compute Gi* and CV is also found.

4.1.4. Extraction of Pseudo Invariant Features (PIF)

As mentioned earlier in section 3.3.2, PIFs are spatially homogeneous and of low local spatial variability. Section 4.1.2 and 4.1.3 states the extraction of spatially homogeneous clusters and regions with low spatial variability respectively in the time period of 1992-2006. These two regions are integrated (using Boolean AND between output shown in Figure 4.6 and Figure 4.11) to extract the areas that are both spatially homogeneous and of low spatial variability. Such region is addressed as PIF in the study. Three different PIF are generated because of three different CV thresholded outputs are combined with the bright clusters generated from implementing Gi* on the images of 1992-2006. The following table 4.2 shows the number of pixels associated in each of the three PIF.

PIF_CV threshold	Number of points in PIF
CV < 3%	394
CV <5%	972
CV <10%	3956

Table 4.2: Number of pixels in PIF obtained from various CV thresholds throughout 1992-2006

Figure 4.12a, 4.12b and 4.12c shows the three different PIFs generated respectively and also a zoomed view around Delhi to see the arrangement of points around heavily urbanized areas.

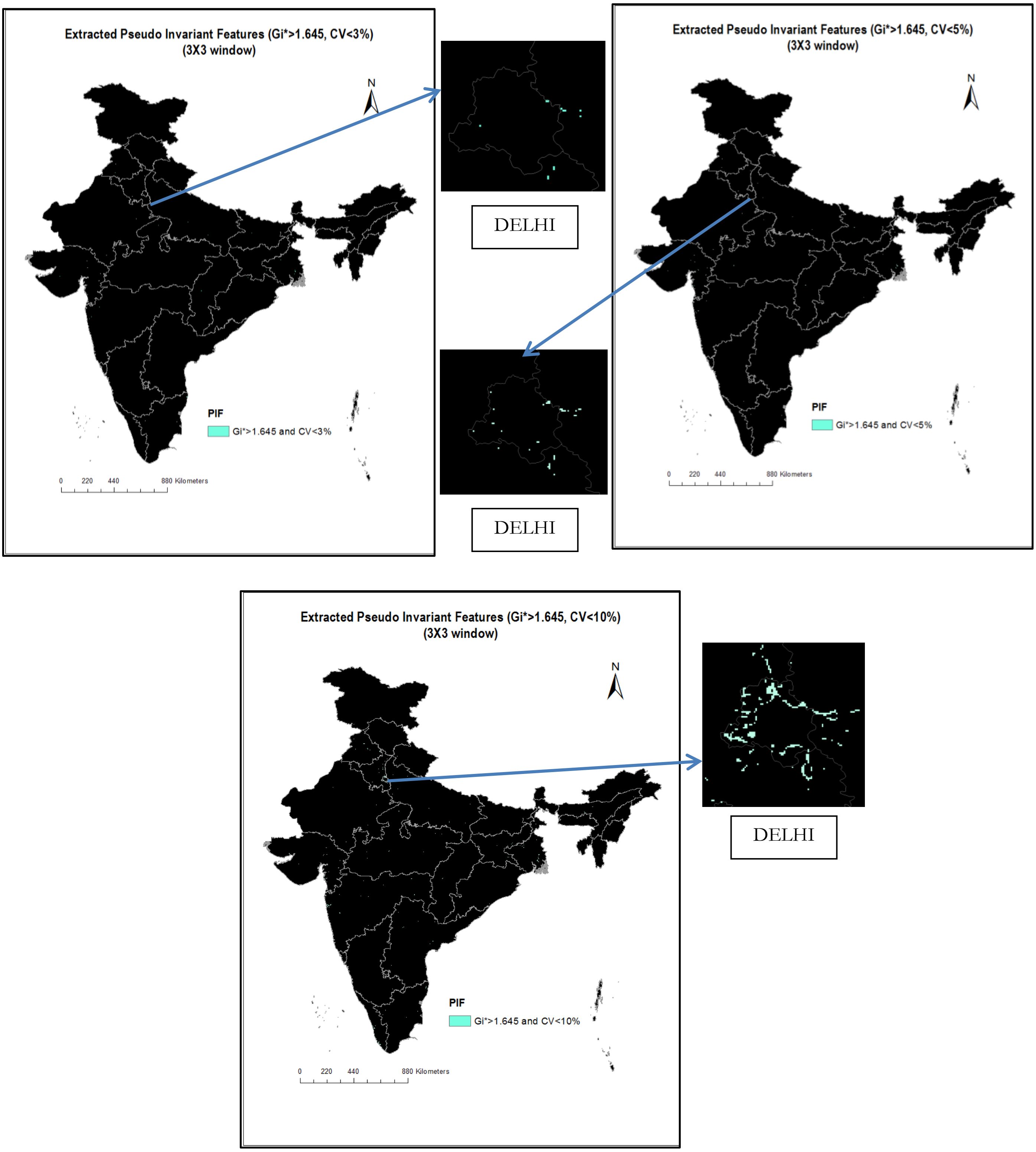


Figure 4.12: Spatially homogeneous and low spatially variable regions (PIF) throughout 1992-2006 for DMSP-OLS NTI of India for a) CV < 3%, b) CV < 5%, c) CV < 10%

From table 4.2 and Figure 4.12a, 4.12b and 4.12c the number of points associated with CV <3% and CV<5% are seen to be very less. The zoomed view in Figure 14a and 14b shows very low number of points around Delhi whereas In Figure 14c, threshold of CV<10% is seen to consist of many points around Delhi as PIF. To select the optimal threshold, regression is performed among the PIFs. PIF of 2001 is used as reference and 1992 PIF as an early year PIF and 2006 as a late year PIFs are used to check the value range of PIFs and also to perform regression in order to select optimal threshold in context of calibration. The results of regression of the PIFs of 2006 and 1992 with respect to PIF of 2001 are noted below in table 4.3.

Year	Window size	Gi* threshold	CV threshold	Regression equation with respect to reference year (2001)	R^2	Figure ref.	No. of obs.
1992	3×3	Gi*>1.645	CV<3%	$(DN_{1992})_i=13.79+0.772*(DN_{2001})_i$	0.29	Fig. 25	123
		Gi*>1.645	CV<5%	$(DN_{1992})_i=-1.67+1.005*(DN_{2001})_i$	0.88	Fig. 27	456
		Gi*>1.645	CV<10%	$(DN_{1992})_i=-7.34+1.049*(DN_{2001})_i$	0.83	Fig. 29	3101
	5×5	Gi*>1.645	CV<10%	$(DN_{1992})_i=-2.56+1.018*(DN_{2001})_i$	0.97	Fig. 31	437
2006	3×3	Gi*>1.645	CV<3%	$(DN_{2006})_i=-1.19+1.014*(DN_{2001})_i$	0.45	Fig. 26	149
		Gi*>1.645	CV<5%	$(DN_{2006})_i=0.28+0.985*(DN_{2001})_i$	0.95	Fig. 28	474
		Gi*>1.645	CV<10%	$(DN_{2006})_i=-1.98+1.008*(DN_{2001})_i$	0.91	Fig. 30	2928
	5×5	Gi*>1.645	CV<10%	$(DN_{2006})_i=0.29+0.976*(DN_{2001})_i$	0.98	Fig. 32	457

Table 4.3: Regression results of various combinations of Gi* and CV thresholded values obtained in different window sizes (3×3, 5×5) with significance level of 0.05; $(DN_{1992})_i$ represents the i_{th} pixel in to be estimated of the year 1992, $(DN_{2006})_i$ represents the i_{th} pixel in to be estimated of the year 2006 and $(DN_{2001})_i$ represents the corresponding i_{th} pixel in the year 2001 used as reference; the Figure numbers point to the regression line fitted scatterplots of the corresponding year mentioned in 1st column of the table; No. of observations show the number of observations in each regression

Table 4.3 shows the result of implementing OLS regression on PIF of 1992 and PIF of 2006 with respect to PIF of 2001. Regression is implemented on PIFs generated by integration of results achieved by Gi* and CV thresholded outputs. The coefficient of determination (R^2) determines the goodness of fit of the line with the trend of the data (precision). As the Gi* threshold is observed to be constant in all the combinations (Gi* > 1.645), the difference in results seem to occur for the variable threshold of CV. R^2 and number of observations are used as the metric to select the optimal threshold for CV.

PIFs obtained with threshold of CV <3% are observed to have a poor R^2 value in both the regressions mentioned in table 4.3. Low R^2 value signifies poor fit of the regression line with the data. In Figure 4.13a and 4.13b the visual interpretation of the regression also shows poor linear fit. Besides, the range of the values in regression is also very narrow. The number of observations are also observed very low in both the years with CV <3%.

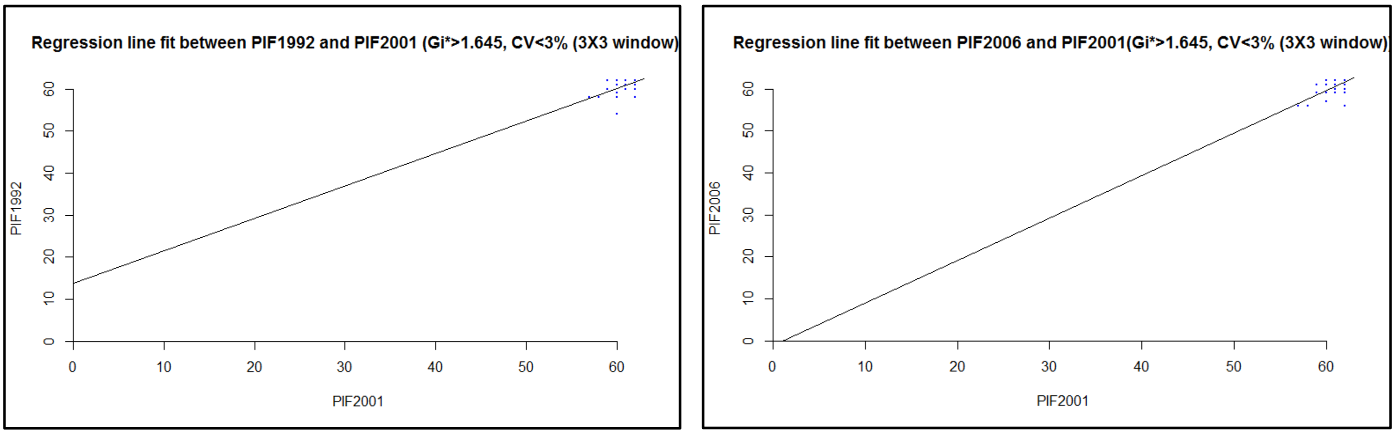


Figure 4.13:OLS implemented with a significance level 0.05 for a) PIF of 1992, b) PIF of 2006 keeping PIF of 2001 as reference; PIFs are generated with Gi*>1.645, CV<3% under 3×3 window

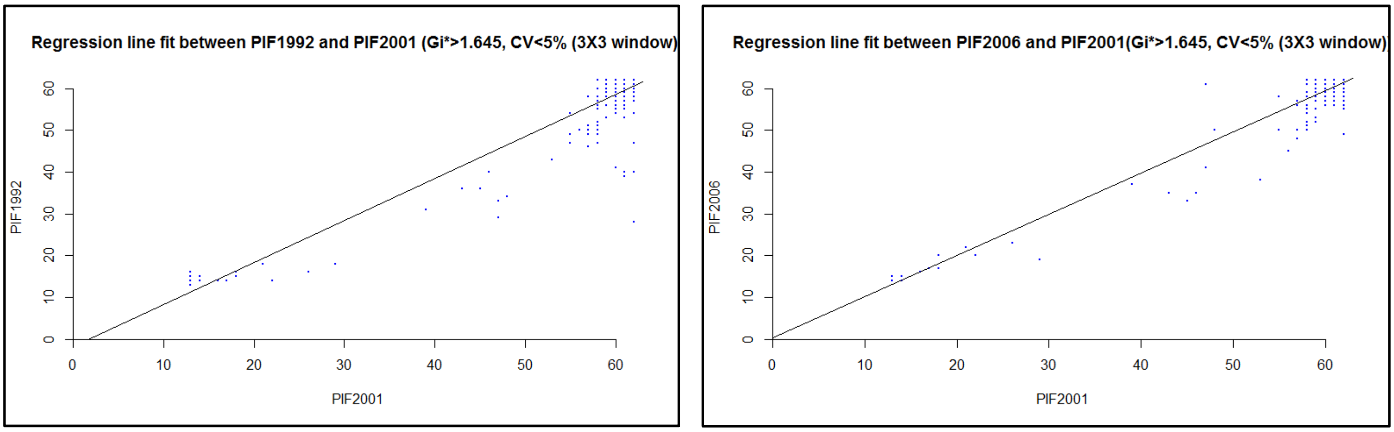


Figure 4.14: OLS implemented with a significance level 0.05 for a) PIF of 1992, b) PIF of 2006 keeping PIF of 2001 as reference; PIFs are generated with Gi*>1.645, CV<5% under 3×3 window

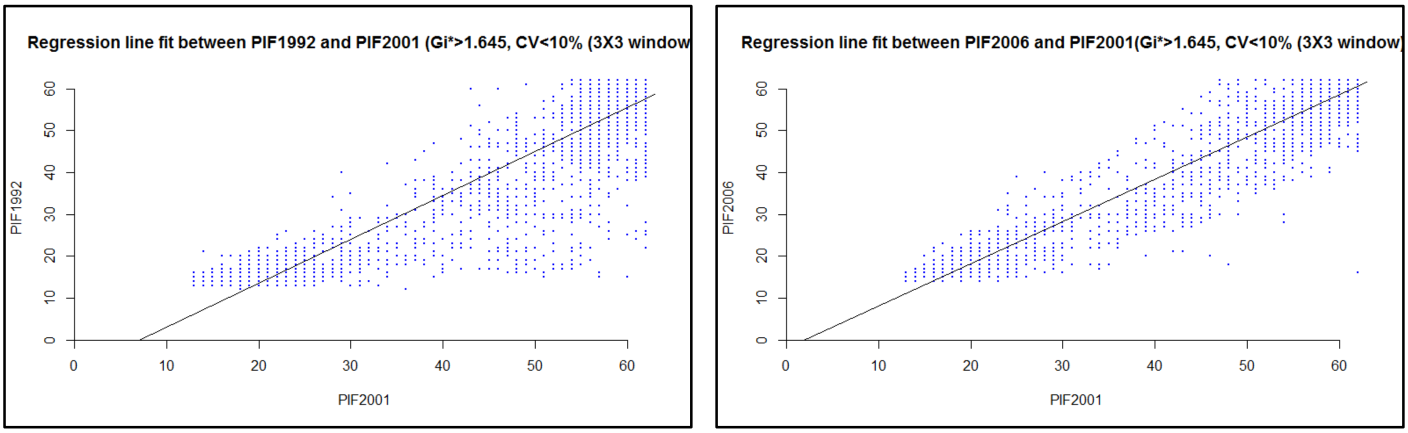


Figure 4.15: OLS implemented with a significance level 0.05 for a) PIF of 1992, b) PIF of 2006 keeping PIF of 2001 as reference; PIFs are generated with Gi*>1.645, CV<10% under 3×3 window

For the PIFs generated with CV<5% threshold (see Figure 4.14a and 4.14b) the range of pixels does increase than of PIFs thresholded with CV<3%. In table 4.3 the R^2 value in regression between PIFs generated with

CV<5% are also noted to be higher than PIFs with CV<3%. The number of observation for 1992 is 456 and for 2006 it is noted to be 474 which is more than of the same of PIFs with CV<3%. However, in the visualization of the regressions in Figure 4.14a and 4.14b it is seen that maximum points in scatter plot are lying between a lower range (approx. between 10 to 25) and a comparatively higher range (approx. between 55-62). In the middle values of these two ranges very few points can be seen. It refers to the fact that the PIFs extracted with CV<5% threshold do not consist of many points in the DN value range of 30 to 55. Such distribution of DN values is not suitable to implement regression. The data span should be continuous in terms of modelling regression. But in PIFs extracted with CV<5% the pixel distribution resembles discrete as maximum numbers of pixels are lying within a defined range and creating two clusters of comparatively low and high values respectively. The fitted regression line is generated considering the clustered values mainly. So it can be assumed from the result that though the R^2 is high in both the years in regression but the lack of points in a certain range and discrete clustering of values are not suitable to implement inter-calibration.

In the PIFs generated with CV<10% do also have a broad range of DNs alike PIFs with CV<5%. The distribution of points in the pixel range is seen to be well distributed all over the range and adequate number of points can also be seen in the scatterplots shown in Figure 4.15a and 4.15b. The R^2 for the regression with PIF of 1992 is observed to be of 0.83 and for 2006 it is observed to be of 0.91 which denote good linear fit of the regression line with the data. The number of observations are also much greater (see in table 4.3) than of the regression of PIFs with CV<3% and CV<5%.Thus, in a comparison among all the PIFs with different CV threshold in terms of number of observations and goodness of fit, CV<10% is considered as the optimal CV threshold to extract PIF from the different yearly image composites.

Besides selecting the optimal threshold for CV, the optimal window size is also found. PIFs for 1992, 2001 and 2006 are again generated in a similar way with section 4.1.2 and 4.1.3. The threshold for Gi* used is Gi*>1.645 and for CV it is used as CV<10%. Both CV and Gi* is implemented in 5×5 window. To select optimal window size, regression is again plotted between the PIFs of 1992 and 2006 with PIF of 2001 obtained in 5×5 window. Figure 4.16a and 4.16b below shows the regression line fit for both the PIFs (1992 and 2006) extracted in 5×5 window.

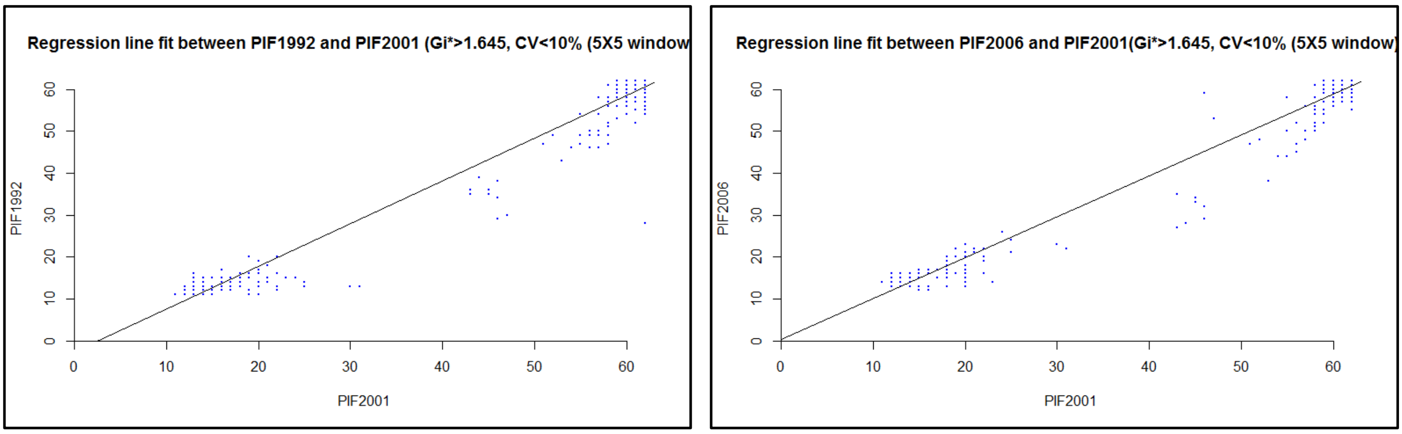


Figure 4.16: OLS implemented with a significance level 0.05 for a) PIF of 1992, b) PIF of 2006 keeping PIF of 2001 as reference; PIFs are generated with Gi*>1.645, CV<10% under 5×5 window

In regression between PIF generated in 5×5 window the problem of clustering of points in specific low and high range is observed (Figure 4.16a and 4.16b). It resembles the clustering seen in regression of PIFs obtained with CV<5% in 3×3 window. The number of observations are also significantly low than of PIFs generated with same threshold (Gi*>1.645, CV<10%) in a 3×3 window (see table 4.3). It is observed that with increase in the size of window PIFs generated with exact thresholds consist of fewer points. Less number of points produces redundant result in implementing regression. Thus, in terms of generating PIF 3×3 window size is selected as optimal. The finalized thresholds for generating PIF are mentioned below:

1. Gi* > 1.645 (3×3 window)
2. CV < 10% (3×3 window)

The generated PIF is used as a mask to extract DN values of corresponding PIF from yearly images from 1992 to 2006. These extracted values for each year are spatially homogeneous and of low local variability. The DN value of a particular pixel inside the PIF does not differ drastically from other year PIF pixel value of same co-ordinate.

The spatial variability of the extracted PIFs of different years from 1992 to 2006 is also calculated to check whether the spatial variability of the whole PIF varies drastically throughout the years or not. As discussed earlier, spatial variability of the whole PIF should also not vary drastically through the years. Figure 4.17 shows the spatial variability of whole PIF values for all the years considered in research.

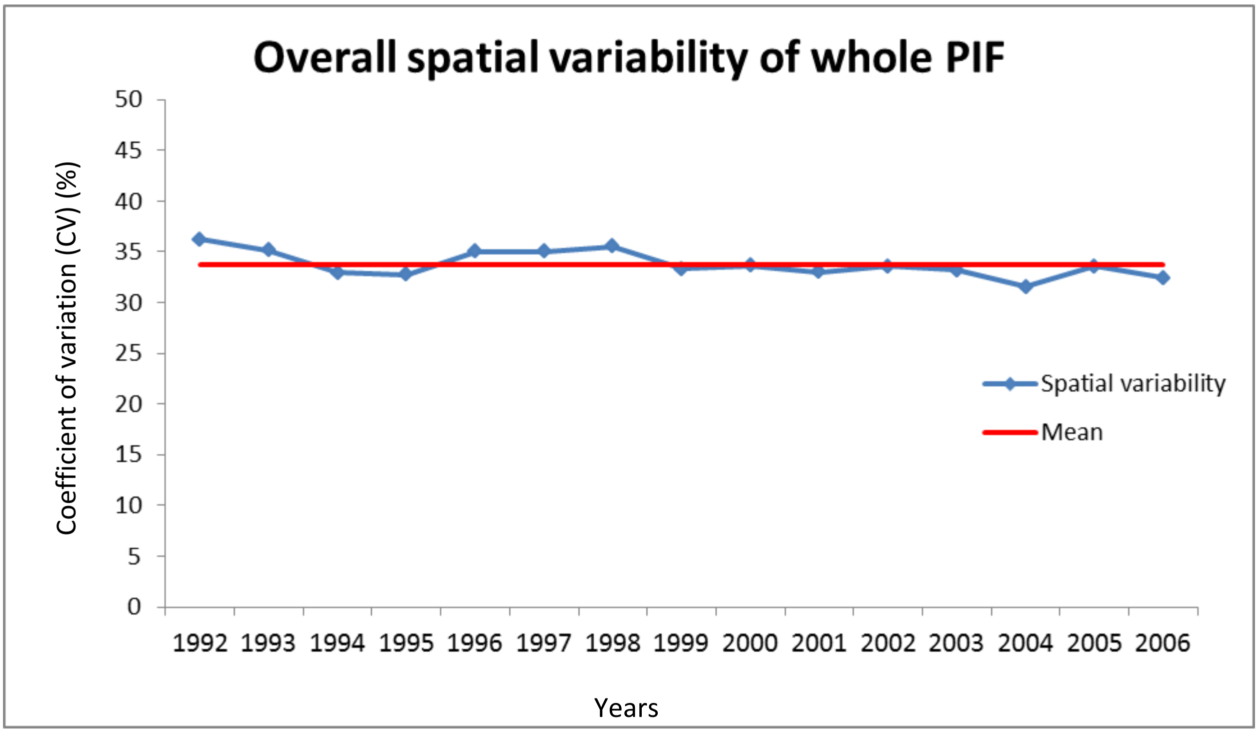


Figure 4.17: Spatial variability of all the points in PIF from 1992-2006. A very narrow deviation from mean spatial variability is observed in all the years.

Spatial variability is measured with calulating CV for the all the points in the PIF in a particular year instead of caculating in a fixed window. Minimum CV for the whole PIF is observed for the year 2004 with 31.53% and maximum variability is noticed for the year 1992 with a value of 36.23%. PIF of 1992 holds highest variability as the variabiity in the DN values is maximum among all the years. In intial years, comparatively large number of low-lit pixels can be noticed. As a result, the variation among DN values is also high which leads to a higher CV value than of late years. With advancement of time, low-lit pixels are turned into bright pixels and the variation among DN values gradually decreases which leads to a lower overall variability in PIFs in later years. However, it can also be seen that the spatial variability does not varry drastically over time. The mean spatial variability of all the years is 33.78%. The red line in Figure 4.17 denotes the mean. The spatial variability of each year is also observed to lie arround the mean which leads to the realization that the inter-year spatial variability is also low in the PIFs.

Thus, it can be inferred from the above results that the extraction of PIF is succesful and the extracted PIFs do obey the criteria mentioned in section 3.3.2. PIF of every year is used to implement regression with respect to PIF of 2001 in order to peroform inter-calibration of the different yearly images. The next section 4.2 states the details of implemention of OLS regression and robust regression for better removal of outliers in fit.

4.2. Comparison of regression to perform inter-calibration

With the extracted PIFs from the DMSP-OLS yearly composites for 1992 to 2006, regression is modeled to perform an empirical calibration or inter-calibration between the images. Linear regression is implemented in the study. OLS is implemented initially on all the PIFs to generate the coefficients of linear regression with which calibration can be performed. Outliers in OLS are defined as per equation 1 in the study. However, OLS is a poor estimator in context of outlier removal in the fit. OLS estimator has a breakdown point of 0% which means even in presence of a single outlier the fit of regression line can be redundant (Massart et al., 1986). Thus in the study, two robust linear regression methods; LMedS and LTS are also used. Coefficient of determination (R^2), Root mean square error (RMSE) and execution time are used in the study to select the optimal regression technique that reduces effect the outliers most effectively. Section 4.2.1 states the implementation of the mentioned regressions and comparison for selection of optimal regression among them. Section 4.2.2 shows the validation of inter-calibration performed with optimal regression technique. A comparison with the coefficients given by Elvidge et al. (2009) is also performed in terms of SOL index.

4.2.1. Comparison of regression techniques

To implement linear regression, PIF of 2001 is used as the reference image in all the estimation techniques. The following Table 4.4 shows the results of OLS, LMedS and LTS implemented for all the years in 1992 to 2006 where image of 2001 is used as reference. All regressions are performed with a significance level of 0.05.

	OLS					Lmeds					LTS				
Year	Slope	Intercept	R^2	RMSE	Execution time(sec)	Slope	Intercept	R^2	RMSE	Execution time(sec)	Slope	Intercept	R^2	RMSE	Execution time(sec)
F101992	1.049	-7.34	0.83	7.92	30.5	0.977	-0.11	0.84	7.13	18.97	1.051	-3.71	0.84	7.09	17.44
F101993	1.017	-6.34	0.84	7.58	17.19	1.000	-2.53	0.85	6.92	9.26	1.037	-3.84	0.86	6.7	8.86
F121994	0.960	0.15	0.88	5.84	23.66	1.045	-1.10	0.9	5.7	9.34	0.958	-2.12	0.89	5.52	9.42
F121995	1.016	-2.48	0.9	5.4	16.86	1.000	-0.55	0.92	5.36	10.11	1.000	-1.33	0.92	5.23	9.34
F121996	1.012	-2.90	0.92	5.35	20	1.000	-0.50	0.93	5.22	9.26	1.034	-0.70	0.93	5.17	9.15
F141997	1.017	-1.93	0.94	6.21	17.53	1.010	-0.68	0.95	5.44	9.75	1.021	-0.62	0.93	5.43	9.83
F141998	1.017	-0.43	0.95	5.51	18.72	1.000	-0.78	0.95	4.81	10.29	1.026	-0.40	0.96	4.73	9.44
F141999	0.978	2.12	0.96	3.94	17.65	1.049	1.03	0.97	3.84	10.44	0.957	0.11	0.97	3.7	10.74
F142000	1.072	-2.82	0.96	3.83	15.75	1.000	-0.51	0.96	3.71	11.55	1.044	-0.89	0.98	3.46	11.35
F152001					R	E	F	E	R	E	N	C	E		
F142002	1.033	-2.14	0.97	2.96	15.2	1.000	0.02	0.98	2.95	10.67	1.020	0.05	0.98	2.85	9.64
F142003	1.063	-2.80	0.95	4.01	16.94	1.040	0.05	0.96	3.8	10.03	1.030	0.04	0.95	3.75	8.85
F162004	1.006	-1.48	0.95	3.78	17.51	0.955	2.30	0.96	3.82	10.09	0.979	0.52	0.96	3.55	9.62
F162005	1.076	-1.13	0.93	5.03	25.89	1.045	-0.38	0.94	4.46	10.38	1.036	0.40	0.95	4.32	8.99
F162006	1.008	-1.98	0.91	4.96	22.3	0.989	-0.50	0.93	4.88	15.4	0.979	0.83	0.94	4.69	12.59

Table 4.4: Regression coefficients (Slope and intercept), corresponding R^2, RMSE and execution time for all the years from 1992 to 2006 for OLS, LMedS and LTS estimator keeping 2001 as reference; R^2 measures precision of fit, RMSE measures error in fit and execution time measures time to complete the process; slope is approximated to three digits after decimal and for intercept it is set as 2 digits after decimal; regressions are performed with significance level of 0.05.

The regression coefficients (Slope and Intercept) obtained from the regression technique are used to calibrated the night-time images. The estimator that handles the outlier best and linearly fits the data most precisely is selected as optimal calibration technique. R^2 here gives the measure of relative error or precision, RMSE gives the measure of accuracy and execution time measures the time to model the regression. Optimal technique is considered in all of these metrics and form combination of all the three best results, optimal regression technique is selected.

It can be seen in table 4.4 and Figure 4.18 that LMedS and LTS both estimators are more precise than of OLS as R^2 is observed to be higher than OLS in maximum number of years. In case of LMedS and LTS, R^2 values lie very close to each other in all the years.

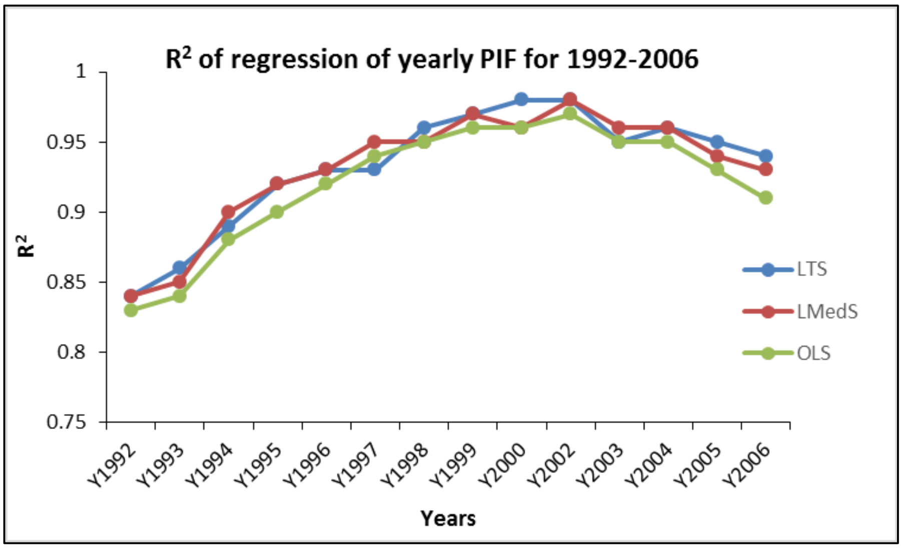


Figure 4.18: R^2 of the regression of PIF of each year with PIF of 2001; Higher R^2 denotes higher precision in the fit of trend line

The mean of the R^2 for LMedS is noted as 0.931 and for LTS it is 0.932 which shows a very little difference in precision between these two models. Thus, the other metrics (RMSE and execution time) are used to select the better estimator.

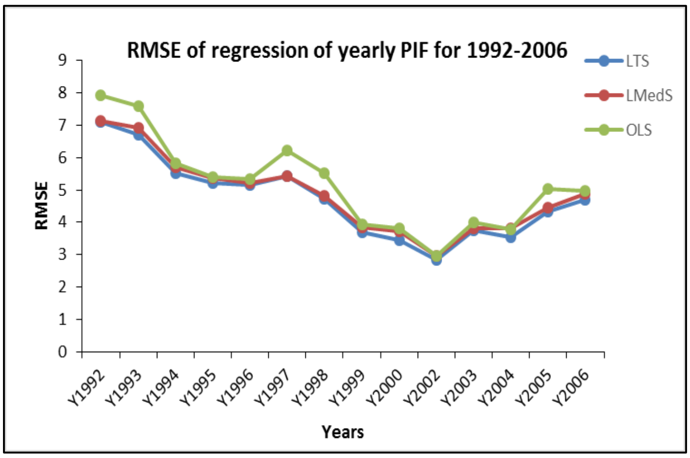


Figure 4.19: RMSE of the regression of PIF of each year with PIF of 2001; Higher RMSE denotes higher accuracy in the fit of trend line

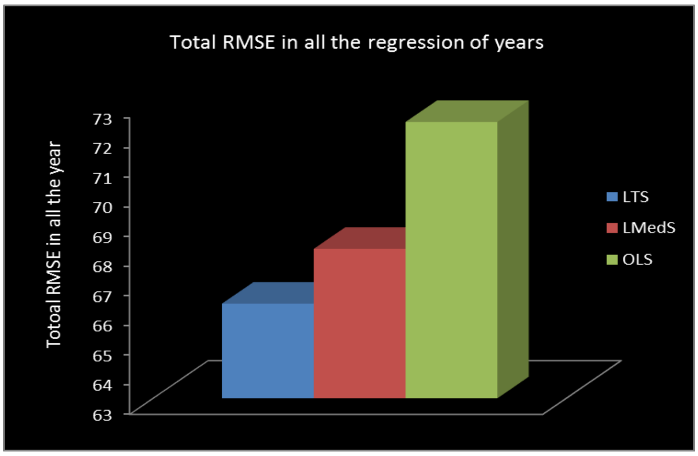


Figure 4.20: Total RMSE of the regression of PIF of each year with PIF of 2001

In Figure 4.19, RMSE of yearly regressions is observed to be least in LTS estimator. Lower the RMSE is, more the accuracy is in the fit. Figure 4.20 besides shows the sum of all the obtained RMSE values for all the years for OLS, LMedS and LTS. LMedS and LTS have much lower sum of RMSE values than of OLS. It states that robust linear regression methods are more accurate (less error) than of OLS in context of inter-calibration. Between LMedS and LTS, LTS has a lower value of sum of RMSE and also in the yearly regression. Thus, LTS is the most accurate estimator among the estimators in the study.

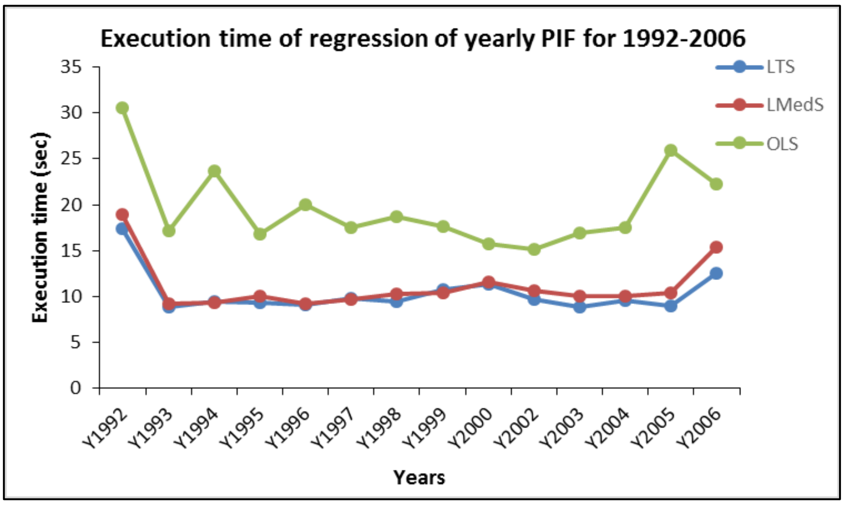


Figure 4.21: Execution time of the regression of PIF of each year with PIF of 2001; Lower computational time denotes lower time complexity in execution.

From Figure 4.21, the execution time for robust methods are observed to be of much low than of OLS. It signifies low process time overhead of the robust estimators. In Figure 4.22 and 4.23, total execution time of all the yearly PIFs and mean execution time for the three estimators in the research are shown respectively. In both the Figures robust methods are observed to execute faster than OLS. LTS is observed to have lowest execution time than OLS and LMedS.

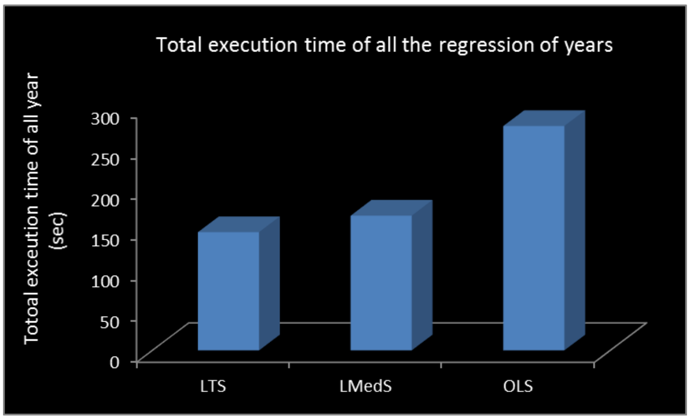


Figure 4.22: Total execution time of the regression of PIF of each year with PIF of 2001

Figure 4.23: Mean execution time of the regression of PIF of each year with PIF of 2001

To select the optimal regression estimator the results of the comparison in terms of R^2, RMSE and execution time are combined. Both the robust estimators are noted to have a high R^2 but a very less difference between them. However, in terms of RMSE and execution time, LTS is the optimal selection as it takes less time to run and also the most accurate one in terms of estimation. Thus, LTS is selected as the optimal regression estimator in the study. The estimated calibrated images are generated using the coefficients retrieved from LTS for every year and SOL has been generated for all the years.

4.2.2. Validation of inter-calibration

The calibrated images are used to generate SOL index to validate the inter-calibration. SOL value for all the years for both calibrated and uncalibrated images are plotted and a trend line is fit in the series. The line with more precise fit is selected as better one. Figure 4.24a and 4.24b shows the SOL plot of uncalibrated and calibrated images respectively.

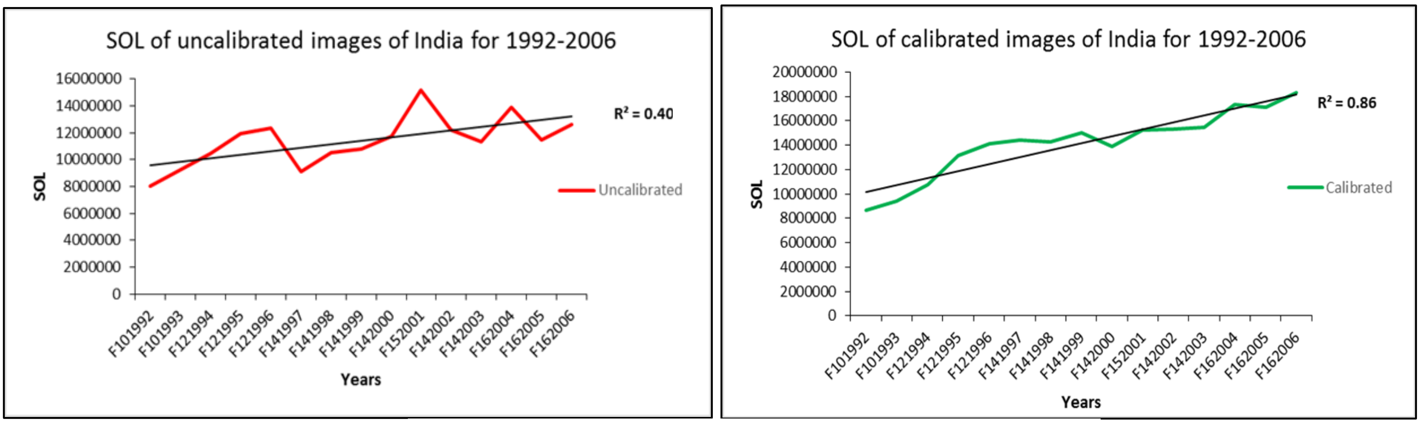


Figure 4.24: Fit of linear trend line in a) SOL of uncalibrated images of 1992-2006 b) SOL of calibrated images of 1992-2006.

The yearly SOL of uncalibrated images in Figure 4.24a shows fluctuation in value with progression in time. However, with progression in time, in case of country like India, the total illumination should be higher with progress in time due to growth in industry and urbanization. The spikes in uncalibrated SOL values are generated due to lack of onboard calibration and multi-sensor acquisition of data. A increasing trend is noticed in the SOL of uncalibrated images but the R^2 is observed to be 0.40 which is considered as low. It denotes low precision in the linear fit of SOLs. The SOL generated from the calibrated images in Figure 4.24b does not suffer from fluctuations. A positive linear trend is noticed in the SOL with a R^2 of 0.86 which states high precision in the fit of line. The SOL values are also observed to increase with time which phenomenon is expected in a developing country like India. The degree of convergence among SOL is also observed to be higher in calibrated SOLs than of uncalibrated SOL. The degree of convergence along with R^2 is mentioned as metrics to validate the inter-calibration. Calibrated SOL is observed to be of both high R^2 value in fit of linear trend and also a higher degree of convergence between the SOL of different years. Thus, the inter-calibration is considered as successful in current research.

In Figure 4.24b, the SOL index in different years do not increase continuously. Drop in value, small though can be seen in some years; thus, increasing linear trend may not be the optimal curve to fit in the SOL. A second order polynomial curve is also fitted in the calibrated SOL data for better fit of the trend line in data. The polynomial fit is shown in Figure 4.25 below.

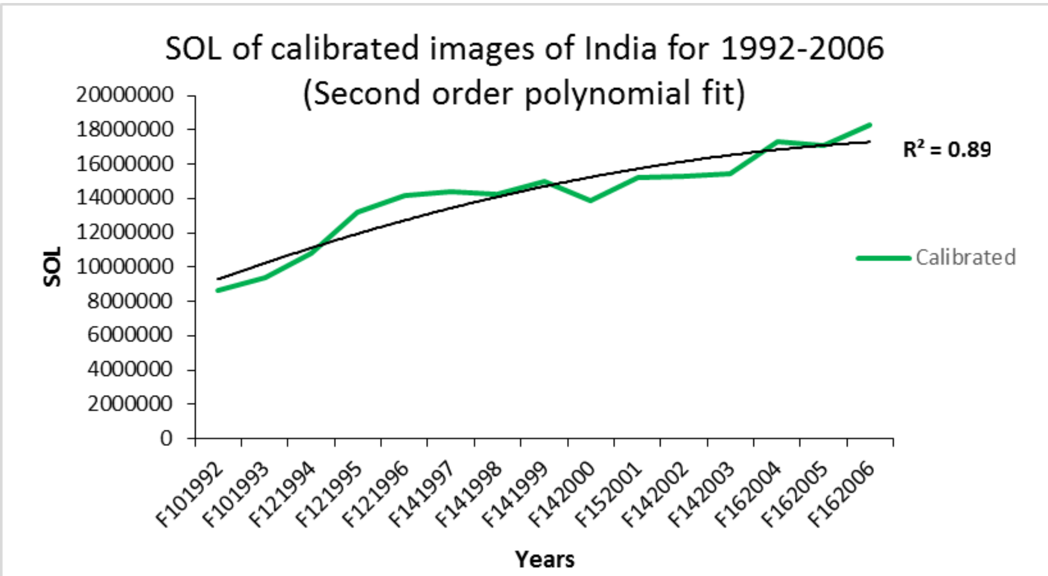


Figure 4.25: Polynomial of second order curve fit in the yearly SOL of calibrated data for 1992-2006

The polynomial curve fits the SOL of calibrated data better than of the linear fit. The fit has a R^2 of 0.89 which is higher than of the linear fit. The drops in SOL value in some years show more of a polynomial nature than of the increasing linear trend.

The DMSP-OLS NTI of 1992-2006 are also calibrated with the coefficients given by Elvidge et al. (2009). These coefficients are obtained by a second order regression with manually selected regions. The SOL values for 1992-2006 are generated from images calibrated with coefficients given by Elvidge et al. (2009) and trend is checked. The trend is comapred with the trend generated from the images calibrated with the coefficients obtained in the research. Figure 4.26a shows linear fit in SOL obtained with Elvidge et al. (2009) given coefficient. and 4.26b shows second order polynomial fit respectively.

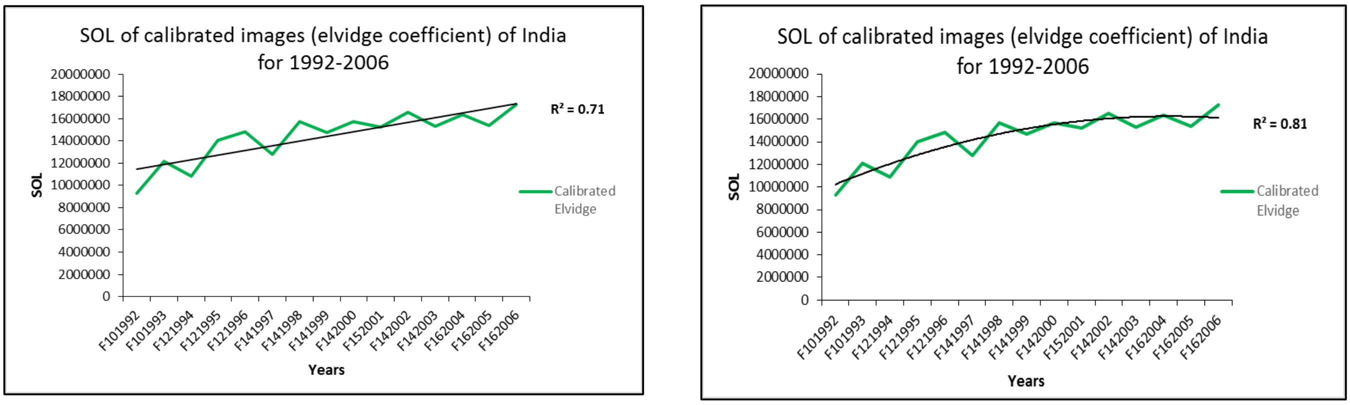


Figure 4.26: SOL generated with Elvidge et al. (2009) given coefficients and a) linear trend line fit, b) second order polynomial fit

The SOL of images calibrated with Elvidge et al. (2009) given coefficient does also have a good degree of convergence than of uncalibrated images. The R^2 is noted to be 0.71 in linear trend fit whereas polynomial fit produce a higher R^2 of 0.81. Higher R^2 denotes better fit of line in the data. The R^2 obtained from the polynomial fit in SOL generated from images calibrated by the established LTS coefficients is 0.89, which is more of than the Elvidge et al., (2009). Thus, the coefficients obtained from current research seems to generate more precise result than of the coefficients given by Elvidge et al., (2009).

The coefficients obtained in the study are used to calibrate the DMSP-OLS NTI of India. The current study also checks the feasibility of obtained coefficients in calibrating NTI of othe countries than India also. As India is a rapidly developing country, the coefficients are also of same manner. Thus, these coefficients are assumed to perform a good inter-calibration in NTI of developing countries. Bangladesh, one of India's neightbouring country is used as such a developing country. Initially the SOL of the DMSP-OLS NTI of Bangladesh for the time period of 1992 to 2006 is generated. These images are calibrated with the obtained coefficients in the study. Figure 4.27a and 4.27b show the SOL of uncalibrated and calibrated SOL of Bangladesh respectively.

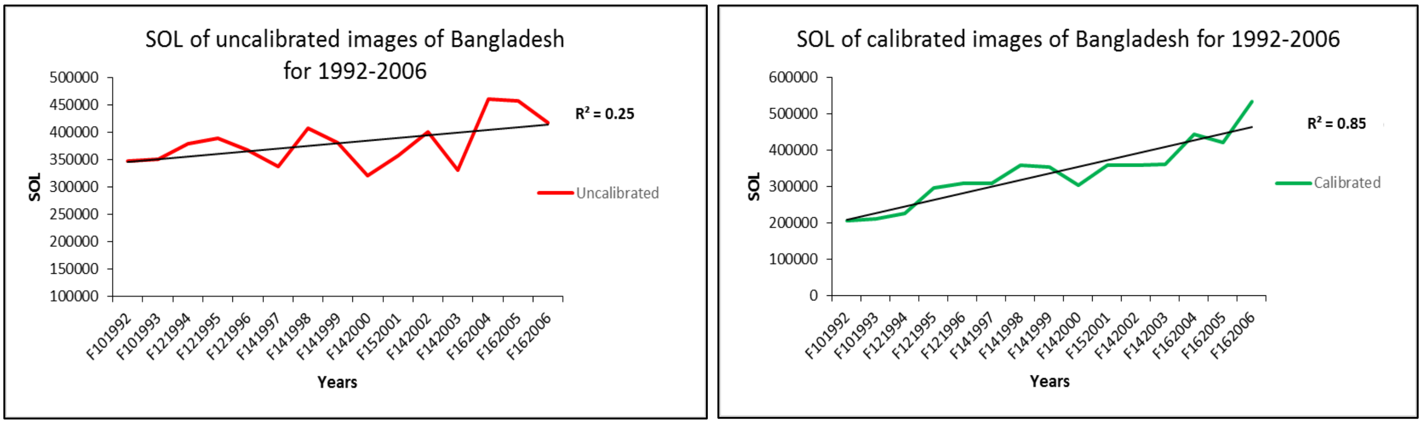


Figure 4.27: SOL of Bangladesh for 1992-2006 for a) uncalibrated image, b) calibrated image; linear fit is performed to check the improvement after calibration

The yearly SOL of uncalibrated images of Bangladesh shows fluctuation with progression in time. These spikes in SOL values are generated due to lack of onboard calibration and multi-sensor acquisition of data. A positive trend is noticed in the SOL of uncalibrated images but the R^2 is observed to be 0.25 which is quite low. It denotes low precision in the linear fit of uncalibrated SOLs. For the SOL generated from the calibrated images in Figure 4.27b does not suffer from steep fluctuations. A strong positive linear trend is noticed in the SOL with a R2 of 0.85 which states high precision in the fit of line. The SOL values are also observed to increase with time. The degree of convergence is also observed to be higher in calibrated SOLs than of uncalibrated SOL alike the SOL of India. In case of Bangladesh, calibrated SOL is observed to be of both high R^2 value in fit of linear trend and also a higher degree of convergence. Thus, the inter-calibration is considered as successful . It refers to the assumption that the obtained coefficients in the research can be used to calibrated the NTI of developing country where luminosity is observed to change at a increasing rate with time.

4.3. Exploration of relationship between DMSP-OLS NTI and socio-economic parameters

It is observed from the calibrated SOL values of different yearly DMSP-OLS NTI composites of India that the night-time light has gradually increased over a period of time at a national level. As mentioned in Section 3.3.1, version 4 "stable light" products have been used. Hence the pixels mainly represent the stable lights emitted from the surface. Such stable lights are mostly emitted from human settlements; therefore it can be assumed that the trend in change in night-time light can also be correlated with socio-economic or demographic indices/parameters. This section describes the results of the efforts made in research to find the relationship between change in DMSP-OLS NTI and change in the socio-economic parameters of India for the time period of 1992 to 2006. Gross Domestic Product (GDP) and Urban Population (UP) are used as the socio-economic parameters in the study. Section 4.3.1 states the relationship between DMSP-OLS NTI and GDP and section 4.3.2 states the relationship between DMSP-OLS NTI and UP.

4.3.1. Relationship between DMSP-OLS NTI and GDP of India

Yearly GDP data for the time period of 1992 to 2006 is used in the study. SOL values for the same time period is plotted with the GDP data. Linear and second order polynomial regression with a significance level of 0.05 are performed between the SOL and GDP data for the time period of 1992 to 2006 keeping SOL as the independent variable. The fitted regression line is used to check the trend in relationship. The quality of relationship is determined with coefficient of determination (R^2). Better the R^2 is, better the relationship is to be assumed. Regression is plotted for SOL of both uncalibrated and calibrated images of India. It gives the effect of inter-calibration in context of relationship between DMSP-OLS NTI and socio-economic parameters (GDP). Figure 4.28 shows the linear relationship between SOL and GDP and Figure 4.29 shows the second order polynomial relationship between SOL and GDP, both for India for time period of 1992 to 2006.

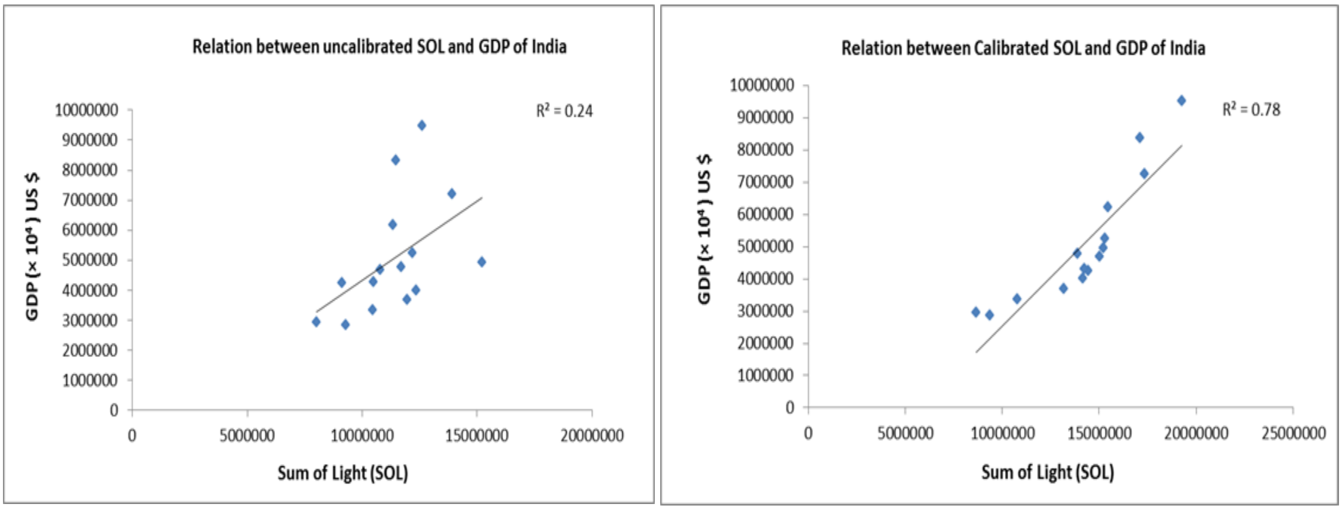


Figure 4.28: Linear relation between SOL and GDP for 1992-2006 of a) uncalibrated data, b) calibrated data.

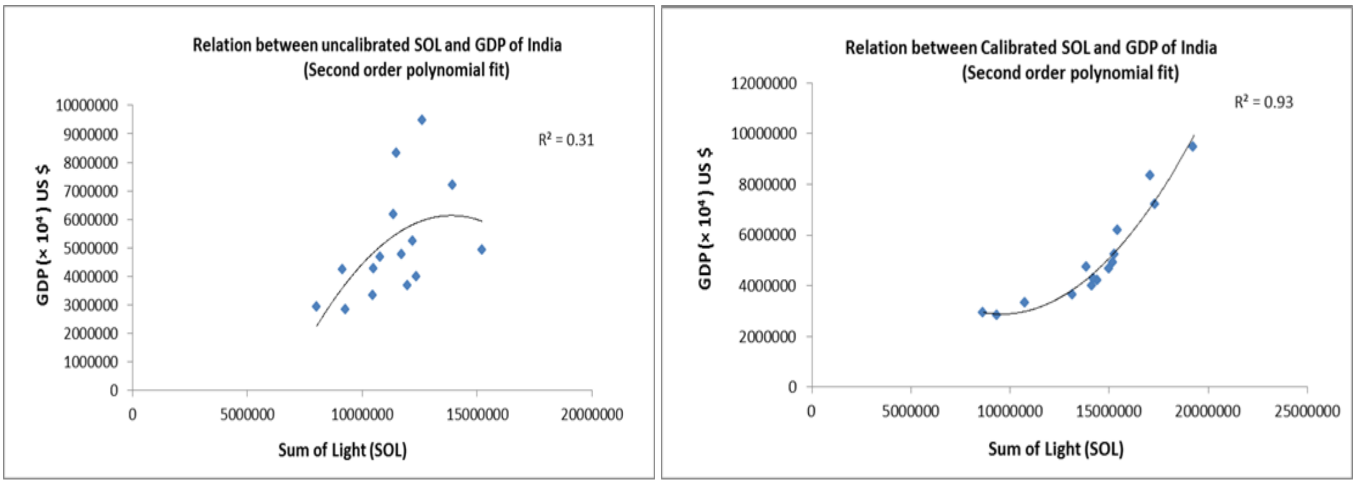


Figure 4.29: Polynomial relation between SOL and GDP for 1992-2006 of a) uncalibrated data, b) calibrated data.

From Figure 4.28a it is seen that SOL of uncalibrated images of different years do have a linear relationship with the yearly GDP data but R^2 is observed to be 0.24 in uncalibrated SOL which is considered low and subsequently a poor fit in the data. In Figure 4.28b, SOL of calibrated image does also have a linear relationship with the yearly GDP data and the R^2 is observed to be 0.78 which is considered as good. Both the Figures (Figure 4.28a and 4.28b) show positive increasing trend. In Figure 4.29a and 4.29b, the polynomial model is observed to fits better than linear fit. In both uncalibrated and calibrated SOL the fit is precise than of the linear one. However, in uncalibrated SOL, a decreasing trend is found in estimating GDP which is completely redundant in case of India as India's GDP is observed to increase at a very high rate. In calibrated SOL a much better increasing trend is found. The GDP of India can be estimated from change in SOL over time using the polynomial relationship among them. Thus, the inter-calibration is observed to increase the R^2 and the goodness of fit in the plot respectively because post calibration the brightness fluctuations are reduced among the SOL values of different years generating an increasing trend in data.

4.3.2. Relationship between DMSP-OLS NTI and Urban Population (UP) of India

Yearly UP data for the time period of 1992 to 2006 is used in the study. Initially UP data for the time period of 1996-2006 was available. Yearly Urban Population data for the years 1992 to 1995 are generated using the following exponential equation:

$$(\boldsymbol{Population})_{\boldsymbol{targe\_year}} = (\boldsymbol{Population})_{\boldsymbol{reference\_year}} \times \boldsymbol{e}^{\boldsymbol{\propto * t}} \qquad (6)$$

Where, *(Population)* $_{target\_year}$ is the population of the target year that will be generated, *(Population)* $_{reference\_year}$ is the population of the reference year to whose respect population of the target year will be computed, e is the Euler's number, α is the coefficient of population (needs to be calculated) and t is time difference of years between target year and reference year. To extrapolate the urban population of 1992-1995 at first α is calculated. To calculate α, base year is selected as 1996 and target year is set as 2006. t is set as (2006-1996)=10 in the study. Obtained α value is 0.02662 for UP of India. The UP for the years 1992 to 1995 are thus calculated with the equation 6.

On generation of UP for 1992 to 2006, SOL values for the same time period is plotted with the UP data. Linear and second order polynomial regression with a significance level of 0.05 are performed between the SOL and UP data for the time period of 1992 to 2006 keeping SOL as the independent variable. The fitted regression line is used to check the relationship. The quality of relationship is determined with coefficient of determination (R^2). Better the R^2 is, better the relationship is to be assumed. Regression is plotted for SOL of both uncalibrated and calibrated images of India. It shows the effect of inter-calibration in context of relationship between DMSP-OLS NTI and socio-economic parameters (UP). Figure 4.30 shows the linear relationship between SOL and UP and Figure 4.31 shows the polynomial relationship between calibrated SOL and UP, both for India for time period of 1992 to 2006.

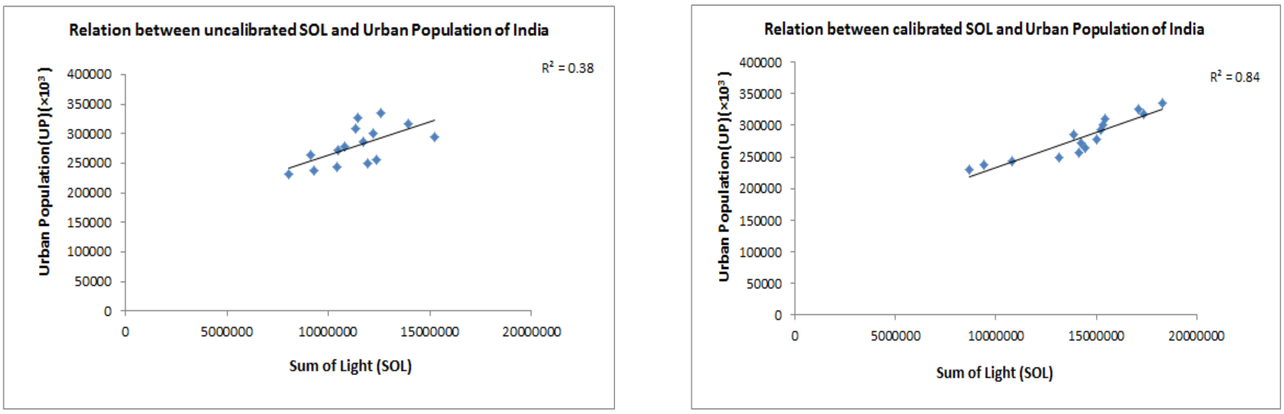


Figure 4.30: Linear relation between SOL and UP for 1992-2006 of a) uncalibrated data, b) calibrated data.

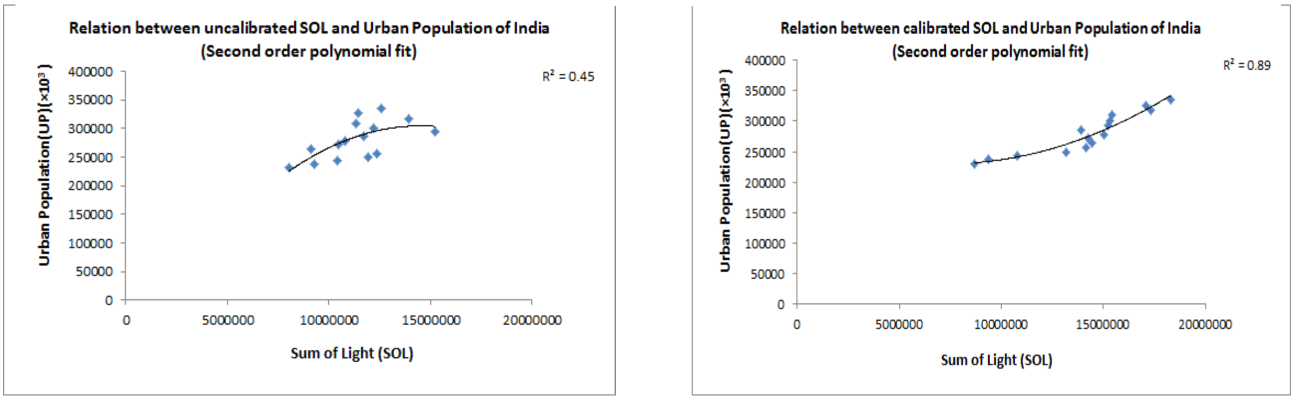


Figure 4.31: Polynomial relation between SOL and UP for 1992-2006 of a) uncalibrated data, b) calibrated data

From Figure 4.30a it is seen that SOL of uncalibrated images different years do have a linear relationship with the yearly UP data but R^2 is observed to be 0.38 which is considered low and subsequently a poor fit in the data. In Figure 4.30b SOL of calibrated image does also have a linear relationship with the yearly UP data and the R^2 is observed to be 0.84 which is considered as quite good. Both the Figures show positive increasing trend. These results resemble the results of Figure 4.28and 4.28b respectively. It can also be concluded in case of UP that the inter-calibration is observed to increase the R^2 and the goodness of fit in the plot respectively because post calibration the fluctuations are reduced among the SOL values of different years generating more of a linear trend in data. In polynomial fit, a fit with better precision is noticed in Figure 4.31a and 4.31b. The R^2 is higher than of the obtained in linear fit. However, in uncalibrated SOL, a decreasing tend can be seen which is redundant. India's urban population has increased sharply over time which is actually shown in Figure 4.31b. An increasing trend is noticed for polynomial fit in calibrated SOL and UP of India.

From studying the relationship between SOL and socio-economic parameters, it can be concluded that change in DMSP-OLS NTI does have relationship with the change in socio-economic parameters over time. In case of India, linear and nonlinear polynomial relationships are found in both GDP and UP. However, the degree of relationship is better and relevant in case of SOL of calibrated images than SOL of uncalibrated ones.. It can also be inferred that inter-calibration also plays a vital role in establishing relationship between DMSP-OLS NTI and socio-economic parameters over time.

4.4. Discussion

This section converses about the quality analysis of different inter-calibration techniques and the method taken to extract the PIF for the inter-calibration. In this study, an automatic method to extract PIF is developed where no prior demographic knowledge is necessary. With the extracted PIFs a comparative implementation of linear ordinary and robust regression is performed in order to reduce the effect of outliers. The degree of relationship between socio-economic parameters (GDP and UP) is also found in the study.

DMSP-OLS NTI for the time period of 1992 to 2006 are taken in the study. Initially the low valued DN ($DN<5$) and saturated pixels ($DN=63$) are removed from all the yearly composites. It removes the blooming and saturation problem respectively. To implement inter-calibration, regions where luminosity does not change drastically are found. Such regions are called PIF. PIFs are set to be spatially homogeneous and of low local spatial variance. Getis statistic (Gi^*) is implemented in a window to select spatially homogeneous areas. Pixels with $Gi^* > 1.645$ are selected as the spatially homogeneous areas as positive output of Gi^* denotes bright clusters and a value more than of 1.645 denotes clustering of pixels at a significance level of 0.10 which restricts random clustering by only 10%. To find regions with low local variance, CV is implemented also in a fixed window. A threshold of CV is computes within which if the CV of a pixel remains that can be considered as of low local spatial variability. A variable threshold of $CV<3\%$, $CV<5\%$ and $CV<10\%$ is implemented to select optimal threshold of CV. To find the threshold, CV output of every threshold is merged with $Gi^*>1.645$ output pixels to extract PIF with variable CV thresholds. These PIFs are used to mask the corresponding pixels values from DN thresholded image of 1992, 2001 and 2006. The PIF with $CV<10\%$ is observed to have highest number of observations and also a good distribution of points over the whole range of pixel values in the PIF. The range is observed to vary from 10 to 63 whereas in case of $CV<3\%$, the range is limited to 55 to 62 only. $CV<5\%$ has a big range of values in regression among PIFs but the distribution of points are clustered instead of continuous. Thus, $CV<10\%$ does seem the optimum threshold for extraction of PIF. On generation of optimal CV threshold, optimal window size to implement Gi^* and CV is found. PIFs are generated with a threshold of $Gi^*>1.645$ and $CV<10\%$ for both 3×3 and 5×5 windows. The better window size is also tested with implementing OLS between PIF value of 1992, 2001 and 2006 for both window sizes. It is seen that for 5×5 window the number of observations are quite low compared to of 3×3 window and alike the scenario of $CV<5\%$, the distribution of points are more cluster oriented than of continuous spread over the whole range. Thus it is inferred that with increase in window size, the number of points decrease and produce redundancy in regression. By comparing all the mentioned criterions, a final threshold and window size is defined to extract the PIF from the images. The threshold is as following:

1. Gi* > 1.645 (3×3 window)
2. CV < 10% (3×3 window)

On generation of PIF for all the years, inter-calibration is performed. Inter-calibration is performed by implementing regression among the years to estimate the calibrated DN values of various years. The PIF of 2001 is used as reference image to all the other images as it is situated in the middle of the whole time period of the study and also it consists of highest number of unique lit pixels among the yearly composites. OLS is implemented along with two robust regression techniques; LMedS and LTS. Robust regression is implemented to remove the effect of outliers in the linear fit. The optimal regression technique is selected in terms of precision, RMSE and execution time. All the yearly composites PIF in the study period are regressed with PIF of 2001. From comparing the precision, RMSE and execution time, LTS is found with lowest RMSE and execution time. The high R^2 observed in LTS also signifies high precision in the linear fit of the model. LMedS and LTS do have a high breakdown range which enables them to fit the trend line irrespective of outliers up to a limit. In OLS the breakdown point is noticed to be of 0% whereas in LMedS the breakdown point is of 50%, much higher than of OLS. In LTS, the breakdown point is very little bit less than of LMedS. Such high breakdown points do play a vital role in reducing the contamination of data by outliers. Higher the breakdown point is, more the fit of regression line has less contamination from outliers in data. Thus, LTS is selected as the optimal regression technique for the calibration.

On generation of the calibration coefficients, all the calibrated images are generated and SOL index is computed for both uncalibrated and calibrated images. Positive increasing trend is seen for both calibrated and uncalibrated SOL but in uncalibrated the R^2 is observed to be of 0.40 which signifies poor precision in the fit. Whereas in calibrated SOL a strong positive linear trend is observed with a R^2 of 0.86. However, SOL is also observed to be less than the previous years in a couple of occasions between 1992 and 2006. Thus, a non-linear fit; polynomial second order fit is also tested and it produced better result in curve fitting in data. Polynomial fit does also have a higher R^2 value than of linear fit in calibration. Calibrated SOL also has a higher degree of convergence among the yearly SOL. It infers that after calibration the fluctuations in the brightness of data due to variable gain acquisition over different satellites are resolved. The results of the calibrated SOL is also compared with SOL generated with images calibrated with coefficients provided by Elvidge et al. (2009). SOL of differnet years from these calibrated images also reduce the fluctuation between the yearly SOL and a positive lnear trend is also shown. However, the R^2 of the linear fit is observed to be of 0.71 and 0.81 for the polynomial fit whereas the R^2 of the fit achieved in SOL images calibrated with coefficients obtained from the study fo linea fit is 0.86 and for polynomial it is observed to be 0.89; in both the fit precision is more than of Elvidge et al. (2009) . Thus, it can be inferred that coefficients obtained from the study are better than of Elvidge et al. (2009).

Yearly GDP and UP data of India for the relevant time period of the research is also plotted with both uncalibrated and calibrated SOL. A positive linear trend is observed that in both the parameters with the change in SOL. However, in case of uncalibrated SOL, there is again a low precision is noticed due to fluctuating nature of uncalibrated SOL. In case of calibrated SOL, R^2 of 0.78 and 0.84 are observed for linear fit with GDP and UP respectively. For polynomial fit, R^2 of 0.93 and 0.89 are observed in estimation of GDP and UP respectively. Between linear and polynomial relationship, polynomial relationship estimates the change in GDP and UP in more of an actual scenario. It shows a steep increase in estimating GDP and UP which do resemble the real life trend of change of GDP and UP for India. Besides, the polynomial fit is found to be more precise which indicates good fit of the curve with the data. Thus it can be inferred that a strong second order polynomial relationship exists between the calibrated images of India and socio-economic parameters taken in the study.

5. CONCLUSIONS AND RECOMMENDATIONS

In this chapter, the conclusions attained through a detailed study of the existing and the developed method is presented. Section 5.1 presents the conclusions on the basis of the research objectives and questions. Section 5.2 shows the answers to the research questions and section 5.3 presents the recommendations for future work.

5.1. Conclusions

The inter-calibration is a relative radiometric normalization technique to create time series compatible datasets. In order to perform inter-calibration on DMSP-OLS NTI regression is plotted to generate coefficients of calibration. The regression is implemented on regions that have stable light intensity over time. In this research, a method has been developed to extract PIF without any prior need of demographic and surface knowledge of the study area. Spatial homogeneity and low local variability are set as the criteria to select PIF. Gi*> 1.645 and CV <10% thresholds are set through a comparative study in the research. Regions that have a pixels consisting Gi* >1.645 and CV<10% for the whole time of study are considered as PIF.

Ordinary and Robust linear Regression are implemented and compared to select optimal estimator for inter-calibration. Robust regression is introduced to check the improvement of outlier handling than of Ordinary linear regression. By comparing the results of regression on basis of outlier removal, precision, accuracy and execution time LTS is selected as the optimal regression technique.

On calibration of DMSP-OLS NTI with coefficients obtained from implementing LTS on all years, SOL index of corresponding years are generated for both calibrated and uncalibrated images. Socio-economic parameters like GDP and UP are checked for any relation with DMSP-OLS NTI. GDP and UP each shows positive linear trend associated with high precision in the fit for the calibrated SOL.

As conclusion of the research, the method to extract PIF without prior knowledge of study area seems successful and the inter-calibration performed with the selected optimal estimator also produces successful calibrated products that have a precise second order polynomial relationship with socio-economic parameters/indices.

5.2. Answers to the research questions

- **What are the spatio-temporal assessment criteria for selecting PIFs in DMSP-OLS nighttime image for performing inter-calibration?**

  Answer: PIF is a region where light intensity does not change drastically through time. Such region in DMSP-OLS NTI consists of pixels that have a less variation in magnitude locally and also a low local spatial variability. Less difference in magnitude of pixels in PIF signifies that the pixels are of close values and have high spatial association in them (spatial homogeneity). Local spatial variation is also noted to be less for the PIF. PIF should also be spatially homogeneous and of low local spatial variability in all the yearly composites. Apart from local spatial variability, overall spatial variability of all the pixels in the extracted PIF should also not vary randomly with drastic changes in magnitude. Spatial homogeneity is checked from Gi* and local spatial variability is checked from CV. These conditions are mentioned as criteria below in numeric bullets.

1. Luminosity of the pixels inside the PIF should remain stable throughout the years.
2. Pixels inside the PIFs should be bright.
3. PIFs should not contain saturated pixels (DN = 63).
4. Pixels inside the PIF should be locally spatially homogeneous in nature. Gi* > 1.645 is the minimal threshold to select statistically significant (α=0.10, α is significance level) bright and locally spatially homogeneous pixel clusters. The optimal window size to calculate Gi* is 3×3 for DMSP-OLS NTI of India of 1992-2006.
5. Pixels should have low local spatial variability. CV <10% is the threshold to select spatially low variable regions. The optimal window size to calculate CV is 3×3 for DMSP-OLS NTI of India of 1992-2006.
6. The overall variability of all the pixels in PIFs should not vary drastically in magnitude with time.

- **Which is the optimal parameter estimation method associated with linear regression for inter-calibration of PIFs?**

  Answer: Research consists of three parameter estimation techniques; OLS, LMedS and LTS. Optimality of these estimators is checked by comparing R^2, RMSE and execution time. LTS is observed to have high R^2 value, low RMSE and less execution time than the rest of the estimators in context of inter-calibration of PIFs of DMSP-OLS NTI. Hence, LTS is the optimal parameter estimator associated with linear regression for inter-calibration of PIF.

- **What is the improvement in DMSP-OLS night-time images after inter-calibration?**

  Answer: The inter-calibration is observed to increase the R^2 and the goodness of fit in the plot of SOL because post calibration the total brightness fluctuations are reduced among the SOL values of different years generating an increasing trend line in data.A positive trend in change in SOL throughout the years is noticed after inter-calibration. For India the positive trend is expected due to its bloom in economy and urbanization over the last two decades which eventually leads to increase in man-made lighting.

- **What is the relationship between inter-calibrated images and socio-economic variables like GDP and UP?**

Answer: DMSP-OLS NTI represents the man-made lighting on the surface at night. Man-made lighting is expected to grow with increase in population, boost in economy or increase in urbanization. So it is obvious that NTI can be an estimator of the indices that measure the social development like GDP and UP. SOL quantifies the total brightness of a whole year composite. So SOL is checked for relationship with GDP and UP. Uncalibrated SOL does have a very relation with both GDP and UP whereas, calibrated SOL has Polynomial second order relationship with a high precision. It reflects strong positive relation between calibrated NTI and socio-economic parameters.

5.3. Recommendations

There are some additional steps that can help to revisit or update the inter-calibration methodology in the research. Some of them are as follows:

- ✓ Instead of removing DN=63 to remove saturation problem radiance calibrated products can be generated for each year composite to rescale the lamination of saturated pixels.
- ✓ To find the optimal window size for the LISA, variogram can also be used instead of regression.
- ✓ Temporal stability of each point of the PIF can also be tested to add the low temporal variability in PIF.
- ✓ Regression modelling of each year can be run as parallel processes to speed up the computation.
- ✓ Calibrated DMSP-OLS NTI can be a used to study the loss of property or wealth from the wars. It can help governmental agencies to estimate the rebuild cost of the affected zone.

REFERENCES

Amaral, S., Monteiro, a. M. V., Camara, G., & Quintanilha, J. a. (2006). DMSP/OLS night-time light imagery for urban population estimates in the Brazilian Amazon. *International Journal of Remote Sensing*, *27*(October 2014), 855–870. https://doi.org/10.1080/01431160500181861

Anselin, L. (1995). Local indicators of spatial association — LISA. *Geographical Analysis*, *27*(2), 93–115. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x

Bailey, T. C. (1994). *A review of statistical spatial analysis in geographical information systems*. *Spatial Analysis and GIS*. https://doi.org/10.1017/CBO9781107415324.004

Bannari, A., Omari, K., Teillet, P. M., & Fedosejevs, G. (2005). Potential of getis statistics to characterize the radiometric uniformity and stability of test sites used for the calibration of earth observation sensors. *IEEE Transactions on Geoscience and Remote Sensing*, *43*(12), 2918–2925. https://doi.org/10.1109/TGRS.2005.857913

Bannari, A., Teillet, P. M., & Richardson, G. (1999). Nécessità de l'étalonnage radiométrique et standardisation des images numériques de télédétection. *Canadian Journal of Remote Sensing*, *25:1*, 45–59. https://doi.org/10.1080/07038992.1999.10855262

Chen, X., & Nordhaus, W. D. (2011). Using luminosity data as a proxy for economic statistics. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(21), 8589–94. https://doi.org/10.1073/pnas.1017031108

Croft, T. A. (1973). Burning Waste Gas in Oil Fields. *Nature*, *245*(5425), 375–376. https://doi.org/10.1038/245375a0

de Vries, C., Danaher, T., Denham, R., Scarth, P., & Phinn, S. (2007). An operational radiometric calibration procedure for the Landsat sensors based on pseudo-invariant target sites. *Remote Sensing of Environment*, *107*(3), 414–429. https://doi.org/10.1016/j.rse.2006.09.019

Derksen, C., Wulder, M., LeDrew, E., & Goodison, B. (1998). Associations between spatially autocorrelated patterns of SSM/I-derived prairie snow cover and atmospheric circulation. *Hydrological Processes*, *12*(15), 2307–2316. https://doi.org/10.1002/(SICI)1099-1085(199812)12:15<2307::AID-HYP798>3.0.CO;2-0

Doll, C. N. H. (2008). *CIESIN Thematic Guide to Night-time Light Remote Sensing and its Applications*. Palisades, NY: Center for International Earth Science Information Network of Columbia University. Retrieved from http://sedac.ciesin.columbia.edu/tg/

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

Elvidge, C. D., Imhoff, M. L., Baugh, K. E., Hobson, V. R., Nelson, I., Safran, J., … Tuttle, B. T. (2001). Night-time lights of the world: 1994-1995. *ISPRS Journal of Photogrammetry and Remote Sensing*, *56*(2), 81–99. https://doi.org/10.1016/S0924-2716(01)00040-5

Elvidge, C. D., Sutton, P. C., Ghosh, T., Tuttle, B. T., Baugh, K. E., Bhaduri, B., & Bright, E. (2009). A global poverty map derived from satellite data. *Computers & Geosciences*, *8*(35), 1652–1660. https://doi.org/10.1016/j.cageo.2009.01.009

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

Getis, A., & Ord, J. K. (1992). The Analysis of Spatial Association. *Geographical Analysis*, *24*(3), 189–206. https://doi.org/10.1111/j.1538-4632.1992.tb00261.x

GDAL. 2017. GDAL - Geospatial Data Abstraction Library: Version 2.1.3, Open Source Geospatial Foundation, http://gdal.osgeo.org

Hamm, N. A. S., Milton, E. J., & Odongo, V. O. (2011). Indicators of spatial autocorrelation for identification of calibration targets for remote sensing. In C.C. Fonte, L. M. . Gonçalves, & G. Gonçalves (Eds.), *Proceedings of the 7th international symposium on spatial data quality* (pp. 51–56). Coimbra, Portugal: Instituto de Engenharia de Sistemas e Compu- tadores de Coimbra (INESC Coimbra).

Hara, M., Okada, S., Moriyama, T., Shigehara, K., & Sugimori, Y. (2004). Progress for Stable Artificial Lights Distribution Extraction Accuracy and Estimation of Electric Power Consumption By Means of Dmsp / Ols Nighttime Imagery. *International Journal of Remote Sensing*, *I*(1), 31–42.

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. https://doi.org/10.3390/rs6086844

Kneubhuler, M., Schaepman, M., & Thome, K. (2006). Long-term vicarious calibration efforts of MERIS at railroad valley playa (NV) - An update. *European Space Agency, (Special Publication) ESA SP*, (615). https://doi.org/10.5167/uzh-97036

Kulkarni, R., Haynes, K., Stough, R. R., & Riggle, J. (2011). Revisiting Night Lights as Proxy for Economic Growth: A Multi-Year Light Based Growth Indicator (LGBI) for China, India and the U.S. *GMU School of Public Policy Research Paper*, (March), 1–21. https://doi.org/10.2139/ssrn.1777546

Letu, H., Hara, M., Tana, G., & Nishio, F. (2012). A saturated light correction method for DMSP/OLS nighttime satellite imagery. *IEEE Transactions on Geoscience and Remote Sensing*, *50*(2), 389–396. https://doi.org/10.1109/TGRS.2011.2178031

Li, C., Ye, J., Li, S., Chen, G., & Xiong, H. (2016). Study on radiometric intercalibration methods for DMSP-OLS night-time light imagery. *International Journal of Remote Sensing*, *37*(16), 3675–3695. https://doi.org/10.1080/01431161.2016.1201232

Li, X., Chen, X., Zhao, Y., Xu, J., Chen, F., & Li, H. (2013). Automatic intercalibration of night-time light imagery using robust regression. *Remote Sensing Letters*, *4*(June), 45–54. https://doi.org/10.1080/2150704X.2012.687471

Liu, Z., He, C., Zhang, Q., Huang, Q., & Yang, Y. (2012). Extracting the dynamics of urban expansion in China using DMSP-OLS nighttime light data from 1992 to 2008. *Landscape and Urban Planning*, *106*(1), 62–72. https://doi.org/10.1016/j.landurbplan.2012.02.013

Lo, C. P. (2001). Modeling the Population of China Using DMSP Operational Linescan System Nighttime Data. *Photogrammetric Engineering & Remote Sensing*, *67*(9), 1037–1047. Retrieved from http://info.asprs.org/publications/pers/2001journal/september/2001_sep_1037-1047.pdf

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

Massart, D. L., Kaufman, L., Rousseeuw, P. J., & Leroy, A. (1986). Least median of squares: a robust method for outlier and model error detection in regression and calibration. *Analytica Chimica Acta*, *187*(C), 171–179. https://doi.org/10.1016/S0003-2670(00)82910-4

Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silverman, R., & Wu, A. Y. (2014). On the least trimmed squares estimator. *Algorithmica*, *69*(1), 148–183. https://doi.org/10.1007/s00453-012-9721-8

Odongo, V. O., Hamm, N. A. S., & Milton, E. J. (2014). Spatio-temporal assessment of tuz golu, Turkey as a potential radiometric vicarious calibration site. *Remote Sensing*, *6*(3), 2494–2513. https://doi.org/10.3390/rs6032494

Python Software Foundation. Python Language Reference, version 2.7. Available at http://www.python.org

R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, http://www.R-project.org.

Robin Willson. (2011). RTWTools. Retrieved August 23, 2016, from http://www.rtwtools.rtwilson.com/

Rondeaux, G., Steven, M. D., Clark, J. A., & Mackay, G. (1998). La Crau: A European test site for remote sensing validation. *International Journal of Remote Sensing*, *19*(14), 2775–2788. https://doi.org/10.1080/014311698214505

Rousseeuw, P. (1984). Least Median of Squares Regression. *Journal of the American Statistical Association*, *79*(388), 871–880. https://doi.org/10.1080/01621459.1984.10477105

Scott, K. P., Thome, K. J., & Brownlee, M. R. (1996). Evaluation of Railroad Valley playa for use in vicarious calibration. In B. Huberty, J. B. Lurie, J. A. Caylor, P. Coppin, & P. C. Robert (Eds.), *SPIE* (Vol. 2818, p. 158). https://doi.org/10.1117/12.256090

Stéfan van der Walt, S. Chris Colbert and Gaël Varoquaux. *The NumPy Array: A Structure for Efficient Numerical Computation*, Computing in Science & Engineering, **13**, 22-30 (2011), doi:10.1109/MCSE.2011.37

Stuart, C. (2011). Robust Regression. *Technical Report*. https://doi.org/10.1007/978-3-642-33765-9_44

Sutton, P. (1997). Modeling Population Density With Night-Time Satellite Imagery and Gis. *Comput., Environ. and Urban Systems*, *214*(3), 227–244. https://doi.org/10.1016/S0198-9715(97)01005-3

Sutton, P., Elvidge, C., & Ghosh, T. (2007). Estimation of gross domestic product at sub-national scales using nighttime satellite imagery. *International Journal of Ecological Economics & Statistics*, *8*(S07), 5–21. Retrievedfromhttp://urizengeography.nsm.du.edu/~psutton/AAA_Sutton_WebPage/Sutton/Publications/Sut_Pub_13.pdf

Sutton, P., Roberts, D., Elvidge, C., & Baugh, K. (2001). Census from Heaven: an estimate of the global human population using night-time satellite imagery. *Int. J. Remote Sensing*, *22*(16), 3061–3076. https://doi.org/10.1080/01431160010007015

Wei, Y., Liu, H., Song, W., Yu, B., & Xiu, C. (2014). Normalization of time series DMSP-OLS nighttime light images for urban growth analysis with Pseudo Invariant Features. *Landscape and Urban Planning*, *128*, 1–13. https://doi.org/10.1016/j.landurbplan.2014.04.015

Wu, J., He, S., Peng, J., Li, W., & Zhong, X. (2013). Intercalibration of DMSP-OLS night-time light data by the invariant region method. *International Journal of Remote Sensing*, *34*(20), 7356–7368. https://doi.org/10.1080/01431161.2013.820365

Zhang, Q., Schaaf, C., & Seto, K. C. (2013). The Vegetation adjusted NTL Urban Index: A new approach to reduce saturation and increase variation in nighttime luminosity. *Remote Sensing of Environment*, *129*, 32–41. https://doi.org/10.1016/j.rse.2012.10.022

Ziskin, D., Baugh, K., Hsu, F., Ghosh, T., & Elvidge, C. (2010). Methods used for the 2006 radiance lights. *Proceedings Of The Asia-Pacific Advanced Network*, *30*, 131–142. Retrieved from http://journals.sfu.ca/apan/index.php/apan/article/view/88

APPENDIX A

Ordinary regression parameter estimator:

Ordinary least square (OLS): OLS is a method to estimate the unknown parameters in a linear regression model. It works by minimizing the sum of square of the differences between observed values in the dataset and the predicted values (residual) by the linear function (Stuart, 2011). Mathematically it can be expressed in the following:

$$\text{minimize} \left(\sum_{i=1}^{n} r_i^2 \right) \tag{7}$$

where *i* is the number of observations in data and r_i is the residual.

The linear regression in the research context can be written as,

$$(DN)_{Li} = C_1\,(DN)_{Ri} + C_0 \tag{8}$$

Where $(DN)_{Li}$ denotes the *i*th pixel value in Uncalibrated image Y ; $(DN)_{Ri}$ denotes the ith pixel in chosen reference image X ; *i* varies from the value *n* denotes the number of pixels in image X or Y; C_1 denotes the gain value and C_0 denotes the bias value.

Robust regression parameter estimators:

Least trimmed square(LTS): LTS is quite similar to OLS method but it does not take all the residuals obtained from the fit. It only takes a finite amount of number of residuals into account. Thus the extreme values(outliers) does not affect the model fit in data. It is mathematically mentioned below in (9):

$$\text{minimize} \left(\sum_{i=1}^{h} (r_i^2)_{i:n} \right) \tag{9}$$

Where $(r_i^2)_{i:n} \leq \ldots \leq (r_i^2)_{n:n}$; are ordered squared residuals and $h=[n/2]+1$, n is the number of observations.

Least median of squares(LMedS): The LMedS method is also a random parameter estimation algorithm that estimates the parameters of regression by minimizing the median of the square of residuals obtained (Rousseeuw, 1984). LMedS is mathematically written below in (10):

$$\text{minimize} \left(median\ r_i^2 \right) \tag{10}$$

Where, r is the obtained residual from the fit of model in data.

Massart et al. (1986) showed in his work that LMedS performs poorly in presence of Gaussian noise. A tolerance value was introduced by Li et al. (2016) to compensate this problem with the standard deviation to make it robust. The robust standard deviation is mentioned below in (11)

$$\boldsymbol{\sigma = 1.4826\left[1 + \frac{5}{(n-t)}\right]\sqrt{M_J}} \tag{11}$$

Where M_J is the minimal median, 1.4826 is a constant used to obtain similar efficiency as least squares should obtain in presence of Gaussian noise. t is the essential observation number and n is the total number of observations.

The weight function is as following in (12)

$$\boldsymbol{p_i = \begin{cases} 1 & if\ r_i^2 \leq (2.5\sigma)^2 \\ 0 & otherwise \end{cases}} \tag{12}$$

Where p_i is the weight to assign and r is the obtained residual from the model. $p_i = 0$ refers to an outlier.

Accuracy Evaluation:

Coefficient of determination (R^2): R^2 is commonly used to measure precession and it can be obtained in the research wok as following (13):

$$\boldsymbol{R^2 = 1 - \frac{SS_E}{SS_r}} \tag{13}$$

Where,

$$\boldsymbol{SS_r = \sum_i (y_i - \bar{y})^2\ , SS_E = \sum_i (y_i - \hat{y}_i)^2} \tag{14}$$

Where, y_i , $\hat{y}_i$ are the original data values and estimated values respectively, That is, SS_r is the total sum of squares and SS_E is the sum of squared errors.

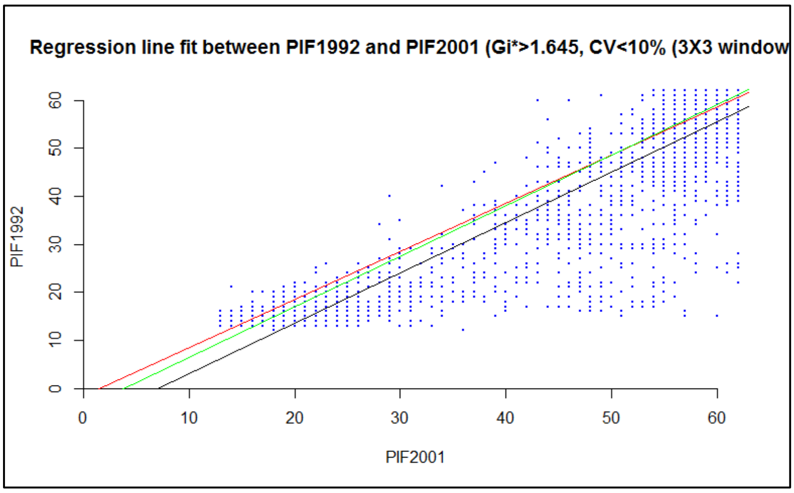
Root mean square error (RMSE): RMSE is a measure of error in estimation, to measure the deviation between the referenced true value and the calibrated value. The expression of the RMSE is mentioned in following (15):

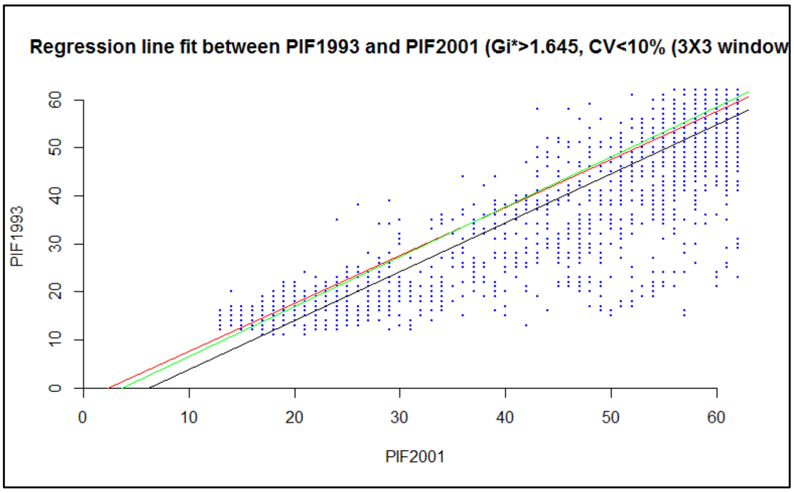
$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(X_{obs,i} - X_{model,i}\right)^2}{n}} \tag{15}$$

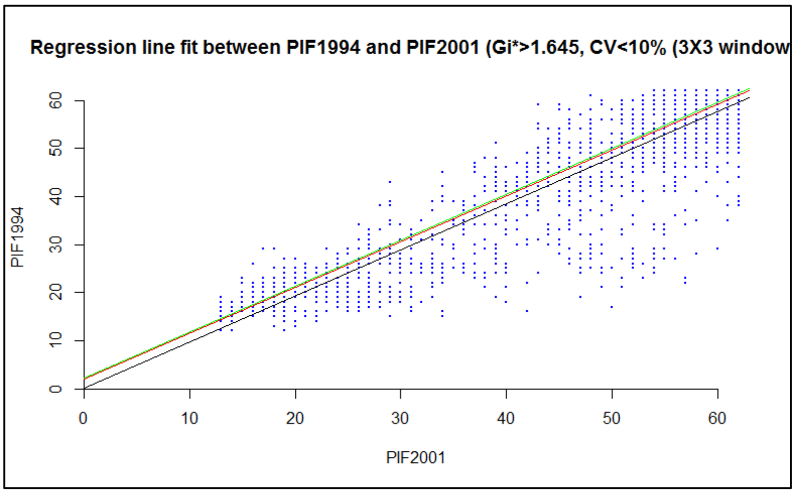
Where X_{obs} is the observed values and X_{model} is estimated values at ith observation. n is total number of observations in data.

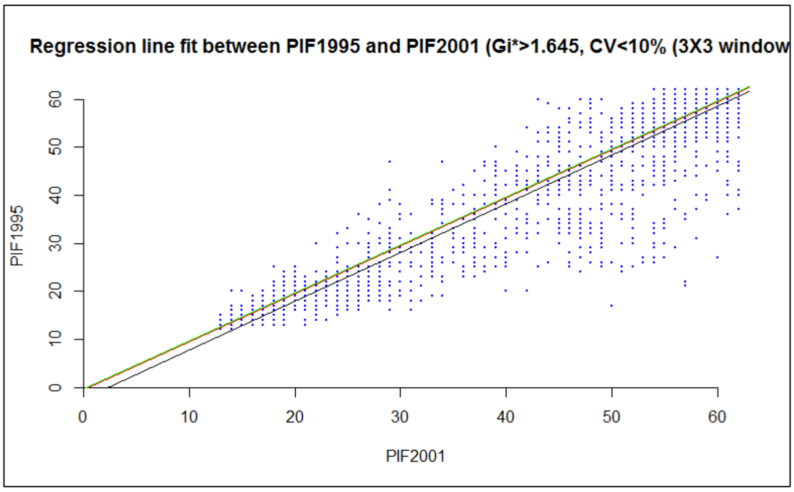
APPENDIX B

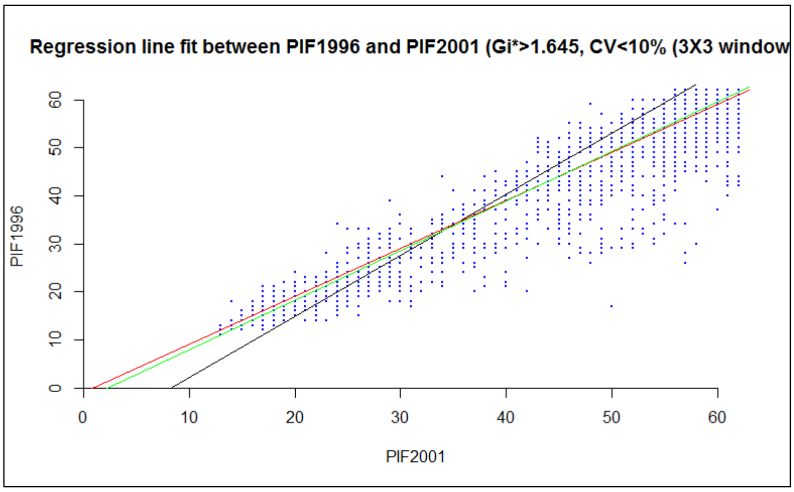
In all the linear regression diagnostics in Appendix B, X axis denotes PIF of 2001 as reference and Y axis denotes PIF of the calibration awaiting year from 1992 to 2006. The black line shows fit of OLS in data, red line shows LMedS fit and green line shows the fit of LTS. All regression ignores the presence of NA values in PIF and all regression are performed with a significance level of 0.05.

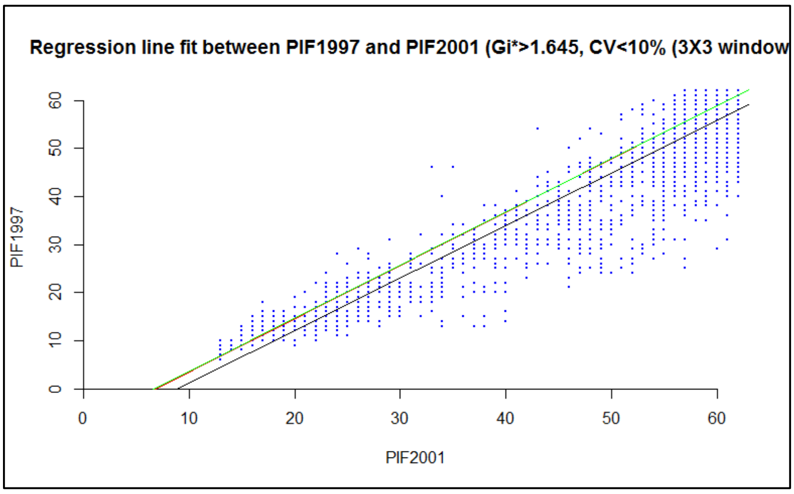


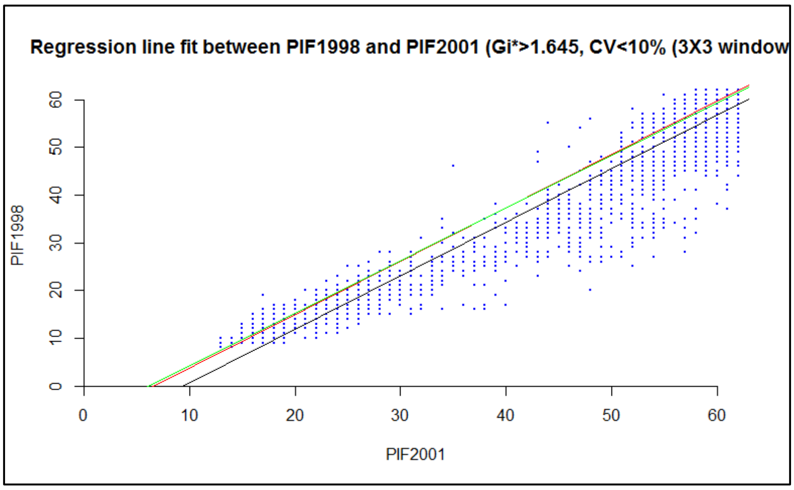


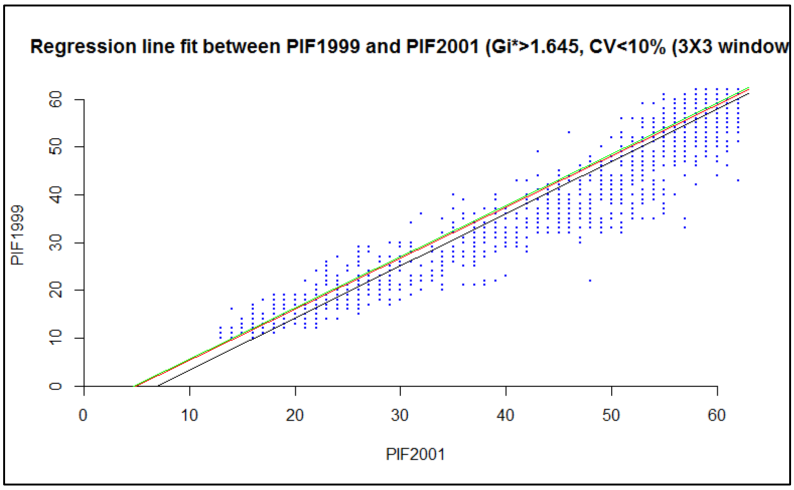

Regression line fit between PIF1994 and PIF2001 (Gi*>1.645, CV<10% (3X3 window
PIF1994
PIF2001

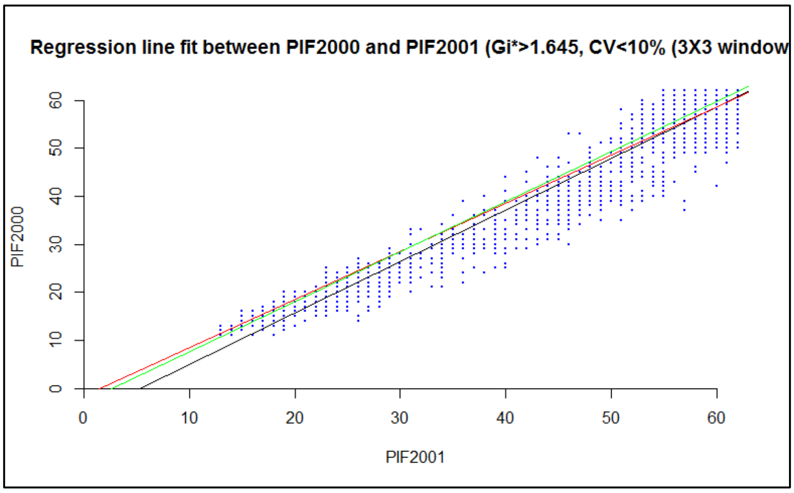

Regression line fit between PIF1995 and PIF2001 (Gi*>1.645, CV<10% (3X3 window
PIF1995
PIF2001

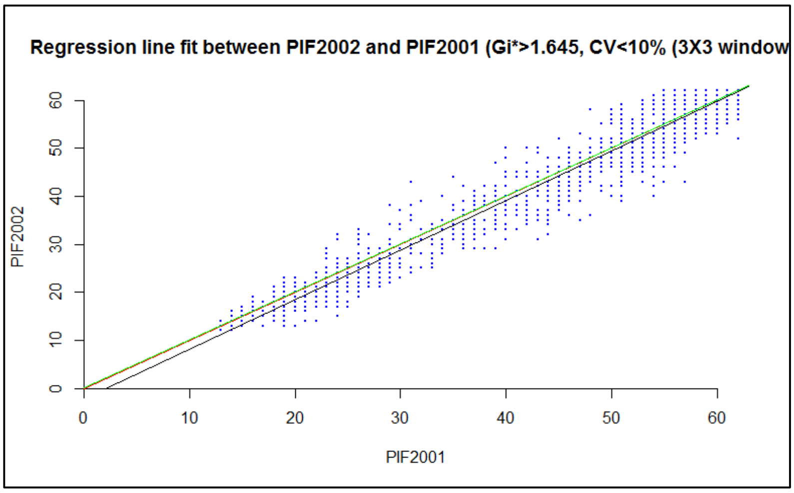

Regression line fit between PIF1996 and PIF2001 (Gi*>1.645, CV<10% (3X3 window
PIF1996
PIF2001

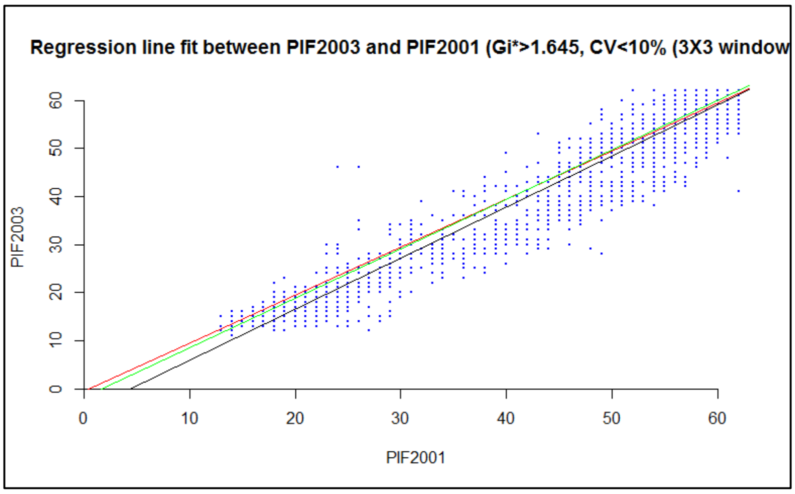

Regression line fit between PIF1997 and PIF2001 (Gi*>1.645, CV<10% (3X3 window
PIF1997
PIF2001
0
10
20
30
40
50
60

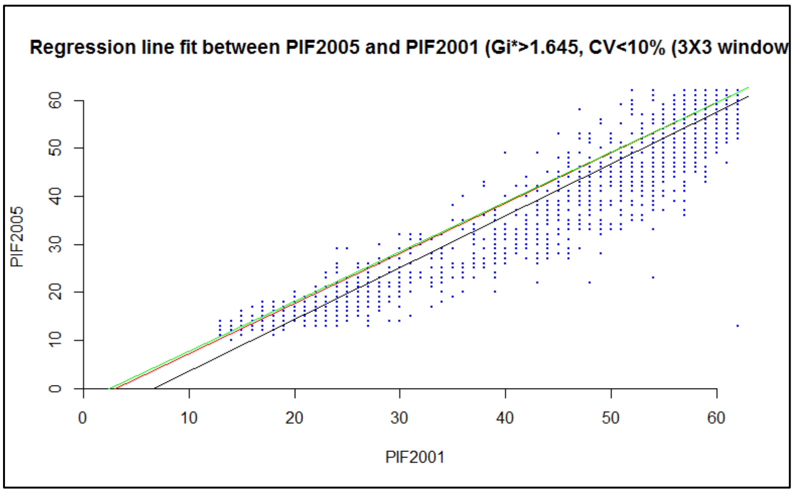

Regression line fit between PIF1998 and PIF2001 (Gi*>1.645, CV<10% (3X3 window
PIF1998
PIF2001
0
10
20
30
40
50
60

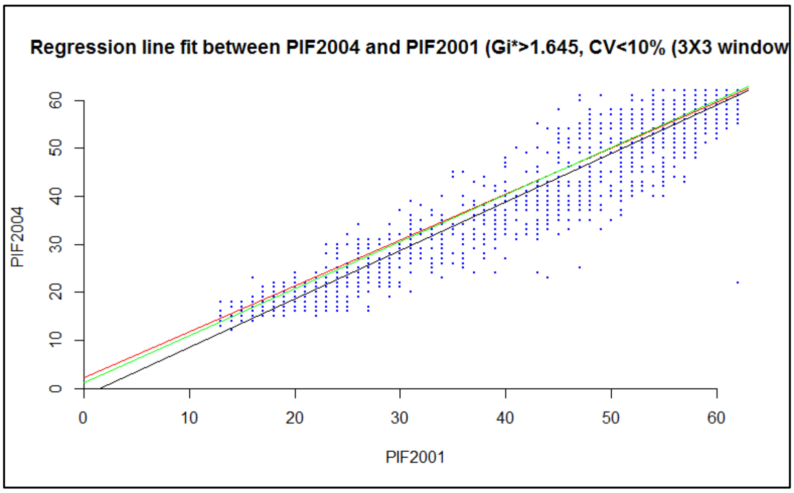

Regression line fit between PIF1999 and PIF2001 (Gi*>1.645, CV<10% (3X3 window
PIF1999
PIF2001
0
10
20
30
40
50
60

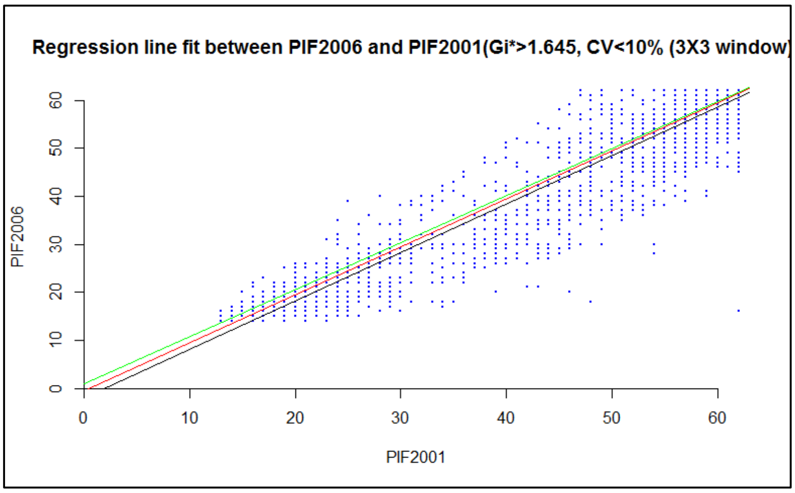

Regression line fit between PIF2000 and PIF2001 (Gi*>1.645, CV<10% (3X3 window
PIF2000
PIF2001


Regression line fit between PIF2002 and PIF2001 (Gi*>1.645, CV<10% (3X3 window
PIF2002
PIF2001


Regression line fit between PIF2003 and PIF2001 (Gi*>1.645, CV<10% (3X3 window
PIF2003
PIF2001


Regression line fit between PIF2005 and PIF2001 (Gi*>1.645, CV<10% (3X3 window
PIF2005
PIF2001
0
10
20
30
40
50
60


Regression line fit between PIF2004 and PIF2001 (Gi*>1.645, CV<10% (3X3 window
PIF2004
PIF2001
0
10
20
30
40
50
60


Regression line fit between PIF2006 and PIF2001(Gi*>1.645, CV<10% (3X3 window)
PIF2006
PIF2001
0
10
20
30
40
50
60

APPENDIX C

The following Figure C-1 and C-2 shows the comparison of uncalibrated SOL and calibrated SOL for 1992-2006 for India and Bangladesh respectively. Uncalibrated SOL does not show consistency in SOL through years in both the countries; the total brightness is observed to fluctuate through years. Whereas for the calibrated SOL the consistency is observed to be much better and also the degree of convergence among the SOL is observed to be very less. For India, calibration is also performed with Elvidge et al. (2009) given coefficients also. From visual interpretation of SOL obtained from Elvidge et al. (2009) given coefficients and the SOL obtained from coefficients obtained from study it is seen that the fluctuation is even less in LTS generated SOL than of the Elvidge et al. (2009) one. However, LTS coefficient generated SOL seems to overestimate the brightness values a bit but consistency is also observed to be more than uncalibrated image SOL for both India and Bangladesh.

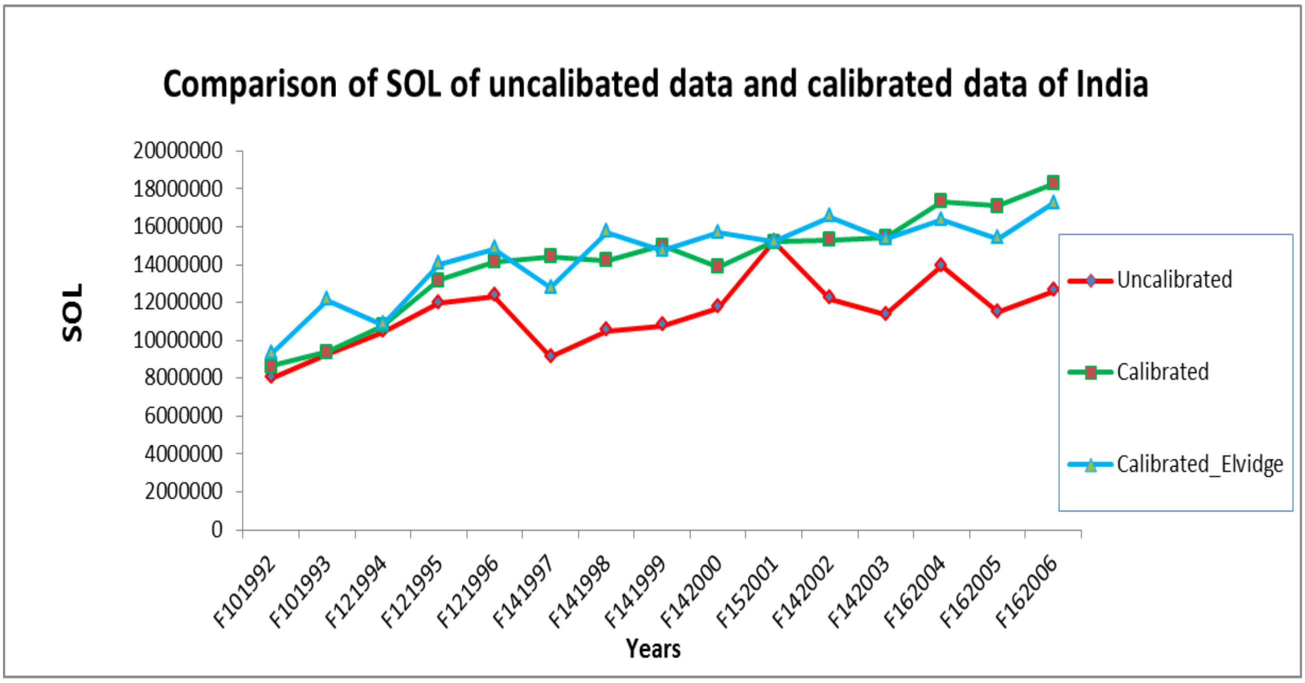


Figure C-1: Comparison of inter-calibrated SOL with uncalibrated SOL for India

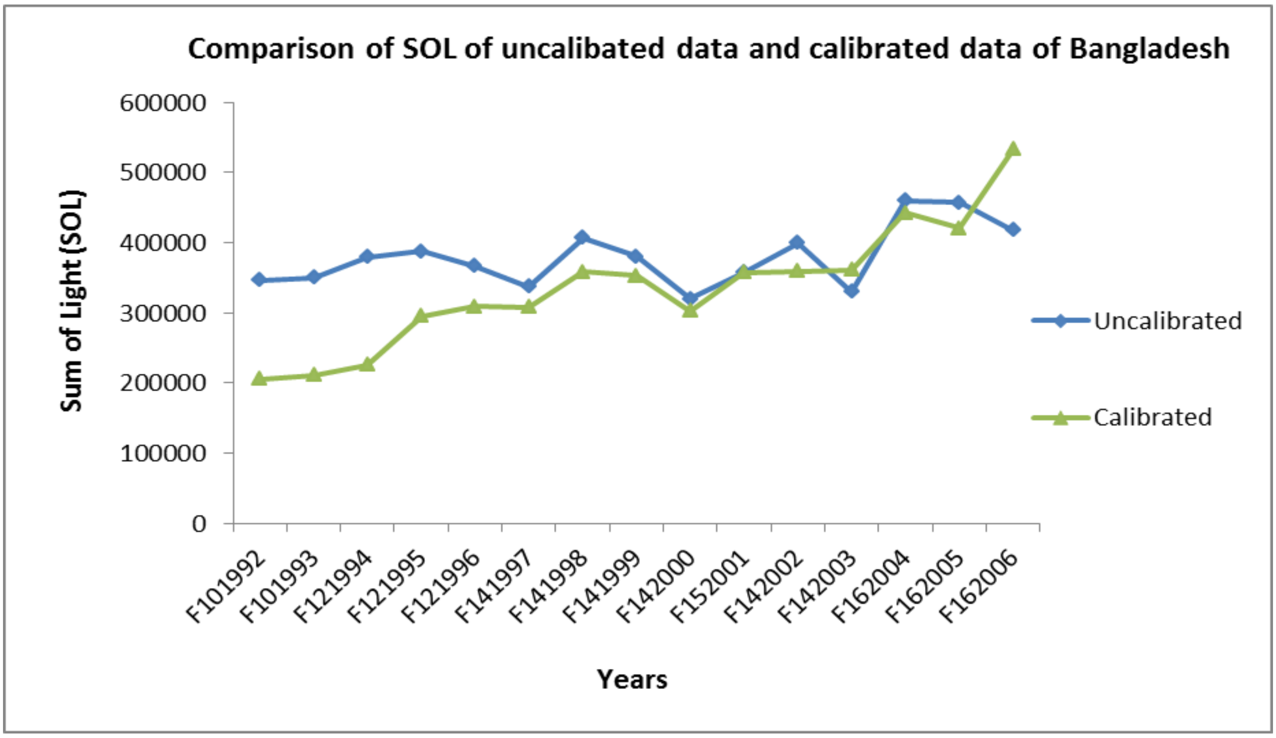


Figure C-2: Comparison of SOL of inter-calibrated SOL with uncalibrated SOL of Bangladesh