Temporal Gap Filling of Nighttime Light Composites

Nalin Sharma [a,*], Prasun Kumar Gupta [a], Prabhakar Alok Verma [a]
[a] Indian Institute of Remote Sensing, ISRO, Dehradun, India
*Corresponding author (email: sharmanalin97@gmail.com)

(Received on 15 April 2024; In final form on 25 January 2025)

DOI: *https://doi.org/10.58825/jog.2025.19.1.152*

Abstract: The temporal nighttime light (NTL) data generated by DMSP-OLS sensors was discovered to have large gaps (missing values) over time. The research aims to provide a scientifically valid gap-filling mechanism for having consistent DMPS-OLS time series data (1992-2013) and predicting the historic NTL (1991-1985) for long-term studies. A deep learning neural network, Long Short Term Memory (LSTM) has been proposed in the study for temporal gap filling and historic NTL prediction. The developed LSTM model is being tested in a time distributed wrapper way having window size (3-7) for the temporal gap filling and prediction of the historic NTL. According to the accuracy evaluation, the developed model has a testing accuracy of R^2 = 0.96 with a window size of 5. The historic population, Gross Domestic Product (GDP), and Electric Power Consumption per capita (EPC) data are utilized to validate the gap-filled and historic NTL. R^2 = 0.91 w.r.t population, R^2 = 0.71 w.r.t GDP, and R^2 = 0.69 w.r.t EPC, is been found during the assessment of these parameters with the sum of light of the year (1985-2013). The historic & gap-filled NTL data can be used in various studies to monitor temporal development.

Keywords: Nighttime light, DMSP-OLS, Deep Learning, LSTM.

1. Introduction

The capture of nighttime light (NTL) by remote sensing satellites has demonstrated an intriguing approach to gather information about human activity (Combs & Miller, 2023). The US Department of Defence developed the Defence Meteorological Satellite Program (DMSP) in August 1962 to gather worldwide weather data for various military purposes. The satellite sensor has been collecting precise data to examine the worldwide spread of clouds and the top temperature of clouds for more than five decades. The DMSP carried an oscillating scan radiometer, the Operational Linescan System (OLS), to acquire the Visible Near-Infrared (VNIR) images of Earth twice a day since 1976. The illuminance sources of the capture include metropolis lights, fires, lightning, gas flares, Aurora, and so forth (Elvidge et al., 1997). From 1992 through 2013, the National Oceanic and Atmospheric Administration (NOAA) processed and archived DMSP NTL data. From 1990 through 2013, the NTL products were developed using a series of satellites (F10 to F18) by the Earth Observation Group (EOG) and is available in the public domain.

NTL data acquisition by remote sensing is a one-of-a-kind benefit entity that spans a wide range of industries and applications, making the need for consistent and exact NTL data even more critical. Among them are the computation of socioeconomic indicators (Sahoo et al., 2020), the monitoring of urban and regional economic development (Tan et al., 2018), the effects of light pollution and population density (Bennie et al., 2015), and the evaluation of significant event. NTL imagery and its uses have spawned a plethora of diverse points of view; one form of research study concentrates on database upgrades and obtaining qualitative and quantitative data from them (Ma et al., 2012); parallelly other application-oriented studies include the study of how NTL and socioeconomic indices, energy consumption, and distribution (Zhu et al., 2021), growth of community (Ma et al., 2012), and other variables interact with one another.

The DMSP-OLS scans the entire globe in the VNIR and Thermal Infrared spectral bands using 14 orbits (Krishna Prasad et al., 2002). Using a sequence of six satellites, NTL data is obtained in DN values ranging from 0 for no light to 63 for maximum brightness. A DMSP-OLS time series analysis found an anomaly in the total aggregate of Digital Number (DN) (Doll, 2008). The data also has pixel saturation issues, which can be resolved by using fixed-gain data from OLS sensor pre-flight calibration (Hsu et al., 2015). The blooming effect results in a fuzzy haze in the urban areas which extends to the non-lit rural areas due to which it becomes difficult to estimate the growth of an area using NTL data (Hsu et al., 2015). According to (Doll, 2008) there is an average of 60% overlap between the pixels, which causes certain pixels to appear to be more lit than they are. The initial resolution of the satellite sensor is 2.7 km and is provided at a resampled resolution of 1 km by EOG. There is salt and pepper noise in some sections of the backdrop when there is no apparent light (Jindal et al., 2022).

EOG has provided an NTL data set that has been widely used in the literature for socioeconomic and demographic studies at the regional scale. When utilized for distributed modelling, it was revealed to have large gaps (missing values) over time. Several attempts have been made to correct for saturation errors in the dataset (Mukherjee et al., 2017), but few studies have been reported on filling up these missing values (Pandey et al., 2017). These gaps could be filled with regression or interpolation, but a scientifically valid gap filling methodology is yet to be suggested.

The Long Short Term Memory (LSTM) neural network is a type of RNN that is used to process data with long-term dependencies. The RNN has a short term memory (vanishing gradient problem) in which it stores the data for one process and delete it after the successful execution of the process and holds new data (Wang et al., 2017). LSTM models are designed to capture long-term dependencies in sequential data. They have a memory cell that allows them to remember information from earlier time steps, making them capable of learning and modelling complex temporal patterns over longer time intervals. This is crucial for accurately filling in missing values in a time series (Selvin et al., 2017). Temporal gap filling often involves dealing with irregularly spaced time series data where the gaps can have variable lengths. LSTM models can handle this variability effectively because they do not rely on fixed-size windows or time intervals. Instead, they can process and learn from sequences of any length, making them flexible in dealing with missing data points at different time intervals (Sherstinsky, 2020). LSTM models are capable of learning non-linear relationships in the data. They can capture and model complex patterns and dependencies between past and future time steps. This flexibility is essential when dealing with real-world time series data, where the underlying patterns can be non-linear and contain various temporal dynamics (Staudemeyer & Morris, 2019a).

Remote sensing frequently involves the collection of time-varying data, such as vegetation indices, sea surface temperatures and precipitation patterns. LSTM models are particularly good at analysing sequential data and can learn complicated temporal patterns from time series remote sensing data (Sherstinsky, 2020). LSTM models can combine data from numerous remote sensing sources, such as satellite imagery and meteorological data, or they can integrate diverse sensor modalities (Xu et al., 2022) . The models can successfully capture the spatiotemporal interactions between multiple data sources by combining LSTM layers, producing more complete and accurate results (Shen et al., 2020). LSTM models may learn patterns and identify places where major changes have occurred by training on historical image sequences (Ma et al., 2012). This is especially important for tracking changes in land use (Zhu et al., 2021), urban growth and environmental conditions (Xu et al., 2022). In addition, LSTM models can be used to reconstruct lost or distorted remote sensing data. LSTM models can improve the resolution of remote sensing data by creating high-quality images from low-resolution inputs (Shen et al., 2020).

Long-term series night light data is valuable for various applications and analyses due to its ability to capture temporal trends, patterns, and changes in nighttime illumination. It provides insights into urbanization, economic development, environmental changes, disaster monitoring, social dynamics, and climate change. It facilitates a comprehensive understanding of the spatiotemporal dynamics and impacts of human activities, making it valuable for diverse applications in research, policy-making, and sustainable development.

The DMSP-OLS satellite is operational since the 1970s but the archiving of data started in 1992 therefore an important use of such a model would also be to predict historic NTL. This study will attempt to (1) build a LSTM model and generate gap-filled NTL data, (2) predict the historic NTL data and (3) test their applicability for socio-economic parameters. The results of the proposed methodology can be correlated with socio-economic datasets and validated.

2. Study Area

The study is conducted over conterminous India, using inter-calibrated NTL data of EOG with ancillary datasets for validation. The summary of the datasets used is given in Table 1 and a brief description is provided below.

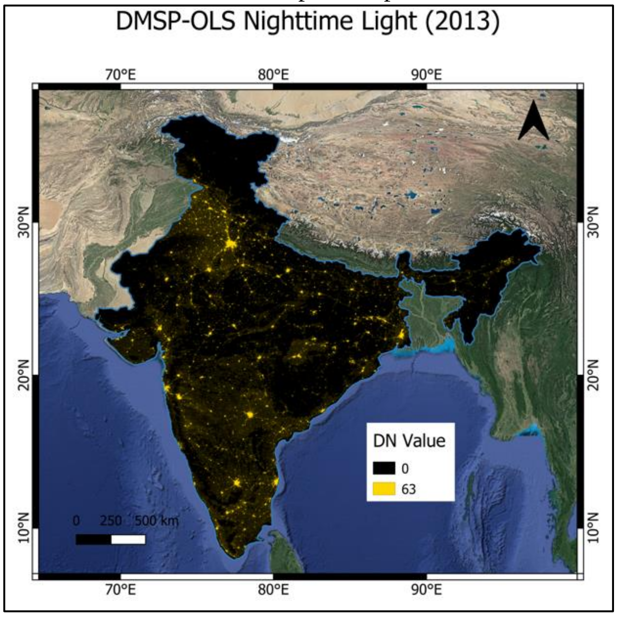

Figure 1. Study area showing DMSP-OLS NTL for 2013 over conterminous India

Conterminous India is chosen to be the study area for the research work. India lies between latitudes 8° 4'N and 37° 6'N, and longitudes 68° 7'E and 97° 25'E as shown in figure 1. It is the second most populous country in the world after China. With 28 states and 8 union territories, it is surrounded from three sides by water with the Indian Ocean in the south, the Bay of Bengal in the southeast, and the Arabian sea in the southwest. Pakistan, China, Nepal, Bhutan, and Myanmar are India's neighbours. It shares a marine boundary with Sri Lanka, the Maldives, Thailand, and Indonesia.

As per the latest Indian State of Forest Report (ISFR) the India's forest cover increased from 19.53% in the 1980s to 21.71% in 2021, while the country's overall green cover, including tree cover, is presently at 24.62% (Agarwal et al., 2023) . The population of the metropolitan cities of India is more than 5 million which is more than the population of some nations of the world. There is a rapid boost in the Indian economy in the last two decades with a growth rate of 6.5-7.0% (Taneja et al., 2023). With a more productive workforce, India's GDP might rise to $9 trillion by 2030 and $40 trillion by 2047 (George, 2023).

Therefore, India is chosen as to be the study area for this reach work. The last decade India is the area of interest for various studies related to NTL data to analyze rural electrification (Kiran Chand et al., 2009), economic activities at night (Henderson et al., 2009), urban sprawl

(Pandey et al., 2013), and effects of COVID-19 on the Indian subcontinent. India is the rapidly growing economy in the world and is an interest area of research for many studies on NTL.

2.2 DMSP-OLS Nighttime Light Data

The NTL data acquired by DMSP-OLS satellite had a major contribution in monitoring the cloud cover and various climatic changes occurring across the globe. With the 3,000 Km swath it covers the globe two times a day (Elvidge et al., 1997). The NTL data is reported in DN Values as it lacks on-board calibration to convert the raw DN into absolute radiance or luminosity values. This makes it challenging to directly compare light intensities between different images or satellites. The DN values provided by DMSP-OLS are relative measurements, which can vary across time and space (Doll, 2008).

The process of inter-calibration involves various steps such as selection of the correct reference data, preprocessing of data for removal of errors, radiometric scaling to ensure that the NTL data from various sources or versions is scaled to a single reference scale, cross-calibration to compare the radiometric values of the reference and target datasets, adjustment and normalisation of NTL images and validation using independent reference sources or ground truth data (Mukherjee et al., 2017). EOG has created an inter-calibrated "Stable Lights Average Digital Number Dataset" (1992-2013) which has been used in the study work (URL: https://eogdata.mines.edu/wwwdata/dmsp/v4composites_rearrange/)

Table 1. Summary of dataset used in the study

Data	Timespan	Temporal Resolution	Spatial Resolution	Source
Stable Lights Average Digital Number Dataset	1992-2013	Annual	~1 km Near Equator	Earth Observation Group[1]
Population	1985-2013	Annual	Country-wise	Census of India[2]
Gross Domestic Product	1985-2013	Annual	Country-wise	The World Bank[3]
Electric Power Consumption per capita	1985-2013	Annual	Country-wise	The World Bank[4]

2.3 Ancillary Data

The NTL can be used to measure and validate various socio-economic parameters such as urbanization, infrastructure development, population growth. electricity consumption, economic growth etc. There are various parameters that can be utilized for the validation of the gap-filled and predicted NTL out of which three major parameters has been chosen in the present study. The annual population, GDP and Electric Power Consumption per capita (EPC) for time series (1985-2013) is taken as ancillary data for validation. A summary of their specification is given in table 1.

3. Data used

The section details the methodology adopted to carry out the study. The overall procedure given in figure 2 demonstrates the overall procedure adopted in the research work. The placeholder box describes the objectives addressed in the research work i.e. temporal gap filling, historic prediction & validation of the gap-filled and predicted historic NTL data. The procedure followed is further explained in sections 3.1, 3.2 & 3.3.

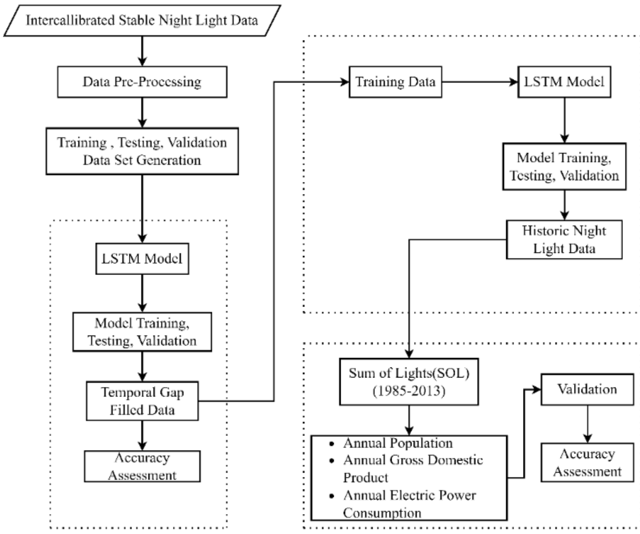

Figure 2. Flow chart of followed methodology

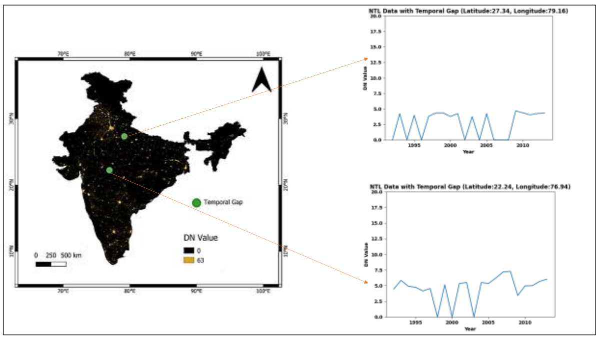

Figure 3. NTL data with temporal gaps

3.1 Data Preprocessing:

The EOG has provided an inter-calibrated DMSP-OLS dataset that has been used in this study. The datasets are available in GeoTIFF format and are a pre-processed inter-calibrated version of "Global Stable Light" products from 1992 to 2013. The annual data of the study area is extracted by masking from the global data and temporal gaps across are identified which are required to be addressed while gap filling. Figure 3 shows sample points in the study area found to have temporal gaps.

A temporal gap in night light data refers to a period when data on nighttime illumination levels is absent or unavailable in lit pixels. In the study pixels, the nightlight illumination is observed for the time series 1992-2013 and found to have many gaps across the series. Figure 4 manifests the number of pixels having zero value and corresponding percentage in the study area. Figure 4 (a) describes a decline in the number of pixels having zero value across the time series (1992-2013). The percentage area with 'zero' DN value ranges from 65% to 35% in time series (1992-2013) which shows a decrease in the temporal gaps in the observed data provided by EOG.

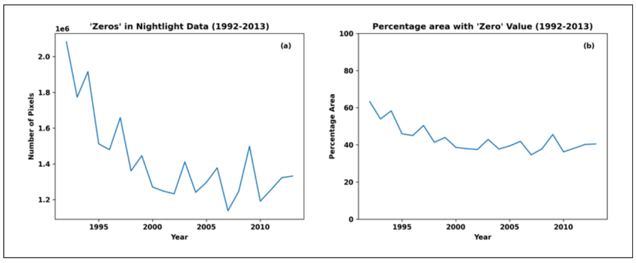

Figure 4. Analysis of temporal gap in India

The data extracted for India is processed and latitude-longitude locations with corresponding DN values are extracted from the annual data images (1992-2013). From the extracted data the training and testing dataset are prepared which will be the input for the LSTM model. Training and testing dataset preparation is done in three main stages (1) NTL data analysis, (2) The window size selection for the processing of NTL data, (3) Splitting of data into training and testing dataset.

The data extracted for the study area is analyzed for temporal gaps. During the analysis of NTL data, it is found to have sudden dips in the data across the time series for various location in the study area as shown in Figure 3. For such locations due to no data value the data extracted show NaN. To avoid the non-lit pixels in the NTL data, DN < 4 are discarded from the extracted data. The pixels with DN Value as 'NaN' are replaced with 0. In this study the aim is to fill such pixels with DN Value as '0'.

The input to the prepared model is given in the time distributed wrapper layer in many to one way which helps to slice the time series of data and generate, a consistent time series as output for an input window size in order to find the appropriate window size for LSTM model. The model can learn patterns and relationships over a period of time by choosing a suitable window size. The model might not be able to accurately reflect long-term dependencies if the window size is too narrow. The model might not be able to capture short-term relationships, if the window size is set to a value that is too large. The time series NTL data has been arranged as:

$$X_{[n,\,w]} = \{k_t, k_{t-1}, \ldots, k_{t-w-1}\}$$
$$y_{[n,\,1]} = \{k_{t-w}\}$$

where k = the DN value of pixel considered for gap filling; t = year from 2013 to (1992 +w) in descending order; w = window size of time distributed wrapper layer; and n = number of pixels considered for gap-filling.
The model is tested for window size between 3 to 7 and accuracy assessment of the model is done to find the correct window size. In the prepared model the original data is separated in the ratio 70:30 with 70 % as the training data and 30 % as the testing data for the model.

3.2 LSTM Model Settings:

The LSTM model is a RNN capable of handling data with long-term dependencies. The model can be segregated in four levels. The LSTM model have memory cell comprised of three gates, the input, output and forget gate that permits the data to added or removed from the cell. Rectified Linear Units (ReLU) has been used as an activation function with a property of non-linearity to resolve the vanishing gradient issues so that the built model can retain data for long time which helps the forget gate to take decision as which information needs to discarded from the previous cell state (Staudemeyer & Morris, 2019b). The "input gate" maintains the point-by-point multiplication of the activation function that adds a stream of data to the present state. The output gate determines which bits of the cell state to retrieve and use to generate the output. It can decide whether to keep or discard information from memory based on its importance to the current time step and the task at hand. The LSTM model is developed by using the appropriate hyperparameters in order to get the accuracy in the generated results. The hyperparameters of LSTM model comprised of the batch size, dropout rate, number of hidden layer (Aslam, 2018). The number of hidden layers and the number of LSTM units in each layer are two crucial hyperparameters that impact the model's ability to learn complicated patterns in input. While adding more hidden layers or LSTM units may help the model recognize more complex patterns, doing so runs the risk of overfitting. On the other side, limiting the amount of hidden layers or LSTM units may lead to under fitting, where the model is unable to fully capture the complex nature of the data.

The hyperparameter batch size define how frequently the model's parameters are altered during the initial training of model (Aufa et al., 2020). A small batch size can lead to noisy gradients and delayed convergence, whereas a large batch size can lead to overfitting. Due to randomly dropping out neurons, the dropout rate hyperparameter is utilized to prevent overfitting during training. Under fitting can arise when the model is unable to capture the whole complexity of the data due to a high dropout rate. In contrast, a low dropout rate may result in overfitting. The activation function in an LSTM model is crucial in bringing non-linearity to the network and enabling sophisticated computations. Activation functions are used to the inputs and outputs of several LSTM cell components such as the input gate, forget gate, output gate, and cell state. The process flow of model can be referred from figure 5.

The model has been tested for different activation functions such as Sigmoid (σ), Hyperbolic Tangent Activation (tanh), and ReLU . The model fitting has been done by tuning different hyperparameter such as batch size, dropout rate, number of hidden layer, activation function, optimizer in order to obtain optimized result.

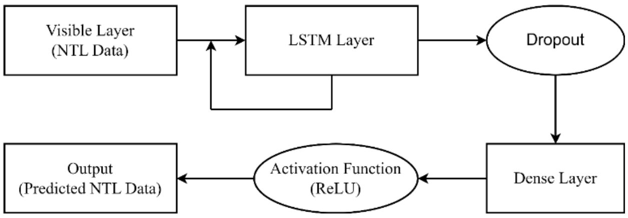

Figure 5. Process flow of developed LSTM model

3.3 Accuracy of Gap-Filled and Predicted NTL

It is vital to quantify a model's performance in order to use it as feedback and comparison. There are various error metrics available using which the accuracy of the results can be verified. The four error metrics that are used to evaluate the developed models accuracy are: R^2 score, Mean Square Error (MSE), Root Mean Square Error (RMSE) and Mean Absolute Error (MAE).

In various research on urbanisation and development, Sum of Light (SOL) is frequently employed as a stand-in for or indicator of socioeconomic characteristics. This association is explained by the fact that nighttime artificial lighting is more prevalent in metropolitan, economically advanced, and inhabited places. SOL data is extracted from the updated annual images of NTL data. SOL refers to a measurement of the total amount of light emitted by a geographic area. The brightness values of each pixel inside a chosen area of interest in a NTL image are added to obtain this measurement. R^2 value is determined between the extracted SOL and the annual population, GDP and EPC data for year (1985-2013).

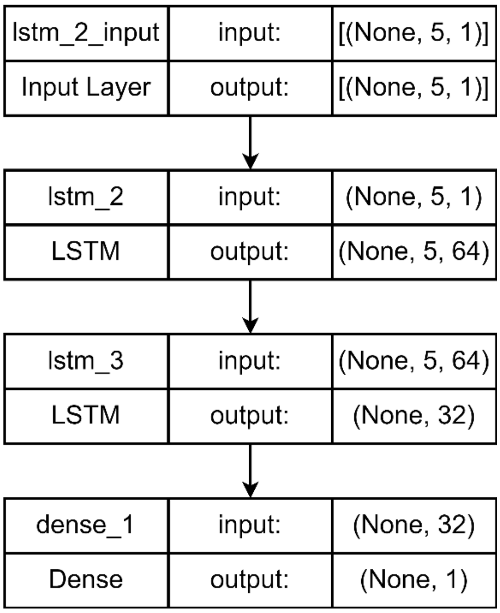

Figure 6. Developed LSTM model architecture

4. Methodology and Analysis

This section has been divided into four parts. The first part (4.1) discusses the analysis of the developed model, second part (4.2) discuss the before and after comparison of the observed and gap-filled NTL for the years (1993-2013). The third part (4.3) discusses the analysis of predicted historic NTL (1985-1991). The fourth part (4.4) discusses the validation of gap-filled and predicted NTL using socio-economic parameters.

4.1 Analysis of Developed LSTM Model

The LSTM model developed as a part of the project has been used to fill the temporal gaps in NTL data. The LSTM model architecture, includes the number of layers, the number of LSTM units in each layer, and the connectivity between layers.

Table 2. Accuracy assessment of LSTM model with varying window sizes

Window Size	R^2 Score	MAE	MSE	RMSE
3	0.91	1.58	4.76	2.18
4	0.92	1.37	4.13	2.03
5	**0.96**	**1.03**	**3.98**	**1.99**
6	0.93	1.13	4.37	2.09
7	0.94	1.17	4.43	2.10

Figure 6 describes the architecture of the developed LSTM model. The first layer of an LSTM network is in charge of processing the input sequence and extracting pertinent characteristics. The input sequence is separated into distinct time steps and each time step is processed by the LSTM layer. The prepared model consists of two hidden layers which are made up of LSTM units or cells, which process the input sequence over time and capture long-term dependencies. The last layer of model is the dense layer that applies a linear transformation to the outputs of the LSTM layers, which could be the hidden states from the last LSTM layer or the output sequence. Each neuron in the dense layer is connected to every neuron in the previous layer, allowing information to propagate across the network.

The developed model has been tested on different window sizes (3, 4, 5, 6, 7) and epochs (25 - 200) to get the accurate results. 'Window Size = 5' and epoch=100 has been found most suitable for filling the temporal gap. The results of accuracy assessment are shown in table 2.

4.2 Analysis of Temporal Gap Filled NTL Images

The DMSP-OLS NTL data is found to have temporal gaps in the time series 1992-2013 and to address the issue LSTM model developed as a part of this study work has been used to fill the NTL data for locations having temporal gaps across the time series (1992-2013). There are many locations identified in the study area with temporal gaps as shown in figure 7. The temporal gaps are being filled with the generated values of the LSTM model as shown in figure 8. The model is able to fill the missing values in the time series, even when there are 1 or more gaps. Minimal or no change has been done to the other pixel values.

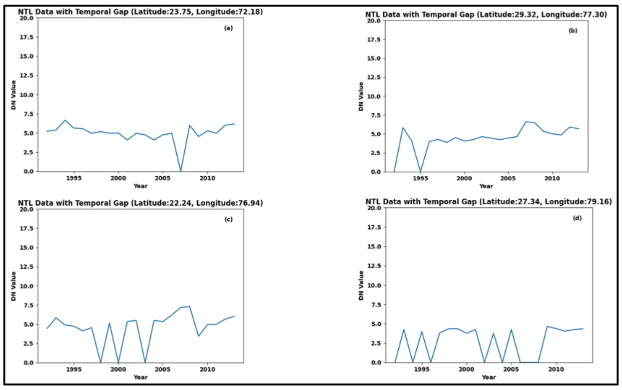

Figure 7. NTL data with temporal gap

Figure 7 describes the sample locations taken in the study area with temporal gaps for analysis. Inconsistent temporal gaps are found across the year for different periods. For sample location in Figure 7 (a), one gap is identified in the time series for the year 2008, and Figure 7 (b) shows two gaps for the years 1990 and 1985; similarly Figure 7 (c) and Figure 7 (d) shows multiple temporal gaps for different locations for varying periods.

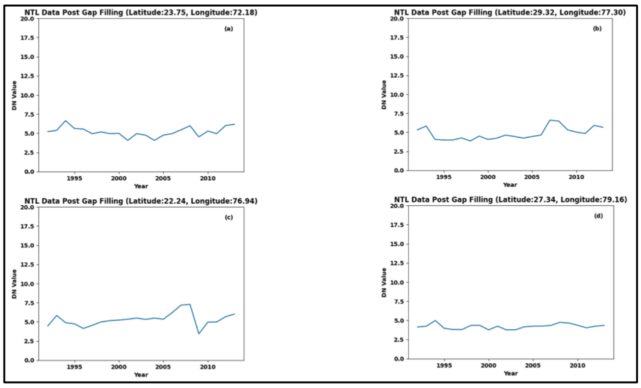

Figure 8. NTL data post gap filling

The temporal gaps identified in the Indian mainland are filled using the developed model and the consistent series of NTL is obtained as output. Figure 8 describes the consistent time series obtained post-gap filling of sample locations shown in figure 7. The trend of the NTL data remains unaltered and only the ‘zeros’ identified in the observed NTL data is provided with the corresponding DN value generated by the model.

A qualitative and quantitative comparison between the observed and gap-filled NTL data for the years 1992, 1997, 2002, 2007, and 2012 is done. In research on urbanization and development, SOL is frequently employed as an indicator of socioeconomic characteristics. This association is explained by the fact that nighttime artificial lighting is more prevalent in metropolitan, economically advanced, and inhabited places. Analysis of SOL of observed and gap-filled NTL data is done for the years 1992-2013 and there is an increase in SOL for each year is found. For 1992 the comparison is shown in figure 9, for 1997 is shown in figure 10, for 2002 is shown in figure 11, for 2007 is shown in figure 12, for 2012 is shown in figure 13. Spatially gaps varying from a few pixels to large holes have been filled by the model. This improves the data and reduces inconsistencies.

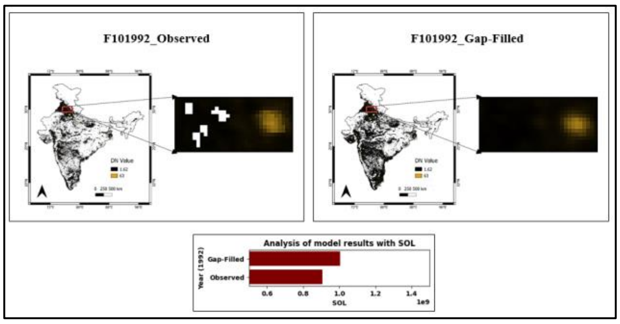

Figure 9. Analysis of observed and gap-filled NTL image (1992)

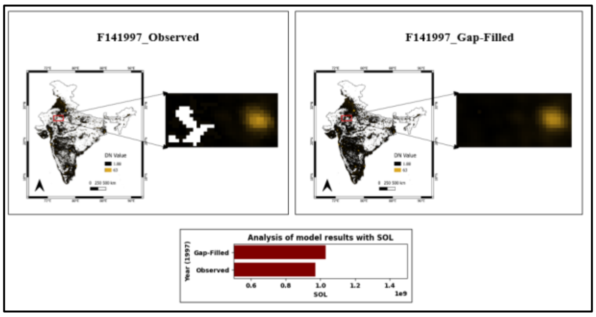

Figure 10. Analysis of observed and gap-filled NTL image (1997)

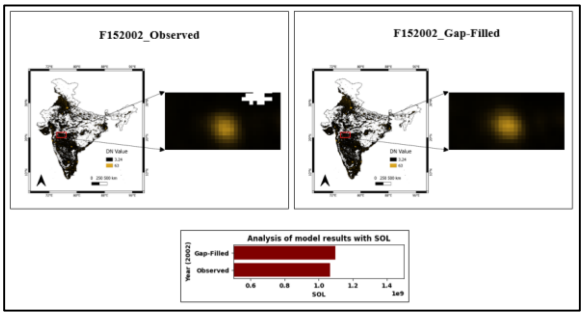

Figure 11. Analysis of observed and gap-filled NTL image (2002)

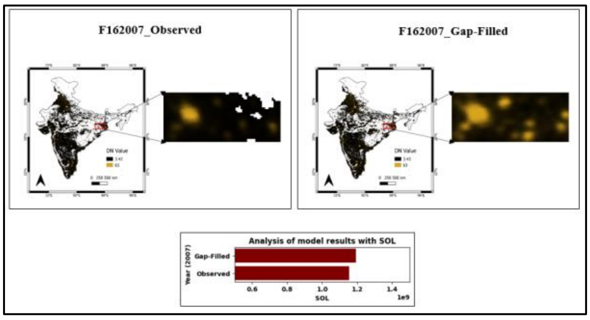

Figure 12. Analysis of observed and gap-filled NTL image (2007)

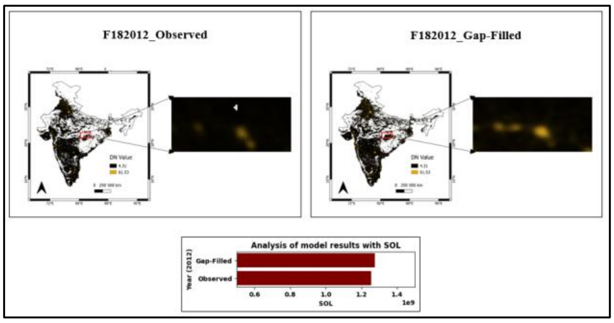

Figure 13. Analysis of observed and gap-filled NTL image (2012)

4.3 Analysis of Historic Predicted NTL Images

The absence of historic NTL reduces its applicability of the data for studying long-term or decadal changes. In the present study using the developed model the prediction of historic NTL data has been carried out for year (1985-2013). The output for 4 sample locations is shown in figure 14.

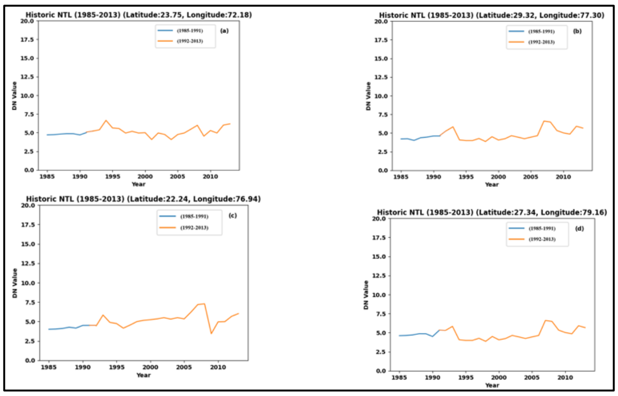

Figure 14. Historic NTL data (1985-2013)

Figure 15 depicts the NTL data predicted by the LSTM model used to generate the historic NTL Images. An increase in NTL can be observed for time series (1985-1991) that correlates with the on-ground development and expansions of economic actives across the study area.

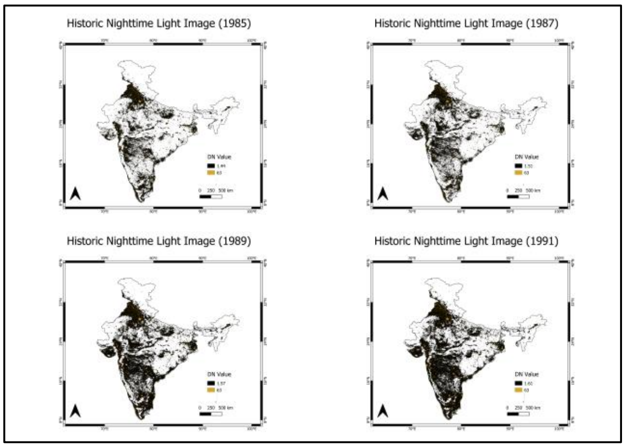

Figure 15. Historic nightlight images

Analysis of predicted historic NTL data is done with SOL and an increase trend is found across the year as shown in figure 16. The predicted NTL can be used to study the ground development of Indian in a specific period and validation of various socio-economic parameters.

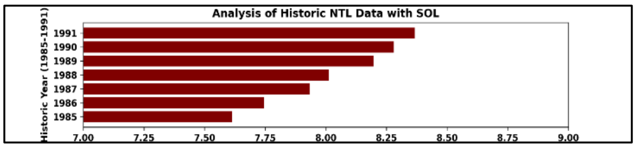

Figure 16. Analysis of predicted historic NTL with SOL

4.4 Validation of Gap-Filled and Predicted NTL Images

The temporal gap-filled and historic NTL data is validated with socioeconomic parameters to determine the accuracy of the NTL data generated using the LSTM model. There are three parameters population, GDP, and EPC used in the present study for validation.

The observed and gap-filled NTL data (1992-2013) were analysed as shown in figure 17 and are found to follow the trend. The notches in the observed data as shown in Figure 17(a) are due to the temporal gaps which are identified and filled by the LSTM model. Post gap filling as shown in Figure 17(b) the NTL trend has improved in time series and for the year 2005-2008 a flattening in the curve can be observed in Figure 17(b) that describes minimal change observed and gap-filled NTL data.

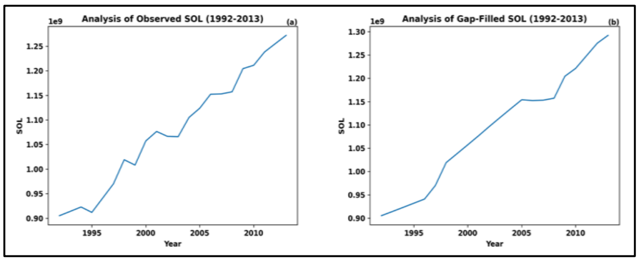

Figure 17. Analysis of observed and gap-filled SOL of NTL data (1992-2013)

R^2 = 0. 83 w.r.t Population, R^2 = 0.69 w.r.t GDP, and R^2 = 0.61 w.r.t EPC data is found as shown in figure 18, which shows a positive correlation between these socio-economic parameters and the historic NTL data predicted by the LSTM model.

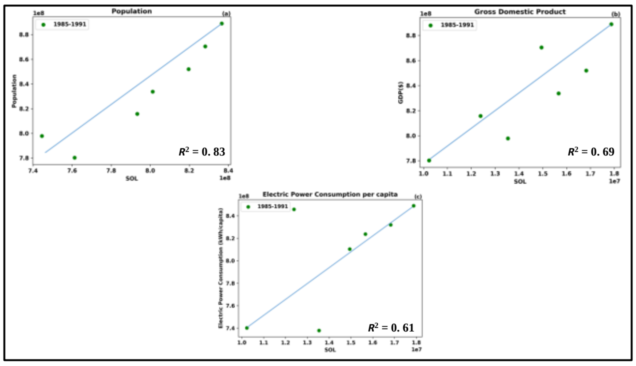

Figure 18. Relation between SOL of historical NTL (1985-1991) and socio-economic parameters

R^2 = 0. 91 w.r.t Population, R^2 = 0.71 w.r.t GDP, and R^2 = 0.69 w.r.t EPC data is found as shown in figure 19, which shows a positive correlation between these socio-economic parameters and the NTL gap-filled and predicted historic data by the LSTM model.

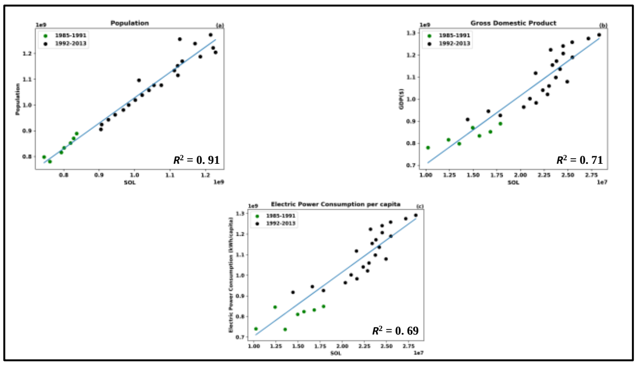

Figure 19. Relation between SOL of historical (1985-1991) and gap filled NTL data (1992-2013) with socio-economic parameters

By analysing figure 18 and 19, it can be concluded that the relationship trend for SOL of historical NTL (1985-1991) and socio-economic parameters is similar to 1992-2013.

5. Discussions

The DMSP-OLS annual composites (1992-2013) generated by EOG have been extensively used in literature for socio-economic and demographic studies at the regional scale. However, while using the same for distributed modeling, it has been found to have large gaps (missing values) across time. Several attempts have been made to correct for saturation errors in the dataset, but few studies have been reported on filling up these missing values. At the beginning of the research work, the temporal gaps in the DMSP-OLS NTL data were identified and a methodology is proposed for filling these temporal gaps using the LSTM network.

In this study, the use of LSTM neural network for filing the temporal gaps is proposed for DMSP-OLS NTL data (1992-2013). After testing with several hyperparameters a robust LSTM model was developed for the study. The developed LSTM model filled the temporal gaps and using the consistent time series, historic NTL is predicted for (1991-1985). The gap-filled and predicted NTL data is validated with the socio-economic parameters, which shows a high degree of positive correlation. The study can be further extended by choosing a different study area and by applying other geo-statistical or machine learning based models for temporal gap-filling.

Consistent NTL data prepared by temporal gap-filling and historic prediction will promote consistency, comparability, accuracy in temporal analysis, trend detection, policy evaluation, disaster response, and climate change research. It will serve as the foundation for rigorous research, effective policy making, and informed decision making, allowing for a thorough understanding of the dynamics and repercussions of nighttime illumination.

6. Conclusion

The DMSP-OLS annual composites (1992-2013) generated by EOG have been extensively used in literature for socio-economic and demographic studies at the regional scale. However, while using the same for distributed modeling, it has been found to have large gaps (missing values) across time. Several attempts have been made to correct for saturation errors in the dataset, but few studies have been reported on filling up these missing values. At the beginning of the research work, the temporal gaps in the DMSP-OLS NTL data were identified and a methodology is proposed for filling these temporal gaps using the LSTM network.

In this study, the use of LSTM neural network for filing the temporal gaps is proposed for DMSP-OLS NTL data (1992-2013). After testing with several hyperparameters a robust LSTM model was developed for the study. The developed LSTM model filled the temporal gaps and using the consistent time series, historic NTL is predicted for (1991-1985). The gap-filled and predicted NTL data is validated with the socio-economic parameters, which shows a high degree of positive correlation. The study can be further extended by choosing a different study area and by applying other geo-statistical or machine learning based models for temporal gap-filling.

Consistent NTL data prepared by temporal gap-filling and historic prediction will promote consistency, comparability, accuracy in temporal analysis, trend detection, policy evaluation, disaster response, and climate change research. It will serve as the foundation for rigorous research, effective policy making, and informed decision making, allowing for a thorough understanding of the dynamics and repercussions of nighttime illumination.

Acknowledgements
This research work was carried out as a part other M.Tech. dissertation by the first author (NS) from Indian Institute of Remote Sensing (IIRS), Dehradun. The authors are grateful to the head of IIRS for the necessary facilities, support and encouragement. Thanks to the EOG for providing inter-calibrated NTL data. We sincerely thank the anonymous reviewer for critical comments and suggestion which helped to improve the manuscript significantly.

Credit authorship contribution statement
Nalin Sharma: Formal Analysis, Methodology, Software, Writing; **Prasun Kumar Gupta**: Conceptualization, Investigation, Visualization, Review; **Prabhakar Alok Verma**: Supervision, Validation, Editing.

Declaration of Competing Interest
The authors declare that they have no known competing financial interest or personal relationship that could have appeared to influence the work reported in this paper.

Data availability statement
These data were derived from the annual night-time inter-calibrated DMSP products available in the public domain by EOG:
https://eogdata.mines.edu/wwwdata/dmsp/v4composites_rearrange/

References

Agarwal P., D. Sahoo, Y. Parida, P. K. Ranjan and C. J. Roy (2023). Land use changes and natural disaster fatalities: Empirical analysis for India. Ecological Indicators, 154(March), 110525. https://doi.org/10.1016/j.ecolind.2023.110525

Aslam M. (2018). Design of sampling plan for exponential distribution under neutrosophic statistical interval method. IEEE Access, 6, 64153–64158. https://doi.org/10.1109/ACCESS.2018.2877923

Aufa B. Z., S. Suyanto and A. Arifianto (2020). Hyperparameter Setting of LSTM-based Language Model using Grey Wolf Optimizer. 2020 International Conference on Data Science and Its Applications, ICoDSA 2020. https://doi.org/10.1109/ICoDSA50139.2020.9213031

Bennie J., J. P. Duffy, T. W. Davies, M. E. Correa-Cano and K. J. Gaston (2015). Global trends in exposure to light pollution in natural terrestrial ecosystems. Remote Sensing, 7(3), 2715–2730. https://doi.org/10.3390/rs70302715

Combs C. L. and S. D. Miller (2023). A Review of the Far-Reaching Usage of Low-Light Nighttime Data. Remote Sensing, 15(3). https://doi.org/10.3390/rs15030623

Doll C. N. H. (2008). CIESIN Thematic Guide to Night-time Light Remote Sensing and its Applications. Earth Science, 1–41.

Elvidge C. D., K. E. Baugh, E. A. Kihn, H. W. Kroehl and E. R. Davis (1997). Mapping city lights with nighttime data from the DMSP Operational Linescan System. Photogrammetric Engineering and Remote Sensing, 63(6), 727–734.

George M. (2023). The Indian economy : The journey to real five trillion GDP and above. 4(2), 46–62.

Henderson J. V., A. Storeygard and D. N. Weil (2009). Nber Working Paper Series Measuring Economic Growth From Outer Space. National Bureau of Economic Research. http://www.nber.org/papers/w15199

Hsu F. C., K. E. Baugh, T. Ghosh, M. Zhizhin and C. D. Elvidge (2015). DMSP-OLS radiance calibrated nighttime lights time series with intercalibration. Remote Sensing, 7(2), 1855–1876. https://doi.org/10.3390/rs70201855

https://censusindia.gov.in/census.website/

https://data.worldbank.org/indicator/EG.USE.ELEC.KH.PC?locations=IN

https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?locations=IN

https://eogdata.mines.edu/wwwdata/dmsp/v4composites_rearrange/

Jindal M., P. K. Gupta and S. K. Srivastav (2022). Improved DMSP nighttime light monthly products over India. International Journal of Remote Sensing, 00(00), 1–30. https://doi.org/10.1080/01431161.2022.2152758

Kiran Chand T. R., K. V. S. Badarinath, C.D. Elvidge and B. T. Tuttle (2009). Spatial characterization of electrical power consumption patterns over India using temporal DMSP-OLS night-time satellite data. International Journal of Remote Sensing, 30(3), 647–661. https://doi.org/10.1080/01431160802345685

Krishna Prasad V., Y. Kant, P. K. Gupta, C. D. Elvidge and K.V. S. Badarinath (2002). Biomass burning and related trace gas emissions from tropical dry deciduous forests of India: A study using DMSP-OLS data ground-based measurements. International Journal of Remote Sensing, 23(14), 2837–2851. https://doi.org/10.1080/01431160110109598

Ma T., C. Zhou, T. Pei, S. Haynie and J. Fan (2012). Quantitative estimation of urbanization dynamics using time series of DMSP/OLS nighttime light data: A comparative case study from China's cities. Remote Sensing of Environment, 124, 99–107. https://doi.org/10.1016/j.rse.2012.04.018

Mukherjee S., S. K. Srivastav, P. K. Gupta, N. A. S. Hamm and V. A. Tolpekin (2017). An Algorithm for Inter-calibration of Time-Series DMSP/OLS Night-Time Light Images. Proceedings of the National Academy of Sciences India Section A - Physical Sciences, 87(4), 721–731. https://doi.org/10.1007/s40010-017-0444-8

Pandey B., P. K. Joshi and K. C. Seto (2013). Monitoring urbanization dynamics in india using DMSP/OLS night time lights and SPOT-VGT data. International Journal of Applied Earth Observation and Geoinformation, 23(1), 49–61. https://doi.org/10.1016/j.jag.2012.11.005

Pandey B., Q. Zhang and K. C. Seto (2017). Comparative evaluation of relative calibration methods for DMSP/OLS nighttime lights. Remote Sensing of Environment, 195, 67–78. https://doi.org/10.1016/j.rse.2017.04.011

Sahoo S., P. K. Gupta and S. K. Srivastav (2020). Comparative analysis between VIIRS-DNB and DMSP-OLS night-time light data to estimate electric power consumption in Uttar Pradesh, India. International Journal of Remote Sensing, 41(7), 2565–2580. https://doi.org/10.1080/01431161.2019.1693077

Selvin, R. Vinayakumar, E. A. Gopalakrishnan, V. K. Menon and K. P. Soman (2017). Stock price prediction using LSTM, RNN and CNN-sliding window model. 2017 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2017, 2017-Janua, 1643–1647. https://doi.org/10.1109/ICACCI.2017.8126078

Shen X., B. Liu, Y. Zhou and J. Zhao (2020). Remote sensing image caption generation via transformer and reinforcement learning.

Sherstinsky A. (2020). Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network. 404(March), 1–43.

Staudemeyer R. C. and E. R. Morris (2019a). a tutorial into Long Short-Term Memory Recurrent Neural Networks. 1–42.

Staudemeyer R. C. and E. R. Morris (2019b). Understanding LSTM -- a tutorial into Long Short-Term

Memory Recurrent Neural Networks. 1–42. http://arxiv.org/abs/1909.09586

Tan M., X. Li, S. Li, L. Xin, X. Wang, Q. Li, W. Li, Y. Li and X. Xiang (2018). Modeling population density based on nighttime light images and land use data in China. Applied Geography, 90(January 2017), 239–247. https://doi.org/10.1016/j.apgeog.2017.12.012

Taneja S., M. Bhatnagar, P. Kumar and R. Rupeika-Apoga (2023). India's Total Natural Resource Rents (NRR) and GDP: An Augmented Autoregressive Distributed Lag (ARDL) Bound Test. Journal of Risk and Financial Management, 16(2). https://doi.org/10.3390/jrfm16020091

Wang Q., Y. Guo, L. Yu and P. Li (2017). Earthquake Prediction based on Spatio-Temporal Data Mining : An LSTM Network Approach. 6750(c), 1–11. https://doi.org/10.1109/TETC.2017.2699169

Xu T., D. Zhou and Y. Li (2022). Integrating ANNs and Cellular Automata – Markov Chain to Simulate Urban Expansion with Annual Land Use Data. 1–15.

Zhu Y., C. Geiß, E. So and Y. Jin (2021). Multitemporal Relearning With Convolutional LSTM Models for Land Use Classification. 14, 3251–3265. https://doi.org/10.1109/JSTARS.2021.3055784