Relationship between Night-Time Lights detected by VIIRS on board SNPP Satellite and Atmospheric Anthropogenic Emissions

Prabhat Kumar Trivedi

Central University of Rajasthan, Bandarsindri, NH-8, Ajmer, Rajasthan-305817

Vikas Singh, Akash Biswal, Amit P. Kesarkar

National Atmospheric Research Laboratory, Department of Space : Government of India, Gadanki, Andhra Pradesh-517112

Table of Contents

Relationship between Night-Time Lights detected by VIIRS on board SNPP Satellite and Atmospheric Anthropogenic Emissions

Prabhat Kumar Trivedi

Central University of Rajasthan, Bandarsindri, NH-8, Ajmer, Rajasthan-305817

Vikas Singh, Akash Biswal, Amit P. Kesarkar

National Atmospheric Research Laboratory, Department of Space : Government of India, Gadanki, Andhra Pradesh-517112

Abstract

The rapid economic development and pollution growth have led to an increase in pollution and hence climate change. Most of the pollution is attributed to anthropogenic activities. These activities are responsible for the change in the global land use pattern, mainly related to the human settlement, which continues to increase as evident by the satellite imagery. These signals can clearly be seen in the global night-lights datasets derived from the Visible Infrared Imaging Radiometer Suite (VIIRS) on board Suomi National Polar-orbiting Partnership (Suomi NPP or SNPP) satellite. The VIIRS Day/Night Band (DNB) collects global low-light imaging data. One of the prominent features of DNB data is the detection of electric lighting present on the Earth's surface. Most of these lights are from human settlements. VIIRS collects source data that could be used to generate monthly and annually science grade global radiance maps of human settlements with electric lighting. These night-time light products have emerged as one of the widely used geospatial data as it shows the locations where artificial lighting is present and hence the human settlement and activities. The low-light imaging data collectd by VIIRS are processed using Geographic Information System (GIS). One of the most important applications of these products is that they help us to understand the relationship between environmental impacts and the human activities on a temporal and spatial scale without relying on the national statistics. As most of the atmospheric emissions are of anthropogenic origin, there may exist a relation between the night-time light data and the atmospheric anthropogenic emissions. In this work, we attempt to explore and examine the relationship between the anthropogenic emissions with the night-time light data. We use the Emissions Database for Global Atmospheric Research (EDGAR) data of anthropogenic emissions. The statistical analysis results show a positive correlation between the night-time light data and the emission data. The correlation coefficient is found to be varying from 0.34 to 0.55 which is minimum for NH_3 and highest for CO. The sector-wise analysis for the CO emission shows emission from Energy for Buildings sector to be better correlated compared to other sectors. The initial results of the statistical analysis suggests that night-time light data can be a good proxy to estimate the anthropogenic emissions where detailed national statistics are not available and will help to develop an appropriate statistical model to estimate the anthropogenic emissions for air quality modelling.

Keywords or phrases: VIIRS, Night-Time Light, Statistics, Emission, Air Quality

1 Introduction

The rising pollution and deteriorating air qualitty are the main concern of today's world as it impacts human health and earth's climate. As per the WHO report (WHO Global Ambient Air Quality Database 2018, Kota, et al, 2018, Sun, et al, 2018), most of the polluted cities are in China and India. The rising levels of pollution are linked to reduced life expectancy (Cohen, et al, 2017) and mortality (Chen, et al, 2013). It has been estimated that 4.2 million deaths in the world (WHO Ambient air pollution: Health impacts) and 1.2 million deaths in India are attributed to the air pollution (Balakrishnan, et al, 2019).

In India, industrialization and urbanization have affected the environment to a great extent. Most of the pollution is attributed to anthropogenic activities. These man-made activities pollute the air by releasing gases, smoke, fumes and dust (Sharma, et al, 2013). Currently, greenhosue gas emissions from the combustion of fossil fuels have reached unprecedented highs (Ma & Jiang, 2019). There is a strong linkage with the man-made activities and the amount of emissions of the pollutants. The human activities continue to increase because of the increase in the population and hence the pollution. These activities and change in the global land use patterns are maily related to the human settlement, which continues to increase as evident by the satellite imagery (Proville, et al, 2017), mainly night-time light data. The presumption is that most social and economic activities at night require light; hence the intensity of NIght-Time Lights and the area they cover shiould correlate with socio-economic indicators, economic development and emissions. These signals are very clearly seen in the global night-time lights datasets derived from Visible Infared Imaging Radiometer Suite (VIIRS) on board Suomi National Polar-orbiting (SNPP). Night-time images provide a continuous, relatively accurate, affordable and direct way to identify human activities and their spatial characteristics (Zhang, et al, 2015). One of the most important applications of these products is that they help us to understand the relationship between environmental impacts and the human activities on a temporal and spatial scale without relying on the national statistics. As most of the atmospheric emissions are of anthropogenic origin, there may exist a relation between the night-time light data and the atmospheric anthropogenic emissions.

The main objective of this work is to study the relationship between the night-time light and the anthropogenic emissions, so we can use the night-time light data as a proxy to study the anthropogenic emissions which can be used to improve the emission estimates where detailed national statistics are not available.

2 The study area and data used

This work is focused on India as the data from the state pollution control board suggest that the pollution levels in most of the cities of India are above the National Ambient Air Quality (NAAQ) standards. Study of the relation between the satellite products and anthropogenic emissions can provide environmentalist and the scientific community to better estimate the spatial heterogeneity in the emissions due to human activities.

In this study, we use the night-time lights as recorded by Visible Infrared Imaging Radiometer Suite (VIIRS) Day/Night Band and the anthropogenic emissions from EDGAR (Emissions Database for Global Atmospheric Research). We have analyzed and visualized the data using univariate and bivariate statistics to study the relationship between the night-time light and anthropogenic emissions.

Fig 1 Global Satellite Image of Night-Light for the year 2012

2.1 VIIRS Night-Time Light Data

Satellite -observed Night-Time lights have emerged as one of the most important geospatial data products to study human activities on Earth (Elvidge, et al, 2017). These products show the locations where artificial lighting is present and a measure of the brightness as observed from space (fig.1). From 1992 to 2013, there is a consistently processed annual time series of night-time lights processed from low-light imaging data collected by the US Air Force Defense Meteorological Satellite Program (DMSP) Operational Linescan System (OLS). DMSP data are the longest-running time series of night-time lights, dating back to 1992. At a large scale, DMSP was useful for most of the thing from generating detailed CO_2 emission maps to creating innovative development indices, to estimating natural gas flaring trends. The global studies based on the DMSP data have explored the basic links and correlation between these data and other well-documented variables, such as population, CO_2, GDP and electric power consumption (Proville, et al, 2017). These relationships signify the insight into the value of using night time lights as the proxies for anthropogenic activities-both on the economical ground and environmental ground.

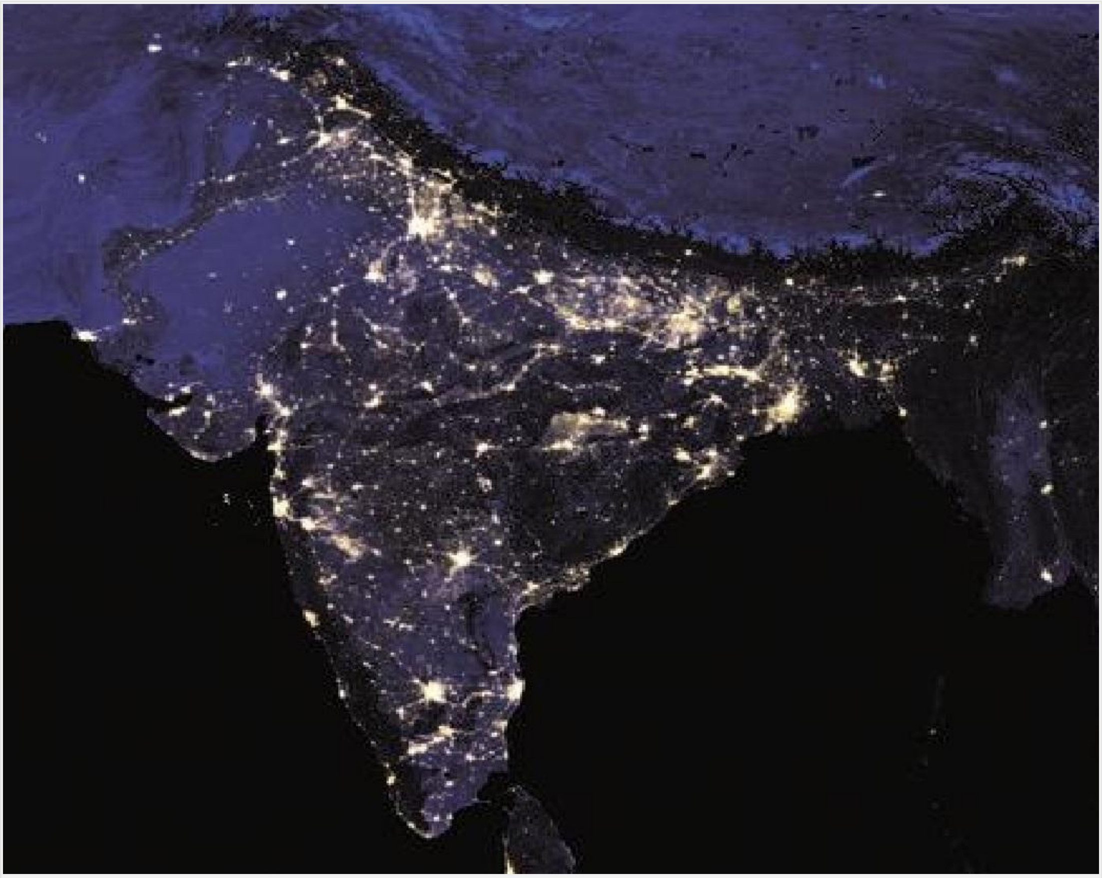

Fig 2 Night-Light Image of India for the year 2012

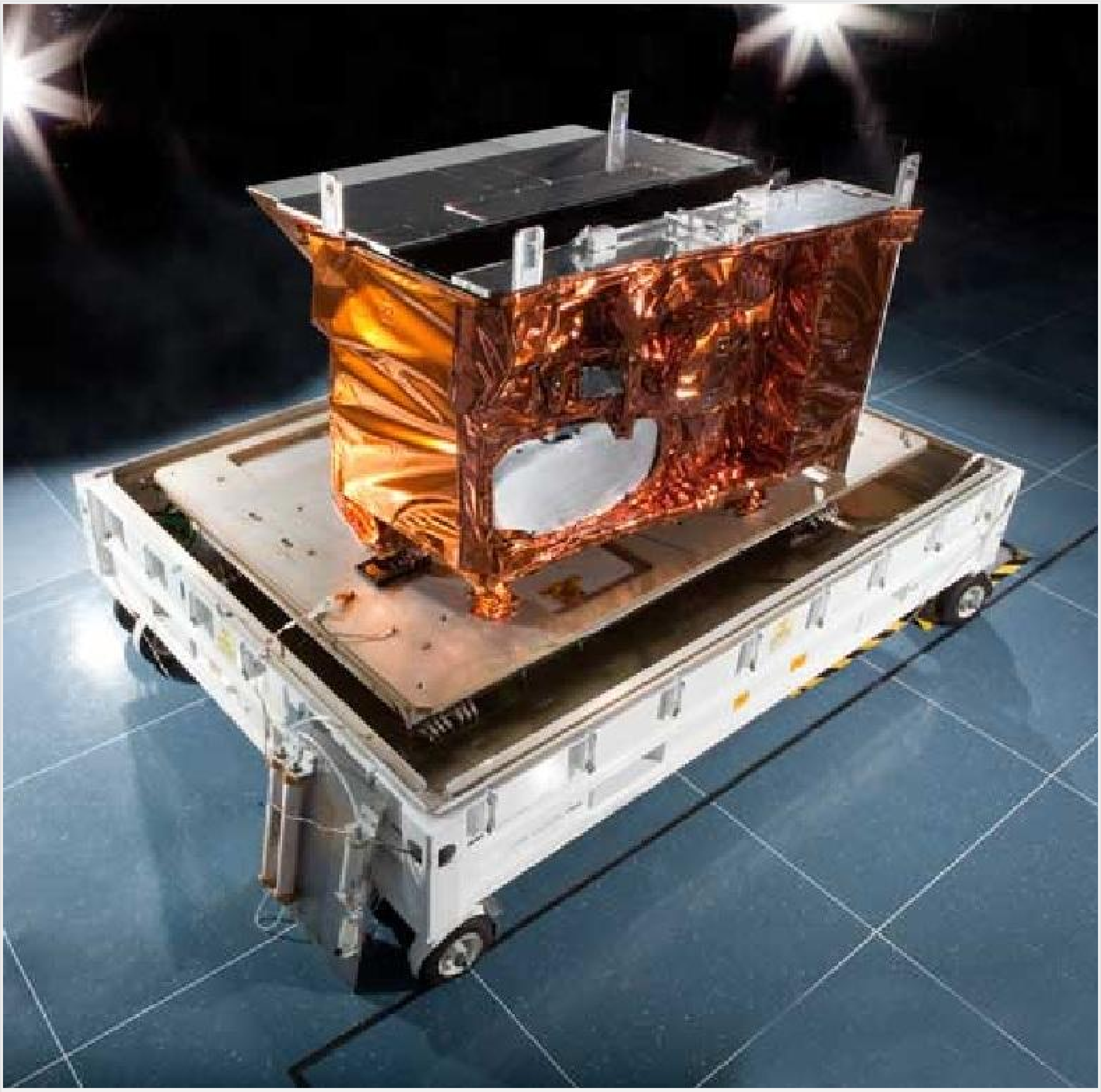

Fig 3 Visible Infrared Imaging Radiometer Suite

The follow on to DMSP for global low-light imaging of the Earth at night is the Visible Infrared Imaging Radiometer Suite (VIIRS) Day/Night Band (DNB), flown jointly by NASA and NOAA. It is the predecessor of DMSP and was launched in 2011, while DMSP was launched in the year 1992. The VIIRS sensor was designed to extend and improve upon the series of measurements initiated by its predecessor, DMSP.

The Visible Infrared Imaging Radiometer Suite (VIIRS) is a sensor designed and manufactured by the Raytheon Company on board the Suomi National Polar-Orbiting

Partnership (Suomi NPP) and NOAA-20 weather satellites. VIIRS is one of five key instruments onboard Suomi NPP, launched on October 28, 2011.

VIIRS is a whiskbroom scanning radiometer that collects imagery and radiometric measurements of the land, atmosphere, cryosphere, and oceans in the visible and infrared bands of the electromagnetic spectrum.

Suomi NPP's orbit allows VIIRS to collect new night-time light data for almost all of the Earth every night. With each orbit, it adds to an ever-growing archive of data that is allowing scientists and geographers to track changes in artificial lights, fishing practices, economic activity, development patterns, the movement of goods and people, and many other research areas in innovative ways and on a global scale.

Table 1 Specifications of VIIRS

Orbit	*830km, 1:30pm mean local solar time, sun-synchronous, polar*
Repeat Cycle	*16 days*
Swath Dimensions	*3000 km, nearly global coverage every day*
Weight	*275 kg*
Spatial Resolution	*750 m*
Data Rate	*5.9 Mbps*
Quantization	*12 bits*
Field of View	*deg*
Wavebands	*9 Visible/NIR bands plus day/night band* *8 mid IR* *4 LW IR*
Design Life	*7 years*
Duration	*Operational*

The VIIRS instrument is a 'whiskbroom' radiometer with 22 channels ranging from 0.412 µm to 12.01 µm. Five of these channels are high-resolution imagery bands or 'I-bands,' as they are commonly referred to, and sixteen are designed as moderate-resolution bands or 'M-bands'. One of these M-bands is the Day/Night band, or DNB, that is a panchromatic band sensitive to visible and near-infrared wavelengths, and with this band VIIRS is able to observe nighttime lights on Earth with better spatial and temporal resolution compared to DMSP. It allows for satellite views in the visible portion of the electromagnetic spectrum at night. It has unprecedented night observation capability. It is superior to its predecessor Operational Line

Scanner (OLS) on the DMSP in both spatial and radiometric performance because it has a finer spatial resolution (Schueler, et al, 2013).

VIIRS collects a complete set of night-time images of the Earth every 24 hours. VIIRS's main objectives include the monitoring and investigation of changes and properties in vegetation, land cover/use, the hydrologic cycle, and the earth's energy budget over both global and regional scales. This information is useful in furthering our understanding of global climate change.

2.2 Emission Database for Global Atmospheric Research (EDGAR)

The Emissions Database for Global Atmospheric Research (EDGAR) provides global past and present-day anthropogenic emissions of greenhouse gases and air pollutants by country and at 0.1 deg X 0.1 deg spatial grid resolution. Anthropogenic emissions are the emissions of greenhouse gases and aerosols due to human activities and their settlements.

EDGAR aims to inform scientists and policymakers on the evolution of the emission inventories over time for all world countries. The total includes emissions from different sectors such as road transportation, industrial sector, the agricultural sector, etc. As an example, the global CO2 emission for the year 2009 from different sectors is shown in Figure 4.

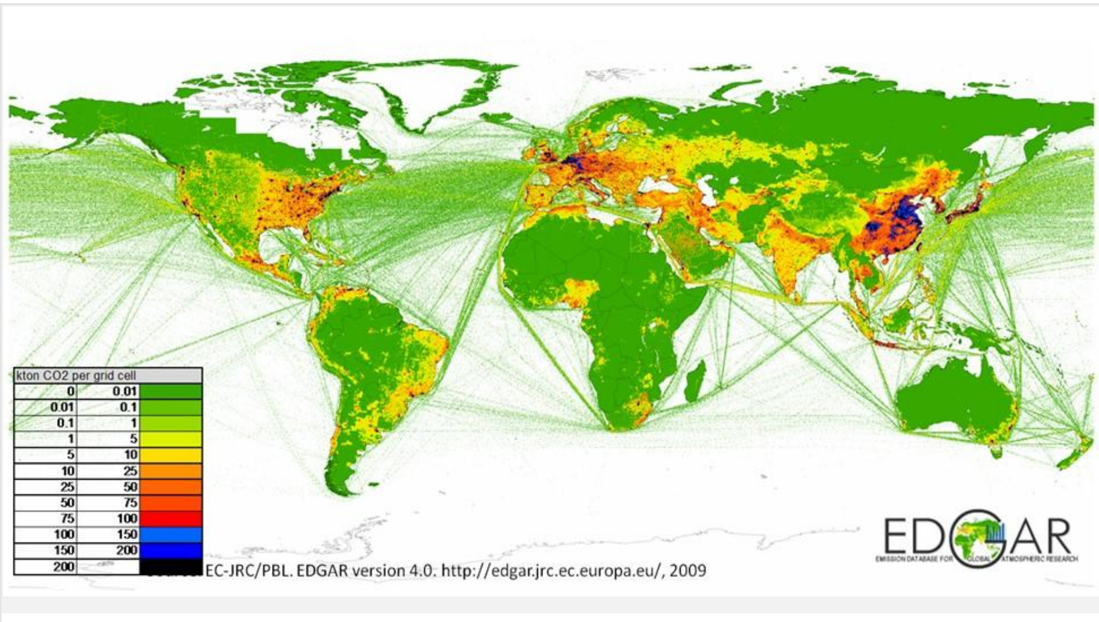

Fig 4 Global Image of CO emisssion for the year 2009

2.3 Atmospheric Pollutants

The following EDGAR anthropogenic emissions are considered to study the relationship between emissions and night-time light data. The spatial distribution of the emissions is shown in Figure 5. The highest emission is seen over the populated area such as megacities and the Indo-Gangetic plain (IGP) region.

- **Black Carbon (BC):**

It is a component of fine particulate matter ($\leq$2.5 µm). It consists of pure carbon in several linked forms. It is emitted due to the combustion of fossil fuels, biomass and biofuel.

- **Carbon Monoxide (CO):**

It is a colorless, odorless, and tasteless flammable gas that is slightly less dense than air. About half of the carbon monoxide in Earth's atmosphere is from the burning of fossil fuels and biomass.

- **Ammonia (NH_3):**

It is a compound of nitrogen and hydrogen. A colorless gas with a pungent smell. The largest source of NH_3 emissions is agriculture, including animal husbandry and NH_3- fertilizer applications. Other sources of NH_3 include industrial processes, vehicular emissions and volatilization from soils and oceans. It plays a significant role in the formation of atmospheric particulate matter.

- **Non-Methane Volatile Organic Compound (NMVOC):**

It is a collection of organic compounds that differ widely in their chemical composition but display achiral behavior in the atmosphere. It gets emitted into the atmosphere majorly from a large number of sources including combustion activities, solvent use and production processes.

- **Nitric Oxide and Nitrogen Dioxide (NO_x):**

It is produced when nitrogen and oxygen gases react in the air during combustion, at high temperature. Around 40% of NO_x emission is due to road transport.

- **Organic Compound (OC):**

It is a particulate aerosol formed by incomplete combustion. OC is formed when biomass fuels, such as wood, are incompletely combusted and fossil fuels, such as oil and coal are incompletely combusted. It has potentially significant impacts on climate change.

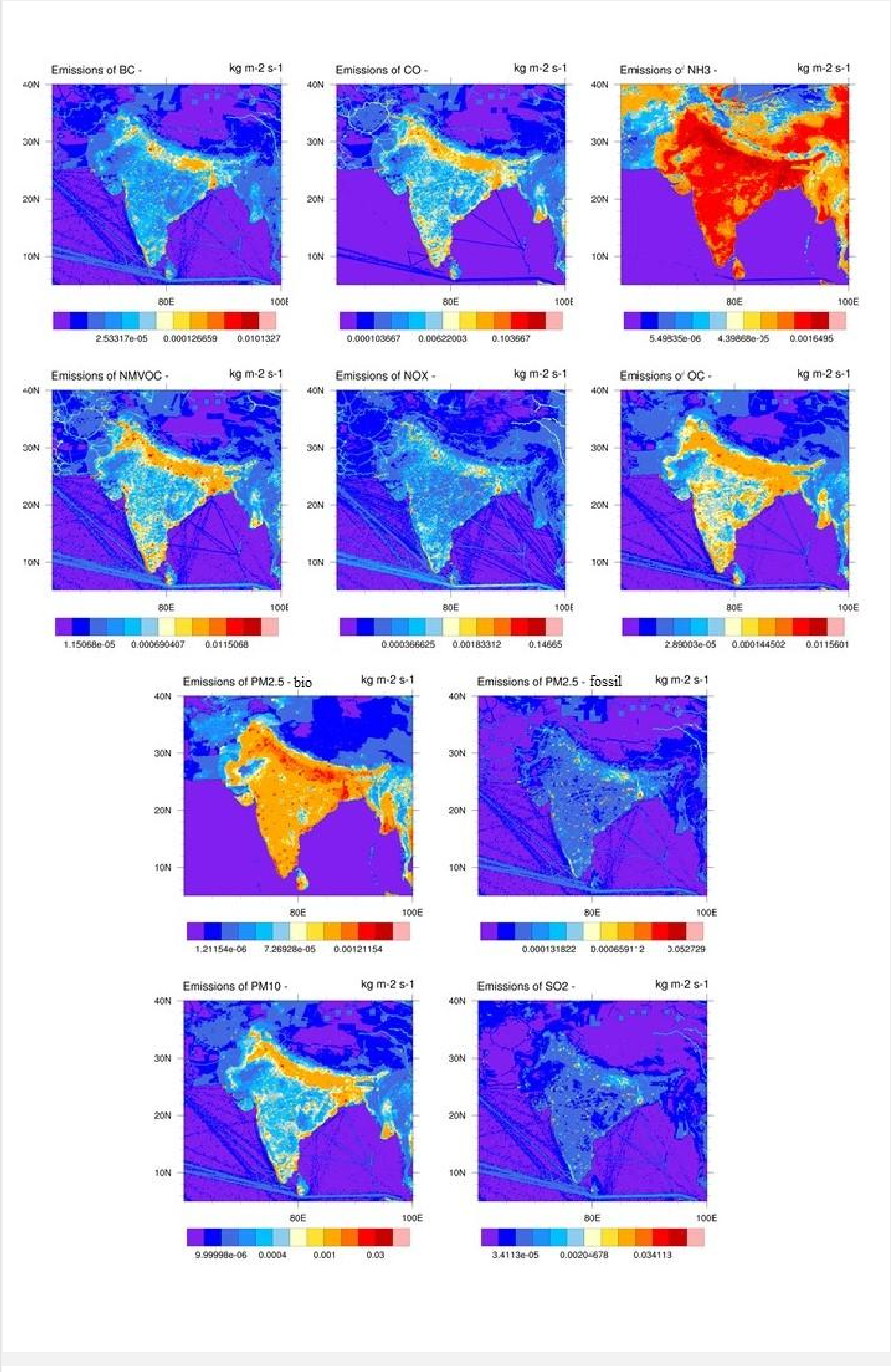

Fig 5 The Emissions of different pollutants over India for the year 2012

- **$PM_{2.5}$:**

It is an atmospheric particulate matter (PM) that have a diameter of fewer than 2.5 micrometers. Emissions of $PM_{2.5}$ from road vehicles are an important source. These fine particles are known to trigger or worsen chronic diseases such as asthma, heart attack and other respiratory problems.

- **PM_{10}:**

It is a particulate matter 10 micrometers or less in diameter. High levels of PM_{10} particles in the air can irritate the eyes and throat. Common sources of PM_{10} particles include sea salt, pollen and combustion activities such as motor vehicles and industrial processes. Dust from unsealed roads is a major source of PM_{10} particles.

- **Sulphur Dioxide (SO_2):**

It has a nasty sharp smell. It reacts easily with other substances to form harmful compounds, such as sulfuric acid, sulfurous acid and sulfate particles. About 99% of the sulfur dioxide in air comes from human sources. The main source of sulfur dioxide in the air is an industrial activity that processes materials that contain sulfur, e.g. the generation of electricity from coal, oil or gas that contains sulfur.

3 **Methodology**

3.1 Data analysis algorithm

We have carried out the data analysis and developed the algorithm using the R programming language (Ihaka & Gentleman, 1996). It is a software environment for statistical analysis, graphic representation and reporting. R provides a wide variety of statistical analysis and graphical techniques and is highly extensible.

We read both the night-time light and emission data, and arranged them with respect to longitude and latitude. Then we arranged the data for the collocated latitudes and longitudes. Pre-processing of the data has been carried out by cleaning the data that include removing missing values, if any, removing zero values from the data and removing outliers above 99 percentile of the data for unbiased results.

From the resulted co-located data sets we find the univariate statistics (such as mean, median, min, max, standard deviation, etc.) and plot their distribution to understand the variability in the data, its shape and its spread. Then we computed the bivariate statistics (such as covariance, correlation, etc.) to examine the relationship between the night-time light and emission. The pollutants showing higher correlation exhibits better relation with the night-time light data. This process is repeated for each pollutant and for the different sectors of the emission. The flowchart of the developed algorithm is shown in Figure 6.

While using R, the following libraries have been used to analyze and investigate the data.:

- **ncdf4** – It is an Unidata's netCDF library. Using this package, netCDF files (NC files) can be opened and data sets can be read in easily. It is also easy to create new netCDF dimensions, variables, and files (Michna & Woods, 2013).

- **raster** – This package is used for reading, writing, manipulating, analyzing and modeling of gridded spatial data. The package implements basic and high-level functions. Processing of very large files is supported. It helps to compute the coefficient of variation of a data just by simple command "cv" (Hijmans, et al, 2015).

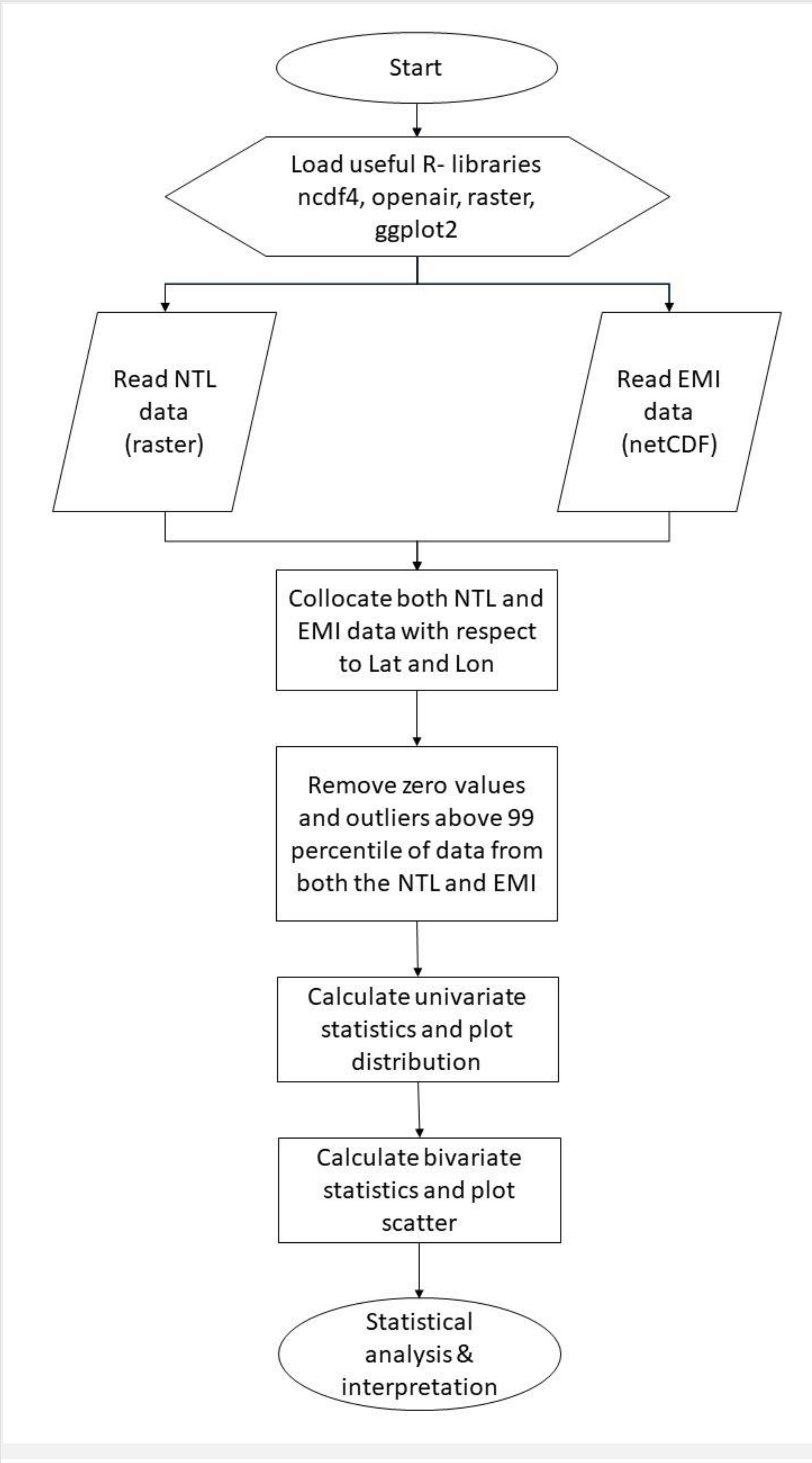

Fig 6 Flowchart of the algorithm

- **openair** – This package is used for the purpose of analyzing air quality data. This helps to plot density scatters with proper color labels (Ropkins & Carslaw, 2012).

- **ggplot2** – It is an enhanced package for 'declaratively' creating graphics. It provides better visualization of the data. It is a powerful package to draw graphics (Heer & Shneiderman, 2012).

In addition to R, we also use QGIS (Geographical Information System) to process the data (Steiniger & Hay, 2009).

3.2 Statistical Indexes

For this study, univariate statistics are used to investigate the data and to visualize it and bivariate statistics are used to establish the relationship between the night-time light and the emission data. The following are the statistics used with their interpretation:

- **Maximum and Minimum:**

The maximum is the largest data value. The minimum is the smallest data value.
We use the maximum & minimum to identify a possible outlier or a data entry error.

- **Range:**

The range is the difference between the largest and smallest data values in the sample. The range represents the interval that contains all the data values. We use the range to understand the amount of dispersion in the data. A large range value indicates greater dispersion in the data. A small range value indicates that there is less dispersion in the data. Because the range is calculated using only two data values, it is more useful with small data sets. We use the maximum & minimum to identify a possible outlier or a data entry error.

$$Range = Max - Min$$

- **Mean:**

The mean is the average of the data, which is the sum of all the observations divided by the number of observations. We use the mean to describe the sample with a single value that represents the center of the data.

$$\mu = \frac{\sum_i^n x_i}{n}$$

- **Median:**

The median is the midpoint of the data set. The median and the mean both measure central tendency. But unusual values, called outliers, affect the median less than they affect the mean. If your data is symmetric, the mean and median are similar. And since our data is not symmetric we have to take use of both the central tendencies.

- **Standard Deviation:**

The standard deviation is the most common measure of dispersion, or how spread out the data are about the mean. We use the standard deviation to determine how spread out the data are from the mean. A higher standard deviation value indicates greater spread in the data. The standard deviation can also be used to establish a benchmark for estimating the overall variation of a process.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}}$$

- **Variance:**

The variance measures how spread out the data are about their mean. The variance is equal to the standard deviation squared. The greater the variance, the greater the spread in the data. Because variance (σ^2) is a squared quantity, its units are also squared, which may make the variance difficult to use in practice. The standard deviation can be easier to use because it is a more intuitive measurement.

$$\sigma^2 = \frac{\sum_{i=1}^{n} |x_i - \overline{x}|}{n}$$

- **Coefficient of variation:**

The coefficient of variation is a measure of spread that describes the variation in the data relative to the mean. The coefficient of variation is adjusted so that the values are on a unitless scale. Because of this adjustment, you can use the coefficient of variation instead of the standard deviation to compare the variation in data that have different units or that have very different means. The larger the coefficient of variation, the greater the spread in the data.

$$CV = \frac{\sigma}{\mu} \times 100$$

- **Covariance:**

Covariance indicates the relationship between two variables whenever one variable changes. If an increase in one variable results in an increase in the other variable, both variables are said to have positive covariance. Decreases in one variable also cause a decrease in the other. Both variables move together in the same direction when they change. Decreases in one variable resulting in the opposite change in the other variable are referred to as negative covariance. These variables are inversely related and always move in different directions. When a positive number is used to indicate the magnitude of covariance, the covariance is positive. A negative number represents an inverse relationship. The concept of covariance is commonly used when discussing relationships between two economic indicators or terms.

$$Cov(x, y) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{n}$$

- **Correlation:**

Correlation refers to a technique used to measure the relationship between two or more variables. When two things are correlated, it means that they vary together. A positive correlation means that high scores on one are associated with high scores on the other, and that low scores on one are associated with low scores on the other. Negative correlation, on the other hand, means that high scores on the first thing are associated with low scores on the second.

$$Correlation = \frac{Cov(x,y)}{\sigma_x \sigma_y}$$

- **Coefficient of Determination:**

The coefficient of determination is a key output of regression analysis. It can be interpreted as the proportion of variation in the dependent variable that is predictable from the independent variable.

$$R^2 = (correlation\ coefficient)^2$$

4 Results and Discussion

The numbers of co-located data points used to calculate the statistics between the night-time light data for different pollutants are shown in table 2:

Table 2 Co-located data points after data pre-processing

Pollutants	**Data_Points**
BC	*25105*
CO	*25135*
NH_3	*25040*
NMVOC	*25128*
NO_x	*25071*
OC	*25111*
$PM_{2.5bio}$	*25111*
$PM_{2.5\ fossil}$	*25061*
PM_{10}	*25089*
SO_2	*25061*

The statistical distribution of each pollutant emissions is shown in Figure 7. It can be seen that distribution of night-time light data and emission data are positively skewed and there is a good similarity between the distribution of night-time light and distribution of BC, CO, NMVOC, $PM_{2.5\ bio}$, PM_{10} emissions, which is helpful to examine the relationship between the night-time light and emissions. The calculated mono-variate statistics are tabulated in Table 3.

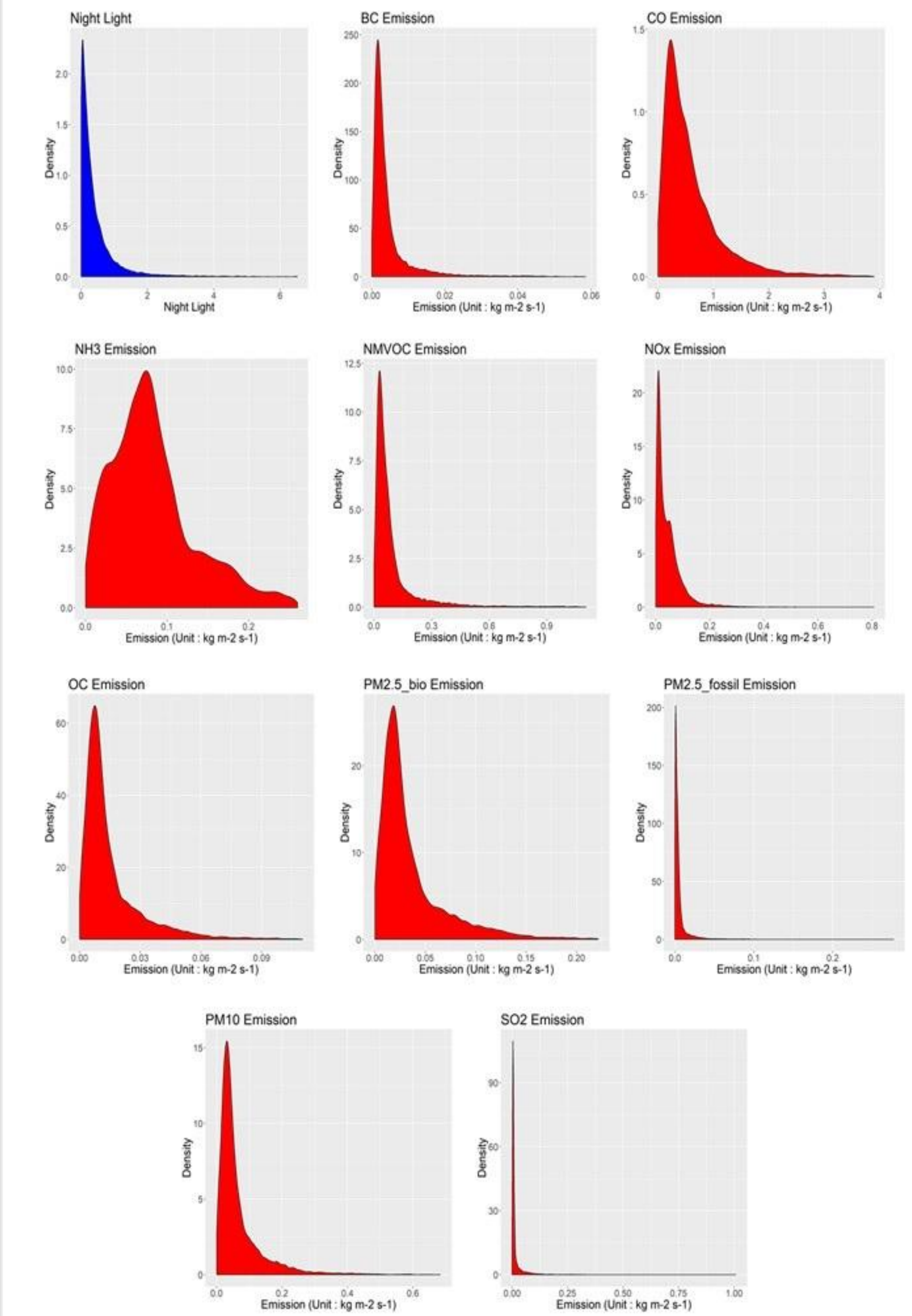

Fig 7 Distribution of Night Light and different emissions

Table 3 Mono-variate statistics for different pollutants

Pollutants	**Mean_NTL *(Radiance)***	**Mean_EMI *(kg m-2 s-1)***	**Median_EMI *(kg m-2 s-1)***	**SD_NTL *(Radiance)***	**SD_EMI *(kg m-2 s-1)***	**Cov_EMI**
BC	*0.44*	*0.005*	*0.003*	*0.67*	*0.007*	*139.8*
CO	*0.44*	*0.603*	*0.441*	*0.67*	*0.553*	*91.6*
NH3	*0.45*	*0.083*	*0.075*	*0.69*	*0.053*	*63.4*
NMVOC	*0.44*	*0.095*	*0.053*	*0.67*	*0.135*	*142.2*
NOx	*0.44*	*0.047*	*0.031*	*0.67*	*0.056*	*119.5*
OC	*0.44*	*0.015*	*0.010*	*0.67*	*0.015*	*99.7*
PM2.5 bio	*0.44*	*0.036*	*0.024*	*0.67*	*0.034*	*95.8*
PM2.5 fossil	*0.44*	*0.008*	*0.003*	*0.67*	*0.020*	*258.6*
PM10	*0.44*	*0.068*	*0.043*	*0.67*	*0.076*	*111.6*
SO2	*0.44*	*0.023*	*0.006*	*0.67*	*0.067*	*292.4*

SD – Standard Deviation; CoV – Coefficient of variation

From table 3 and figure 7 it can be observed that for night-time light data is centered at 0.44. The center of the distribution of each pollutant is shown in table 3. It can be seen that the data is not symmetric for any of the pollutant so using the values of mean and median for each pollutant from table 3 and their distribution curves from figure 7 it can be seen that there is a good similarity between the distributions of night-time light and emission data. It can also be analyzed from the values of standard deviation that night-time light data has 67% variation from its mean and the data for CO emission is more deviated from its mean compared to other pollutants, however using the values of coefficient of variations, data for SO_2 emission and $PM_{2.5\ fossil}$ have the greatest spread from their means, NH_3 is showing less variation (i.e. around 63%) variation from its relative mean in comparison of other pollutants.

Table 4 Bivariate statistics for different pollutants

Pollutants	**Covariance *(Radiance kg m^{-2} s^{-1})***	**Correlation**	***R^2**
BC	*0.002*	*0.51*	*0.26*
CO	*0.201*	*0.55*	*0.30*
NH_3	*0.012*	*0.34*	*0.11*
NMVOC	*0.046*	*0.51*	*0.26*
NO_x	*0.020*	*0.52*	*0.27*
OC	*0.005*	*0.50*	*0.25*
$PM_{2.5\ bio}$	*0.011*	*0.46*	*0.22*
$PM_{2.5\ fossil}$	*0.006*	*0.42*	*0.18*
PM_{10}	*0.026*	*0.51*	*0.26*
SO_2	*0.018*	*0.40*	*0.16*

*R^2 – Coefficient of Determination

The bivariate statistics, covariance, correlation and coefficient of determination of the considered pollutant emissions are shown in table 4. The corresponding scatter is shown in Figure 8. The covariance and correlation of all the pollutants are positive, it implies that night-time light and emissions both are varying in the same direction whenever one quantity changes. The correlation coefficient varies from 0.34 to 0.55. It is found to be minimum for NH_3 and highest for CO. The pollutants of anthropogenic origin show correlation more than 0.5. The coefficient of determination suggests that 30% of the variations in the CO emissions can be predicted by the night-time light and only 11% of the variations in the NH_3 emissions can be predicted by the night light using linear regression model.

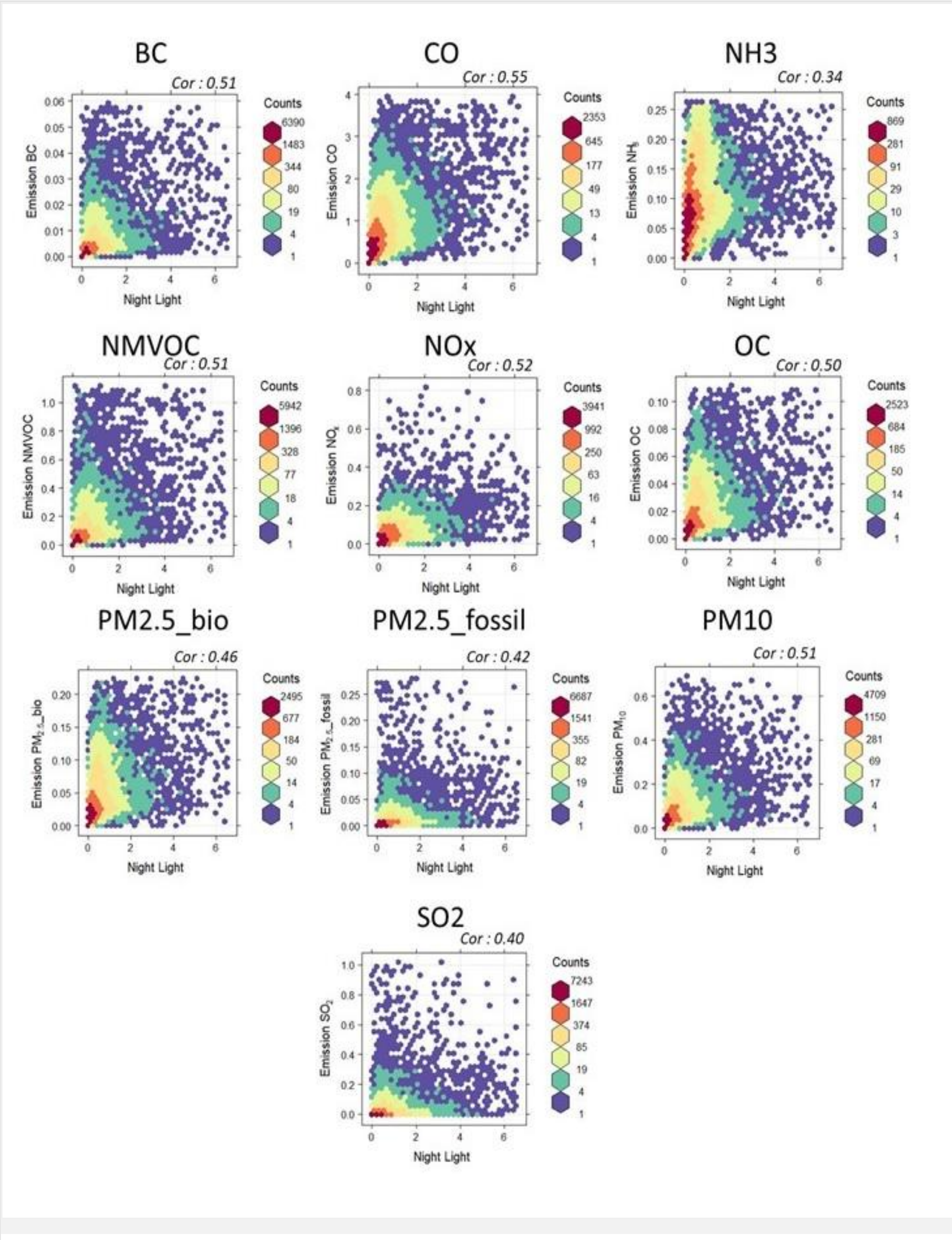

Fig 8 Scatter Plots of different pollutants

We also study the relation of night-time light data emission from different sectors of the CO. The sectors under consideration and the number of valid data points along with bivariate statistics are shown in Table 5. The highest correlation is found for the Energy for Buildings emission sector.

Table 5 Bivariate Statistics for different sectors of CO emission

Sector_Name	**Data_Points**	**Covariance *(Radiance kg m^{-2} s^{-1})***	**Correlation**	**$^*R^2$**
Combustion for manufacturing	*5306*	*0.392*	*0.44*	*0.19*
Road Transportation	*16332*	*0.051*	*0.29*	*0.08*
Energy for Buildings	*25114*	*0.082*	*0.50*	*0.25*
Chemical Processes	*4694*	*0.000*	*0.49*	*0.24*
Agricultural Waste Burning	*24494*	*0.017*	*0.20*	*0.04*

$^*R^2$ – Coefficient of Determination

5 Conclusions

Statistical analysis has been carried out to study the relationship between human settlements as identified from the space and the anthropogenic emissions. The VIIRS Night-Time lights have been considered as a proxy for the human settlement. The EDGAR anthropogenic emissions of the pollutants over India has been processed. The pollutants considered here are Black Carbon (BC), Carbon Monoxide (CO), Ammonia (NH_3), Non-Methane Volatile Organic Compound (NMVOC), Nitric Oxide and Nitrogen Dioxide (NO_x), Organic Compound (OC), Particulate matter less than 2.5 μm ($PM_{2.5}$), Particulate matter less than 10 μm (PM_{10}) and Sulphur Dioxide (SO_2). A data analysis algorithm has been written in the R statistical programming language. The algorithm collocates the data according to their geographical coordinates and calculates various univariate and bivariate statistics which are used to examine the relationship between the night-time light and the anthropogenic emissions.

The statistical analysis suggests both the night-time light data and pollutants are positively skewed and similar in nature. The covariance and correlation of all the pollutants are positive with the night-time light, it implies that night-time light and emissions both vary in the same direction whenever one quantity changes. The correlation coefficient varies from 0.34 to 0.55. The correlation is found to be minimum for NH_3 and highest for CO. The pollutants of anthropogenic origin show correlation more than 0.5. The coefficient of determination suggests that 30% of the variations in the CO emissions can be predicted by the night-time light and only 11% of the variations in the NH3 emissions can be predicted by the night-time light using linear regression model.

We also study the relation of night-time light data emission from different sectors of the CO. The sectors considered here are Combustion for manufacturing, Road Transportation, Energy for Buildings, Chemical Processes and Agricultural Waste Burning. Among all, the highest correlation is found for the emissions from Energy for Buildings sector. Initial results suggest that night-time light data can be a good proxy to estimate the anthropogenic emissions where detailed national statistics are not available.

References

1. WHO Global Ambient Air Quality Database 2018 , accessed 11 July 2019, <https://www.who.int/airpollution/data/cities/en/>
2. Kota, S.H., Guo, H., Myllyvirta, L., Hu, J., Sahu, S. K., Garaga, R., ... & Zhang, H. (2018). Year-long simulation ofgaseous and particulate air pollutants in India. *Atmospheric environment*, *180*, 244-255.
3. Sun, C., Luo, Y., & Li, J. (2018). Urbantraffic infrastructure investment and air pollution: Evidence from the 83cities in China. *Journal of cleaner production*, *172*,488-496.
4. Cohen, A.J., Brauer, M., Burnett, R., Anderson, H. R., Frostad, J., Estep, K., ... &Feigin, V. (2017). Estimates and 25-year trends of the global burden of diseaseattributable to ambient air pollution: an analysis of data from the GlobalBurden of Diseases Study 2015. *The Lancet*, *389*(10082),1907-1918.
5. Chen, Y.,Ebenstein, A., Greenstone, M., & Li, H. (2013). Evidence on the impact ofsustained exposure to air pollution on life expectancy from China's Huai Riverpolicy. *Proceedings of the National Academy of Sciences*, *110*(32),12936-12941.
6. WHO Ambient air pollution: Health impacts, accessed 11 July 2019, <https://www.who.int/airpollution/ambient/health-impacts/en/>
7. Balakrishnan, K., Dey, S., Gupta, T., Dhaliwal,R. S., Brauer, M., Cohen, A. J., ... & Sabde, Y. (2019). The impact of airpollution on deaths, disease burden, and life expectancy across the states ofIndia: the Global Burden of Disease Study 2017. *The Lancet PlanetaryHealth*, *3*(1), e26-e39
8. Sharma, S.B., Jain, S., Khirwadkar, P., & Kulkarni, S. (2013). The effects of airpollution on the environment and human health. *Indian Journal ofResearch in Pharmacy and Biotechnology*, *1*(3), 391-396.
9. Ma, X.,& Jiang, Q. (2019). How to Balance the Trade-off between EconomicDevelopment and Climate Change? *Sustainability*, *11*(6),1638.
10. Proville,J., Zavala-Araiza, D., & Wagner, G. (2017). Night-time lights: A global,long term look at links to socio-economic trends. *PloS one*, *12*(3),e0174610.

11. Zhang, Q.,Levin, N., Chalkias, C., & Letu, H. (2015). Nighttime light remotesensing–monitoring human societies from outer space. *PS Thenkabail(Ed.)*, 289e310.
12. Elvidge, C. D., Baugh, K., Zhizhin, M., Hsu, F.C., & Ghosh, T. (2017). VIIRS night-time lights. *InternationalJournal of Remote Sensing*, *38*(21), 5860-5879.
13. Schueler,C. F., Lee, T. F., & Miller, S. D. (2013). VIIRS constant spatial-resolutionadvantages. *International journal of remote sensing*, *34*(16),5761-5777.
14. Ihaka, R.,& Gentleman, R. (1996). R: a language for data analysis and graphics. *Journalof computational and graphical statistics*, *5*(3), 299-314.
15. Michna, P.,& Woods, M. (2013). RNetCDF–A packagefor reading and writing NetCDF datasets. *The R Journal*, *5*(2),29-36.
16. Hijmans, R.J., van Etten, J., Cheng, J., Mattiuzzi, M., Sumner, M., Greenberg, J. A., ...& Hijmans, M. R. J. (2015). Package 'raster'. *R package*.
17. Ropkins,K., & Carslaw, D. C. (2012). openair-Data Analysis Tools for the AirQuality Community. *R Journal*, *4*(1).
18. Heer, J.,& Shneiderman, B. (2012). Interactive dynamics for visual analysis. *Queue*, *10*(2),30.
19. Steiniger,S., & Hay, G. J. (2009). Free and open source geographic information toolsfor landscape ecology. *Ecological informatics*, *4*(4),183-195.