ORIGINAL ARTICLE

Environmental risk factor assessment for major respiratory disorders in metropolitan cities of India using VIIRS Suomi Aerosol data and Google Trends

Diptarshi Mitra[1] · Shiva Reddy Koti[1] · Prabhakar Alok Verma[1] · Sameer Saran[1]

Abstract

This study has investigated the association between the amount of atmospheric aerosols and the occurrences of Asthma, Chronic Obstructive Pulmonary Disease (COPD) and Lung Cancer in Delhi, Mumbai, Chennai, Kolkata and Bengaluru. Aerosol Optical Thickness (AOT) data of Visible Infrared Imaging Radiometer Suite (VIIRS) and Google Trends (GT) have been used to acquire information regarding the abundance of atmospheric aerosols and the occurrences of the respiratory diseases respectively. The result of Granger causality test between AOT and GT has shown that Delhi, Mumbai and Chennai were quite vulnerable to the three respiratory diseases whereas Bengaluru did not display so much vulnerability to these ailments. Kolkata was not so much vulnerable to Asthma but did exhibit susceptibility to the other two diseases. GT is validated by correlating with Annual Morbidity data of Delhi. The result of Granger causality test between Particulate Matter (diameter ≤ 10 μm) (PM_{10}) data and GT validates the result of Granger causality between AOT and GT, and shows the trustworthiness of GT and AOT. Thus, this study also proves the usefulness of VIIRS AOT and GT as dependable sources of information on atmospheric aerosols and prevalence of the respiratory diseases respectively, and the effectiveness of Granger causality test as a tool of analysis in health and geographic information systems (GIS).

Introduction

A risk factor increases the likelihood of developing a disease (Parritz and Troy 2018; Dovjak and Kukec 2019). Environmental risk factor assessment involves the study and analysis of an entity, associated with the environment, which can increase the vulnerability to a disease. Chronic respiratory diseases are one of the major causes of concern about human health. Chronic respiratory diseases account for 4 million deaths across the globe annually. Major preventable chronic respiratory diseases include chronic obstructive pulmonary disease (COPD), asthma, lung cancer, Sleep Apnea Syndrome and pulmonary hypertension. (World Health Organization 2007).

Among the considerable number of respiratory ailments, COPD, asthma and lung cancer impose a significant burden on India (Bhome 2012; Jindal et al. 2012; Noronha et al. 2016). The abundance of atmospheric aerosols is one of the causative factors for COPD, asthma and lung cancer (Mowli and Subbayya 1989; Longo et al. 2011a; Kim and Criner 2013; Davidson 2014). Major risk factors for COPD include tobacco, and indoor and outdoor air pollution (Atkinson et al. 2014; DeVries et al. 2016; Li et al. 2016). Asthma is a very common disease across the world. Although deaths due to asthma have decreased in the past 25 years, still it is a matter of concern because of the unavailability of therapeutic regimens, and the fact that the reason for its prevalence is not yet well known (Beasley et al. 2015). Some of the research studies have investigated the risk factors for asthma and found them to be genetic, environmental, air pollutants, dietary factors and abnormal immunological responses (Oland et al. 2017; Törmänen et al. 2018; Ochoa-Avilés et al. 2020). COPD and

✉ Diptarshi Mitra
  diptarshimitra@yahoo.co.in

1   Geoinformatics Department, Indian Institute
    of Remote Sensing (IIRS), 4 Kalidas Road, Dehradun,
    Uttarakhand 248001, India

lung cancer are common in same patients; however, major risk factor for lung cancer is cigarette smoking, some of the other factors being genetic susceptibility, poor diet, occupational exposures and air pollution (Petty 2005; Malhotra et al. 2016; Akhtar and Gupta Bansal 2017).

Atmospheric aerosols are minute particles suspended in the atmosphere, caused by various natural and anthropogenic activities (National Aeronautics and Space Administration 1996; National Oceanic and Atmospheric Administration and United States Department of Commerce 2017). There is association between atmospheric aerosol and satellite derived Aerosol Optical Thickness (AOT). AOT is the degree to which aerosols prevent the transmission of light by absorption and/or scattering, and is a function of the wavelength of the light involved (Godin 2015). If the amount of aerosols in the atmosphere is more, the AOT value will be higher, and if the amount is less, the AOT value will be lower.

Generally, studies on respiratory disease surveillance use patient level data. However, collecting such data is quite difficult due to many legal hurdles and privacy concerns. To overcome this problem, we have used Google Trends (GT) as a proxy for patient level data. GT is a normalized score about the number of searches for a keyword, relative to the total number of searches for all possible keywords, in Google, from a region, in a particular time interval. Since Internet is a very common thing in metropolitan cities, therefore, if there is an increase in the occurrence of a disease in a region, people are expected to search more extensively for keywords related to that disease and/or its symptoms, in Google. By searching for keywords related to a specific disease and/or its symptoms, people want to know more about that disease. Accordingly, people take preventive measures like visiting doctor, changing lifestyle etc. Hence, the Google Trends data corresponding to a keyword related to a particular disease and/or its symptoms, can be expected to give an idea about the prevalence of this disease, in a particular region, during a specified time interval.

In this article, we have studied about the presence of atmospheric aerosols and its effect on the occurrences of COPD, asthma and lung cancer, in five metropolitan cities of India, during 2012–2017.

Here, PM_{10} (particulate matter with diameter ≤ 10 μm) and annual morbidity data have been used for validation purpose; the relevant details have been discussed in the following sections.

Materials and methods

Study area

Five Indian cities viz., Delhi, Mumbai, Chennai, Kolkata and Bengaluru, comprise the study area for this research. Delhi is the capital of India. Delhi faces influx of large number of people throughout the year because of employment, business, medical facilities and tourism. Mumbai is one of the densely populated cities of India, situated at the shores of the Arabian Sea. It is also known as the financial capital of India. Mumbai is also a hub for many industries like textile, petrochemical etc. Chennai is located along the coast of the Bay of Bengal, with many chemical, pharmaceutical, textile and machinery industries. Kolkata is situated on the bank of the Hooghly River, on the north of the Bay of Bengal. Kolkata is acknowledged as the cultural capital of India and its friendliest metropolis (Mouli et al. 2004). Bengaluru is the IT hub of India; it is one of the most traffic-congested cities in the world. Each selected city is a metropolitan city with population of more than 10 million. These cities have well developed infrastructure to cater to the demand for services of such huge population. Because of huge population and infrastructure, pollution due to traffic, powerhouse combustion, electricity generators etc. is very high in these cities. Due to the high level of pollution, it is appropriate to investigate the relationship between aerosols and respiratory diseases in these cities.

Data

The details of the data used in this study are given in Table 1:

The AOT data (corresponding to 550 nm wavelength) of Visible Infrared Imaging Radiometer Suite (VIIRS) have been obtained by ordering at National Oceanic and Atmospheric Administration—Comprehensive Large Array-data Stewardship System (NOAA-CLASS). The VIIRS aerosol data are processed from the corresponding Sensor Data Records (SDRs) (containing calibrated radiance/reflectance and brightness temperatures, along with geolocation) and ancillary data, on a granule by granule basis (US Department of Commerce and National Oceanic and Atmospheric Administration 2017; National Oceanic and Atmospheric Administration 2021). One VIIRS granule generally consists of 768×3200 (along-track by cross-track) 750 m pixels (National Oceanic and Atmospheric Administration 2021). The processing yields the pixel-level Intermediate Products (IPs) consisting of AOT and other data, and also the associated quality flags (National Oceanic and Atmospheric Administration 2021). Environmental Data Records (EDRs) are created by the aggregation of 8×8 IPs for AOT and other data, and using various combinations of the quality flags (National Oceanic and Atmospheric Administration 2021). EDR AOT data have been employed in this work.

Google Trends data have been collected with the help of 'google-trends-api', on the basis of the keywords mentioned in Table 2. In Google Trends, search result is normalized (as already indicated) to the time and location of a query by dividing each search data item by the total searches in the

Table 1 The datasets and their technical details

Dataset	Source	Remarks	Temporal Extent of Collected Data			PM_{10} Data Collection Station/s (CPCB Station/s) with Coordinates (Latitude in °N, Longitude in °E)		
			Delhi	Mumbai & Chennai	Bengaluru & Kolkata	Delhi	Bengaluru	Kolkata
VIIRS-Suomi AOT	National Oceanic and Atmospheric Administration (NOAA)	Spatial Resolution: 6 km; Wavelength: 550 nm	02.05.12–31.08.17	02.05.12–31.08.17	02.05.12–31.08.17	–	–	–
GT	Google	Spatial Resolution: Regional	01.01.12–31.08.17	01.01.15–31.08.17	01.01.15–31.08.17	–	–	–
Annual Morbidity	Directorate General of Health Services, Delhi	Spatial Resolution: Regional	2012–16 (Asthma), 2014–16 (COPD), 2012 & 2014–16 (Lung Cancer)	–	–	–	–	–
PM_{10}	Central Pollution Control Board (CPCB) of India	–	02.05.12–31.08.17	–	01.01.15–31.08.17	Anand Vihar (28.65, 77.32), ITO (28.63, 77.24), Mandir Marg (28.64, 77.20), Punjabi Bagh (28.67, 77.13), R K Puram (28.56, 77.19)	City Railway Station (12.98, 77.57)	Rabindra Bharati University (22.63, 88.38), Victoria (22.54, 88.34)

geography and time range it represents, and then scaling the resulting number on a range of 0 to 100 (Trends Help 2021).

Directorate General of Health Services, Government of National Capital Territory of Delhi, publishes annual morbidity data for Delhi in Annual Reports. The annual morbidity data comprises patient records from the Out-Patients Departments (OPDs) and the In-Patients Departments (IPDs) of some of the hospitals of Delhi [Directorate General of Health Services (Government of National Capital Territory of Delhi) 2017]. However, the annual morbidity data for COPD, for the years, 2012–2013, and those for lung cancer, for the year 2013, have not been found in the respective Annual Reports.

PM_{10} data have been collected from the website (http://cpcb.nic.in/) of Central Pollution Control Board (CPCB) of Government of India. CPCB collects PM_{10} data, using ground based measuring instruments, installed at automatic air quality monitoring stations, situated at certain locations (Central Control Room for Air Quality Management—All India 2021). PM_{10} data could not be collected for Mumbai and Chennai as none of the CPCB station/s of either of these two cities has any PM_{10} value. The stations, corresponding to which PM_{10} data have been collected, are: Delhi: Anand Vihar, ITO, Mandir Marg, Punjabi Bagh, R K Puram; Bengaluru: City Railway Station; Kolkata: Rabindra Bharati University, Victoria.

Table 2 Keywords corresponding to respiratory diseases (Source: (Davidson 2014; Longo et al. 2011a, b))

Keyword	Asthma	COPD	Lung Cancer
Asthma	+	–	–
Chest discomfort	+	+	+
Chronic bronchitis	–	+	–
Chronic cough	+	+	+
Chronic cough at night	+	–	–
COPD	–	+	–
Coughing up blood	–	–	+
Cyanosis	–	+	–
Dyspnea	+	+	+
Emphysema	–	+	–
Fatigue	–	+	+
Lung cancer	–	–	+
Noisy breathing	+	+	+
Rapid breathing	+	+	–
Rust coloured sputum	–	–	+
Shortness of breath	+	+	+
Tachypnea	+	+	–
Wheezing	+	+	+

(+: the keyword corresponds to the disease; –: the keyword does not correspond to the disease)

Causality and Correlation:

Granger causality test can be elucidated with the help of Autoregressive Distributed Lag (ADL) model. According to this model, the dependent variable, at any point of time, can be explained with the help of its own lagged values as well as the lagged values of the independent variable.

The basic equation of Granger causality test, using ADL model, is given in Eq. 1 (Koop 2005):

$$Y_t = \alpha + \delta t + \varphi_1 Y_{t-1} + \cdots + \varphi_p Y_{t-p} + \beta_1 X_{t-1} + \cdots + \beta_q X_{t-q} + e_t \tag{1}$$

where, X is the independent variable, Y is the dependent variable, t is time, e_t is error, q is the no. of lags and α, δ, φ's and β's are constants. X can be said to Granger cause Y if any or all of $\beta_1, \ldots, \beta_q$ are statistically significant. This implies that, if one tests the hypothesis that $\beta_1 = \beta_2 = \ldots = \beta_J = 0$ and, if the hypothesis is rejected, one can say that X Granger causes Y. Here, J ($\leq$ q) is the number of restrictions, i.e., the number of β's that are considered zero, in the hypothesis.

In this study, it is required to check whether the AOT (independent variable) data from VIIRS sensor are Granger causing the disease data obtained from Google Trends (dependent variable). For that, the Regression Sum of Squares (SSR) based F-test has been employed. The SSR based F statistic which has been used for hypothesis testing at the significance level of 5%, is given in Eq. 2 (Greene 2003):

$$F = \frac{(R_U^2 - R_R^2)/J}{(1 - R_U^2)/(T - q)} \tag{2}$$

where, $R_U^2 = R^2$ from unrestricted regression model, calculated using SSR, $R_R^2 = R^2$ from restricted regression model, calculated using SSR.

In unrestricted regression model, none of the β's is taken as zero and in the restricted regression model, J β's are taken as zero.

Coefficient of determination, $R^2 = \frac{SSR}{SST}$ where, SSR: Regression Sum of Squares, SST: Total Sum of Squares.

And J = no. of restrictions, T = no. of observations, q = no. of lags, T-q = degrees of freedom (denominator). Besides, J = degrees of freedom (numerator).

Now, p = probability of getting the F statistic value for particular degrees of freedom.

If, for certain number of lags, the calculated value of p is less than the specified value of the level of significance, α (here, $\alpha = 0.05$), and the calculated value of F is more than the table value of F, it can be said that the AOT dataset is Granger causing the GT dataset at that particular level of significance.

Generally, for any keyword, corresponding to a disease, and any particular metro-city, Granger causality test yields

multiple positive results, but the result with the minimum p value, with the condition that $p < 0.05$ and $F_{calculated} > F_{table}$, has been recorded for each keyword-metro-city combination. In some cases, the Granger test has produced no result for the condition, $p < 0.05$ and $F_{calculated} > F_{table}$. In these situations, the result with the minimum possible p value has been recorded.

Now, it is prudent to check how much reliable the Google Trends data are, as a source of information about the occurrences of the three respiratory diseases. For that purpose, this study has used the annual morbidity records.

To find out the relationship between the GT and the annual morbidity data, Pearson's correlation coefficient (r) has been calculated. A higher positive value of r indicates a better correlation.

For validating the outcome of the Granger causality test between AOT and Google Trends datasets, and as an alternative way of checking the reliability of the Google Trends data, Granger causality test has been performed (in the same way as mentioned above) between PM_{10} data, obtained from CPCB, and the corresponding Google Trends data. As hinted earlier, PM_{10} data give the amount of Particulate Matter (i.e., atmospheric aerosols) with diameter ≤ 10 μm, present in unit volume of the atmosphere; the importance of PM_{10} lies in the fact that these are inhalable because of their sizes. The trustworthiness of the Google Trends data and also of the result of the Granger causality test between AOT and Google Trends datasets, increases if PM_{10} data are found to Granger cause Google Trends data in a similar way as AOT data have Granger caused Google Trends data.

Method

The main outline of the method is as follows. AOT data have been obtained from NOAA-CLASS. The GT data have been collected by employing JavaScript programming, and subsequently, suitably processed to get metro-city-wise lists of daily GT values for a period of approximately six years, for Delhi, and around three years, for the other four cities, and for each keyword mentioned in Table 2. Since Google Trends is a single value having one-day temporal resolution for entire region, daily AOT is also averaged spatially to get a single value for each selected metro-city. The annual morbidity data have been collected, manually analyzed, processed and converted into disease-wise lists of Annual Morbidity values for Delhi. The PM_{10} data have been collected and appropriately processed to get metro-city-wise lists of daily average PM_{10} values. Next, each GT dataset (i.e., list of daily GT values, for a particular metro-city and a particular keyword) has been tallied date-wise with the corresponding AOT dataset to create the respective combined AOT-GT dataset. Granger causality has been checked,

with the help of SSR based F-test, for each combined AOT-GT dataset. Subsequently, the GT data, for Delhi, have been averaged for each year for suitable keywords, as indicated by the result of Granger causality test and the availability of annual morbidity data, and tallied year-wise with the annual morbidity values, for each disease, to create the respective combined Annual Morbidity-GT dataset. The daily Google Trends data have been averaged over a year since annual morbidity is a single value for a year. Next, coefficient of correlation has been calculated for each combined Annual Morbidity-GT dataset. Subsequently, Google Trends data (i.e., daily GT values, for appropriate metro-cities and appropriate keywords, as indicated by the result of Granger causality test and the availability of PM_{10} data) have been tallied date-wise with the corresponding PM_{10} data to create the respective combined PM_{10}-GT datasets. Next, Granger causality has been checked, with the help of SSR based F-test, for each combined PM_{10}-GT dataset.

The method can be described pictorially with the help of the flowchart shown in Fig. 1.

Results and discussion

The result of Granger causality test between AOT and Google Trends is shown in Table 3.

According to Table 3, Delhi, Mumbai and Chennai exhibited high to very high degrees of causality for all the three diseases whereas, Bengaluru displayed medium level of causality, for each of them. In the case of Kolkata, degree of causality was medium for asthma, but high for the other two diseases.

This result implies that during the period of study, if we take into account atmospheric aerosols as the only causative factor: (a) Delhi, Mumbai and Chennai were quite vulnerable to the selected respiratory diseases, (b) Bengaluru was not so much vulnerable to these three diseases, and (c) Kolkata was also not so much vulnerable to asthma, but did display vulnerability to the other two diseases.

Unique keywords uniquely identify a disease, e.g., the keyword, "Asthma" uniquely refers to the disease asthma; others (i.e., non-unique keywords) signify two or three diseases at a time, e.g., the keyword, "Fatigue" refers to two diseases viz., COPD and lung cancer. When causality has been detected in a metro-city for a unique keyword, it can be precisely stated that the city is vulnerable to only a particular disease, to which the keyword refers; otherwise, it can be said that the city is vulnerable to any or all of the diseases indicated by the respective non-unique keyword.

The map in Fig. 2 qualitatively displays the vulnerabilities of the metro-cities to the respiratory diseases, due to the
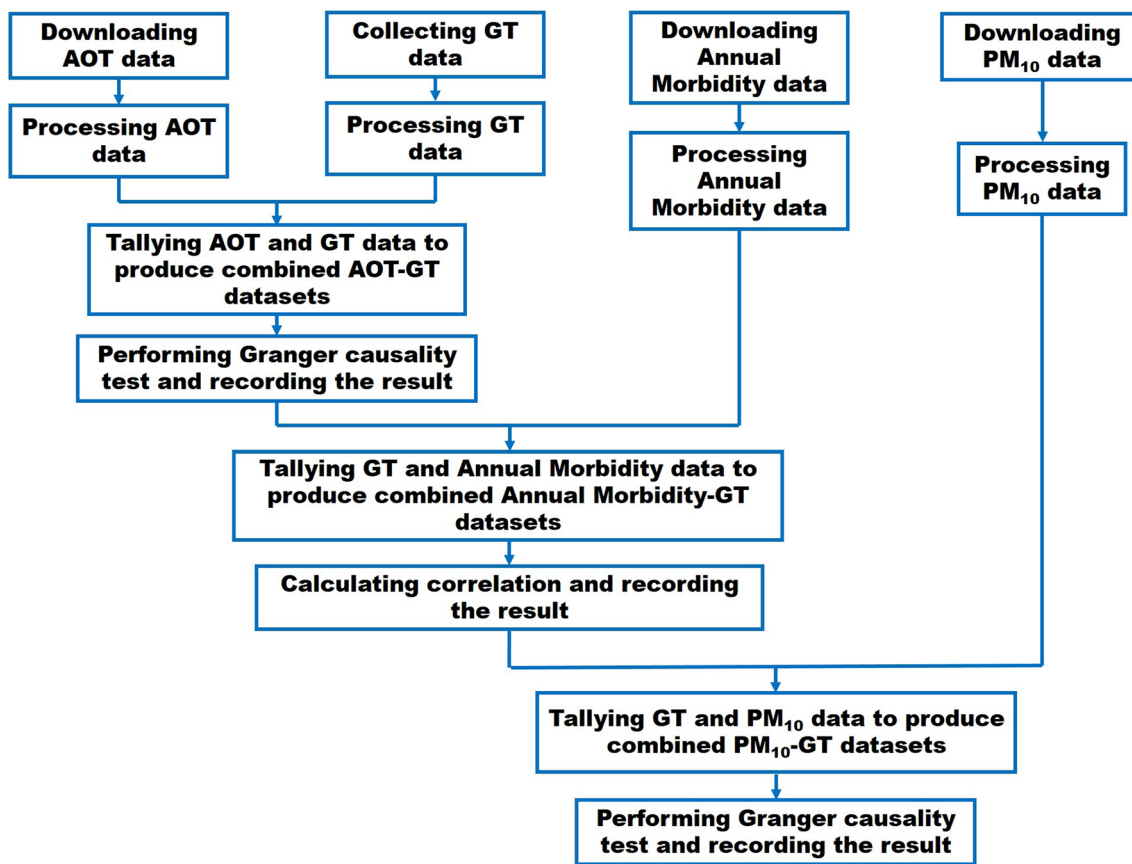
Fig. 1 Flowchart depicting the method

Table 3 The nature of causality between aerosols and diseases, using AOT and Google Trends data

City	Disease	Period	Keyword	Whether Unique Keyword	F_{table}	$F_{calculated}$	p	Degree of Causality
Delhi	Asthma	2012–2017	Asthma	Yes	3.8463	9.3494	0.0023	Very high
Delhi	COPD	2012–2017	COPD	Yes	2.6096	3.6504	0.0121	High
Delhi	Lung cancer	2012–2017	Lung Cancer	Yes	3.8463	11.9361	0.0006	Very high
Mumbai	Asthma	2015–2017	Asthma	Yes	3.8526	4.2513	0.0395	High
Mumbai	COPD/ lung cancer	2015–2017	Fatigue	No	1.2268	1.4786	0.0008	Very high
Chennai	Asthma	2015–2017	Asthma	Yes	1.4079	1.6948	0.0051	Very high
Chennai	COPD/ lung cancer	2015–2017	Fatigue	No	1.3633	1.4797	0.0184	High
Kolkata	Asthma	2015–2017	Asthma	Yes	1.3019	1.1979	0.1300	Medium
Kolkata	COPD/ lung cancer	2015–2017	Fatigue	No	3.8567	4.2201	0.0404	High
Bengaluru	Asthma	2015–2017	Asthma	Yes	1.5264	1.3575	0.1169	Medium
Bengaluru	COPD/ lung cancer	2015–2017	Fatigue	No	1.3384	1.3146	0.0614	Medium

(p < 0.01: *Very high* causality; $0.01 \leq p < 0.05$: *High* causality; $p \geq 0.05$: *Medium* causality)

presence of atmospheric aerosols only, during the respective study periods.

The plots depicted in the map in Fig. 2, have been generated with the help of the expression:

$$[\text{diff\_p} - (\text{diff\_p}_{min} - 20)] \quad \text{for each disease and each city}$$

where, $\text{diff\_p} = (1 - p) * 1000$ and $\text{diff\_p}_{min} = 870$ (diff_p_{min} is the minimum value of diff_p)

Fig. 2 Vulnerabilities of the selected metro-cities in India to the three respiratory diseases, due to the presence of atmospheric aerosols
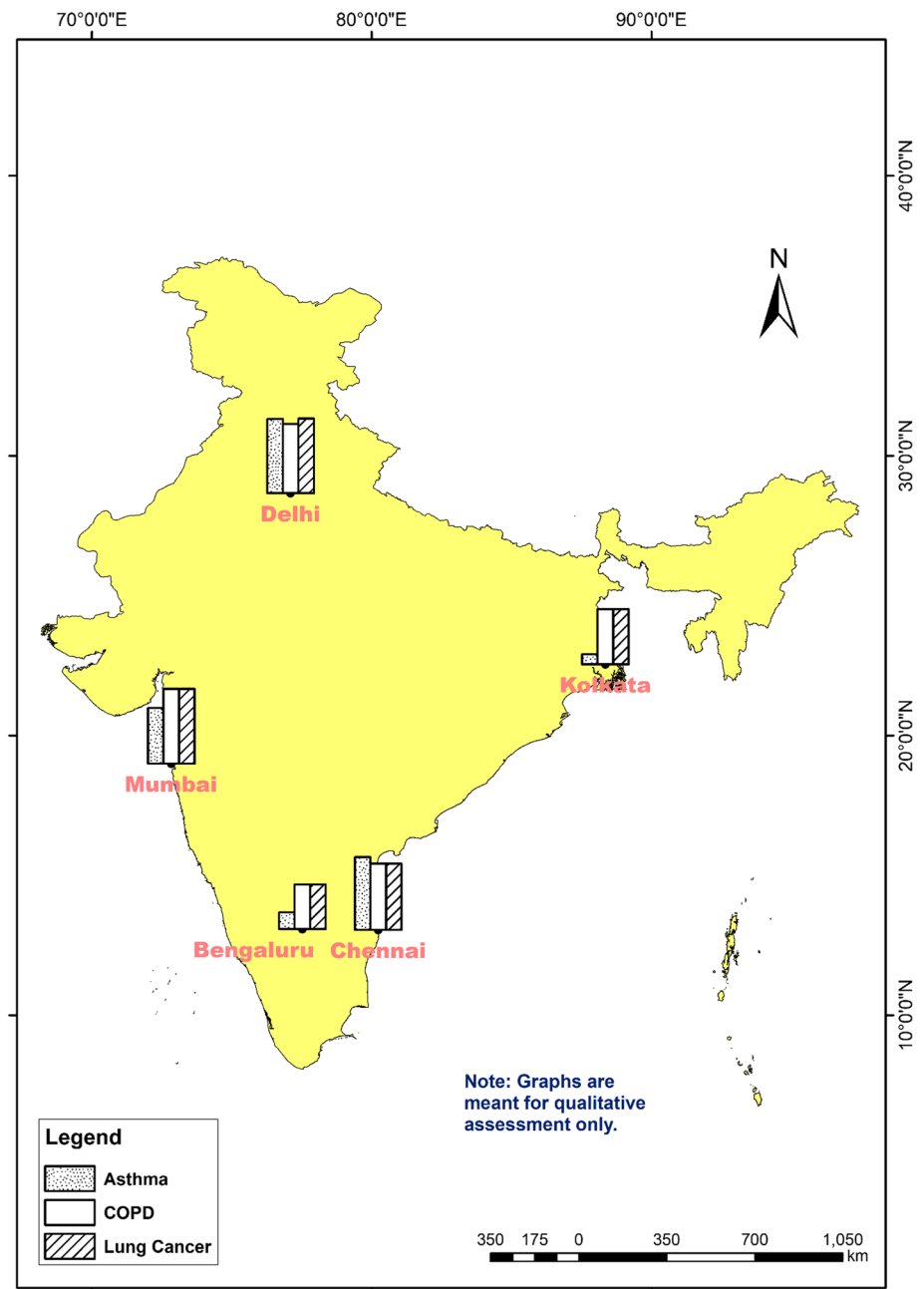
For calculating diff_p, the p values given in Table 3, have been used. The number, 20, in the above expression, has been chosen as an arbitrary offset value, so that (diff_p – diff_p$_{min}$), for diff_p = 870, does not become zero.

It may be mentioned here that many researchers have studied the relation between respiratory diseases and aerosols (De Miguel-Diez et al. 2019). As for example, De Souza et al. (2017) have shown the linkage between aerosols and prevalence of Asthma in Brazil, and Cao et al. (2021) have discussed about how COVID-19 (COronaVIrus Disease 2019) can be transmitted through aerosols.

Some scientists have worked on the use of GT data in disease surveillance. For example, Tkachenko et al. (2017) have found that GT data can help in Type 2 Diabetes surveillance, Bousquet et al. (2017) have shown the association between GT and the epidemiology of allergic rhinitis, and Bhattacharya et al. (2013) have used GT and similar other tools (HealthMap and Kazemill) for acquiring information about impending outbreaks of certain diseases (which, however, do not include asthma, COPD and lung cancer), and have concluded that, despite certain limitations, Google Trends has a lot of potential in this regard. As another instance, it can be stated that, Ricketts and Silva (2017) have revealed

the association between GT data related to mortality, and the actual incidents of mortality, in case of a number of diseases including asthma and COPD, in USA.

However, we have found no such report where GT data have been used for the surveillance of asthma, COPD and lung cancer in India. Moreover, we have come across no such study where Granger causality has been checked between AOT and GT datasets, to understand the relation between the presence of atmospheric aerosols and the prevalence of asthma, COPD and lung cancer.

The values of Pearson's correlation coefficient (r) between annual morbidity and GT data are shown in Table 4.

We find in Table 4 that, there is medium to high levels of correlation between annual morbidity and Google Trends. Thus, we can say that Google Trends has been satisfactorily validated by the values of Pearson's correlation coefficient (between annual morbidity and Google Trends). The errors of approximation, caused by the yearly averaging of Google Trends data, may have propagated and accumulated over the years, and may be responsible for lowering the value of r. This seems to be true as longer the period, lower the correlation. In spite of using the average Google Trends, the degree of correlation is reasonably good.

It may be mentioned here that, many researchers have worked on morbidity; as for example, Okoroiwu et al. have investigated the causes of morbidity and mortality in a hospital of Nigeria (Okoroiwu et al. 2020). However, we have

found no such paper where correlation between Annual Morbidity and GT data have been computed to know the effectiveness of GT data for the surveillance of Asthma, COPD and Lung Cancer.

The outcome of Granger causality test between PM_{10} and Google Trends is shown in Table 5.

Table 5 shows that Delhi exhibited very high degree of causality for all the three diseases whereas, Bengaluru exhibited high causality for asthma and medium causality for the other two diseases. Kolkata displayed low level of causality for asthma, but medium level of causality for the other two diseases.

This result implies that during the period of study, if we take into account PM_{10} as the only causative factor: (a) Delhi was considerably vulnerable to all the three diseases, (b) Bengaluru was quite vulnerable to asthma but not so much susceptible to the other two diseases, and (c) Kolkata was not quite vulnerable to any of the three diseases.

The result of the Granger causality test between AOT and Google Trends is agreeing satisfactorily with that of the Granger test between PM_{10} and Google Trends, for most of the relevant city-disease combinations, as shown in Tables 3 and 5. However, following two discrepancies arise in case of asthma for Bengaluru, and COPD and lung cancer for Kolkata.

The first one is that, the result of causality test involving satellite AOT data, shows that Bengaluru was not vulnerable to asthma due to atmospheric aerosols, while the result of causality test involving PM_{10}, indicates that the city was susceptible to the disease due to the presence of PM_{10}. This difference may be due to the fact that some of the atmospheric aerosols/particulate matter are recorded by the ground based measuring instruments but not captured by the satellite sensor.

The second discrepancy is that, the outcome of Granger causality using satellite AOT data, specifies that Kolkata was vulnerable to either or both of COPD and lung cancer due to atmospheric aerosols, whereas the outcome of

Table 4 Correlation between annual morbidity and GT data

Disease	Period	r	Degree of Correlation
Asthma	2012–2016	0.55	Medium
COPD	2014–2016	0.96	High
Lung Cancer	2012 & 2014–2016	0.80	High

($r \geq 0.75$: *High* correlation and $0.75 > r \geq 0.50$: *Medium* correlation)

Table 5 Causality between particulate matter and diseases, using PM_{10} and Google Trends data

City	Disease	Period	Keyword	Whether Unique Keyword	F (table)	F (calculated)	p	Degree of Causality
Delhi	Asthma	2012–2017	Asthma	Yes	3.8523	9.1988	0.0025	Very high
Delhi	COPD	2012–2017	COPD	Yes	3.0063	5.3910	0.0047	Very high
Delhi	Lung Cancer	2012–2017	Lung cancer	Yes	1.4268	1.7487	0.0043	Very high
Bengaluru	Asthma	2015–2017	Asthma	Yes	8.5444	24.2177	0.0112	High
Bengaluru	COPD/ Lung Cancer	2015–2017	Fatigue	No	3.8584	2.8692	0.0909	Medium
Kolkata	Asthma	2015–2017	Asthma	Yes	3.8662	0.1767	0.6744	Low
Kolkata	COPD/ Lung Cancer	2015–2017	Fatigue	No	1.3575	1.2895	0.0855	Medium

($p < 0.01$: *Very high* causality; $0.01 \leq p < 0.05$: *High* causality; $0.05 \leq p \leq 0.50$: *Medium* causality; $p > 0.50$: *Low* causality)

Granger causality using PM_{10} data, shows that the city was not susceptible to either or both of the two diseases due to the presence of PM_{10}. This difference has occurred probably due to the fact that PM_{10} constitutes only a part of the total atmospheric aerosol content.

The overall result of the Granger causality test between PM_{10} and Google Trends validates the outcome of the Granger causality test between AOT and Google Trends, besides showing the trustworthiness of not only Google Trends but also AOT.

A number of researchers have worked on the relation between PM_{10} and respiratory diseases and disorders. For instance, Choi et al. (2020) have shown that inhalation of PM_{10} increases airway inflammation, and Jo et al. (2017) have found that increase in PM_{10} and $PM_{2.5}$ (particulate matter with diameter ≤ 2.5 μm) is one of the factors that lead to higher prevalence of acute bronchitis, allergic rhinitis and asthma.

However, we have seen no such study where Granger causality between PM_{10} and GT data has been used as a means to investigate the association between the presence of particulate matter (diameter ≤ 10 μm) and the occurrences of asthma, COPD and lung cancer.

Finally, the following inferences can also be drawn from the outcomes of the Granger causality tests and the values of Pearson's correlation coefficient: (a) VIIRS AOT and Google Trends can be effectively used as sources of information regarding the amount of aerosols in the atmosphere, and the trends of the occurrences of asthma, COPD and lung cancer, and (b) Granger causality test is appropriate for detecting causality between relevant datasets in the field of geographic information systems (GIS) and Health.

It may be mentioned here that there are some studies on the usefulness of VIIRS data. For instance, Huang et al. (2016) have validated VIIRS AOT and Ångström Exponent (AE) data by comparing them with the data from AErosol RObotic NETwork (AERONET), and found that VIIRS products are authentic sources of information on aerosols.

Granger causality test has been used as a tool of analysis in different fields by several scientists. For example, Hamta et al. (2021) have successfully applied it to find out how some climatic variables are affecting the prevalence of visceral leishmaniasis, and Ivanovski et al. (2020) have employed it to infer that capital expenditures cannot be used for accurately forecasting the gross domestic product (GDP) in the Republic of North Macedonia (Ivanovski et al. 2020; Hamta et al. 2021). However, we have come across no such paper where Granger causality test has been used as a tool of analysis in the field of Health GIS, which has been done in this study.

Conclusions

This work judges the vulnerabilities of Delhi, Mumbai, Chennai, Kolkata and Bengaluru, to asthma, COPD and lung cancer, during the specified time periods, due to the presence of atmospheric aerosols. Besides, this study reveals that, AOT and GT can be considered as effective sources of information about the amount of atmospheric aerosols, and the nature of prevalence of the three aforesaid respiratory diseases, respectively, and Granger causality test can be regarded as a suitable tool of analysis in the field of Health GIS. The study highlights the importance of correlating environmental pollution with human health and controlling the diseases related to environmental factors through GIS.

Declarations

Conflict of interest On behalf of all the authors, the corresponding author states that there is no conflict of interest.

References

Akhtar N, Gupta Bansal J (2017) Risk factors of Lung Cancer in non-smoker. Curr Probl Cancer 41:328–339

Atkinson RW, Kang S, Anderson HR et al (2014) Epidemiological time series studies of PM2.5 and daily mortality and hospital admissions: a systematic review and meta-analysis. Thorax 69:660–665

Beasley R, Semprini A, Mitchell EA (2015) Risk factors for Asthma: is prevention possible? Lancet 386:1075–1085

Bhattacharya I, Ramachandran A, Bhattacharya J, Dogra NK (2013) Google Trends for formulating GIS mapping of disease outbreaks in India. Int J Geoinformatics 9:9–19

Bhome AB (2012) COPD in India: iceberg or volcano. J Thorac Dis 4:298–309

Bousquet J, Agache I, Anto JM et al (2017) Google Trends terms reporting Rhinitis and related topics differ in European countries. Allergy 72:1261–1266

Cao Y, Shao L, Jones T et al (2021) Multiple relationships between aerosol and COVID-19: a framework for global studies. Gondwana Res 93:243–251

Central Control Room for Air Quality Management - All India (2021) Continuous Stations Status. https://app.cpcbccr.com/ccr/#/caaqm-dashboard-all/caaqm-landing

Choi J, Sim JK, Oh JY, et al (2020) Relationship between Particulate Matter (PM10) and airway inflammation measured with exhaled nitric oxide test in Seoul, Korea. Can Respir J

Davidson S (2014) Davidson's principles and practice of medicine. Churchill Livingstone (Elsevier), London

De Miguel-Diez J, Hernández-Vázquez J, López-de-Andrés A, et al (2019) Analysis of environmental risk factors for Chronic

Obstructive Pulmonary Disease exacerbation: a case-crossover study (2004–2013). PLoS One 14

De Souza A, Santos DA da S, Ikefuti PV, (2017) Association between climate variables, pollutants, aerosols and hospitalizations due to Asthma. O Mundo da Saúde 41:359–367

DeVries R, Kriebel D, Sama S (2016) Low level air pollution and exacerbation of existing COPD: a case-crossover analysis. Environ Heal 15

Directorate General of Health Services (Government of National Capital Territory of Delhi) (2017) Annual Report 2016–2017

Dovjak M, Kukec A (2019) Creating healthy and sustainable buildings: an assessment of health risk factors. Springer Nature, Cham (Switzerland)

Godin R (2015) Joint Polar Satellite System (JPSS) VIIRS Aerosol Optical Thickness (AOT) and Particle Size Parameter Algorithm Theoretical Basis Document (ATBD). Greenbelt, Maryland

Greene WH (2003) Econometric analysis, 5th edn. Prentice Hall, Hoboken, New Jersey

Hamta A, Saghafipour A, Farahani LZ et al (2021) The Granger causality analysis of the impact of climatic factors on Visceral Leishmaniasis in Northwestern Iran in 1995–2019. J Parasit Dis 45:17–23

Huang J, Kondragunta S, Laszlo I et al (2016) Validation and expected error estimation of Suomi-NPP VIIRS Aerosol Optical Thickness and Ångström Exponent with AERONET. J Geophys Res Atmos 121:7139–7160

Ivanovski Z, Narashanov Z, Korunovska V (2020) Granger causality test for the government's capital expenditures on the GDP of the Republic of North Macedonia in VAR environment. UTMS J Econ 11:183–201

Jindal SK, Aggarwal AN, Gupta D et al (2012) Indian study on epidemiology of Asthma, respiratory symptoms and chronic bronchitis in adults (INSEARCH). Int J Tuberc Lung Dis 16:1270–1277

Jo E-J, Lee W-S, Jo H-Y et al (2017) Effects of Particulate Matter on respiratory disease and the impact of meteorological factors in Busan, Korea. Respir Med 124:79–87

Kim V, Criner GJ (2013) Chronic Bronchitis and Chronic Obstructive Pulmonary Disease. Am J Respir Crit Care Med 187:228–237

Koop G (2005) Analysis of economic data. John Wiley & Sons, Inc., Hoboken, New Jersey

Li M-H, Fan L-C, Mao B et al (2016) Short-term exposure to ambient fine Particulate Matter increases hospitalizations and mortality in COPD: a systematic review and meta-analysis. Chest 149:447–458

Longo DL, Kasper DL, Jameson JL et al (eds) (2011a) Harrison's principles of internal medicine Vol.-2. McGraw Hill Education, New York

Longo DL, Kasper DL, Jameson JL et al (eds) (2011b) Harrison's principles of internal medicine Vol.-1. McGraw Hill Education, New York

Malhotra J, Malvezzi M, Negri E et al (2016) Risk factors for Lung Cancer worldwide. Eur Respir J 48:889–902

Mouli PC, Kumar MP, Reddy SJ, Mohan SV (2004) Monitoring of air pollution in Indian metropolitan cities: modelling and quality indexing. Int J Environ Pollut 21:365–382

Mowli PP, Subbayya NV (1989) Air pollution and control. Divyajyoti Prakashan, Jodhpur

National Aeronautics and Space Administration (1996) Atmospheric aerosols: what are they, and why are they so important? https://www.nasa.gov/centers/langley/news/factsheets/Aerosols.html

National Oceanic and Atmospheric Administration (2021) STAR/SMCD/VIIRS aerosol calibration and validation. https://www.star.nesdis.noaa.gov/smcd/emb/viirs_aerosol/products.php

National Oceanic and Atmospheric Administration, United States Department of Commerce (2017) Sea spray aerosols (Western Atlantic Climate Study: WACS2 (2014)). https://saga.pmel.noaa.gov/pmel-theme/sea-spray-aerosols

Noronha V, Pinninti R, Patil VM et al (2016) Lung Cancer in the Indian subcontinent. South Asian J Cancer 5:95–103

Ochoa-Avilés C, Morillo D, Rodriguez A, et al (2020) Prevalence and risk factors for Asthma, Rhinitis, Eczema, and atopy among preschool children in an Andean city. PLoS One 15

Okoroiwu HU, Uchendu KI, Essien RA (2020) Causes of morbidity and mortality among patients admitted in a tertiary hospital in Southern Nigeria: a 6 year evaluation. PLoS One 15

Oland AA, Booster GD, Bender BG (2017) Psychological and lifestyle risk factors for Asthma exacerbations and morbidity in children. World Allergy Organ J 10

Parritz RH, Troy MF (2018) Disorders of childhood: development and psychopathology, 3rd edn. Cengage Learning, Boston

Petty TL (2005) Are COPD and Lung Cancer two manifestations of the same disease? Chest 128:1895–1897

Ricketts CF, Silva CG (2017) An analysis of morbidity and mortality using Google Trends. J Hum Behav Soc Environ 27:559–570

Tkachenko N, Chotvijit S, Gupta N, et al (2017) Google Trends can improve surveillance of Type 2 Diabetes. Sci Rep 7

Törmänen S, Lauhkonen E, Riikonen R et al (2018) Risk factors for Asthma after Infant Bronchiolitis. Allergy 73:916–922

Trends Help (2021) FAQ about Google Trends data. https://support.google.com/trends/answer/4365533?hl=en#:~:text=Google Trends normalizes search data to make comparisons between terms easier.&text=Each data point is divided, would always be ranked highest

US Department of Commerce, National Oceanic and Atmospheric Administration (2017) Visible Infrared Imaging Radiometer Suite (VIIRS) Sensor Data Record (SDR) user's guide. Washington, D.C

World Health Organization (2007) Global surveillance, prevention and control of chronic respiratory diseases: a comprehensive approach. Geneva