

ASSESSING POPULATION SENSITIVITY TO URBAN AIR POLLUTION USING GOOGLE TRENDS AND REMOTE SENSING DATASETS

Prakhar Misra^{1,*}, Wataru Takeuchi¹

¹ Institute of Industrial Science, The University of Tokyo, Tokyo - 153-8505, Japan - (mprakhar, wataru)@iis.u-tokyo.ac.jp

Commission VI, WG VI/4

KEY WORDS: MODIS AOD, AirRGB, time series, sensitivity, perception of air pollution, developing countries

ABSTRACT:

This study demonstrates relationship between remote sensing satellite retrieved fine aerosol concentration and web-based search volumes of air quality related keywords. People's perception of urban air pollution can verify policy effectiveness and gauge acceptability of policies. As a serious health issue in Asian cities, population may express concern or uncertainty for air pollution risk by performing search on the web to seek answers. A 'social sensing' approach that monitors such search queries, may assess people's perception about air pollution as a risk. We hypothesize that trend and volume of searches show impact of air pollution on general population. The objectives of this research are to identify those atmospheric conditions under which relative search volume (RSV) obtained from Google Trends shows correlation with measured fine aerosol concentration, and to compare search volume sensitivity to rise in aerosol concentration in seven Asian megacities. We considered weekly relative search volumes from Google Trends (GT) for a four year period from January, 2015 to December, 2018 representing diverse PM_{2.5} concentrations. Search volumes for keywords corresponding to perception of air quality ('air pollution') and health effects ('cough' and 'asthma') were considered. To represent PM_{2.5} we used fine aerosol indicator developed in an earlier research. The results suggest that tendency to search for 'air pollution' and 'cough' occurs when AirRGB R is in excess and temperature is below the baseline values. Consistent with this, in cities with high baseline concentrations, sensitivity to rise in AirRGB R is also comparatively lower. The result of this study can be used as an indirect measure of awareness in the form of perception and sensitivity of population to air quality. Such an analysis could be useful for forecasting health risks specially in cities lacking dedicated services.

1. INTRODUCTION

1.1 Background

In most Asian cities outdoor fine particulate matter air pollution are increasing (Misra et al., 2017). Studies have shown that fine particulate matter results in severe health impacts that include pulmonary disorders, irritation in eyes as well as cardiac issues. As per WHO, particulate matter has well documented and strongest evidence of health risks that leads to harmful impacts on the cardiovascular, respiratory and pulmonary functionality. In India alone, outdoor air pollution was responsible for more than 670,000 deaths in 2016. To deal with these health effects, several countries have launched large-scale policies, e.g. China's Action Plan on Prevention and Control of Air Pollution and India's National Clean Air Program. In other countries, like Thailand, there is a growing demand for 'Clean-air Act' to overcome the issue of polluted air. However to move in the direction for achieving air quality and environmental management objectives, public perception about air pollution as a risk needs to be studied (Bickerstaff, Walker, 2001). An understanding of what levels of air pollution are perceived as acceptable among the public, can verify the effectiveness of policies as well as help improve the communication framework for achieving the desired change in attitude and behavior (Eden, 1996, Elliott et al., 1999). Perception of people regarding exposure and health effects is critical to gauge response of people and acceptability of related policies (Egondi et al., 2013). Although it is known environmental factors affect the health impact, (Elliott et al., 1999) suggested that perception of air pollution affects the relation between air pollution and health.

*Corresponding author

Since epidemiological datasets are not publicly available for many cities, web-based social sensing datasets may be useful in reflecting impact of pollution on human health. However it is not clear to what extent information derived social sensing dataset, such as Google Trend, is related to air pollution concentration. Furthermore, as cities differ in baseline concentrations and social parameters, whether a unit rise in pollutant concentration results in a similar response on social sensing datasets across all cities is also unclear.

1.2 Objective of this research

The objective in this paper is to: 1) assess whether Google Trends for air quality related search queries shows correlation and seasonality with measured fine aerosol concentration, and 2) estimate health sensitivity of general population to measured fine aerosol concentration.

2. METHODOLOGY

The flowchart of dataset and methodology employed for studying the relationship between Google Trends based search activity and pollution concentration is shown in Figure 1. Data location consist of 7 Asian megacities.

Seven Asian megacities were chosen to study their perception and sensitivity. In order of their increasing per capita gross national income (GNIpc), these cities are: Dhaka (\$3,677), Karachi (\$ 5,311), Delhi (\$ 6,353), Bangkok (\$ 12,376) Taipei (\$ 24,318), Seoul (\$ 35,945) and Tokyo (\$ 38,986). These cities were chosen based on their population size, known air pollution issues and popularity of Google as the dominant search engine.

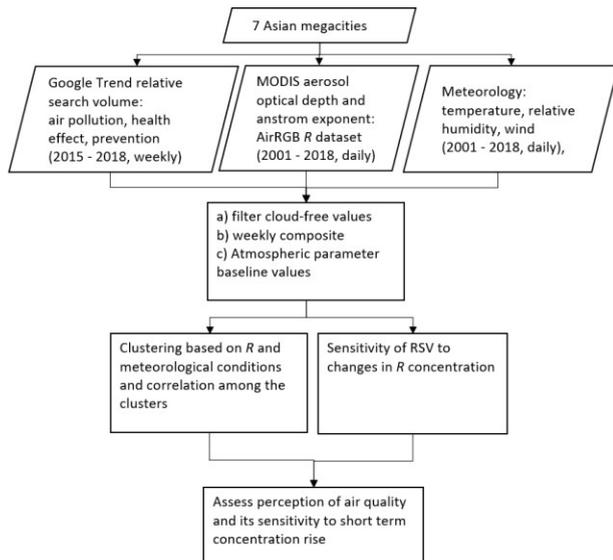


Figure 1. Flowchart studying relationship between Google Trends based search activity (*RSV*) and fine aerosol concentration *R*

Each of them (except Tokyo) sees PM_{2.5} concentrations above the WHO specified 24 hour limits although to a much variable degree.

2.1 Theoretical background

Within social sciences, risk is defined as (Rosa, 2003) “a situation or an event where something of human value (including humans themselves) is at stake and where the outcome is uncertain” while perception of risk is defined as ‘the subjective assessment of the probability of a specified type of accident happening and how concerned we are with the consequences’ (Sjöberg et al., 2004). There are many studies that have studied public perception of air pollution, but few have probed the correlation between perception and measured air pollutant concentrations. Even fewer studies have focused on developing countries (Saksena, 2011). This is important to explain why there is a gap between the perception of general public and an expert who relies on physical information. The first such study found that in Los Angeles County air quality was perceived as ‘smoggy’ when visibility (as a surrogate indicator of pollution), O₃ and SO₂ concentrations exceeded their baseline values (Flachsbar, Phillips, 1980).

Measuring perception is difficult as it is a subjective phenomenon and lacks any related open data. In absence of real-time data, social sensing and remote sensing data could be used to understand perception and sensitivity of humans to air pollution to some extent. Social sensing is defined as ‘any source of information that can be identified in modern social networking and Web tools that expresses some situation or fact about users (e.g., their preferences or scheduled activities) and their social environments’ (Rosi et al., 2011). Popular data sources include Google Trends, Twitter API and Sina Weibo. Researches have shown impact of air pollution on expressed happiness on Sina Weibo (Kahn et al., 2019). Google Trends has been used to assess the impact of meteorological environment on human behavior. Although search data is not significantly better than autoregressive model, however in absence of other data sources search data could be useful (Goel et al., 2010). COPD

symptoms are triggered with chronic particulate air pollution (Schikowski et al., 2014) as well as short-term exposures (Li et al., 2016). In recent years, use of Google Trends in health care research has been increasing (Nutti et al., 2014), such as for COPD symptoms (Boehm et al., 2019), asthma (Bousquet et al., 2017), influenza (Polgreen et al., 2008) as well as flu (Eysenbach, 2006). (Maurer, Holbach, 2016) showed that search requests represent not only the concern for the issue concern (also known as salience) but also the public uncertainty regarding the issue, where uncertainty can be defined as absence of knowledge on an issue. Since cough is a common symptom of COPD (Smith, Woodcock, 2006), it was chosen as a search keyword indicating health impact. Based on this we hypothesized that among these keywords, search for ‘air pollution’ corresponds to perception of air quality, while ‘cough’ and ‘asthma’ correspond to short term health-effect.

2.1.1 Perception of fine aerosol concentration Perception of air quality is subjective to each individual. Perception of indoor air quality is established on the basis of temperature, humidity and velocity of the air as well as the quotient of the amount of air pollution and fresh air supplied (Roelofsen, 2018). Since perception of air quality is ultimately related with health effects (Elliott et al., 1999), search keywords *RSV* keywords may inform how varying outdoor pollutant concentration affects the health. Perception of air quality was assessed by identifying correlation of *GT* search volume relative to *R*. Clustering methods were used to group observations with similar atmospheric conditions. Correlation between mean observation in atmospheric parameters influences *RSV*. Following the perception index of (Roelofsen, 2018), *R*, *temp*, and *wind* were considered for clustering the data. As past studies have suggested the role of atmospheric visibility in perception of air pollution (Flachsbar, Phillips, 1980), AOD values were also for cluster correlation included as inverse indicator of visibility (Kaufman, Fraser, 1983). It should be noted however that if the light attenuation is high, retrieved AOD will be high irrespective of the particle being coarse or fine. Hierarchical Agglomerative Clustering (HAC) algorithm was used to generate the clusters. HAC is a bottom-up algorithm where each object is initially a single-element cluster and at each step, similar clusters are combined to progressively bigger cluster. Two set of clusters, *C*₁ and *C*₂, were generated using HAC, to obtain the correlation between among cluster components as shown in Equation 1.

$$C_1 : HAC(R, temp, wind)$$

$$C_2 : HAC(\Delta R, \Delta temp, \Delta wind) \quad (1)$$

R, *temp*, *wind* and *RSV* for each location was standardized from 0 to 1 and pooled into a single dataset. Cluster set, *C*₁, were computed over this dataset. Another cluster set, *C*₂ was computed over ΔR , $\Delta temp$ and $\Delta wind$ such that, ΔX refers to $X - X_{2015-2018}$, where $X_{2015-2018}$ is the baseline obtained as mean *X* from 2015 to 2018.

2.1.2 Sensitivity of perception We are interested in identifying how sensitive is *RSV* in responding to rise in *R*. Sensitivity is commonly defined as a measure of the proportion of actual positives that are identified as such. This increase or decrease is quantified through a relative change metric (Preis et al., 2013). The relative change metric Δ/n at time *t* within a time-frame of δt for *R* or *RSV* is defined as Equation 2.

$$\Delta/n(t, \delta t) = n(t) - N(t-1, \delta t) \quad (2)$$

where, $N(t-1, \delta t) = (n(t-1) + n(t-2) + \dots + n(t-\delta t))/\delta t$. For this research, N was computed at δt of both 1 week and 2 weeks, which means that relative change of current search volume or concentration was computed in comparison with its mean over the previous 2 and 3 weeks respectively. Based on this, we define the time-frame sensitivity, s , of relative change in RSV , Δ/RSV , to rise in R , Δ/R , as Equation 3. Only those cases were considered when Δ/R was greater than 10 units so as to ensure Δ/R is large enough to be perceivable.

$$s(\delta t) = \frac{\Sigma TP_{\delta t}}{\Sigma TP_{\delta t} + \Sigma FN_{\delta t}} \quad (3)$$

where, $TP_{\delta t}$ (true positive) is considered 1 when within a time-frame δt , both ΔR and the co-occurring ΔRSV are positive, $FN_{\delta t}$ (false negative) is considered 1 when Δ/R is positive but co-occurring Δ/RSV is negative. We also define weighted sensitivity $sw(\delta t)$, that takes into not only the direction of change, but also the magnitude of relative change as shown in Equation 4. A high value (≥ 1) of $\frac{\Delta RSV}{\Delta R}$ is a result of larger change in RSV relative to R . This could be on account media-generated hype or signify the high sensitivity of population.

$$sw(\delta t) = \frac{\Sigma(TP_{\delta t} \cdot \frac{\Delta RSV}{\Delta R})}{\Sigma(TP_{\delta t} \cdot \frac{\Delta RSV}{\Delta R}) + \Sigma(FN_{\delta t} \cdot \frac{\Delta RSV}{\Delta R})} \quad (4)$$

2.2 Data used

2.2.1 Google Trends (GT) Relative search volume (RSV) for any search term is available from the web search engine Google, since 2004 at daily scale on a geographical basis of a metropolitan region or a state. If the specified duration is bigger than 9 months then the RSV is available at weekly scale. These relative search volumes are available as integers on a relative scale of 0 to 100, 100 being the maximum RSV within the specified time period duration and 0 being the minimum. RSV was considered between January, 2015 to December to obtain data free of Google's previous algorithmic adjustments. The RSV values are affected by any new relative maxima or minima in search volume. Due to this, in some locations a very high RSV for a few dates along with very low values (5) for other dates was obtained, masking the subtler daily changes in RSV . To partly overcome this, smaller overlapping periods were downloaded and then joined by inter-calibration (Challet, Bel Hadj Ayed, 2014). In addition RSV was further scaled between 0 percentile and 95 percentile to remove extreme value RSV . The keywords probed were the search terms: 'air pollution', 'cough' and 'asthma'. These keywords are collectively referred to as GT henceforth. For Bangkok, Tokyo, Taipei and Seoul corresponding translated terms in Thai, Japanese, Chinese and Korean were used. The Google Trend time series for GT were batch downloaded using a Python module 'pytrends' available from GitHub. The RSV also exist in the form search 'Topics' in Google Trends, which allow search volume of all words related to that topic but since the full list of related words is not disclosed, this option was not considered.

2.2.2 MODIS based fine aerosol indicator Moderate Resolution Imaging Spectro-radiometer (MODIS) retrieved AOD (aerosol optical depth) and AE (angstrom exponent) global satellite products were decomposed into components using a decomposition scheme following (Misra et al., 2017). The decomposition scheme provides 3 components on a scale of 0 to 100 for high AOD and AE, high AE and low AOD, and low AOD and AE. The first component (labeled R) corresponds to $PM_{2.5}$ levels (Misra et al., 2017). To account for aerosol hygroscopic growth during presence of precipitable water, R was corrected for relative humidity using an empirical relationship (Chin et al., 2002) as shown in Equation 5.

$$R = \frac{R_{obs}}{rhum} \quad (5)$$

Where, R_{obs} refers to the AirRGB R corrected by relative humidity. Mass extinction coefficient (σ_{rhum}) characterizes hygroscopic growth in aerosol at different relative humidity ($rhum$) values (Chin et al., 2002). Relative humidity values for the locations were obtained from the reanalysis weather data, described below in 2.2.3. This dataset was composited at weekly level from 2011 to 2018 and was gridded at 10 km resolution.

2.2.3 Reanalysis weather data Daily local weather climatological information for each location was obtained between 2012 to 2019 from the NCEP/NCAR Reanalysis 1 data set from NOAA Earth System Research Laboratory (NOAA ESRL, n.d.). It is globally available at 2.5 degree mesh grid. Minimum temperature, maximum temperature, wind speed vectors and relative humidity were obtained at daily scales. Daily

mean temperature ($temp$) was calculated as average of minimum and maximum temperature, which along with relative humidity ($rhum$) and wind speed magnitude (wnd) was averaged to weekly level for further analysis.

3. RESULTS AND DISCUSSION

3.1 Trend of R and RSV

Figure 2 shows difference of R from 2015 and 2001 highlighting the study locations. Detailed four-year weekly trend for AirRGB R and RSV of 'air pollution', 'cough' and 'asthma' over some study locations is shown in Figure 6 in Appendix (Section 4). Taipei has the highest missing R due to frequent cloud cover (80.08% missing) followed by Bangkok (67.69% missing). Within the observable days, the highest R baseline values between 2010 and 2018 are seen in Bangkok (42.94), followed by Delhi (41.58) and Dhaka (39.13). In Delhi, Dhaka and Karachi, the baseline R between 2010 to 2018 is higher than the baseline between 2001 to 2010 by 1.92, 1.12 and 0.92 respectively and shows existence of seasonal trends. This clearly shows a rising fine mode aerosol in these locations. During the same period, R decreased by in Taipei, Seoul, Bangkok and Tokyo respectively. The factors affecting the seasonality are meteorological conditions like boundary layer inversion and rainfall and anthropogenic factors like annual crop-burning cycles. It appears that 'air pollution' searches also follow a seasonal regime and are only partially dependent on R values as higher value of R do not always translate into higher searches. At the same time there is a very clear but non-overlapping seasonality in 'cough' and 'asthma' RSV for Karachi, Delhi, Taipei, Seoul and Tokyo. The coefficient of variation (defined

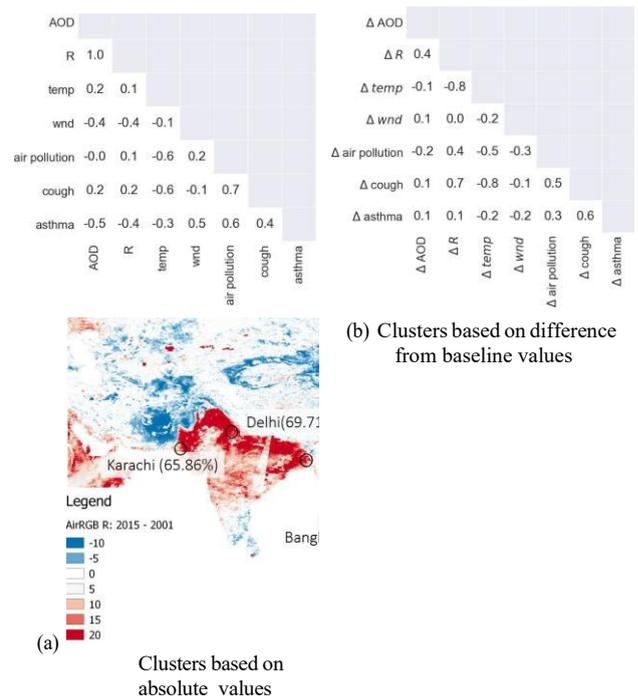


Figure 2. Difference of R from 2015 and 2001 highlighting the study locations and annual percentage cloud-free observable days using MODIS.

as ratio of σ and μ) indicates the standardized dispersion of frequency distribution, the high value of which can suggest strong presence of seasonality. Coefficient of variation for ‘cough’ RSV is highest in Delhi and for ‘asthma’ RSV it is highest in Tokyo. In some cities like Delhi there is overlap in seasonality of ‘cough’ and R . Specially for Delhi ‘cough’ shows seasonality with high R values in months of November, December and January. Also in Delhi, three exceptional peaks were noted in the December of 2016, 2017 and 2018 where ‘air pollution’ searches coincided very well with high R . This is likely a result of rising mass media news articles (Negi et al., 2017) and social campaigns since 2015. This could correspond to the ‘issue-attention cycle’ (Downs, 1996) wherein public attention is transformed from pre-discovery of problem stage to alarmed discovery.

3.2 Correlation between R and Google Trends RSV

Figure 3 shows the result of correlation between mean value of clusters derived from absolute values and deviance from baseline values of R , $temp$, and $wind$. Correlation among cluster set C_2 (comprising difference from baseline values) is higher than the cluster based only on absolute values (C_1 in Equation 1). In the C_2 case, correlation of ΔR with ‘air pollution’ ΔRSV (0.4), ‘cough’ ΔRSV (0.7) and ‘asthma’ ΔRSV (0.1) is higher than the C_1 case of correlation of R with ‘air pollution’ RSV (0.1), ‘cough’ RSV (0.2) and ‘asthma’ RSV (-0.4). Similar but smaller improvements are seen for correla-

tion of RSV with $temp$ and $wind$. It is interesting to note that compared to C_1 , C_2 results in more correlation improvement in ‘cough’ and ‘asthma’ RSV than ‘air pollution’. It is also seen that AOD has lower correlation compared to R with RSV . This implies that although visibility of the air could be a visual indicator of perception, it is the presence of fine aerosol that governs perception.

There also exists an inverse relationship between absolute values of R and $wind$ speed as well as ΔR and $\Delta temp$. This is a well-studied phenomena where high wind speed leads to pollutant dispersion and dilution. Higher than usual temperat-

ures cause to surface-to-air convection currents also leading to dilution. However it is interesting to note that both low *temp* (Figure 3 (a)) as well as negative $\Delta temp$ (Figure 3 (b)) lead to high 'air pollution' *RSV*. This correlation of *temp* with 'air pollution' *RSV* is stronger than that of *R* with 'air pollution' *RSV*. It suggests absolute and relative differences of temperatures govern the perception of air quality more than the *R*. This is shown more clearly in Figure 4. It is seen that a $\Delta temp$ lower than -3° and a ΔR larger than -2 than leads to positive

Figure 3. Heat map showing correlation between mean values of the clusters based on the (a) absolute values and (b) and difference from mean(Δ) for concentration R , temperature $temp$ and wind speed wnd .

'air pollution' ΔRSV . For both 'air pollution' and 'cough', very high positive RSV also occur when $\Delta temp$ is lower than

-7° and ΔR is higher than 11. For 'asthma', although positive ΔRSV occur with lower $\Delta temp$ and higher ΔR , it is not as high as 'air pollution' and 'cough'. It could be because apart from air pollution, asthma is also triggered by allergens. Further asthma affects children more than adults which could be another reason.

3.3 Sensitivity of GT to short-term rise in R

The sensitivity and weighted sensitivity in terms of relative change in search volume of keywords to rise in R in the 7 cities between 2015 and 2018 is shown in Figure 5. Compared to other cities in Figure 5 (a), in Tokyo 'air pollution' Δ/RSV has

the highest 1-week sensitivity ($s(1) = 0.60$) to Δ/R . Also, a lower sensitivity to 2-week ($s(2) = 0.50$) suggests that 'air pollution' RSV is lower in the second week (compared to 'air pollution' RSV in the first week) when continuous rise of R takes place. While in Seoul and Bangkok, 'air pollution' Δ/RSV has the lowest 1-week sensitivity ($s(1) = 0.45$). Dhaka has the highest sensitivity of 'air pollution' Δ/RSV to 2-week Δ/R

($s(2) = 0.65$), showing that continuous 2-week Δ/R lead to more positive Δ/RSV than just 1-week rise. There is a large gap between $s(1)$ (0.46) and $s(2)$ (0.64) in Seoul and between

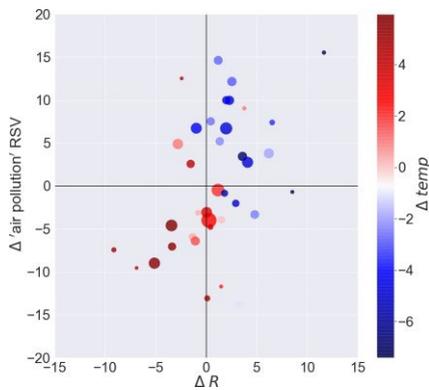
$s(2)$ (0.44) and $sw(2)$ (0.65) in Delhi, signifying that less than 50% positive Δ/R events are noticed, but when noticed they lead to larger than usual 'air pollution' Δ/RSV .

A high 1-week sensitivity of 'cough' Δ/RSV ($s(1) > 0.55$) is found in Dhaka, Karachi and Delhi in Figure 5 (b). 2-week sensitivity is highest in Tokyo ($s(2) = 0.66$) followed by Delhi ($s(2) = 0.6$). Low weighted sensitivity ($sw(1) = 0.37$, $sw(2) = 0.35$) in Delhi compared to non-weighted sensitivity ($sw(1) = 0.54$, $sw(2) = 0.60$) suggest that positive 'cough'

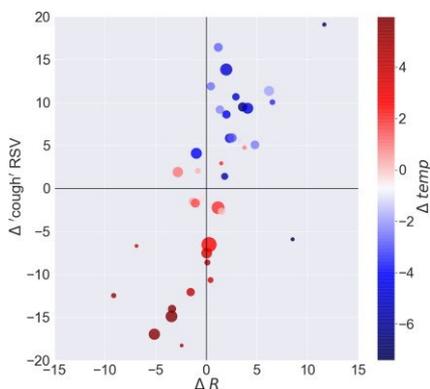
Δ/RSV takes place during positive Δ/R , however there is low correspondence between their magnitudes. In Figure 5 (c), Karachi has the highest 1-week sensitivity of asthma Δ/RSV ($s(1) = 0.66$) followed by Seoul ($s(1) = 0.56$) and the lowest

in Delhi ($s(1) = 0.34$) and Tokyo ($s(1) = 0.34$). Apart from Karachi and Seoul, $s(1)$ is lower than 0.5 for other cities, which means less than 50% of positive Δ/R correlates with increase

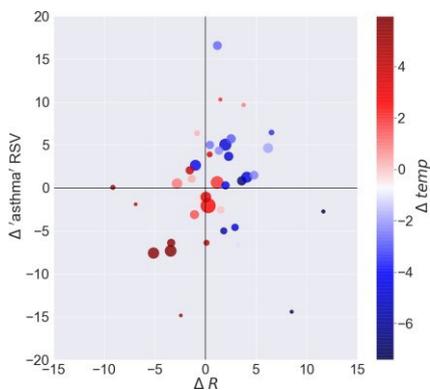
in asthma RSV . Sensitivity of 'cough' to Δ/R is more sensitive than 'asthma' in Dhaka, Delhi and Tokyo in contrast to 'asthma' which is more sensitive than 'cough' in Karachi and Seoul. This could be due to difference in the constituents of the local air pollutants, leading to differences in health impacts.



(a) 'air pollution'

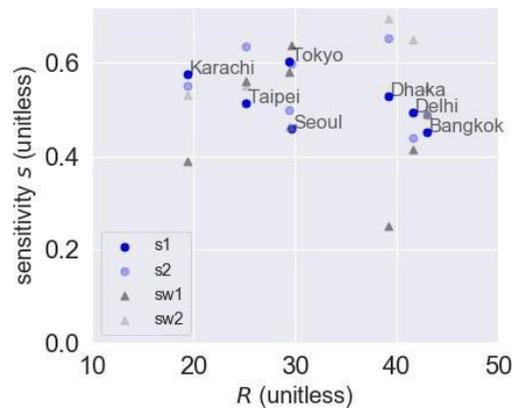


(b) 'cough'

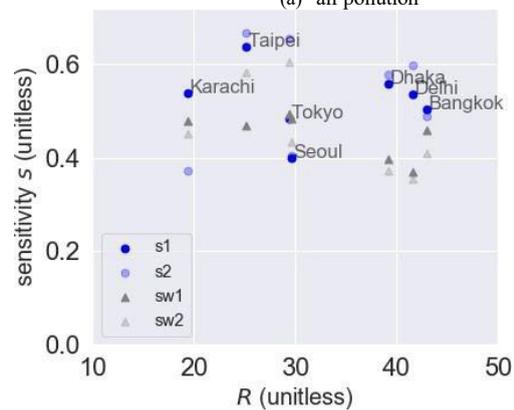


(c) 'asthma'

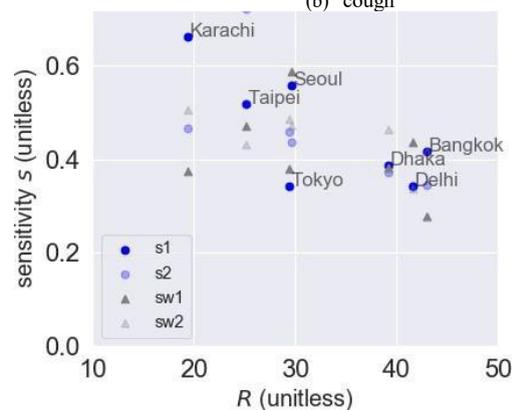
Figure 4. Plot of mean cluster values based on difference from baseline of concentration indicator (ΔR) and temperature ($\Delta temp$) showing difference from mean relative search volume (ΔRSV) values of (a) 'air pollution', (b) 'cough' and (c) 'asthma'. Generally positive ΔR and negative $\Delta temp$ leads to positive ΔRSV of 'air pollution', 'cough' and 'asthma'. Bubble size represents relative cluster size.



(a) 'air pollution'



(b) 'cough'



(c) 'asthma'

Figure 5. Mean sensitivity of the change in (a) 'air pollution', (b) 'cough' and (c) 'asthma' Δ/RSV with respect to change in AirRGB Δ/R for the 7 mega cities. s1 and s2 refer to sensitivity for the period of 1 and 2 weeks while sw1 and sw2 refer to their respective weighted sensitivity.

As asthma is also affected by biogenic particles such as pollen grains, it is possible that the pollen season and high pollutant period coincides in Karachi and Seoul, leading to high sensitivity in ‘asthma’.

3.4 Discussion

In this study we established that people’s perception of air quality and its health risk, cough is related with difference of fine mode aerosol level from its baseline. At same R but a different $temp$, perception as shown by RSV could be different. Since COPD diseases, whose common symptom is cough, could be caused to presence of certain viruses which spike in winter season, this confounds the influence of air pollution on health. A similar observation was also regarding COPD search terms (Boehm et al., 2019). Policy implication of this finding could be that for the success of policies that target people behavior, current approaches should also contains different provisions for awareness campaigns based on weather conditions. In fact, warmer seasons may needs extra measures and campaigns to spread awareness regarding pollution. This also implies that in places with climate warming, perception of air pollution as a risk may decrease. These findings are partially similar to that of climate change perception where frequent anomalies lead to lowered expectations and less chances of the events being perceived as remarkable (Moore et al., 2019).

From the perspective remote sensing this result has some bias due to consideration of only non-cloudy observations. A known limitation is that compared to surveys which provide a direct indication of public concern based in individual response, internet search data can only provide active collective public response (Bromley-Trujillo, Poe, 2018), and by extension cannot inform individualistic inclinations that may govern the reason for concern. Further internet searches represents only active interest, so when air pollution may not be searched for on a clean air day, it is unclear whether perception regarding air pollution has changed. Although studies using Google Trends dataset are increasing, currently set standards for data collection and an optimal search strategy is lacking (Nutti et al., 2014). Our analysis was limited by a narrow selection of Google Trend keywords. There are two issues here. First, a broader selection of keywords is needed that to identify effect of polluted air on people, e.g. ‘smog’, ‘sick’, ‘mask’, etc. However these words are also carry several other connotations which make can confound the interpretation of search volume leading to lower content validity (Mellon, 2014). Second, when there is a preference to search in a local informal language, the interpretation of the word may not always coincide with the direct English translation. e.g. in Hindi language, the same word is used to inquire about wind and air quality. A systematic approach could be scraping the most frequently used words in air pollution related news articles in the local language and obtaining search volumes corresponding to those terms. Another limitation is with using Google Trends unavailability of daily-level data and absolute search frequencies (Maurer, Holbach, 2016). So far we assumed that the most popular method of obtaining information is by searching for ‘air pollution’ or ‘asthma’. However with rise in smartphone app based alert systems in many places, the apparent concern for these topic as measured on Google Trends may decrease leading to erroneous conclusions. Further, this analysis was performed at the scale of one week with a dataset of 4 years. For a much smaller time period, Google Trends is also available at daily and hourly scales. Temporally denser data can be obtained and studied with dense concentration readings, e.g. Aeronet retrieved AOD or personal PM2.5

measurements. If behavior of other emission sources like real time traffic congestion status can be obtained, its inclusion can reveal impact of emission sources and rising concentration on human health. This could be useful for preparation of medical facilities and stocking medicinal supplies. Nighttime light from VIIRS is also an important predictor of human activities which shows discrimination with air pollution (Misra, Takeuchi, 2016) and can be included as independent variable to separate perception among diverse study locations.

4. CONCLUSION

In this paper we have shown that geo-located social sensing data can combine with traditional remote sensing data to reveal perception and sensitivity of web-based search volumes with respect to outdoor air pollution. This is the first study to demonstrate relationship between air quality related search volumes and the air quality concentration. The results suggest that tendency to search for ‘air pollution’ and ‘cough’ occurs when Air- RGB R is in excess and temperature is below the baseline values. For the Asian megacities, the highest sensitivity to rise in pollutant concentration is shown by ‘air pollution’ in Delhi, ‘asthma’ and ‘cough’ in Dhaka, ‘headache’ in Seoul and ‘mask’ in Tokyo. Further the level of R that results in increased RSV of GT varied for the cities. It was found that rise in RSV of GT occurred at higher R in low GNIpc cities than the R levels of high GNIpc cities. RSV and remote sensing datasets could be beneficial in assessing perception and health effects during high pollution episodes.

ACKNOWLEDGEMENTS

NCEP Reanalysis data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their web site at <https://www.esrl.noaa.gov/psd/>. The first author is thankful to Tateisi Foundation for providing international conference travel grant.

REFERENCES

- Bickerstaff, K., Walker, G., 2001. Public understandings of air pollution: The ‘localisation’ of environmental risk. *Global Environmental Change*, 11(2), 133–145.
- Boehm, A., Pizzini, A., Sonnweber, T., Loeffler-Ragg, J., Lamina, C., Weiss, G., Tancevski, I., 2019. Assessing global COPD awareness with Google Trends. *European Respiratory Journal*, 1900351.
- Bousquet, J., O’Hehir, R. E., Anto, J. M., D’Amato, G., Mo’sges, R., Hellings, P. W., Van Eerd, M., Sheikh, A., 2017. Assessment of thunderstorm-induced asthma using Google Trends. *Journal of Allergy and Clinical Immunology*, 140(3), 891–893.e7.
- Bromley-Trujillo, R., Poe, J., 2018. The importance of salience: public opinion and state policy action on climate change. *Journal of Public Policy*, 1–25.
- Challet, D., Bel Hadj Ayed, A., 2014. Do Google Trend Data Contain More Predictability than Price Returns? *SSRN Electronic Journal*, 1–17. <http://www.ssrn.com/abstract=2405804>.

- Chin, M., Ginoux, P., Kinne, S., Torres, O., Holben, B. N., Duncan, B. N., Martin, R. V., Logan, J. a., Higurashi, A., Nakajima, T., 2002. Tropospheric Aerosol Optical Thickness from the GOCART Model and Comparisons with Satellite and Sun Photometer Measurements. *Journal of the Atmospheric Sciences*, 59(3), 461–483.
- Downs, A., 1996. Up and down with ecology– the ” issue-attention cycle”. P. Peretz (ed.), *The politics of american economic policy making*, M.E. Sharpe, New York, 48.
- Eden, S., 1996. Public participation in environmental policy: Considering scientific, counter-scientific and non-scientific contributions. *Public Understanding of Science*.
- Egondi, T., Kyobutungi, C., Ng, N., Muindi, K., Oti, S., 2013. Community Perceptions of Air Pollution and Related Health Risks in Nairobi Slums. *International Journal of Environmental Research and Public Health*, 4851–4868.
- Elliott, S. J., Cole, D. C., Krueger, P., Voorberg, N., Wakefield, S., 1999. The Power of Perception: Health Risk Attributed to Air Pollution in an Urban Industrial Neighbourhood. *Risk Analysis*, 19(4), 621–634. <http://doi.wiley.com/10.1111/j.1539-6924.1999.tb00433.x>.
- Eysenbach, G., 2006. Infodemiology: tracking flu-related searches on the web for syndromic surveillance. *Annual Symposium proceedings. AMIA Symposium*, 244–8.
- Flachsbar, P. G., Phillips, S., 1980. An Index and Model of Human Response to Air Quality. *Journal of the Air Pollution Control Association*, 30(7), 759–768.
- Goel, S., Hofman, J. M., Lahaie, S., Pennock, D. M., Watts, D. J., 2010. Predicting consumer behavior with Web search. *Proceedings of the National Academy of Sciences*, 107(41), 17486–17490. <http://www.pnas.org/lookup/doi/10.1073/pnas.1005962107>.
- Kahn, M. E., Zhang, X., Wang, J., Sun, C., Zheng, S., 2019. Air pollution lowers Chinese urbanites’ expressed happiness on social media. *Nature Human Behaviour*. <http://dx.doi.org/10.1038/s41562-018-0521-2>.
- Kaufman, Y. J., Fraser, R. S., 1983. Light Extinction by Aerosols during Summer Air Pollution. *Journal of Climate and Applied Meteorology*, 22(10), 1694–1706.
- Li, J., Sun, S., Tang, R., Qiu, H., Huang, Q., Mason, T., Tian, L., 2016. Major air pollutants and risk of COPD exacerbations: a systematic review and meta-analysis. *International Journal of Chronic Obstructive Pulmonary Disease*, Volume 11, 3079–3091. <https://www.dovepress.com/major-air-pollutants-and-risk-of-copd-exacerbations-a-systematic-review-peer-reviewed-article-COPD>.
- Maurer, M., Holbach, T., 2016. Taking Online search queries as an indicator of the public agenda: The role of public uncertainty. *Journalism and Mass Communication Quarterly*, 93(3), 572–586.
- Mellon, J., 2014. Internet Search Data and Issue Salience: The Properties of Google Trends as a Measure of Issue Salience. *Journal of Elections, Public Opinion and Parties*, 24(1), 45–72. <http://dx.doi.org/10.1080/17457289.2013.846346>.
- Misra, P., Fujikawa, A., Takeuchi, W., 2017. Novel decomposition scheme for characterizing urban air quality with MODIS. *Remote Sensing*, 9(8), 1–19. <http://www.mdpi.com/2072-4292/9/8/812>.
- Misra, P., Takeuchi, W., 2016. Analysis of air quality and nighttime light for Indian urban regions. *IOP Conference Series: Earth and Environmental Science*, 37(2), 012077.
- Moore, F. C., Obradovich, N., Lehner, F., Baylis, P., 2019. Rapidly declining remarkability of temperature anomalies may obscure public perception of climate change. *Proceedings of the National Academy of Sciences*, 116(11), 4905–4910. <http://www.pnas.org/lookup/doi/10.1073/pnas.1816541116>.
- Negi, N., Onyon, L., Murukutla, N., Puri, P., Mullin, S., 2017. Online media coverage of air pollution risks and current policies in India: A content analysis. *WHO South-East Asia Journal of Public Health*, 6(2), 41.
- NOAA ESRL, n.d. NCEP/NCAR Reanalysis 1. <https://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.pressure.html>.
- Nuti, S. V., Wayda, B., Ranasinghe, I., Wang, S., Dreyer, R. P., Chen, S. I., Murugiah, K., 2014. The use of google trends in health care research: A systematic review. *PLoS ONE*, 9(10).
- Polgreen, P., Chen, Y., Pennock, D., Nelson, F., 2008. Using Internet Searches for Influenza Surveillance. *Clinical Infectious Diseases*, 47(11), 1443–1448.
- Preis, T., Moat, H. S., Stanley, H. E., 2013. Quantifying Trading Behavior in Financial Markets Using Google Trends. *Scientific Reports*, 3(1), 1684. <http://www.nature.com/articles/srep01684>.
- Roelofsen, P., 2018. A new methodology for the evaluation of the perceived air quality depending on the air pollution , caused by human bioeffluents , the temperature , the humidity as well as the air velocity. 8975.
- Rosa, E. A., 2003. The logical structure of the social amplification of risk framework (SARF): Aferatheoretical foundations and policy implications. *The Social Amplification of Risk*.
- Rosi, A., Mamei, M., Zambonelli, F., Dobson, S., Stevenson, G., Ye, J., 2011. Social sensors and pervasive services: Approaches and perspectives. *2011 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)*, 17number 1, IEEE, 525–530.
- Saksena, S., 2011. Public Perceptions of Urban Air Pollution Risks. *Risk, Hazards & Crisis in Public Policy*, 2(1), 19–37. <http://doi.wiley.com/10.2202/1944-4079.1075>.
- Schikowski, T., Mills, I. C., Anderson, H. R., Cohen, A., Hansell, A., Kauffmann, F., Kramer, U., Marcon, A., Perez, L., Sunyer, J., Probst-Hensch, N., Kunzli, N., 2014. Ambient air pollution: a cause of COPD? *European Respiratory Journal*, 43(1), 250–263. <http://erj.ersjournals.com/cgi/doi/10.1183/09031936.00100112>.
- Sjøberg, L., Moen, B., Rundmo, T., 2004. Explaining risk perception. An evaluation of the psychometric paradigm in risk perception research. Technical report, Trondheim.
- Smith, J., Woodcock, A., 2006. Cough and its importance in COPD. *International Journal of COPD*, 1(3), 305–314.

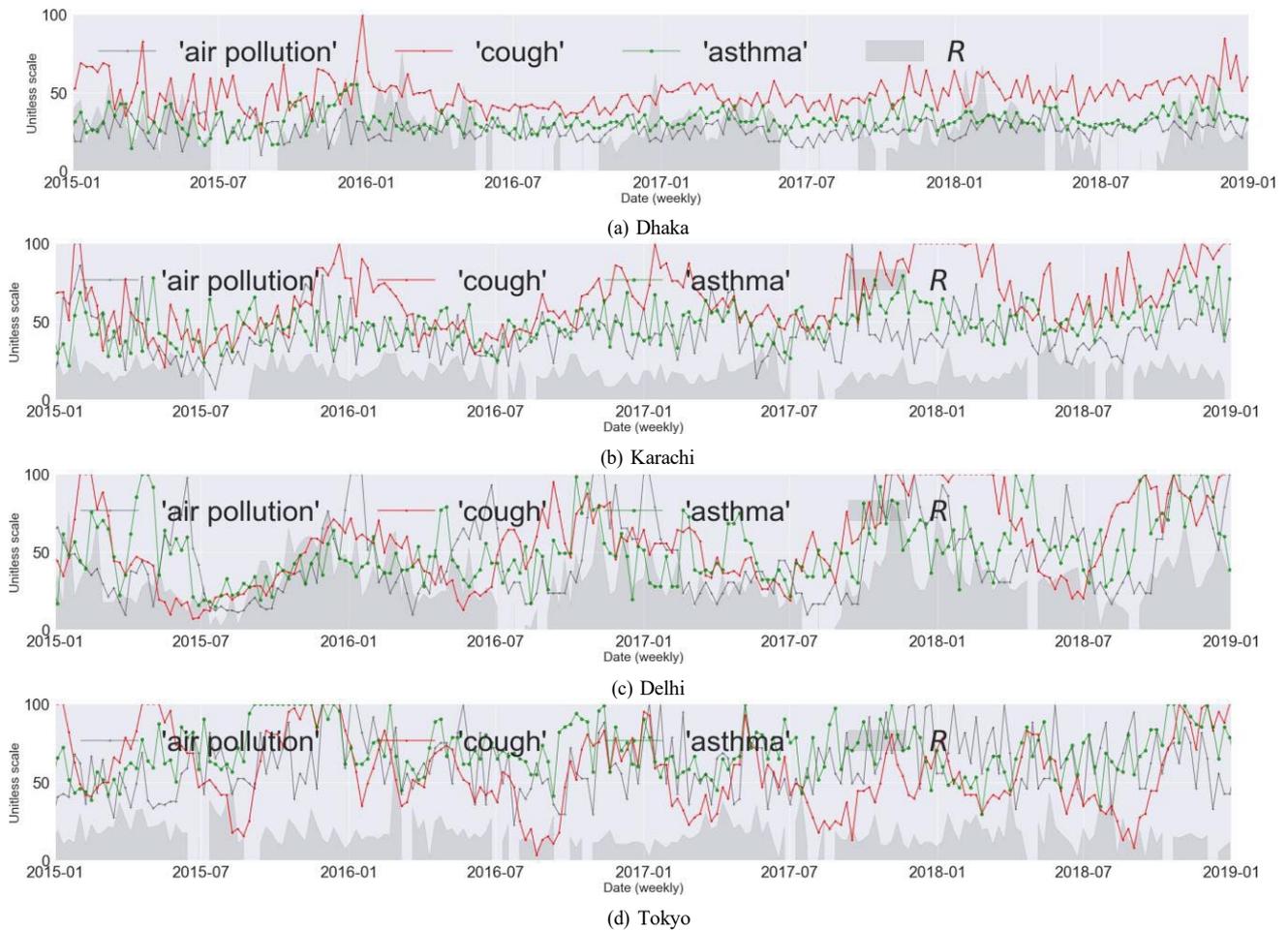


Figure 6. Weekly relative search volume (RSV) of the search terms: 'air pollution' (black), 'cough' (red) and 'asthma' (green) are shown along with the fine aerosol component R (shaded gray) for (a) Dhaka, (b) Karachi, (c) Delhi and (d) Tokyo under study.

APPENDIX

Weekly relative search volume (RSV) of the search terms is shown in Figure 6.