What matters in public perception and awareness of air quality? Quantitative assessment using internet search volume data

Young-Hee Ryu and Seung-Ki Min

Division of Environmental Science and Engineering, Pohang University of Science and Technology (POSTECH), Pohang, Republic of Korea

E-mail: yhryu02@postech.ac.kr

Keywords: air quality, perception, awareness, internet search volume data, google trends

Abstract

Recently, the issue of air quality in South Korea reached an unprecedented level of social concern regarding public health, quality of life, and environmental policies, even as the level of particulate matter less than 10 µm (PM$_{10}$) showed a decreasing trend. Why have social concerns emerged in recent years, specifically after 2013–2014? This study aims to understand how people perceive air quality apart from the measured levels of airborne pollutants using internet search volume data from Google and NAVER. An empirical model that simulates the air quality perception index (AQPI) is developed by employing the decay theory of forgetting and is trained by PM$_{10}$, visibility, and internet search volume data. The results show that the memory decay exponent and the accumulation of past memory traces, which represent the weighted sum of past perceived air quality, play key roles in explaining the public’s perception of air quality. A severe haze event with an extremely long duration that occurred in the year 2013–2014 increased public awareness of air quality, acting as a turning point. Before the turning point, AQPI is more influenced by sensory information (visibility) due to the low awareness level, but after the turning point it is more influenced by PM$_{10}$ and people slowly forget about air quality. The retrospective AQPI analysis under a low level of awareness confirms that perceived air quality is indeed worst in the year 2013–2014. Our results provide a better understanding of public perception of air quality, and will contribute to the creation of more effective regulatory policies. It should be noted, however, that the proposed model is primarily meant to diagnose historic public perception and that more sophisticated models are needed to reliably predict perception of air quality.

1. Introduction

Recently, the air quality in South Korea has received an unprecedented level of attention. According to internet search volume data from Google and NAVER (the most popular search engine in the world and South Korea, respectively), public concerns about air quality emerged in the winter of 2013–2014 (see the internet search volume data associated with air quality in figure 1). South Korea and China have made continuous efforts to reduce pollutant emissions; as a result, emissions of particulate matter with a diameter less than 10 µm (PM$_{10}$) in Seoul, the capital of South Korea, also show a decreasing trend during 2003–2008 (Yeo et al 2019). During the winter-spring season, the mean PM$_{10}$ concentration in Seoul shows a decreasing trend from 2001, but no trend in the last six years (figure 1). It is, therefore, mysterious why public concerns emerged specifically in the winter of 2013—spring of 2014 even though the mean PM$_{10}$ concentration was comparable to or even lower than it was in previous years. In the present study, we aim to explore the factors that affect perceptions of air quality and why perceived air quality is not consistent with measured levels of air pollutants.
It is crucial to understand key triggers and dynamic patterns (herein, decay pattern) of public perception of air quality for preparedness and mitigation of severe air pollution episodes. Self-reported health status, such as disease severity and depressive symptoms, is often influenced by perceived air quality, rather than measured levels of air pollutants (Lercher et al. 1995, Yen et al. 2006, Piro et al. 2008). Air quality perceptions have been investigated over the past decades, but the majority of studies rely on survey data (Auliciems and Burton 1973, Kim et al. 2012, Pantavou et al. 2018, Reames and Bravo 2019). Survey-based studies can conduct in-depth research that promotes understanding of cognitive processes of public perception and awareness, but the number of participants who take surveys is limited mainly by cost, which likely results in poor representations of the target community, such as demographic characteristics and spatial coverage. Due to the high costs, surveys are difficult to perform frequently and with regularity.

Along with increasing rates of internet usage in recent decades, internet search volume data have been used to track public interests, as reviewed by Jun et al. (2018), and the data can be regarded as a good indicator of how people respond to social issues including economic activities (Choi and Varian 2012), epidemic disease (Dugas et al. 2013), public policy (Shirky 2011), and natural disasters (Kam et al. 2019). For example, Google Trends provides an unbiased sample of Google search data at various time scales (daily to monthly) in all parts of the world since 2004.

Only a few studies have utilized internet search volume data to investigate how people perceive and respond to air quality (Lu et al. 2018, Dong et al. 2019). Lu et al. (2018) used the search index from Baidu (China’s largest search engine) to investigate the relationship between air quality and public concern monitoring data. They found that the year 2013 was a turning point for public concern about air quality in China and showed a positive correlation between PM2.5 (particulate matter with a diameter less than 2.5 μm) and the search index with a time lag of 0–4 d. However, Lu et al. (2018) did not explain why the year 2013 was a turning point even though the air quality in China in the past was much worse than it
was in 2013 (based on the time series of PM$_{2.5}$ shown in Lang et al. (2017) and Lin et al. (2018)). Another recent study (Dong et al 2019) used the Baidu search index for the city of Shanghai, China, and showed a quicker public response to air quality index, within a day.

In the present study, we develop an empirical model that simulates public perception of air quality (referred to as air quality perception index, AQPI) by employing the decay theory of forgetting and using physical parameters. Previous studies (Lu et al 2018, Dong et al 2019) used only air quality monitoring data (PM$_{2.5}$ or air quality index). Our model considers three main factors: PM$_{10}$ concentration, visibility, and memory decay exponent. This model is designed to assess the relative importance of physical and psychological factors in air quality perception. According to a review by Oltra and Sala (2014) on public perception of air quality, the parameters PM$_{10}$ and visibility are risk-related factors which correspond to the level of air pollution and the sensory characteristics of air pollution, respectively. The memory decay exponent can be regarded as a psychological factor associated with local memory and prior experience. Note that the model developed in the present study is intended to be primarily used to understand public perception of air quality rather than to forecast AQPI.

2. Data and methods

2.1. PM$_{10}$ data

Daily PM$_{10}$ data at 25 Seoul stations (located in urban settings) are averaged. Note that the winter-spring season spanning from November to May is only considered because of the significant seasonality of PM$_{10}$, which is high in the winter and spring seasons (Ghim et al 2015). In this study, the first year of each winter-spring season is used to indicate the period from November of that year to May of the next year: e.g. the 2013 season means November 2013 to May 2014. PM$_{2.5}$ data in South Korea are only available for 2015 and later, so PM$_{10}$ data instead of PM$_{2.5}$ data are used in this study because of the longer record of PM$_{10}$ data (since 2001).

2.2. Meteorology data

Daily visibility, precipitation (only counted if daily precipitation is greater than or equal to 1 mm), and air temperature at the Seoul synoptic surface station (37.57°N, 126.97°E) from November 2001 through May 2019 are used in this study. The vertical profiles of potential temperature and wind speed observed at the Osan rawinsonde station (37.1°N, 127.03°E) are used to identify boundary-layer stagnant days.

2.3. Number of stagnant days

The boundary-layer air stagnation index developed by Huang et al. (2018) is modified in this study to characterize stagnant days in South Korea and to investigate how weak-ventilation days are associated with AQPI. Unlike Huang et al. (2018), the maximum mixing depth is replaced with the boundary layer height at 06 UTC (= 15:00 local time) at Osan rawinsonde station; it is estimated using the vertical profile of potential temperature following the method of Liu and Liang (2010). Ventilation is defined as the vertical integration of wind speeds from the surface to the top of the boundary layer. As in Huang et al. (2018), a day is identified as stagnant when the ventilation is less than 6000 m$^2$ s$^{-1}$; the value of convective available potential energy is less than that of convective inhibition; and daily precipitation is less than 1 mm.

2.4. Internet search volume data and number of internet news

Google Trends and NAVER Datalab provide search activity data since 2004 and 2016, respectively. In this study, public perception of air quality in South Korea is quantitatively measured at the aggregate level of the nation using Google Trends or NAVER Datalab data. The selection of search terms associated with a topic is important when exploring internet search activities. We use the search term ‘fine dust’ (미세먼지 in Korean), which is widely used among the public to represent air quality in South Korea. Figure 1 shows the internet search volume data with the keyword ‘fine dust.’

Google Trends provides relative search volume data, in which the maximum value in a given search period is scaled to 100. The search period can be selected by a user, and the temporal resolution of search volume data is automatically determined depending on the length of the selected search period. We obtain monthly Google Trends data (denoted by GT$(m)$, where $m$ is the index for month, $m = 1, \ldots, 79$) for the period from November 2012 to May 2019 and daily Google Trends data for each month (denoted by GT$_m(j)$, where $j$ is the index for a day in month $m$). Each GT$_m(j)$ has its own maximum value of 100 in month $m$. The weighted daily Google Trends data, GT$(m, j)$, is computed by first dividing the daily Google Trends data, GT$_m(j)$, by its sum, $\sum_n GT_n(j)$ (where $n$ is the number of days in month $m$), and then weighting it by the monthly Google Trends data in the corresponding month, GT$(m)$. The resulting GT$(m, j)$ represents continuous and consistent search volume data at daily time scales over the entire period. To highlight the early stage of increasing public awareness during the 2013 season, the daily GT value on 27 February 2014, GT$(16, 27)$, is scaled to 100 ($GT_scaled(m, j)$).

$$GT_{scaled}(m, j) = GT(m) \times \frac{GT_m(j)}{\sum_n GT_m(j)} \times \frac{100}{GT(16, 27)}.$$

Due to the scaling, the value of $GT_{scaled}(m, j)$ can be higher than 100. We also consider the trend in
the rate of individuals using the internet (World Bank data, data.worldbank.org) and find a gradual increase in the rate during the study period (e.g. 84% in 2012 and 96% in 2018). We apply this rate to $GT_{scaled}(m_j)$ and use the data in this study (see supplementary for details, available online at stacks.iop.org/ERL/15/0940b4/mmedia).

NAVER Datalab provides continuous daily internet search volume data since 2016, so the data can be directly compared to $GT_{scaled}(m_j)$ without the need for post-processing. NAVER Datalab data are obtained using the same search term ‘fine dust’. Due to the popularity of NAVER in South Korea, we chose to use NAVER internet search volume data since 2016. The temporal correlations between the two datasets are high during the overlapping period (0.913 in 2016–2017 and 0.926 in 2017–2018). An analysis of the cross-validation of the data obtained from Google Trends and NAVER Datalab is provided in supplementary.

The daily number of internet news related to air quality is obtained from the NAVER search engine; internet news has been archived in NAVER since 2001. The procedures used to obtain the number of internet news are given in supplementary.

2.5. Air quality perception index model

Empirical models that estimate AQPI are developed using $PM_{10}$, visibility, and temperature (used for the type 2 model) and by introducing the memory decay exponent. The scaled internet search volume data $GT_{scaled}(m_j)$ are assumed to be the true values of AQPI. In psychology, it is traditionally thought that forgetting follows a power law function (Anderson 1982, Wixted and Ebbesen 1997). So, the strength of the memory trace for a single event ($S$) decays with time according to a power law function:

$$S = A_i t^{-d}$$

where $A_i$ is the strength of public memory on day $i$; $t$ is the time ($t = 1$ on day $i$); and $d$ is the memory decay exponent to be estimated, which quantifies how quickly people forget about the event. During the early stage of increasing public awareness, $A_i$ is assumed to be the power of $PM_{10}$ times the power of visibility on day $i$ with exponents $p$ and $v$, respectively.

$$A_i = PM_{10}^p \text{Visibility}_i^v.$$  

As Anderson (1982) proposed that the total strength of memory is the summation of individual strengths (i.e. $\sum S$), the AQPI on the present day is expressed as the summation of the strengths of past memories.

AQPI - type 1: $\text{AQPI} = \sum_{i=0}^{t} PM_{10}^p \text{Visibility}_i^v t^{-d}$

where $t = |i| + 1$ to set the present day as 1. Values with a negative $i$ indicate information from $i$ day(s) ago. This model (hereafter AQPI-type 1) weights recent events more heavily than past events.

The coefficients $p$, $v$, and $d$ are determined by minimizing root-mean-square-error (RMSE) against the internet search volume data $GT_{scaled}(m_j)$. After the 2014 season, the model above is found to be less accurate compared to its performance for the 2013 season. So, air temperature is additionally considered to improve model performance under the assumption that people are more concerned about outdoor air quality when the weather is mild (AQPI-type 2 model). We assume that air quality perception is not sensitive to temperature when it is very cold or very warm outside; thus, we introduce a sigmoid function, $(1 + e^{-\mu(T_i - \tau)})^{-1}$, to the strength of public memory ($A_j$). Here, $\mu$ represents how sensitive people are to a temperature near the temperature at $\tau$, and $T_i$ is the air temperature on day $i$. The AQPI-type 2 model is expressed by

$$\text{AQPI - type2: } \text{AQPI} = \sum_{j=-6}^{0} \frac{PM_{10}^p \text{Visibility}_j^v}{1 + e^{-\mu(T_j - \tau)}} t^{-d}.$$  

The obtained coefficients are given in table 1.

3. Results and discussion

3.1. $PM_{10}$, visibility, and internet search volume

$PM_{10}$ shows decreasing trends in general from 2001 onwards either with or without Asian dust; however, it does not show a significant trend since 2012 (figure 1). Asian dust and its adverse health effects have been recognized by the South Korean public since the 2000 s (Lee et al 2005), so days influenced by Asian dust are filtered out. Unlike $PM_{10}$, visibility does not show any significant trends during the 18-year period, but the 2013 season indeed shows the second-worst visibility (11.390 km without Asian dust). This finding implies that degraded visibility may be the reason for the emerging public concern regarding air quality in the 2013 season. The 2007 season shows the worst visibility during the 18-year period (11.387 km without Asian dust), but few public concerns about air quality were raised. These findings suggest that visibility is not the sole driving factor in public perception of air quality.

In the 2012 season, very low search volume data are seen in Google Trends. Note that there are no or negligible search volume data or internet news associated with ‘fine dust’ prior to 2012 because the term ‘fine dust’ began to emerge in 2012 (figure 1). The volume of Google search data associated with ‘fine dust’ abruptly increases in the 2013 season, decreases in the 2014 season, increases again in the next season, and reaches its peak in the 2016 season. After the 2016 season, the internet search volume declines. So, the
The memory decay exponent and the accumulation of past memory traces (hereafter, memory effects) play key roles in public perception especially when the level of public awareness is high. The AQPI estimated only by PM$_{10}$ and visibility ($A_v$, in section 2.5) is greatly underestimated, particularly during the severe haze event, because it does not account for the duration and accumulation of memory traces (dashed line in figure 2). The correlation coefficient in the early (later) period without the memory effects decreases to 0.84 (0.77) from 0.92 (0.87) with the memory effects. Therefore, the memory effects, which represent the weighted sum of past perceptions of air quality, should be considered in addition to the measured levels of air pollutants (PM$_{10}$) and visibility to understand the public perception of air quality.

The performance of the AQPI type-2 models after the 2013 season shows relatively high correlation coefficients, greater than 0.8 for all years (table 1). The coefficients for the 2015 and 2016 seasons differ markedly from those for other years. This is most likely attributable to severe Asian dust events that occurred in the 2015 and 2016 seasons, which led to a very high daily PM$_{10}$ of ~ 200 µg m$^{-3}$. Note that after the 2013 season the number of days with Asian dust is included when AQPI is estimated by the AQPI type-2 model as people become aware that Asian dust increases PM$_{10}$. Other than the 2015 and 2016 seasons, the exponent $p$ ranges from 0.9 to 1.05 and $v$ ranges from 0.24 to 0.59. Although the coefficients vary yearly, the magnitude of $p$ is consistently larger than that of $v$, indicating that air quality perception is influenced by PM$_{10}$ to a greater extent than visibility after the turning-point season. The decline in PM$_{10}$ threshold over the years also supports the finding that PM$_{10}$ plays a more influential role in air quality perception (table S1). The PM$_{10}$ threshold is defined as the minimum PM$_{10}$ among the PM$_{10}$ values showing daily increase in internet search volume larger than the corresponding seasonal mean. The threshold in the 2013 season is 52.1 µg m$^{-3}$, and it decreases to 40.3 µg m$^{-3}$ in the 2018 season. This suggests that laypeople become more informed about the actual level of pollutants when they have an elevated level

2013 season is also a turning point for South Korea, which is analogous to the situation in China (Lu et al 2018).

### 3.2. AQPI model

In the 2013 season, a severe haze event occurred for 10 consecutive days (22 Feb–3 March, 2014), and it triggered increases in internet search activity and public awareness. The Google Trends data show a peak on 27 February, 2014. This severe and extremely long event is a key factor that makes the 2013 season the turning-point season. This haze event is the longest and worst haze episode during the 18-year study period based on perceived air quality (not based on PM$_{10}$). The 2013 season is divided into early (before the severe haze event) and later (during and after the haze event) periods to quantitatively assess the changes in dynamic characteristics of perception and awareness of air quality. Two AQPI-type 1 models are developed for the early (AQPI$_{low\text{-}awareness}$ model) and later periods (AQPI$_{high\text{-}awareness}$ model) (see table 1 for coefficients). Clearly, either PM$_{10}$ or visibility alone does not reflect the Google Trends data, but the AQPI models perform well. Note that the early 2013 period excludes rainy days when determining coefficients under the assumption that people are not concerned about outdoor air quality when it is raining. Regardless of rainy days, the simulated AQPI is correlated with Google Trends data (for the early period, $r = 0.83$ with rainy days and 0.92 without rainy days).

The ratio of visibility exponent to PM$_{10}$ exponent before the haze event ($v/p = 1.14$) is larger than the $v/p$ ratio after the event (0.58). This implies that people perceive the severity of air quality based more on visibility than PM$_{10}$ when public awareness is low. This finding is consistent with public perception of pollutants related to coal combustion and mining in the northeast European part of Russia (Walker et al 2006). They found that locals evaluate air pollution using direct observations (sensory information, e.g. discoloration of snow) and personal experiences (e.g., respiratory symptoms), and that scientific knowledge plays a minor role in shaping these perceptions. The magnitude of the decay exponent becomes smaller ($d = −1.17$) when public awareness is raised compared to in the early period ($d = −1.62$). This means that people forget more slowly after the turning point event than before.

#### Table 1. Coefficients of AQPI models for each period and performance statistics. The unit of $\tau$ is °C, and the other coefficients are unitless. The last column, $r$, indicates the correlation coefficient (Pearson correlation).

| period   | $p$  | $v$  | $\mu$ | $T$ | $d$     | min RMSE/ max AQPI (%) |
|----------|------|------|-------|-----|---------|------------------------|
| 2013 early | 1.11 | 1.26 | 1.14  | 0.73| −1.62   | 5.1/0.92               |
| 2013 late  | 0.88 | 0.51 | 0.87  | 1.17| −1.17   | 9.4/0.87               |
| 2014      | 1.02 | 0.59 | 0.17  | 5   | −1.7    | 7.8/0.86               |
| 2015      | 1.5  | 0.28 | 0.07  | 40  | −2.0    | 8.3/0.88               |
| 2016      | 1.6  | 0.28 | 0.07  | 45  | −2.1    | 5.9/0.91               |
| 2017      | 1.05 | 0.24 | 0.32  | 3   | −2.1    | 11.8/0.81              |
| 2018      | 0.9  | 0.53 | 0.58  | −2  | −0.9    | 7.3/0.89               |
Figure 2. (a) Timeseries of PM$_{10}$, 100/visibility, and Google Trends data for the period from 1 November 2013 through 31 May 2014. The gray shading in (a) indicates precipitation days with a daily precipitation amount greater than or equal to 1 mm. (b) Timeseries of Google Trends data and estimated air quality perception index by the AQPI$_{low\_awareness}$ and AQPI$_{high\_awareness}$ models. The dashed light blue (magenta) line indicates the estimated air quality perception index by AQPI$_{low\_awareness}$ (AQPI$_{high\_awareness}$) excluding memory effects (see main text for details).
of awareness, and so they respond to a lower level of PM$_{10}$.

The temperature dependency coefficients also show year-to-year variations, but their ranges are quite small except for the 2015 and 2016 seasons: 0.17–0.44 for $\mu$ and $-2$–5 for $\tau$. After the 2013 season, the magnitude of exponent $d$ increases except for the 2018 season, meaning that people lose their memory about air quality faster compared to the turning-point season (the 2013 season). The smaller magnitude of $d$ in the 2018 season is of interest, and the large number of internet news issued in the 2018 season (see figure 1(a)) is speculated to play a role in the retaining of public perception.

To apply our AQPI models to other societies, we recommend that the models be trained with relevant datasets to determine the values of the coefficients because the level of public awareness may be different from the target community in this study (South Korea). The AQPI models developed in this study should be regarded as optimized models that best reproduce the historical AQPI measured by internet search volume to understand public perception and associated dynamics over time. For use in air quality perception forecasting, different approaches (e.g. statistical time series methods) might be required because such methods can include other factors (e.g. the influence of mass media) and complex interactions among factors.

3.3. Historical severity of perceived air quality

Figure 3(a) shows the number of days on which the simulated AQPI given by the AQPI$_{low\_awareness}$ model is greater than 21 (the standard deviation of GT$_{scaled}$ in the 2013 season) (N$_{APQI}$) for the seasons from 2001 to 2018. The 2013 season indeed shows the largest N$_{APQI}$ (40 d) during the 18-year period, followed by the 2002 season (31 d). So, people perceived that poor air quality days occurred most frequently in the 2013 season. Figure 3(b) shows the number of days estimated by the AQPI$_{high\_awareness}$ model to answer the question: “What if the level of public awareness of air quality were high in the past?” The worst (second worst) air quality would have been experienced in the 2002 (2001) season, which is consistent with the highest (second highest) PM$_{10}$ in the 2002 (2001) season when excluding the influence of Asian dust. The third worst year would have been the 2007 season, when the visibility is lowest. Due to a lack of public awareness of air quality in the past, the poor perceived air quality in the 2001, 2002, and 2007 seasons was not well recognized by the public.

3.4. Potential mechanisms for high AQPI

Previous studies have demonstrated that the recent poor air quality in East Asia is closely related to weak wind speed or an increased number of stagnant days (Wang and Chen 2016, Kim et al 2017, Wang et al 2017, Yang et al 2018, An et al 2019, Lou et al 2019). Figure 4 shows the number of stagnant days (N$_{stag}$), N$_{APQI}$, seasonal precipitation, and the ratio N$_{stag}$ to precipitation from the 2001 to 2013 seasons to examine the meteorological conditions that influence the increase in public awareness before the turning-point year. N$_{stag}$ is highest in the 2013 season (112 d). The correlation between N$_{stag}$ and N$_{APQI}$ is rather weak, with a correlation coefficient of 0.47 ($p$-value = 0.11). Note that the correlation between PM$_{10}$ and N$_{APQI}$ is only 0.35 ($p$-value = 0.24), indicating that PM$_{10}$ is not sufficient to explain public perception of air quality. Interestingly, N$_{stag}$ in the 2012 season (111 d) is similar to that in the 2013 season, but N$_{APQI}$ is much lower in the 2012 season (13 d) than in the 2013 season (40 d).

The large amount of precipitation (436 mm) in the 2012 season likely mitigates air quality due to washout effects. N$_{APQI}$ is negatively correlated with precipitation (correlation coefficient = $-0.51$, $p$-value = 0.077), as expected. The ratio of N$_{stag}$ to precipitation shows a high correlation of 0.65 ($p$-value = 0.015) with N$_{APQI}$. So, frequent stagnant days with low precipitation would cause frequent high AQPI days. The amount of precipitation in the 2013 season was indeed low (202 mm). These results are consistent with the findings of Wang and Chen (2016). They showed that, together with a decline in the volume of Arctic sea ice, reduced precipitation and surface wind during winter intensified haze pollution in central North China after 2000.

4. Summary and concluding remarks

We present AQPI models that are developed using PM$_{10}$, visibility, and a memory decay exponent to understand how people perceive air quality differently from measured air quality monitoring data and the potential triggers that influence perception. The AQPI models are trained using internet search volume data from Google and Naver and corresponding model parameters, including the memory decay exponent and the power exponents of PM$_{10}$ and visibility.

We found that the memory decay exponent and the accumulation of past memory traces that account for the duration of prior events should be considered together with PM$_{10}$ and visibility to understand public perception of air quality. The model exhibits different characteristics of air quality perception, depending on the level of public awareness. When the level of awareness is low, people forget quickly and their perceptions depend more on sensory information (i.e. visibility). After experiencing a severe haze event that triggers an increase in public awareness, people forget more slowly and their perceptions are influenced more by PM$_{10}$. The retrospective analysis of AQPI assuming a low level of awareness shows that the 2013 season has the most frequent high AQPI days, in accordance with the
longest and worst perceived air quality episode. This indicates that the severe haze event acted as a turning point, escalating public concerns from the 2013 season onwards. The poorer air quality represented by higher PM$_{10}$ in the past was not recognized by the public because of the low awareness of air quality. Frequent stagnant days with low precipitation likely degrade perceived air quality.

This study attempts to develop a public perception model that is trained by internet search volume...
data for the first time. Our model can serve as a benchmark and can be applied to other societies (e.g. China). However, more sophisticated models, based either on physical parameters or statistical approaches, that consider many other factors (e.g. mass media or social media) will need to be developed to predict air quality perception more reliably. Our results provide a better understanding of air quality perception, which will contribute to bridging the gap between experts and laymen, educating people, and creating effective regulatory policies. Increased public awareness of and concerns about air quality can pressure policy makers and stakeholders to improve air quality and, furthermore, can foster public engagement in reducing pollutant emissions. The proposed method and findings of this study can be applied to public perceptions of other environmental issues, especially those with a low level of public awareness, for example, climate change and natural disasters.

Acknowledgments

We thank Dr. Jonghun Kam for his helpful and valuable comments and suggestions. We also acknowledge three anonymous reviewers for their constructive and clarifying comments that helped improve this article. This study is supported by the Korea Meteorological Administration Research and Development Program under Grant KMI2018-01214.

Data availability

The data used in this study are openly available: PM$_{10}$ from Air Korea (www.airkorea.or.kr), surface meteorology data from the Korea Meteorological Administration (https://data.kma.go.kr), rawinsonde data from the University of Wyoming (www.weather.uwyo.edu/upperair/sounding.html), Google Trends (https://trends.google.com/), and NAVER Datalab (https://datalab.naver.com/). The data directly used in figures and table are available at https://doi.org/10.5281/zenodo.3828344.

ORCID iDs

Young-Hee Ryu  https://orcid.org/0000-0002-7214-6039
Seung-Ki Min  https://orcid.org/0000-0002-6749-010X

References

An Z et al 2019 Severe haze in northern China: A synergy of anthropogenic emissions and atmospheric processes Proc. Natl Acad. Sci. 116 8657–66
Anderson J R 1982 Acquisition of cognitive skill Psychol. Rev. 89 369–406
Auliciems A and Burton I 1973.Trends in smoke concentrations before and after the clean air act of 1956 Atmos. Environ. (1967) 7 1063–70
Choi H and Varian H 2012 Predicting the present with google trends Econ. Rec. 88 2–9
Dong D, Xu X, Xu W and Xie J 2019 The relationship between the actual level of air pollution and residents’ concern about air pollution: evidence from shanghai, China Int J. Environ. Res. Public Health. 16 4784
Dugas A, Jalalpour M, Gel Y, Levin S, Torcaso F, Igusa T and Rothman R 2013. Influenza forecasting with google flu trends PLoS ONE 8 e56176
Ghim Y S, Chang Y-S and Jung K 2015 Temporal and spatial variations in fine and coarse particles in seoul, Korea Aerosol Air Qual. Res. 15 842–52
Huang Q, Cai X, Wang J, Song Y and Zhu T 2018 Climatological study of the boundary-layer air stagnation index for china and its relationship with air pollution Atmos. Chem. Phys. 18 7573–93
Jun S-P, Yoo H S and Choi S 2018 Ten years of research change using google trends: from the perspective of big data utilizations and applications Technol. Forecasting Soc. Change 130 69–87
Kam J, Stowers K and Kim S 2019 Monitoring of drought awareness from google trends: a case study of the 2011–17 california drought Weather Clim. Soc. 11 419–29
Kim H C et al 2017 Recent increase of surface particulate matter concentrations in the seoul metropolitan area, korea Sci. Rep. 7 1–7
Kim M, Yi O and Kim H 2012 The role of differences in individual and community attributes in perceived air quality Sci. Total Environ. 425 20–26
Kurokawa J and Ohara T 2019 Long-term historical trends in air pollutant emissions in asia: regional emission inventory in aSia (REAS) version 3.1 Atmos. Chem. Phys. Discuss. 1–51
Lang J, Zhang Y, Zhou Y, Cheng S, Chen D, Guo X, Chen S, Li X, Xing X and Wang H 2017 Trends of PM2.5 and chemical composition in beijing, 2005–2015 Aerosol Air Qual. Res. 17 412–25
Lee B-E, Ha E-H, Park H, Kwon J-S and Shin K-R 2005 Study on perceived symptoms and preventive behaviors during asian dust storms among children, adults and elders J. Korean Soc. School Health 18 15–25 (www.koreascience.or.kr/article/JAKO200510103497188.pdf)
Lercher P, Schmitzberger R and Kolfer W 1995 Perceived traffic air pollution, associated behavior and health in an alpine area Sci. Total Environ. 169 71–74
Lin C, Li Y, Lai AK H, Li C and Fung J C H 2018 15-year PM2.5 trends in the pearl river delta region and hong kong from satellite observation Aerosol Air Qual. Res. 18 2355–62
Liu S and Liang X-Z 2010 Observed diurnal cycle climatology of planetary boundary layer height J. Climate 23 5790–809
Lou S, Yang Y, Wang H, Smith S J, Qian Y and Rasch P J 2019 Black carbon amplifies haze over the north china plain by weakening the east asian winter monsoon Geophy. Res. Lett. 46 452–60
Lu Y, Wang Y, Zuo J, Jiang H, Huang D and Rameezdeen R 2018 Characteristics of public concern on haze in China and its relationship with air quality in urban areas Sci. Total Environ.. 637–638 1597–606
Oltra C and Sala R 2014 A Review of the Social Research on Public Perception and Engagement Practices in Urban Air Pollution (wwwosti.gov/etdweb/servlets/purl/2225183)
Pantavou K, Psiloglou B, Lykoudis S, Mavrakis A and Nikolopoulou G K 2018 Perceived air quality and particulate matter pollution based on field survey data during a winter period Int. J. Biometeorol. 62 2139–50
Piro F N, Madsen C, Næs s Ø, Næs t P and Clausen s B 2008 A comparison of self reported air pollution problems and GIS-modeled levels of air pollution in people with and without chronic diseases Environ. Health. 7 9
Reames T G and Bravo M A 2019 People, place and pollution: investigating relationships between air quality perceptions, health concerns, exposure, and individual- and area-level characteristics Environ. Int. 122 244–55
Shirky C 2011 The political power of social media: technology, the public sphere, and political change Foreign Affairs 90 28–41
Walker T R et al 2006 Perceived and measured levels of environmental pollution: interdisciplinary research in the subarctic lowlands of northeast European Russia AMBIO J. Human Environ. 35 220–8
Wang H-J and Chen H-P 2016 Understanding the recent trend of haze pollution in eastern China: roles of climate change Atmos. Chem. Phys. 16 4205–11
Wang X, Dickinson R E, Su L, Zhou C and Wang K 2017 PM2.5 pollution in china and how it has been exacerbated by terrain and meteorological conditions Bull. Am. Meteor. Soc. 99 105–19
Wixted J T and Ebbesen E B 1997 Genuine power curves in forgetting: A quantitative analysis of individual subject forgetting functions Mem. Cognit. 25 731–9
Yang Y, Wang H, Smith S J, Zhang R, Lou S, Qian Y, Ma P-L and Rasch P J 2018 Recent intensification of winter haze in China linked to foreign emissions and meteorology Sci. Rep. 8 1–10
Yen I H, Yelin E H, Katz P, Eisner M D and Blanc P D 2006 Perceived neighborhood problems and quality of life, physical functioning, and depressive symptoms among adults with asthma Am. J. Public Health. 96 873–9
Yeo M J, Im Y S, Yoo S S, Jeon E M and Kim Y P 2019 Long-term trend of PM2.5 concentration in seoul J. Korean Soc. Atmos. Environ. 35 438–50
Zheng B et al 2018 Trends in China's anthropogenic emissions since 2010 as the consequence of clean air actions Atmos. Chem. Phys. 18 14095–111