China Air Pollution Study Based on Volunteer Monitoring Stations

Merrick Lex Berman
Harvard CGA
mberman@fas.harvard.edu

Fei Carnes
Harvard CGA
fmeng@fas.harvard.edu

Zhenyang Hua
Harvard CGA
zhua@fas.harvard.edu

ABSTRACT

Recent air pollution incidents in Chinese cities prompted this study, which makes use of air particulate monitoring information being submitted by volunteer-owned equipment to a public website. As an indicator of severe air pollution spikes, such as those seen in February 2014, the daily average readings for all monitoring locations were mapped and animated as a time series.

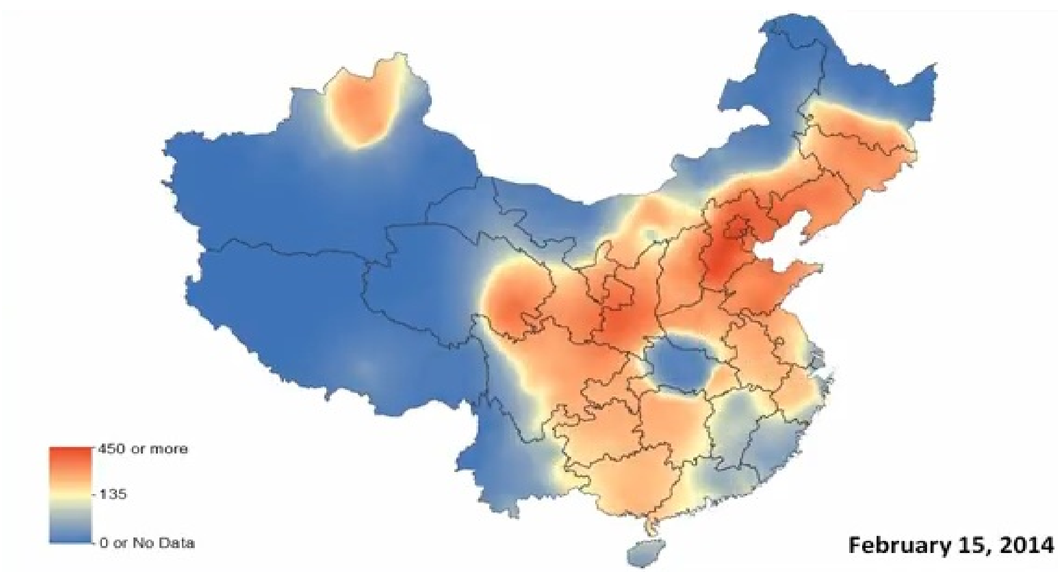

Figure 1: China air pollution PM2.5 Level, Feb 15 2014 [13]

INTRODUCTION

Severe air pollution in Chinese cities has been of major concern for many years. The lack of timely Air Quality Index [AQI] data, in particular for sudden spikes in dangerous levels affecting large populations, was the motivation for this study.

The authors have collected samples from a public website where monitoring information has been submitted in real time by volunteers.[1]

From the aquired data samples, the AQI indicator for 2.5 micron particulate matter [PM 2.5] was normalized into a series of daily average readings for each of the recorded monitor locations. The daily average maps, which reveal the areas suffering from low to high levels of airborne PM2.5, comprise a time series for the collection period.

Subsequently, the time series was animated to provide a concrete visualization [2] of the daily changes in PM2.5. The worst levels detected from the monitoring equipment correspond closely to eyewitness reports from various media outlets.

The raw data samples have also been made publicly available for further analysis. [3]

DATA ACQUISITION

There are a number of existing datasets acquired by remote sensing related to air quality monitoring. For example the data collected from multiple satellites in the NASA AIRS series are based on global imagery at various scales, from 5km pixels to 50km pixels. The datasets, available in near-real time, are processed to provide information about aerosol components and optical depth, or the degree to which the air is penetrable by the sensors. [4]

Similarly, the NASA Ozone and Air Quality website makes available daily captures of remote sensing images processed to study various gases, aerosols and other properties of atmospheric pollution. [5] In addition to these sources there are numerous studies related to the aerosol concentrations of black carbon. The U.S. EPA Report to Congress on black carbon pointed out that remote sensing provides information about spatial variability of aerosol particulates throughout the entire atmospheric column, and needs to be combined with ground based monitors to adequately identify the areas of surface concentrations. [6]

All of the remotely sensed data, in addition to requiring a large investment of pre-processing time, is unable to detect conditions below the cloud cover or to detect the localized severity of pollution levels within larger regions of general smog, or conditions within which the optical depth is too shallow to allow for any meaningful measurements to be acquired about conditions on the ground. In addition, the ground-based monitoring network mentioned in the EPA report consisted of some 200 monitors for the entire United States, while the number of monitors used in China for the same report amounted to a mere handful.

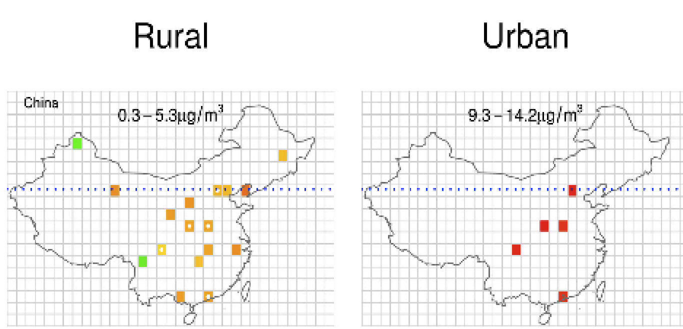

Figure 2: China ground monitoring stations for black carbon *EPA Report*, p.121 [6]

Seeking alternatives to help us understand the recent air pollution events in China, we sought out any information available about air quality indicators on the ground. We found two initiatives that provide near real time maps of Air Quality Indicator levels, one presumably sponsored by the Chinese Environmental Survey Agency, called the National Urban Air Monitoring Platform [NUAMP] [7] and one non-governmental website, called Aqicn.org. [1]

The NUAMP website provides a real-time map of cities and a graph showing the AQI level for cities that are selected by clicking on the map. This site was developed by the Advanced Technology Institute of Zhongshan University and the Guangdong Environmental Survey Center. When using this website, the AQI values shown in the graph, (or in a pop-up window for cities with AQI above 100), do not seem to correspond to the detailed values shown for the specific readings of either PM2.5 or PM10.

The Aqicn.org website provides a real-time map of all monitor locations, arranged by cities, and for which the raw monitor readings are provided. Upon investigation, the readings are based on a mashup of publicly available data being sourced from various providers such as the Weather Underground and various provincial or municipal environmental agencies in China; together with real time data about particulate matter levels being acquired and contributed by volunteers.

According the FAQ of the Aqicn.org website, [8] volunteers who purchased and installed a combination of the Dylos Air Particle Monitor [9] and the Rasberry PI hardware interface, [10] could install free software that automates the upload process from their own monitor to the Aqicn server. [11]

Although the monitoring readings from each Dylos unit being are automatically sent to the Aqicn server, the raw data is only viewable in the form of webmaps or digital images on the Aqicn website. Nonetheless, a closer examination of the published webpages showed that the AQI levels could be found for each monitor location, including values for: PM2.5, PM10, O3, NO2, SO2, CO, as well as other temperature and weather indicators. We also found that by tracing each webpage related to a particular monitor station over time, we could compile a snapshot of the values for a specific time of day.

Based on the publicly available webpages, we then proceeded to automate a method for capturing all the values found on the website under the China category. Even with an optimized process, the assembly of one batch of data took several hours. Therefore we did not attempt to collect a continuous stream of data, but to limit the collection samples to three times per day. In consultation with our colleagues at Harvard School of Public Health, we decided that the three sampling times (eight hours apart), should coincide with (1) the beginning of the morning rush hour, (2) the end of the work day, and (3) a middle-of-the-night sample. The exact time for each batch of AQI data being collected from each station varies slightly each day depending on the server response conditions.

Testing of the data collection was completed by late January 2014, and deployed beginning February 7th 2014. Once launched, our data acquisition method for this project should continue to produce samples three times a day. Each complete batch of monitoring data is being saved in plain-text, comma-delimited files, and compressed into time-stamped archives.   All the data has been made available for free public download. [3]

DATA PROCESSING AND ANALYSIS

Owing to the unverifiable nature of the source data, and the incomplete coverage of monitoring stations across China, the processing and analytical approaches to the data that we could take were limited.   We could not, for example, attempt a gap analysis to fill in values for intervening areas where there were no monitors providing data.   On the other hand, we did have some detailed values to measure the presence of particulate matter several times per day for a set of stations that fluctuated from in number from 700 to over 1000 stations at any given sampling time.   This was a sample on the order of 50 times larger than that EPA Carbon Study for China.

Not being atmospheric scientists, we decided to focus only on the PM2.5 values. Even though many other factors are obviously involved, we felt that the sampling of particulate matter readings spread across more than 200 cities in China -- *not available from any other source* -- justified taking a look at how the PM2.5 readings varied over space and time.

Since we typically had three samples per day, we first calculated daily averages for PM2.5 for each location. From the daily averages, we then performed a variety of spatial analysis methods using geographic information systems software (ArcGIS), to find an outcome most suited to showing the variation over space and time for the PM2.5 values.   We began with what has become commonly referred to as a "heat map," which is a calculation based solely on the kernel density, or distance between locations. However, this type of kernel density analysis shows only the prevalence of *where* monitors are located.   What we needed for this study was a way to show the relative intensity of the PM2.5 values across space.

For our purposes, geostatistical analysis using kriging interpolation presented the most intuitive result to visualize the air pollution concentrations from our dataset. Although the analysis is based on air monitoring stations in China that are not evenly distributed across space, we found other studies that had taken the same approach. [12] Of course, interpolation methods are also going to be primarily skewed towards where the data is concentrated spatially, but it does offer the normalization factor to smooth the resulting classes spatially based on a selected attribute. Therefore, by using the kriging interpolation, despite its drawbacks, the daily average readings from the sensors could be visualized and their relative distribution and concentrations mapped.

For each set of daily averages, the same interpolation with kriging, normalized on the PM2.5 values was run, resulting in a set of maps for each day in February for which data was available. We then had a time series of PM2.5 concentration maps running from February 7th to February 28th.

Finally the maps were animated at the rate of one second per frame, so that the month of changing air pollution (using the proxy of daily average PM2.5 values) is seen in ~ 30 seconds. The resulting video was published on Youtube. [2]

REFERENCES

[1] Aqicn.org. http://aqicn.org/

[2] China Air Pollution (Feb 2014) Time Series Animation http://t.co/K1zCyQe0uW

[3] China AQI Raw Data http://aqi.cga.harvard.edu/china/

[4] NASA AIRS data. http://goo.gl/qGKqts

[5] NASA Ozone and Air Quality http://ozoneaq.gsfc.nasa.gov

[6] EPA Report on Black Carbon (2012): p129. http://www.epa.gov/blackcarbon/2012report/Chapter5.pdf

[7] National Urban Air Monitoring Platform http://113.108.142.147:20035/emcpublish/

[8] Aqicn FAQ http://aqicn.org/faq/2013-09-08/dylos-air-particule-counter-experimentation-part-1/

[9] Dylos Air Particle Monitor http://www.dylosproducts.com/

[10] Rasberry Pi http://www.raspberrypi.org/

[11] Dylos Reader Scripts https://github.com/aqicn/dylos

[12] "GIS to predict air pollution in mining region" *Geospatial World*, (Jul 2012) http://goo.gl/7jXXOt