Enabling Spatiotemporal Analysis and Visualization of Air Pollution in China and India

Merrick Lex Berman[1], Devika Kakkar[1], Wendy Guan[1]*, Fei Carnes[1]
[1]Center for Geographic Analysis, Harvard University, Cambridge, Massachusetts, USA
*Corresponding author, e-mail: wguan@cga.harvard.edu

Keywords: air pollution; China; India; spatiotemporal; PM2.5

Objective: This paper presents the preliminary achievements of an on-going effort to develop a free archive of ground-based air pollution data collected from real-time monitoring stations in China and India. The study also includes the development of a methodology for processing air pollution measurements from streaming raw data into pollution exposure maps. Results are applicable to public health studies at city and county levels.

Background: China's air pollution has become increasingly severe, with recurrent extreme air pollution events that pose serious dangers to public health. The situation has become so dire that Beijing repeatedly declared a "red alert" warning since 2015 [1]. The Chinese government has begun to address the issue, and in 2013 launched a $277 billion investment aimed at improving air quality [2].

Severe air pollution also occurs in India, which prompted the India government to publish hourly Air Quality Index (AQI) monitoring data since April of 2015 [3].

This study aims at building a free, open dataset of AQI for more than 900 ground-based monitoring stations across China, and about 50 stations across India. The result supports multi-disciplinary studies.

Data: In China, AQI data in near real time became accessible on several websites (both official and unofficial) in mid-2013. The Center for Geographic Analysis of Harvard University (CGA) responded to the appearance of these public AQI data sources by developing a method to acquire and preserve the observations for analysis. This data "scraping" process launched in January 2014 stores the air pollution measurements three times per day, and makes the data freely available on CGA website [4]. Each observation snapshot records the location, timestamp, and both physical and chemical variables, including PM2.5, PM10, O3, NO2, SO2, CO, Temperature, Dew Point, Atmospheric Pressure, Humidity, and Wind Speed.

This China AQI data archiving effort has been in near-continuous operation and by the end of 2015 had approximately 2,300,000 total observation snapshots collected at more than 1,500 monitoring locations across China. Since then we have become aware of two similar efforts undertaken at Berkeley Earth [5] and Beijing City Lab [6].

In India, AQI data became publicly available on a website maintained by IIT Kanpur [7]. The Indian Ministry of Environment Central Pollution Control Board published technical documentation [8], which provided details about the monitoring equipment and data collection methodology. The authors developed a prototype to acquire daily average and hourly readings from the Ministry of Environment's official monitor stations. Currently there are about 50 stations registered in the pollution-monitoring network. The data collection prototype identifies the stations, their locations, then acquires hourly readings for AQI measurements such as PM2.5, So2, No2, CO, O3. It also gets synchronous meteorological information from a separate web service [9].

Analysis: Our first attempt to visualize the air pollution data was to apply a universal kriging at country level on China to create an interpolated surface of PM2.5 concentration at a series of time steps. The result was an animated map, effective in visual communication. However, it exposed problems in the overly generalized interpolation model due to the extremely large range of distances between monitoring stations. Further spatial-statistical analysis is needed for the evaluation of spatial autocorrelation and regionalization patterns. Preliminary analysis shows that measurement stations do effectively capture processes at multiple scales, from which local as well as regional patterns can be discovered. The approach is to assess hot spots of pollutants using a multi-scaled statistical process developed at Heidelberg's GIScience Research Group to reveal spatial structures [10], map spatial patterns, and determine parameters. These parameters define the non-random spatial behavior of the pollutants. A geographically weighted regression [GWR] model is needed to estimate local effects of several independent variables, and to interpolate surface grids for the coefficients of those variables. These surface grid maps will reveal spatial variation in relation to the pollution variables.

Although there is no perfect means for interpolation of air pollution measurements in between monitoring stations [11], we feel that GWR model will be more effective than universal Kriging interpolation for our specific use case. This approach accounts for volatility in the variations of AQI values over time, and find "break points" as distances within which regionalization patterns and clustering can be detected. Further, this method can be applied to a particular local study area of interest and the resulting surface grids of cumulative exposure values can then serve as a starting point for further research.

Result and Discussion: Results of this study includes three parts. First is the AQI data harvesting tools capable of integrating multiple web sources, selecting or aggregating multiple times as harvesting target, and processing raw data into standard archiving files. Second is the published AQI archive for China on the Dataverse Network [12]. Third is the analytical model for detecting regionalization from the AQI data, and interpolating from the monitoring point measurements to a reliable air quality field at various scales. The third component is still under development.

Conclusion and Future Work: This project is an on-going effort aimed at providing a platform for applied data science that can answer important questions related to health and sustainability. This work enhances and expands access to ground-based China and India air pollution data by providing an ongoing data collection system for free public download. It also develops a methodology for interpreting from a national level time series of observations to form a space-time continuous air quality field at various scales; and providing a basic visualization platform that shows the current and cumulative observations of air pollution intensity. Building upon these resources, we will create channels for open data exchange, collaborative analysis, and continued discussion on air pollution towards reducing its harmful impact on societies in China, India, and around the world.

ACKNOWLEDGMENT

Zhenyang Hua contributed to the initial data-archiving program during his internship at the CGA in 2014. René Westerholt of the Heidelberg University contributed in the development of a geographically weighted regression [GWR] model for data analysis following his visiting fellowship at the CGA in 2013.

REFERENCES

[1] E. Wong, "Smog So Thick, Beijing Comes to a Standstill," New York Times, December 9, 2015, p. A1.

[2] S. Wee, "China to invest $277 billion to curb air pollution: state media," Reuters, United States Edition, Environment, R. Popeski Ed., Wed Jul 24, 2013, 10:15pm EDT.

[3] Daily News & Analysis, "India's first Air Quality Index launched; will monitor pollution levels across country," India, New Delhi, Apr 6, 2015.

[4] Center for Geographic Analysis, "Air Quality Index Archive," http://aqi.cga.harvard.edu/china/about/, Harvard University, last accessed on March 15, 2017.

[5] R. A. Rohde and R. A. Muller, "Air Pollution in China: Mapping of Concentrations and Sources," Berkeley Earth, 2015, http://berkeleyearth.org/wp-content/uploads/2015/08/China-Air-Quality-Paper-July-2015.pdf , last accessed on March 15, 2017.

[6] Beijing City Lab, "China PM2.5," https://www.beijingcitylab.com/projects-1/13-pm2-5/, Projects, last accessed on March 15, 2017.

[7] Central Pollution Control Board, "National Air Quality Index," http://164.100.160.234:9000/, Ministry of Environment, Forests and Climate Change, Government of India, last accessed on: March 15, 2017.

[8] Central Pollution Control Board, "National Air Quality Monitoring Programme Introduction," http://www.cpcb.nic.in/air.php, Ministry of Environment, Forests and Climate Change, Government of India, last accessed on: March 13, 2017.

[9] OpenWeatherMap, Inc., "Open Weather Map," http://openweathermap.org/, last accessed on: March 15, 2017.

[10] R. Westerholt, B. Resch and A. Zipf, "A local scale-sensitive indicator of spatial autocorrelation for assessing high- and low-value clusters in multiscale datasets," International Journal of Geographical Information Science, vol. 29, pp.868-887, 2015 Issue 5.

[11] M. Bravo, M. Fuentes, Y. Zhang, M. J. Burr and M. Bell, "Comparison of exposure estimation methods for air pollutants: Ambient monitoring data and regional air quality simulation," Environ Res. Vol. 116, pp.1–10, 2012 July.

[12] B. Lex, "China AQI PM25s Archive Dataverse," https://dataverse.harvard.edu/dataverse.xhtml?alias=china_pm25s, Harvard Dataverse, V1, 2017.