Environmental Pollution xxx (2016) 1–4

Contents lists available at ScienceDirect

Environmental Pollution

journal homepage: www.elsevier.com/locate/envpol

Short communication

Quality screening for air quality monitoring data in China☆

Jianzheng Liu [a, b], Weifeng Li [a, b, *], Jie Li [a]

[a] Department of Urban Planning and Design, Faculty of Architecture, The University of Hong Kong, Hong Kong, China
[b] Shenzhen Institute of Research and Innovation, The University of Hong Kong, Shenzhen, China

ARTICLE INFO

Article history:
Received 23 January 2016
Received in revised form
3 June 2016
Accepted 18 June 2016
Available online xxx

Keywords:
Particulate matter
Quality screening and control
Air quality monitoring data

ABSTRACT

Particulate matter data obtained from the national air quality monitoring network in China has become an essential and critical data source for many current and forthcoming studies as well as the formulation and implementation of air pollution regulatory policies on particulate matter ($PM_{2.5}$ and PM_{10}). However, the quality control of this data is dubitable and can affect many future studies and policies. This study identifies and elucidates two significant quality control issues with the data. They are $PM_{2.5}$ levels exceeding concurrent co-located PM_{10} levels and the registration of same concentrations for consecutive hours at some stations. Future studies utilizing particulate matter data need to acknowledge and address these issues to ensure accurate and reliable results.

1. Introduction

Particulate air pollution in China has attracted considerable global attention in the past five years (Hu et al., 2014; Rohde and Muller, 2015; Zhang and Cao, 2015). Strong debates on the air quality and its health implications have ensued among experts, media, and public in the country. Public calls for government action to provide air quality information have been growing. In response to the increasingly severe issue of particulate air pollution, the Ministry of Environmental Protection has established a national air quality monitoring network, and provided detailed information on particulate matter pollution level through national and city-level air quality reporting platforms (http://113.108.142.147:20035/emcpublish/). By the end of 2015, more than 1800 state and city-owned air quality monitoring stations had been established in almost all the prefectural cities, including 31 provincial capitals, with records of hourly air pollutants measurements for the past three years. At each station, automated monitoring systems consisting of the sample collection unit, sample measurement unit, data collection and transport unit, and other accessory equipment were installed and used to measure the six pollutants. The data is rich and reports hourly concentration of six pollutants: particulate matter with aerodynamic diameter less than 2.5 μm ($PM_{2.5}$), particulate matter with aerodynamic diameter less than 10 μm (PM_{10}), sulfur dioxide (SO_2), nitrogen dioxide (NO_2), ozone (O_3), and carbon monoxide (CO). In particular, the measurement of $PM_{2.5}$ and PM_{10} concentration is made with both the micro-oscillating balance method and the β absorption method (Wang et al., 2014).

These data sources have become fundamental and critical to studies on particulate air pollution as well as the formulation and implementation of air pollution regulatory policies in China because of the high temporal resolution and excellent spatial coverage, and the official and legal implication in the data. Currently, dozens of studies based on this data have already been published (Chai et al., 2014; Hu et al., 2014; Rohde and Muller, 2015; Wang et al., 2014; Zhang and Cao, 2015). However, the quality control of the data is questionable. As data are continuously obtained from air quality monitoring networks every hour and many relevant studies and policies based on this data are still in their infancy, this study intends to elucidate and highlight two imperative data quality control issues to the researchers' attention and suggest that researchers using data from any source, whether in China or elsewhere, be expected to do data screening of their own, guided by the objectives of their own research.

2. Quality control issues with air quality monitoring data

2.1. $PM_{2.5}$ measurements exceeding co-located PM_{10} levels measured concurrently

$PM_{2.5}$ is universally defined as the mass concentration of

☆ This paper has been recommended for acceptance by Hageman Kimberly Jill.

* Corresponding author. Department of Urban Planning and Design, Faculty of Architecture, The University of Hong Kong, Pokfulam Road, Hong Kong.
E-mail addresses: jzliu@connect.hku.hk (J. Liu), wfli@hku.hk (W. Li), jessieleepku@hotmail.com (J. Li).

http://dx.doi.org/10.1016/j.envpol.2016.06.037
0269-7491/

particulate with aerodynamic diameter less than 2.5 μm, and PM_{10} is defined as the mass concentration of particulate matter with aerodynamic diameter less than 10 μm. By definition, $PM_{2.5}$ concentration is included in the PM_{10} concentration levels for a site at a given time and is therefore lesser than the PM_{10} concentration, irrespective of the measurement methods employed.

However, a non-negligible proportion of the data from the national air quality monitoring network has exhibited higher $PM_{2.5}$ than concurrent PM_{10} measurements. The study analyzed air quality monitoring data provided by third parties AQISTUDY.cn and EPMAP.org for 2014. Data from the two parties are identical to the air quality data gathered by the aforementioned national air quality reporting platforms. We use data provided by the two third parties because the national air quality reporting platform only reports the air quality of the day and does not show historical data and is unavailable to the public. Fortunately, third parties created by civic efforts such as AQISTUDY.cn and EPMAP.org have been crawling this data since late 2013. As we noticed that there were missing hourly measurements in both two data sources, we combined the two to obtain a more complete 24-h $PM_{2.5}$ measurement dataset for each day of 2014.

Preliminary analysis of the data reveals that 46 monitoring stations, constituting 4.36% of the total number of stations exhibit higher readings of $PM_{2.5}$ in over 5% of the records. Similarly, 367 stations constituting 34.75% of the total number of stations, exhibit higher $PM_{2.5}$ in 1% of the records. While it is impressive to learn that only 1.10% of the total records exhibit such erroneous readings which is acceptable because all measurements contain some error, some stations registered erroneous data of over 20%. For example, 20.44% of the particulate matter measurements at South Ring Road station in Beijing exhibit higher $PM_{2.5}$ than PM_{10} (Fig. 1).

We then looked at which cities or regions are more likely to have such problem. Statistics show that the top 5 cities with highest proportion of erroneous readings are Beijing (8.85%), Jiangyin (6.27%), Jurong (3.69%), Nanchang (3.59%) and Changsha (2.92%). 73 cities, constituting 38.42% of all 190 cities with air monitoring stations in 2014 exhibit higher readings of $PM_{2.5}$ in over 1% of the records.

The discrepancy could have been caused by an instrument failure, such as a malfunction of the sampling units (Air Quality Expert Group, 2012). Similar problems have been recorded in other countries in the world. For example, a report from the Air Quality Expert Group to the UK Department for Environment, Food and Rural Affairs indicated a flaw with the measuring instrument when $PM_{2.5}$ was higher than PM_{10} at locations where both parameters were being measured in parallel (Air Quality Expert Group, 2012). However, UK's particulate matter monitoring program has implemented specific quality assurance and control procedures to ensure data accuracy and quality, while the air quality monitoring program in China does not have similar measures.

Further studies that deal with the comparison or correlation of $PM_{2.5}$ and PM_{10} should be aware of this discrepancy (Hu et al., 2014; Wang et al., 2014; Zhang and Cao, 2015). This study recommends that researchers, with due caution, should consider eliminating particulate matter readings from stations with over 5% erroneous readings, and include readings from stations with less than 5% erroneous readings, after accounting for the deviation.

2.2. Same measurements registered for consecutive hours

Particulate matter values are volatile in nature (Air Quality Expert Group, 2012). Extensive studies demonstrate that meteorological conditions such as wind speed, relative humidity, temperature, and boundary layer height have a major impact on the diurnal patterns of $PM_{2.5}$ (Zhao et al., 2009; Du et al., 2013; Cheng and Li, 2010). Multiple physical and mechanical factors also contribute to variations in the particulate matter readings. As wind speed, temperature, relative humidity, and other factors are unlikely to remain constant for several hours, it is very unlikely that the particulate matter measurements remain unchanged for several hours.

Rohde and Muller (2015) first raised the issue of repeated particulate matter readings. Their study contends that an air quality monitoring instrument rarely registers the same measurement for two consecutive hours. Identical measurements recorded can be

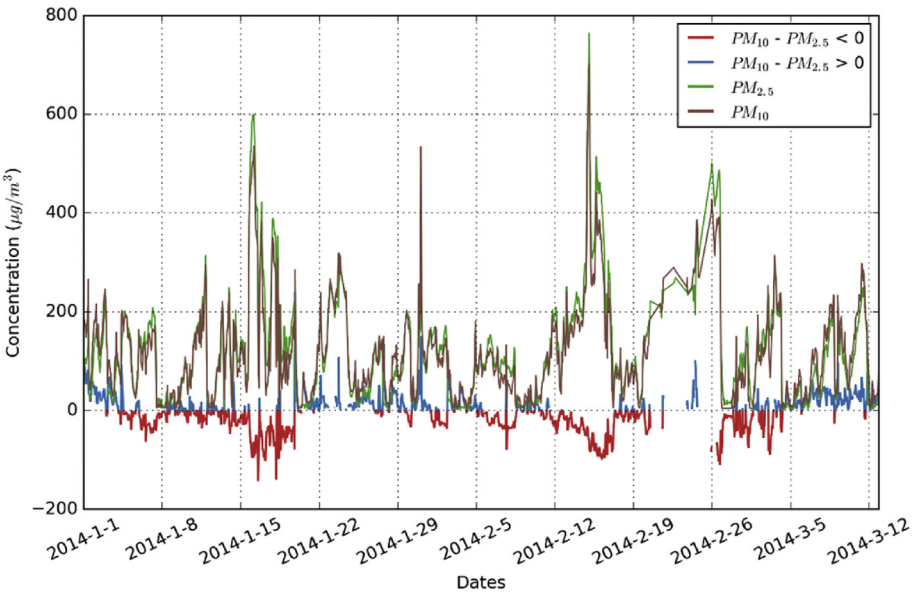

Fig. 1. $PM_{2.5}$ concentration, PM_{10} concentration, and their difference of (PM_{10} minus $PM_{2.5}$) in South Ring Road station in Beijing in 2014. This figure shows only part of the yearly data to highlight the issue of $PM_{2.5}$ exceeding PM_{10}. Red lines refer to the concentration difference where $PM_{2.5}$ exceed co-located PM_{10} measured concurrently. The gaps between lines indicate missing data. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Please cite this article in press as: Liu, J., et al., Quality screening for air quality monitoring data in China, Environmental Pollution (2016), http://dx.doi.org/10.1016/j.envpol.2016.06.037

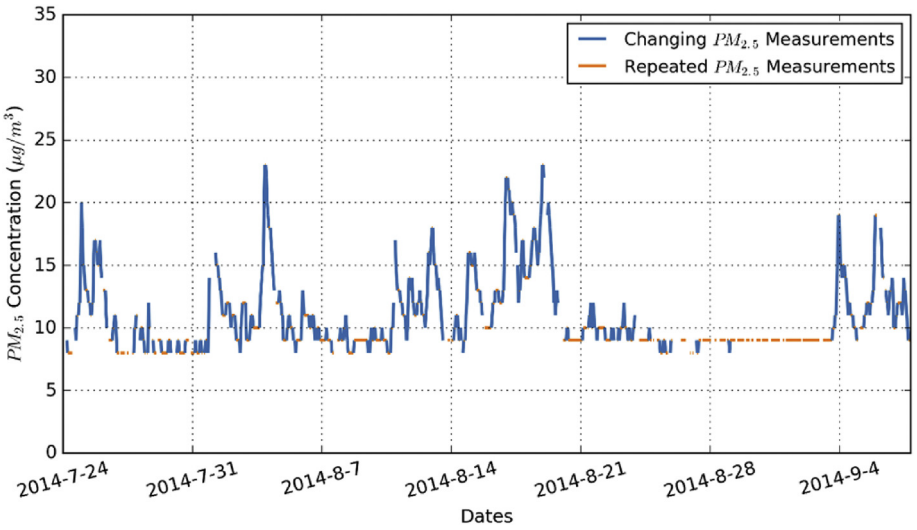

Fig. 2. Repeated readings of $PM_{2.5}$ in the station of Hainan University in 2014. This figure shows the unchanged values in red to highlight the issue of repeated measurement. The gaps between lines indicate missing data. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

attributed to machine failure with the instrument simply repeating the last valid measurements received (Rohde and Muller, 2015).

This study agrees with Rohde and Muller (2015) and pursues with a more cautious approach by considering five consecutive unchanged particulate matter readings as potentially erroneous.

The study used air quality ground monitoring data for 2014 and searched for data records that report five or more identical consecutive measurements. The results show that only 1.03% of the records register repeated measurements for five or more consecutive hours. However, the proportion varies considerably among stations, with Hainan University station in Hainan, exhibiting the maximum proportion of 12.38% (Fig. 2).

34 stations constituting 3.22% of the total number of stations have recorded over 5% of repeated $PM_{2.5}$ measurements. 322 stations constituting 30.49% of the total number of stations have recorded over 1% of repeated $PM_{2.5}$ measurements. Statistics on the city scale shows that 5 cities have over 5% of repeated $PM_{2.5}$ measurements, including Haikou (8.36%), Zhanjiang (6.75%), Yunfu (5.49%), Fuzhou (5.44%) and Sanya (5.22%). 72 cities constituting 37.89% of all 190 cities with air monitoring stations in 2014 have recorded over 1% of repeated $PM_{2.5}$ measurements. All the 5 cities are located near the south China sea, which suggest that coastal cities are likely to have the problem of repeated $PM_{2.5}$ measurements. We speculate that the hot and humid climatic condition might affect the performance of the air quality monitoring instrument.

Similar issue for PM_{10} is comparatively lesser as only 0.17% of the PM_{10} data records registered identical consecutive measurements. Only 79 stations, accounting for 7.48% of total number of stations, registered over 1% of records with repeated PM_{10} measurements. Party school station in Tongchuan, Shaanxi has recorded a maximum proportion of 4.53%. The city-scale statistics show that only 7 cities, accounting for 3.68% of all cities, registered over 1% of repeated PM_{10} measurements. The city of Nanchong recorded the highest proportion of 1.38%.

As the issue is probably caused by mechanical errors or instrument problems such as instrument saturation and zero values being repeated, this study recommends that future research dealing with particulate matter time-series data address this issue by setting erroneous measurements to null and remove them from calculation, unless this makes data insufficient for further analysis. If the proportion of erroneous data points for a station is too high, the researcher should consider excluding data collected from the station.

3. Conclusion

To the best of our knowledge, current literature using the air quality monitoring data in China rarely describe these aforementioned issues although the authors of these literature and agencies might be aware of them. We argue that both two data quality control issues can be addressed by eliminating erroneous data in future studies, unless this would make data unavailable or insufficient.

Some of the data figures discussed in this article may alter if different data sets are analyzed. However, the two data quality control issues highlighted in this article are significant in China and possibly in other countries too. Researchers using data from any source, whether in China or elsewhere, are advised to use particulate matter data with prudence and caution, and do data screening of their own, guided by the objectives of their own research.

Acknowledgements

This research is supported by Urban China Initiative Grant Project “Urban Climate and its Impact on Mental and Social Well-being of Residents: Empirical Evidence from China”. We thank the Urban China Initiative and its co-founders including the McKinsey& Company, the School of Public Policy and Management of Tsinghua University and the Columbia Global Centers| East Asia.

References

Air Quality Expert Group, 2012. Fine Particulate Matter (PM2.5) in the United Kingdom. Department for Environment, Food and Rural Affairs, London, UK.

Chai, F., Gao, J., Chen, Z., Wang, S., Zhang, Y., Zhang, J., Zhang, H., Yun, Y., Ren, C., 2014. Spatial and temporal variation of particulate matter and gaseous pollutants in 26 cities in China. J. Environ. Sci. 26, 75–82.

Cheng, Y.-H., Li, Y.-S., 2010. Influences of traffic emissions and meteorological conditions on ambient PM10 and PM2.5 levels at a highway toll station. Aerosol Air Qual. Res. 10, 456–462.

Du, C., Liu, S., Yu, X., Li, X., Chen, C., Peng, Y., Dong, Y., Dong, Z., Wang, F., 2013. Urban boundary layer height characteristics and relationship with particulate matter mass concentrations in Xi'an, central China. Aerosol Air Qual. Res. 13, 1598–1607.

Hu, J., Wang, Y., Ying, Q., Zhang, H., 2014. Spatial and temporal variability of PM2.5 and PM10 over the North China plain and the Yangtze River delta, China. Atmos.

Environ. 95, 598–609.
Rohde, R.A., Muller, R.A., 2015. Air pollution in China: mapping of concentrations and sources. PLoS One 10, e0135749.
Wang, Y., Ying, Q., Hu, J., Zhang, H., 2014. Spatial and temporal variations of six criteria air pollutants in 31 provincial capital cities in China during 2013–2014. Environ. Int. 73, 413–422.
Zhang, Y.-L., Cao, F., 2015. Fine particulate matter (PM2.5) in China at a city level. Sci. Rep. 5, 14884.
Zhao, X., Zhang, X., Xu, X., Xu, J., Meng, W., Pu, W., 2009. Seasonal and diurnal variations of ambient PM2.5 concentration in urban and rural environments in Beijing. Atmos. Environ. 43, 2893–2900.

Please cite this article in press as: Liu, J., et al., Quality screening for air quality monitoring data in China, Environmental Pollution (2016), http://dx.doi.org/10.1016/j.envpol.2016.06.037