Science of the Total Environment 458–460 (2013) 70–80

Contents lists available at SciVerse ScienceDirect

Science of the Total Environment

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

Spatial and temporal characteristics of particulate matter in Beijing, China using the Empirical Mode Decomposition method

Maogui Hu [a,*], Lin Jia [b], Jinfeng Wang [a], Yuepeng Pan [c]

[a] State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China
[b] State Key Laboratory of Environmental Criteria and Risk Assessment, Chinese Research Academy of Environmental Sciences, Beijing 100012, China
[c] State key Laboratory of Atmospheric Boundary Layer Physics and Atmospheric Chemistry, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China

HIGHLIGHTS

- 27 stations in Beijing are classified into 3 clusters by kernel K-means method.
- Temporal periods and trends were abstracted by EMD method efficiently.
- The pollution levels in different regions differ greatly.
- The trend of PM10 concentration in the City is decreasing since 2008 in general.
- Background mass concentration rises from a regional background monitoring station.

ARTICLE INFO

Article history:
Received 24 December 2012
Received in revised form 27 March 2013
Accepted 2 April 2013
Available online 30 April 2013

Keywords:
Spatiotemporal distribution
Kernel analysis
Empirical Mode Decomposition
Air pollution
Particulate matter

ABSTRACT

Air pollution has become a serious problem in Beijing, China. Daily PM_{10} mass concentration measurements were collected at 27 stations in Beijing over a 5-year period from January 1, 2008 to October 31, 2012. We used a new clustering method (kernel *K*-means) and a new period and trend decomposition method (Empirical Mode Decomposition, EMD) to explore the spatial and temporal characteristics of the PM_{10} mass concentration in the City. The temporal period and trend of each cluster center were decomposed using the EMD method, which is an adaptive data analysis method that requires no prior information. The daily PM_{10} mass concentrations varied greatly from 5 $\mu g/m^3$ to more than 600 $\mu g/m^3$. All of the stations were partitioned into three clusters by the kernel *K*-means method, and which represent the low-, middle- and high-pollution stations, respectively. The first cluster contained nine stations, mainly located in the north suburban area. The second cluster, whose degree of pollution was much more serious than the first cluster, contained 13 stations distributed in urban and peri-urban areas. The pollution level in the southern part of Beijing was much more serious than in the northern part of the City. The third cluster contained five stations located outside the second-cluster stations. The total decreased amplitudes of the three clusters during the whole period were 19 $\mu g/m^3$, 10 $\mu g/m^3$ and 4 $\mu g/m^3$, respectively. Although the global trend of the PM_{10} mass concentration decreased in general, it was not the same for each season and station. The trends in summer and winter declined, while in spring, it has been increasing in recent years. Five types of trends can be found for stations, including monotonic decreasing, rise fall, fall rise fall, fall rise and rise. The rising trend of the regional background air pollution monitoring station, Miyun-reservoir, indicates an increase in the City's background PM_{10} mass concentration.

1. Introduction

Air pollution has become a serious threat to people all over the world. This problem is the focus of worldwide research, especially in developed countries. Historically, the effects of air pollution are well documented, for example, the London smog event in 1952 that caused thousands of deaths. More recently, problems relating to air pollution are increasing in many developing countries due to urbanization, industrial development, increasing use of cars, etc., with China being a prime example. Air pollution is becoming a major problem in China (Chan and Yao, 2008), and the main pollutants include sulfur dioxide (SO_2), nitrogen dioxide (NO_2), ozone (O_3), and particulate matter (PM). PM is one of the most harmful air pollutants to humans, and originates from many sources including natural dust, industrial emissions and traffic gasses (Liu et al., 2004). It is a mixture of solid particles and liquid droplets that vary in size. PM_{10} consists of inhalable particles with an aerodynamic diameter less than 10 μm.

* Corresponding author. Tel.: +86 10 64889055.
E-mail address: humg@Lreis.ac.cn (M. Hu).

0048-9697/$ – see front matter
http://dx.doi.org/10.1016/j.scitotenv.2013.04.005

Many studies have shown that PM_{10} has significant adverse effects on human health (Dockery et al., 1993; Pope et al., 2002). It can go deep into the lung, which not only decreases the function of the respiratory and cardiovascular systems, but also increases pollution-related disease mortality (Qiu et al., 2012; Künzli and Tager, 2005; Nafstad et al., 2003; Zeger et al., 2000). Discovering how the PM concentration is spatially and temporally distributed is an important problem in reducing pollution with effective measures. Spatial characteristics can help to identify interesting spatial patterns from the spatial distribution of PM concentration, while temporal characteristics of time series PM concentration data can be used to discover its period and trend. The trend is an intrinsically fitted monotonic function, or a function in which there can be at most one extremum within a given data span (Wu et al., 2007). However, temporal trend analyses that are often implemented on the original time series data rarely take potentially confounding factors into account, and are prone to being affected by a small number of extremes (Yang, 2002; van der Wal and Janssen, 2000). It is important to select an appropriate method to abstract the natural trend from the observed data.

Beijing, the capital of China, is one of the most seriously air-polluted cities in the country (J. Chen et al., 2009; He et al., 2001; Xie et al., 2005; Zhang et al., 1997; Zheng et al., 2005). The annual average PM_{10} mass concentrations were 180 μg/m^3 and 142 μg/m^3 in 1999 and 2005, respectively (Chan and Yao, 2008). These values are much higher than the air quality guideline (AQG) of 20 μg/m^3 recommended by the World Health Organization (WHO, 2006). As a result of these large PM_{10} concentration values, in 1998 the Beijing government adopted several measures to improve air quality, including adjusting industrial structures, improving heating facilities and limiting the number of privately owned cars. For the 29th Olympic Games held in Beijing in 2008, the government adopted many measures to reduce air pollution (Zhang et al., 2010). The quality of Beijing's air and whether the adopted measures have produced positive effects are of particular interest to people in recent years. Thus, in this research, we used daily data collected over a 5-year period to analyze the spatial and temporal characteristics of PM_{10} mass concentration since 2008 in Beijing.

The objective of this study was to explore the spatial and temporal characteristics of the PM_{10} mass concentration in Beijing since 2008 based on a new trend abstraction method (Empirical Mode Decomposition, EMD) and a high-dimensional clustering method. We collected daily PM_{10} mass concentration data from 27 air pollution monitoring stations over a 5-year period from January 2008 to December 2012 to identify spatial and temporal variations and trends. The remainder of the paper is organized as follows. To explore the spatial clustering and general temporal characteristics of the PM_{10} mass concentration at different locations, we clustered the Beijing Environmental Protection Bureau (B-EPB) observation stations into different clusters. The cluster centers were analyzed to compare their spatial and temporal characteristics. A kernel K-means clustering method was adopted to partition the 27 stations. The temporal period and trend of the daily PM_{10} mass concentration were separated from the cluster centers and stations using the EMD method. We introduce the Materials and methods used in Section 2. Results of interesting spatial and temporal characteristics of PM_{10} mass concentration in Beijing are presented in Section 3. Finally, we discuss the results and draw conclusions in Section 4.

2. Materials and methods

2.1. Data

There are 27 air pollution monitoring stations installed by the B-EPB. The station network covers urban, peri-urban and suburban districts in typical regions, and represents the whole City's air quality very well (Fig. 1; Yizhuang-ETDZ is short for Yizhuang Economic and Technological Development Zone). The air quality data we collected were daily Air Pollution Index (API) and corresponding main pollutant type obtained from the B-EPB during the period January 1, 2008 to October 31, 2012. The pollutants monitored for each station were SO_2, NO_2 and PM_{10}. For every station, each pollutant's daily individual API was calculated from the 24 h average mass concentration. The highest individual API was selected and reported as the daily API by the B-EPB. We collected records for 1806 days at the 27 stations. Among all records, the main pollutant in 38,354 records was PM_{10}, in 728 records, it was SO_2, and in 49 records, it was NO_2. In 9579 records, the main pollutant was not reported because it was at acceptable level (API ≤ 50), as described by the Chinese API standards. In all of the known main pollutant records, 98.01% of records showed the main pollutant to be PM_{10}. Thus, of the three observed pollutants, PM_{10} is clearly the most serious one. We assumed that the unreported main pollutant in the 9579 records where the main pollutant was not reported was PM_{10}. This assumption would include some inaccuracies; however, the proportion would be very small. Finally, the daily PM_{10} mass concentration was calculated from the API according to the following relationship (Beijing Environmental Monitoring Center, 2012).

$$C = \frac{I - I_{\text{low}}}{I_{\text{high}} - I_{\text{low}}}\left(C_{\text{high}} - C_{\text{low}}\right) + C_{\text{low}} \tag{1}$$

where I is the API; C is the pollutant concentration; C_{low} is the concentration breakpoint that is $\leq C$; C_{high} is the concentration breakpoint that is $>C$; I_{low} and I_{high} are the index breakpoints corresponding to C_{low} and C_{high}, respectively. The breakpoints of C and I are displayed in Table 1. Daily PM_{10} mass concentrations of the 777 records whose main pollutant was not PM_{10} were interpolated using the Kriging interpolation method (Isaaks and Srivastava, 1989).

2.2. Spatial and temporal type clustering

For the daily PM_{10} mass concentrations, each station had records for 1806 days, so it had 1806 dimensions. By exploring the daily records for the 27 stations, it was found that a rather large number of records from some stations interwove together, where the concentration at one station was not always higher or lower than another station, and it was hard to separate the stations with a linear plane. Kernel K-means is a clustering method that is good at partitioning high-dimensional and nonlinearly separated data. For a given cluster number, it finds the best appropriate partition by minimizing the within-cluster sum of squares defined as the sum of the squared Euclidean distances between each station and the corresponding cluster center. It is similar to the classical K-means, but all of the operations are performed in a feature space (Dhillon et al., 2004). The kernel K-means method maps the original data to a higher dimensional feature space where the data can be easily separated linearly. In a kernel feature space, the original data $(x_1, \ldots, x_N)$ become $(\varphi(x_1), \ldots, \varphi(x_N))$, where $\varphi(\cdot)$ is the mapping function. Then, the data can be clustered by the following steps:

a. Initialize K-cluster centers m_k ($k = 1, \ldots, K$) in the feature space;
b. assign each station $\varphi(x_i)$ ($i = 1, \ldots, N$) to its nearest center, m_k, as measured by the Euclidean distance in the feature space: $k = \arg\min \|\varphi(x_i) - m_k\|^2$;
c. update the clusters and recalculate the sum of the within-cluster variation E;

$$m_k = \frac{1}{N_k}\sum_{x_j \in \text{class } k}\varphi\left(x_j\right), E = \sum_{k=1}^{K}\sum_{j=1}^{N_k}\left\|\varphi\left(x_j\right) - m_k\right\|^2 \tag{2}$$

where N_k is the station number in cluster k; and

d. repeat steps b and c until E is stable, which can be defined by that the difference between adjacent iterations is less than a small threshold.