Atmospheric Environment 126 (2016) 66–75

Contents lists available at ScienceDirect

Atmospheric Environment

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

Using a new WALSPMF model to quantify the source contributions to PM2.5 at a harbour site in China

Guo-Liang Shi [a, d], Jiao Xu [a], Xing Peng [a], Ying-Ze Tian [a], Wei Wang [b], Bo Han [c], Yu-Fen Zhang [a], Yin-Chang Feng [a, *], Armistead G. Russell [d]

[a] State Environmental Protection Key Laboratory of Urban Ambient Air Particulate Matter Pollution Prevention and Control, College of Environmental Science and Engineering, Nankai University, Tianjin, 300071, China
[b] College of Software, Nankai University, No. 94 Weijin Road, Tianjin, 300071, China
[c] Tianjin Key Laboratory for Air Traffic Operation Planning and Safety Technology, Civil Aviation University of China, Tianjin, 300300, China
[d] School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, 30332-0512, Georgia

HIGHLIGHTS

- WALSPMF model was proposed in this work.
- Synthetic receptor dataset was apportioned by the new model.
- $PM_{2.5}$ samples were collected in a large harbour in China.
- WALSPMF and EPAPMF model were used to apportion $PM_{2.5}$ sources.

ARTICLE INFO

Article history:
Received 8 July 2015
Received in revised form
21 November 2015
Accepted 23 November 2015
Available online 1 December 2015

Keywords:
Source apportionment
WALSPMF
Particulate matter
Receptor model

ABSTRACT

$PM_{2.5}$ variances have adverse impacts on human beings and the environment; therefore, source apportionment is very important and is a hot global topic. In this work, a new model called WALSPMF is proposed and evaluated for its accuracy. First, a synthetic test was carried out to compare the estimated source profile and contributions with the synthetic ones. Average absolute error (AAE) values were also calculated between the estimated and synthetic source contributions; most of the values were low (<15%), which indicated that the results of the WALSPMF model might be acceptable. Next, samples of $PM_{2.5}$ were collected from a large harbour sampling site in China (Tanggu). The $PM_{2.5}$ mean level was 110.63 μg m^{-3}, with a range of 28.67 μg m^{-3}–302.17 μg m^{-3}. The ambient $PM_{2.5}$ dataset was separately introduced into both the WALSPMF and EPAPMF 5.0 models to identify the possible sources and their contributions. Five source categories were extracted by the two models and can be identified in the following consistent order: coal combustion (33% for WALSPMF, 30% for EPAPMF 5.0), secondary nitrate (19% for WALSPMF, 21% for EPAPMF 5.0), crustal dust (18% for WALSPMF, 22% for EPAPMF 5.0), secondary sulphate (16% for WALSPMF, 15% for EPAPMF 5.0), and vehicle exhaust (14% for WALSPMF, 12% for EPAPMF 5.0). The positive results of multiple verifications suggested good performance of the WALSPMF model; thus, it is essential to put this new model forward as a way to potentially enrich the modern source apportionment technique.

1. Introduction

$PM_{2.5}$ (particulate matter with an aerodynamic diameter less than or equal to 2.5 μm) has been found to be the primary pollutant in many cities of the world (Linares and Díaz, 2009). Numerous studies have shown $PM_{2.5}$ to be a cause of direct and indirect adverse effects to human health, atmospheric visibility and global climate change. (Jiménez et al., 2009; Keim et al., 2005; Mimura et al., 2014). Due to the evidence of these adverse effects, the World Health Organization (WHO) recommends a 24-h standard limitation of 25 μg m^{-3} for $PM_{2.5}$ (WHO, 2005). Consequently, there are many calls for increased attention to $PM_{2.5}$, and determining how to control the $PM_{2.5}$ pollution has become a hot global topic.

* Corresponding author.
E-mail addresses: nksgl@hotmail.com (G.-L. Shi), fengyc@nankai.edu.cn (Y.-C. Feng).

http://dx.doi.org/10.1016/j.atmosenv.2015.11.046
1352-2310/

There are many factors that can lead to an increased $PM_{2.5}$ concentration in the air, such as the emission source categories and their associated contributions. Therefore, to better understand its properties and help pollutant management, studies to evaluate $PM_{2.5}$ concentrations, chemical compositions and sources are very necessary. To date, there appear to be many tools to help recognize the source categories of $PM_{2.5}$ in the atmosphere worldwide; for example, factorization models and the CMB model have been the mostly widely used. Among them, factorization models have been important models, including PCA, PMF, Unmix, ME-2 and so on. All of the models mentioned above are based on factorization; PCA is the basic factorization model, while the PMF model is the most widely used (Manousakas et al., 2015; Taiwo et al., 2014; Villalobos et al., 2015).

PMF, as one of the most important receptor models, has been recommended by the US EPA for source apportionment and has been widely used over the years (Al-Dabbous and Kumar, 2015; Liu et al., 2015; Manousakas et al., 2015; Parworth et al., 2015). Similar to the general factor analysis models, the PMF model attempts to apportion the source profile and the source contribution on the basis of the receptor chemical composition dataset, which allows the sources to be identified by its extracted source profile. USEPA PMF 5.0 and PMF2 (two-way PMF), two valuable models throughout the world, have proven to be successful when used in source apportionment studies. In addition to the two models mentioned above, some other PMF solutions have been proposed by the other researchers; for example, Bzdusek and Christensen (2006) developed a variant of PMF with a nonnegative least squares (NNLS) technique. This new method of PMF can enrich the solutions of PMF and can greatly enhance the availability of the PMF model. It is well known that source apportionment results for environmental data were always unpredictable and difficult to judge. Thus, it is essential to perform source apportionment with more than one model, as the validity of a conclusion can be increased after the results from different models are compared.

In this work, a new type of PMF model called the WALSPMF, which combines an eigenvalue-based method and weighted alternating least squares (WALS) process, has been developed to attempt the apportionment of possible sources. This paper can be divided into the following three parts: (1) the principle of the WALSPMF model is described in detail; (2) the synthetic receptor is constructed and apportioned by the new model, and the estimated source contributions by WALSPMF are compared with the simulated contributions to test the performance of the new model; and (3) an ambient dataset analysis from a large harbour site (Tanggu) in China using WALSPMF and EPAPMF 5.0 is described. The results from different PMF models are compared and discussed. The findings of this work can provide useful information for improving the PMF model in further source apportionment works.

2. Methods

2.1. Principle of WALSPMF

The principle of WALSPMF is similar to that of two-way PMF (PMF2). WALS is the weighted alternating least squares, which has been proposed to handle data uncertainties more appropriately and to avoid noise propagation into the estimated parameters. Andrews et al. have studied chemometrics methods based on PCA and WALS, and their results indicated that these methods can provide more reliable results and better estimated parameters when high amounts of noise are present (Andrews et al., 1996; Dadashi et al., 2013; Terrado et al., 2009; Wentzell et al., 1997).

PMF2, developed by Paatero and Tapper (1994), attempts to apportion the source profile and the source contribution on the basis of observations at the receptor site (Al-Dabbous and Kumar, 2015; Habre et al., 2011; Liu et al., 2015; Manousakas et al., 2015; Parworth et al., 2015):

$$x_{ij} = \sum_{h=1}^{p} g_{ih} f_{hj} + e_{ij} \tag{1}$$

where x_{ij} is the concentration of the jth species measured in the ith sample, f_{hj} (μg/μg) is the fraction of the jth element in the hth source, g_{ih} is the contribution of the hth source to the ith sample, e_{ij} are the residuals, and p is the number of factors. Unlike some conventional factor analysis methods, the factor loadings and scores (f_{hj} and g_{ih}) in PMF are constrained to non−negative values (Paatero, 2007).

The task of PMF is minimizing Q (the 'object function'). Q is defined as:

$$Q(E) = \sum_{i=1}^{m} \sum_{j=1}^{n} (e_{ij}/\sigma_{ij})^2 \tag{2}$$

where the value σ_{ij} is the "uncertainty" in the jth species for the ith sample. The uncertainty σ_{ij}, which should be inputted by users, is used to down weight observations that include sampling errors, detection limits, missing data, and outliers (Paatero, 2007).

Similarly, the object of WALSPMF is to find a minimum Q value, as shown in Eq. (2). The solution to finding the minimum Q is using the WALS (weighted alternating least squares) iteration. The detailed process of WALSPMF is described as follows.

(1) The PCA step, which is to determine the initial source profile (F_0) and contribution (G_0) (Thurston and Spengler, 1985).

The receptor matrix X (m × n: m is the number of the sample, and n is the number of the chemical species) would first be normalized:

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_j} \tag{3}$$

where z_{ij} is the normalized concentration value of jth species in ith sample, x_{ij} is the original concentration (μg m^{-3}) of jth species in ith sample, $\bar{x}_j$ is the mean concentration (μg m^{-3}) of jth species for all m samples, and σ_j is the standard deviation of jth species for all m samples. Thus, for the normalized matrix Z (m × n), the mean concentration of each species is 0 and the standard deviation is 1.

Next, the factors would be extracted from matrix Z by PCA with varimax rotation (Thurston and Spengler, 1985):

$$Z_{m\times n} = T_{m\times p} \times L'_{n\times p} \tag{4}$$

where T is the factor score matrix, L′ is the factor loading matrix (L' is the transpose matrix of L), and p is the number of extracted factors. According to related research, the suggested number of extractor factors would include those factors with an eigenvalue greater than 1; this rule is adopted for the WALSPMF model in this work.

The APCS would be calculated by a matrix and an artificial matrixZ_0, as introduced by Thurston and Spengler (1985), and the initial F (called F_0) would be obtained through APCS and receptor matrix X. The detailed process for calculating the APCS and F_0 is described in the Supporting Information (Section S1).

(2) The nonnegative constrained step to obtain the nonnegative profile (F_0^*) and contributions (G_0^*). To obtain the nonnegative