Atmospheric Environment 108 (2015) 151–157

Contents lists available at ScienceDirect

Atmospheric Environment

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

Constraining the factor analytical solutions obtained from multiple-year receptor modeling of ambient $PM_{2.5}$ data from five speciation sites in Ontario, Canada

Uwayemi M. Sofowote [a, *], Yushan Su [a], Ewa Dabek-Zlotorzynska [b], Ankit K. Rastogi [c, 1], Jeff Brook [d], Philip K. Hopke [e]

[a] Air Quality Monitoring and Assessment Unit, Environmental Monitoring and Reporting Branch Ontario Ministry of the Environment and Climate Change, 125 Resources Road, Etobicoke, Ontario M9P 3V6, Canada
[b] Analysis and Air Quality Section, Air Quality Research Division, Atmospheric Science and Technology Directorate, Science and Technology Branch, Environment Canada, 335 River Road, Ottawa, Ontario K1A 0H3, Canada
[c] Faculty of Medicine, University of Toronto, King's College Circle, Toronto, Ontario M5S 1A8, Canada
[d] Process Research Section, Air Quality Research Division, Atmospheric Science and Technology Directorate, Science and Technology Branch, Environment Canada, 4905 Dufferin Street, Toronto, Ontario M3H 5T4, Canada
[e] Institute for a Sustainable Environment, Clarkson University, P.O. Box 5708, Potsdam, NY 13699-5708, USA

HIGHLIGHTS

- A new method for imposing mathematical constraints on multi-site PMF is discussed.
- The concept of softness in factor solutions is discussed.
- An iterative method that uses the measure of softness to select the individual constraints to impose is detailed.

ARTICLE INFO

Article history:
Received 24 September 2014
Received in revised form
11 February 2015
Accepted 16 February 2015
Available online 17 February 2015

Keywords:
$PM_{2.5}$
PMF
Receptor modeling
Speciation
Source apportionment

ABSTRACT

Rotational ambiguity in factor analyses leads to solutions that are not always consistent with reality. The inherent non-negativity constraints in positive matrix factorization (PMF) help to prevent factor solutions from becoming overly unrealistic, but they are not sufficient to prevent unwanted rotations that could manifest in factors that should have similar compositions varying across multiple sites. The Canadian National Air Pollution Surveillance (NAPS) network operates five fine particulate matter ($PM_{2.5}$) speciation sites in Ontario. Data from these sites from 2005 to 2010 were subjected to PMF to obtain factors representing sources of particulate matter. Eight factors were found to be common across these sites. These factors had profiles that varied greatly from one site to the other, suggesting that the PMF solutions were impacted by some rotational ambiguity. New features in the EPA PMF V5 program allow the use of a priori information to impose mathematical constraints that guide the evolution of the factor solutions. These constraints reduce the rotational space. In situations where major emissions sources are known and located in the neighborhood of receptors, or emissions inventories and literature source profiles exist, it is easy to use these profiles to force the factor solutions to conform to the expected signatures. In our case, reported source profiles were neither available nor applicable due to the large spatial span of potential sources and receptor sites. This work describes how such constraints can be generated and used in these complex situations. The fundamental principle explored in this work is the concept of 'stiffness' of PMF solutions to identify the desirable non-rotating factors.

1. Introduction

Factor analytical (FA) models are typically employed to extract the underlying factors and their contributions from ambient air

* Corresponding author.
E-mail address: Uwayemi.Sofowote@ontario.ca (U.M. Sofowote).
[1] Currently a medical student at University of Toronto.

http://dx.doi.org/10.1016/j.atmosenv.2015.02.045

quality monitoring data sets comprising samples (typically in rows) and the concentrations of measured variables (in columns) when minimal or no prior information is available regarding the sources of pollution (Hopke, 1985; Paatero and Tapper, 1993). Strict orthogonality of factors such as offered by principal components analyses (PCA) is generally not physically valid for solutions derived from environmental data sets. Thus, another well-known FA model used in ambient air studies called positive matrix factorization (PMF) produces solutions that are not unique, i.e., there may exist an infinite number of solutions that can be obtained by simple rotations of the factor axes (Paatero and Hopke, 2009; Paatero et al., 2005, 2002; Xie et al., 1999). This indeterminateness defines the problem of rotational ambiguity wherein the non-negativity constraints in PMF by themselves may not be sufficient to guarantee physically relevant factors. To counter this problem, a priori information no matter how minimal can be used to help obtain a better physically meaningful solution. The general framework for applying constraints to PMF solutions from multiple sites in the close neighborhood of potential point sources has already been discussed by Escrig et al. (2009), Amato et al. (2009), and Amato and Hopke (2012). Hitherto, there has been no report on developing constraints for multiple receptor sites that are regionally disperse and are not necessarily affected by the same point sources.

This study is the first to assess the extent of rotational ambiguity in PMF solutions generated from PM speciation sites operated by Canada's National Air Pollution Surveillance (NAPS) program and develop simple diagnostic rules to differentiate between potential source regions (i.e., locally or regionally sourced) based on the resistance to pulling factor elements to target values. Other studies have reported analyses for the NAPS sites in Ontario. Previously, Jeong et al. (2011) had applied PMF to data from five sites across Canada; including two in Ontario. In that study, no attempt was made to use a priori knowledge to make factors in the same general region agreeable. More recently, a follow-up study that included the site at Simcoe has been reported by the same authors but the focus of their work was on an elemental carbon (EC)-rich factor found at Windsor (Jeong et al., 2013).

2. Methods

2.1. Constraints in the PMF receptor model

The object function that is minimized in the PMF analyses Q, is defined by

$$Q = Q_{main} + Q_{aux} \tag{1}$$

where Q_{main} is the main portion of the objective function that is solved by the minimization of the sum of the scaled residuals (Paatero and Tapper, 1993) for a data set which can be represented as a matrix X (n × m) = G F + E (n × m), where n and m are the number of observations and variables; F, G and E are the factor profile, contribution and residual matrices respectively. In PMF, Q is solved by applying non-negativity constraints i.e., $F \geq 0$, $G \geq 0$.

$$Q_{main} = \sum_{i=1}^{m} \sum_{j=1}^{n} \left[\frac{e_{ij}}{s_{ij}}\right]^2 \tag{2}$$

where e_{ij} is the residual value given by:

$$e_{ij} = x_{ij} - \hat{x}_{ij} = x_{ij} - \sum_{k=1}^{p} g_{ik} f_{kj} \tag{3}$$

$\hat{x}_{ij}$ is the factor analytic modeled value of the measured x_{ij} value and s_{ij} is the uncertainty or standard deviation associated with the determination of x_{ij}. The values f_{kj} and g_{ik} are elements of the F and G matrices, respectively.

The expanded relative uncertainties and the method detection limits of speciated PM variables were used to generate equation-based uncertainties s_{ij} and missing data values were replaced with the median value of the species prior to the model run. All model runs were monitored by examining the Q values obtained in the robust mode (Hopke, 2001).

The Multilinear Engine (ME-2) is a program that generically solves all FA models (including the bilinear version that is the typical form of PMF analyses) (Paatero, 1999). It includes the ability to incorporate user-defined scripts for the addition of auxiliary equations that contain expressions of a priori information that help guide the solutions of the FA, thus reducing the total space of possible factor rotations.

Q_{aux} is derived from the auxiliary equations that can be viewed as a measure of ambiguity in the factor matrices in spite of the initial non-negativity constraints of PMF Q_{main}, i.e., it describes the total sum of the residuals of the auxiliary equations scaled by their individual ‘softness’ (Paatero and Hopke, 2009).

$$Q_{aux} = \sum_{v=1}^{v} Q_{aux,v} = \sum_{v=1}^{v} \left[\frac{r_v}{s_v}\right]^2 \tag{4}$$

where r_v are the residuals ($f_v - a_v$) for every target value a_v, and s_v denotes the user-specified ‘softness’ of all the *v* auxiliary equations which are not to be confused with the standard deviations derived from the randomness of the residuals in Q_{main} (note that f_v and a_v can also be created from abundance ratios of factor elements within profiles). It is clear that the larger the value of s_v, the smaller Q_{aux} will be, and by extension the smaller will be the overall change in Q. A ‘stiff’ factor (obtained from solving Q_{main}) will require a relatively strong pull (i.e., a higher Q_{aux}) to achieve a rotation away from its original solution while for a ‘soft’ factor; Q_{aux} can be relatively smaller since the rotation can be achieved relatively easily. The ease of the rotation can also be estimated by the ‘softness’ in the factor solution. A ‘soft’ solution will have relatively larger s_v values and is inherently more subject to rotational ambiguity.

$$s_v = \frac{f_v - a_v}{\sqrt{Q_{aux,v}}} \tag{5}$$

A stiff solution indicates that the factor is resistant to pulling and is fairly unique with minimal rotational ambiguity. The degree of softness or stiffness can be gauged by monitoring the s_v (and Q_{aux}) values for each mathematical constraint applied to each factor. There are no hard and fast rules as to the increment in Q, (i.e., dQ), generated by applying the auxiliary equations that will indicate rotational ambiguity or a lack thereof. The relevant literature has suggested that while dQ in the 100s may be acceptable, dQ in the thousands are probably suspicious (Amato et al., 2009; Paatero and Hopke, 2009). More recently, guidelines of a maximum dQ of about 5% have been suggested (Norris et al., 2014). However, a hard pull in an auxiliary equation may be attempted if the forced rotation of the factor is empirically justified (Paatero and Hopke, 2009). Thus, one of our objectives was to explore the softness/stiffness of factor solutions by pulling factor relationships to specific targets or anchor values that were generated as a priori information from multiple sites across the Province of Ontario, Canada and discuss their implications for constraining PMF source apportionment of particulate matter (PM).

In Canada, the NAPS program, a cooperative program of the federal, provincial, territorial, and some municipal governments, supports various air quality programs across the country, and has