Introduction and Application of the PMF Model to Estimate the Source Apportionment of PM2.5 at Various Sites

InJo Hwang

Department of Environmental Engineering, Daegu University, Daegu 712-714, Korea

Abstract: To manage ambient air quality and establish effective emissions reduction strategies, it is necessary to identify sources and to apportion the ambient PM mass. To do so, receptor models have been developed that analyze various measured properties of the pollutants at the receptor site, identify the sources, and estimate their contributions. Receptor modeling is based on a mathematical model that analyzes the physicochemical properties of gaseous and/or particulate pollutants at various atmospheric receptors. Among the multivariate receptor models used for PM source identification and apportionment, positive matrix factorization (PMF) has been developed by Paatero in 1997. PMF have been developed for providing a new approach to multivariate receptor modeling based on explicit least-squares technique. Also, PMF shown to be a powerful technique relative to traditional multivariate receptor models. PMF has been implemented in two different algorithms: PMF2 (or PMF3) and the multilinear engine (ME). Since the release of PMF2 and ME, these programs have been successfully applied to assess ambient PM source contributions at many locations in the world. In this study, I would like to introduce about outline of the PMF model and application of the PMF model to estimate the source apportionment of ambient $PM_{2.5}$ at various sampling sites in USA and Korea. This study suggests the possible role for maintain and manage ambient air quality and achieve reasonable air pollution strategies.

Introduction

Diverse pollutants from a variety of sources have resulted in serious atmospheric pollution such as visibility degradation, acidic deposition on local and regional scales, and such climate chance on a global scale. Because PM (particulate matter) has adverse human health and welfare effects, many countries developing and implementing environmental policies in order to control PM by establishing ambient air quality standards for PM. To establish the relationship between PM concentration and public health impacts, to understand source/receptor relationships, the STN (speciation trends network) program was started by US EPA in 2000. Moreover, in order to protect visibility in the mandatory Class 1 areas, the IMPROVE program was established in 1985 (Malm et al., 1994).

To manage ambient air quality and establish effective emission reduction strategies, it is necessary to identify sources and to apportion the ambient particulate matter (PM) mass. Quantitative and qualitative analysis of sources is needed to facilitate control policies to reduce ambient air pollutants. To do so, receptor models have been developed that analyze various characteristics of the pollutants at the receptor site and estimate the source contributions (Hwang and Hopke, 2006). Receptor modeling is based on a mathematical model that analyzes the physicochemical properties of gaseous and/or particulate pollutants at various atmospheric receptors. Among the multivariate receptor models used for PM source identification and apportionment, positive matrix factorization (PMF) has been developed by Paatero in 1997. PMF

have been developed providing a new approach to multivariate receptor modeling based on explicit least-squares technique. Also, PMF has been shown to be a powerful technique relative to traditional multivariate receptor models. It has been successfully used to estimate ambient PM source identifications and contributions in the previous studies of air pollution research field during the past decade in the world.

The objective of this study was to introduce about outline of the PMF model and application of the PMF model to estimate the source apportionment of ambient $PM_{2.5}$ at various sampling sites in USA and Korea. This study suggests the possible role for maintain and manage ambient air quality and achieve reasonable air pollution strategies.

Experimental Methods

Sample Collection and Analytical Methods

$PM_{2.5}$ samples were collected at two San Jose STN sites, Kalmiopsis IMPROVE site (Oregon), and Yeongeon-dong, Seoul (PM_{10} samples) (Fig. 1). The $PM_{2.5}$ Samples were collected on teflon, nylon, and quartz filters by SASS sampler and IMPROVE sampler. The Teflon filters were used for the analysis of mass concentrations and elemental analysis by ED-XRF for Fe to Pb or particle-induced X- ray emission (PIXE) for Na to Mn, and proton elastic scattering analysis (PESA) for elemental hydrogen concentration. The nylon filters were used for the analysis of anions (SO_4^{2-}, NO_3^-, and Cl^-) by ion chromatography (IC). The quartz filters were used for organic carbon (OC) and elemental carbon (EC) by TOT (thermal optical transmittance) method or IMPROVE/thermal optical reflectance (TOR) method. In the case of TOR method, total carbon was separated into eight temperature-resolved carbon fractions, defined by temperature and oxidation atmosphere (Hwang and Hopke, 2007).

Samples were collected every 3 or 6 days, and a total of 221 samples were collected from February 2000 to May 2002 at the 4th Street site, and 230 samples were obtained from October 2002 to February 2005 for the Jackson Street site. A total of 493 samples were obtained from March 2000 to May 2004 for the Kalmiopsis IMPROVE site. Also, a total of 164 samples were collected from August 2006 to November 2007 for the Yeongeon-dong, Seoul site (Hwang and Yi, 2010).

PMF model

An explicit least-square approach called PMF can be written as Eq (1) (Paatero, 1997)

$$X_{ij} = \sum_{k=1}^{p} g_{ik}f_{kj} + e_{ij} \qquad i = 1, \cdots, n; \ j = 1, \cdots, m; \ k = 1, \cdots, p \tag{1}$$

where x_{ij} is the measured concentration in the jth species in the ith samples, g_{ik} is the mass concentration of the kth source contributing to the ith sample (i.e., source contributions), f_{kj} is the mass fraction in the jth species from the kth source (i.e., source profiles), e_{ij} is the portion of the measurements that cannot be fitted by the model (residuals), and p is the total number of sources. PMF depends on error estimates for each measured data value (Polissar et al., 1998). Increased error estimates provide a useful method to reduce the weight of below detection limit (BDL) and missing values. Polissar et al. (1998) have suggested an approach for estimating the concentration

values and their associated error estimates including BDL (below detection limit) values and missing data, and this approach were used in this study. The details were reported in previous studies (Hwang and Hopke, 2007). The parameter, FPEAK, was used to control rotations. By setting a nonzero value of FPEAK, the PMF2 is forced to add one g vector to another and subtract the corresponding f factors from each other, thereby yielding more physically realistic solutions (Hopke, 2000). In this study, FPEAK values between -1 and 1 were examined. To select the input variable, we examine the average value of S/N ratio over the data set. Variables with very low S/N values (bad variables, S/N ratio $\ll 0.2$) were excluded from the analysis while the weak variables ($0.2 <$ S/N ratio < 2) were downweighted.
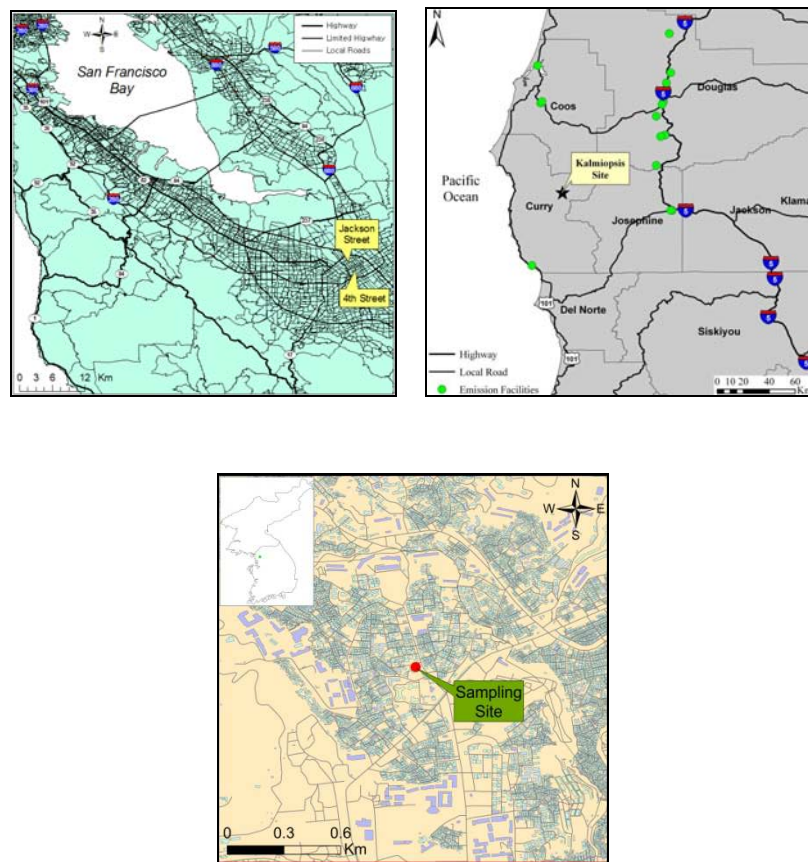

Figure 1: The location of the sampling sites.

Results and Discussion

Source apportionments

The optimal number of sources was determined to be 9 based examinations of the several indicators. To estimate actual unit of source profiles and source contributions, we obtained scaling factors using MLR (multiple linear regression). Figure 2 and 3 shows the source profiles and temporal variation of source contributions for the 4th Street and Jackson Street in San Jose.
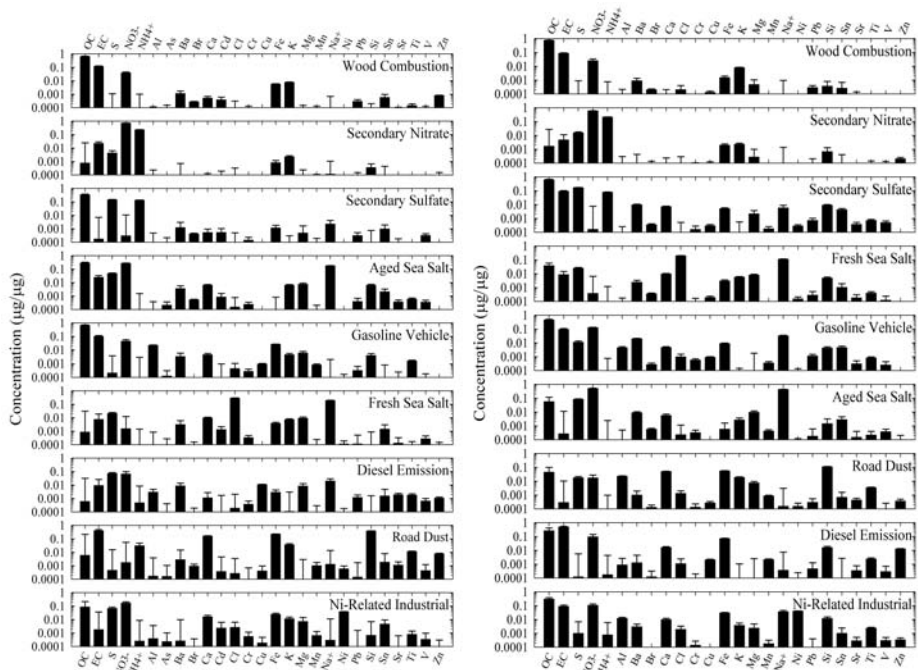
Figure 2: The source profiles for the 4th Street (left) and Jackson Street (right).
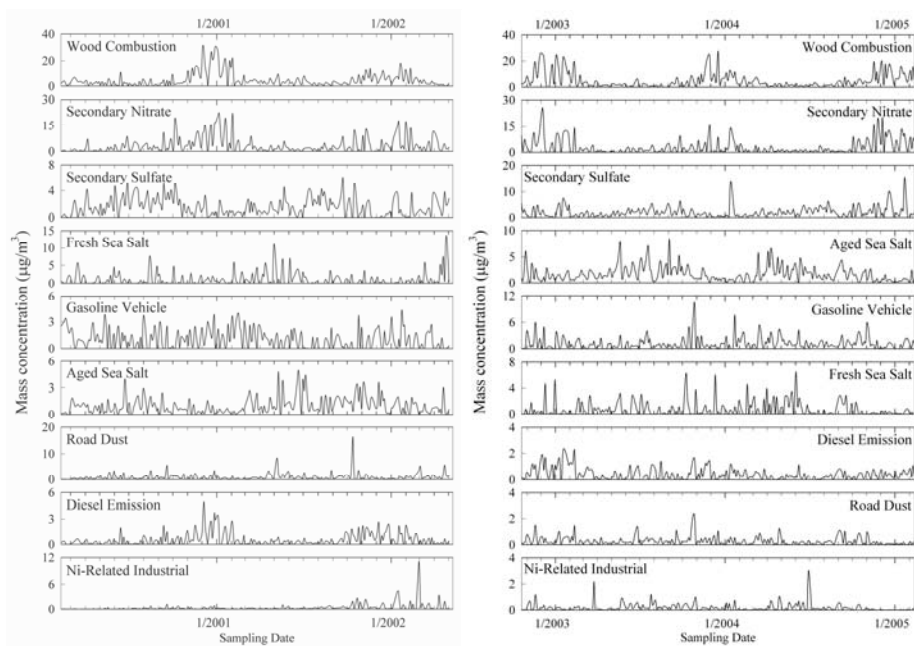

Figure 3: The temporal variation of source contributions for the 4th Street (left) and Jackson Street (right).

Figure 4 shows the comparison of source profiles resolved by PMF for both sites. This comparison used only major marker species and high contributing species. In the case of wood combustion, secondary nitrate, secondary sulfate, fresh sea salt, gasoline vehicle, aged sea salt, and Ni-related industrial source profiles were estimated good agreement in two sampling sites. But, road dust and diesel emission source profiles showed some difference for several species.
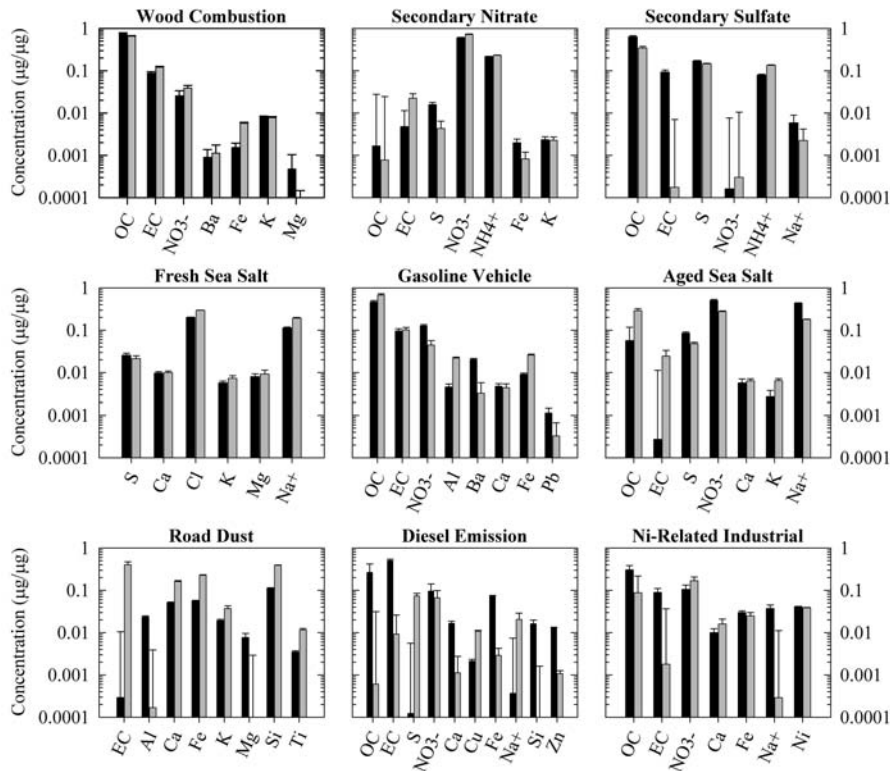

Figure 4: Comparison of source profiles resolved by PMF for the 4th Street and Jackson Street.

To analyze local point source impacts from various wind directions, CPF (conditional probability function) values were calculated using source contribution coupled with the wind direction values measured at the site. CPF is probability that a given factor contribution from a given wind direction will exceed a predetermined threshold criterion. CPF values were defined by Eq (2) and sources are likely to be located in the directions that have high CPF values.

$$CPF = \frac{m_{\Delta\theta}}{n_{\Delta\theta}}$$

(2)

where $m\Delta\theta$ is the number of occurrence from wind sector $\Delta\theta$ that are upper 25 % of source contributions, $n\Delta\theta$ is the total number of occurrence from the same wind sector. The CPF plots in the 4th Street and Jackson Street are presented in Figure 5. The dark line denotes the 4th Street and grey line denote the Jackson Street. For example, CPF plot indicated fresh sea salt source located to the west, Pacific Ocean is located to the west of the sampling site. Therefore, this result supports the sea salt assignments.
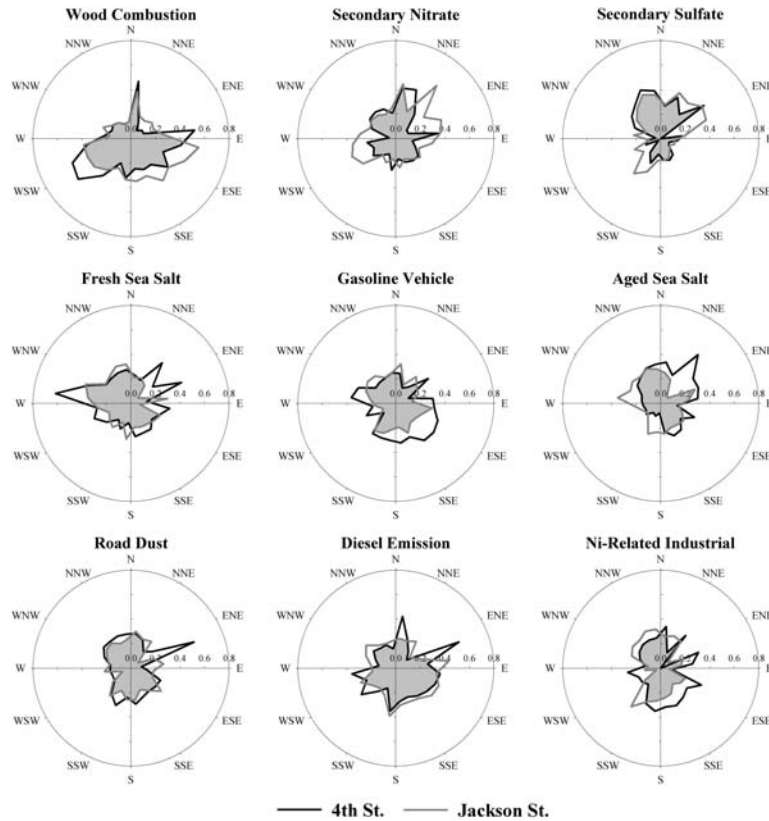
Figure 5: CPF results based on source contributions obtained by the PMF analysis of the data from 4th Street (dark line) and Jackson Street (gray line) in San Jose (gray areas denote overlap area between 4th Street and Jackson Street).

Conclusions

In this study, data from $PM_{2.5}$ samples collected at the various sites such as San Jose STN sites, Kalmiopsis IMPROVE site, and Seoul were analyzed using the PMF2 model. In the case of San Jose STN sites, nine sources were resolved from each sampling site including wood combustion, secondary nitrate, secondary sulfate, fresh sea salt, gasoline vehicle, aged sea salt, road dust, diesel emission, and Ni-related industrial source. Also, contributions of the identified sources showed similar average results at both initial site and relocated site. Therefore, the results from the two sites are quite comparable to one another and could probably be combined into a single analysis.

REFERENCES

[1] Hopke P. K., A guide to positive matrix factorization, in workshop on UNMIX and PMF as applied to PM2.5. Edited by Willis, R.D., RTP, NC, EPA 600/A-00/048, 2000.

[2] Hwang I. J. and Hopke P. K., Comparison of source apportionments of $PM_{2.5}$ at two San Jose STN sites. *J. Air and Waste Manage. Assoc.*, 56, 1287-1300, 2006.

[3] Hwang I. J. and Hopke P. K., Estimation of source apportionment and potential source locations of $PM_{2.5}$ at a west costal IMPROVE site. *Atmos. Environ.*, 41, 506-518, 2007.

[4] Hwang I. J. and Yi S. M., Source apportionment of PM10 using the PMF model in Seoul metropolitan area. Proceeding of the 29th American Association for Aerosol Research Conference (2010. 10. 25), Portland, Oregon, 2010.

[5] Malm W. C., Sisler J. F., Huffman D., Eldred R. A., and Cahill T. A., Spatial and seasonal trends in particle concentration and optical extinction in the United States. *Journal of Geophysical Research*, 99 (D1), 1347–1370, 1994.

[6] Paatero P., Least squares formulation of robust non-negative factor analysis. *Chemom. Intell. Lab. Syst.*, 37, 23-35, 1997.

[7] Polissar A.V., Hopke P. K., Paatero P., Malm W. C., and Sisler J. F., Atmospheric Aerosol over Alaska 2. Elemental Composition and Sources. *J. Geophys. Res.*, *103,* 19045-19057, 1998.