

## RECEPTOR MODELLING OF TOXIC AIR POLLUTANTS IN ANKARA ATMOSPHERE

Öznur Oğuz Kuntasal<sup>1\*</sup>, Deniz Karman<sup>2</sup> and Gürdal Tuncel<sup>1</sup>

<sup>\*1</sup>Middle East Technical University, Department of Environmental Engineering  
06531 Ankara, Turkey oznuro@metu.edu.tr, tuncel@metu.edu.tr

<sup>2</sup>Carleton University, Department of Civil and Environmental Engineering, 1125  
Colonel By Drive, Ottawa, Ontario, K1S 5B6 Canada, Deniz\_Karman@carleton.ca

### ABSTRACT

Positive matrix factorization (PMF) receptor model is applied to volatile organic compound (VOC) data collected at a residential site in Ankara, Turkey during summer of 2003. Data set includes 98 compounds containing, isoprene, halogenated compounds, aromatics, paraffins and olefins. The PMF model explained at least 96% of variation in the data. The sources and computed average source contribution estimates (SCE) are; gasoline vehicle exhaust 42%, diesel vehicle exhaust 30%, architectural coating 11%, biogenic emissions 9% and solvent use 8%. Motor vehicle emissions are the major source of VOCs measured at residential station in Ankara during summer.

**Key Words:** Positive Matrix Factorization, Receptor Modeling, Volatile Organic Compounds, Source Contribution Estimate.

### 1. INTRODUCTION

Receptor models are mathematical procedures for identifying and quantifying the sources of air pollution at a site (i.e., receptor), primarily on the basis of concentration measurements at the receptor site and generally, without need of emission inventories and meteorological data (Watson et al., 2001). Receptor modeling is a critical tool in developing air quality management plans. In this research, Positive Matrix Factorization (PMF) is applied to volatile organic compound (VOC) data collected at a residential site in Ankara, Turkey during summer of 2003.

Positive Matrix Factorization is a new variant of receptor models. Unlike more conventional methods of factor analysis such as principal component analysis (PCA), PMF produces non-negative factors, aiding factor interpretation, and utilizes error estimates of the data matrix. PMF assumes that  $X$  is the matrix of observed data and  $\sigma$  is the known matrix of standard deviations of elements of  $X$ . Both  $X$  and  $\sigma$  have dimensions of  $n \times m$ . The model solves bilinear matrix problem  $X = GF + E$  where  $G$  is the unknown factor scores matrix of dimensions  $n \times p$ ,  $F$  is the unknown factor loadings matrix of dimensions  $p \times m$ , and  $E$  is the matrix of residuals. The problem is solved in the weighted least square sense. Furthermore the solution is constrained so that all the elements of  $G$  and  $F$  are required to be non-negative (Paatero and Tapper, 1994). Over the past few years PMF has been successfully applied in many

atmospheric studies (Hopke, 2003; Paterson et al., 1999; Polissar et al., 1998; Zhao et al., 2004).

The aim of this research is to perform source apportionment of VOCs in Ankara. Source apportionment is an important step in the development of air pollution control strategies. This is the first receptor modeling study conducted to apportion sources of speciated VOCs measured in an urban atmosphere in Turkey.

## 2. MATERIALS AND METHODS

Ambient air measurements were conducted as part of this research to generate VOC data set. The sampling was conducted at consecutive 4-hour intervals over a 24-hour period for two months in summer and two months in winter seasons at a nose-level sampler located on a residential site in Ankara. Samples were also collected at a tunnel and underground garage to generate motor vehicle related source profiles. Samples were collected onto cartridges packed with Tenax TA and Carbopack B resins. Analysis was performed by thermal desorption followed by gas chromatography coupled to a mass selective detector (GC/MSD). Time resolved data provide information on ambient levels of 98 VOCs ranging from C5 to C12, including, isoprene, halogenated compounds, aromatics, paraffins and olefins. This is the first speciated VOC data set generated in Ankara. Detailed information on sampling and analytical methodology is provided by Kuntasal et al. (2004).

Two-dimensional Positive Matrix Factorization (PMF2) receptor model is applied to Ankara VOC data. There are two types of input to the model, namely; i) data matrix and ii) error estimates of the data matrix. A pretreatment of data is required prior to utilize in the model. Although PMF2 can handle incomplete data, very high amount of below detection limit (BDL) or missing values might result in erroneous results. In this study, a method suggested by Paatero and Hopke (2003) for discarding or down weighting of high-noise variables was utilized. In the suggested method, a variable is called “weak” variable if it contains signal (S) and noise (N) in comparable amounts. Similarly, variables containing much more noise than the signal are termed “bad” variables. The element with the S/N larger than 2 and between 0.2 and 2 can be considered as a normal and a weak element, respectively. However, the element with the  $S/N < 0.2$  can be considered as “bad” variable. The bad element should be excluded from analysis, unless it is an important marker for one of the sources.

The model is computed under different initial conditions in order to obtain optimum solution. A critical step in PMF analysis is determination of the number of factors (i.e., sources). The rules suggested by Zhao et al. (2004) are used in this research. Model performance parameters including sum of squares errors (Q) and distribution of scaled residuals helped ascertain the optimum solution. PMF2 runs with 5 factors yield optimum solution.

In general, bilinear factor analysis has rotational ambiguity (Paatero et al., 2002). In the PMF2, FPEAK option is used to control rotation problem. Effect of this parameter is investigated. PMF2 is computed with 5 factors and FPEAK values

ranging from -1.0 to 1.0 for the summer data set. Model runs with FPEAK values of -1.0, -0.5, -0.1, 0.0, 0.1, 0.5, 0.7 and 1.0 yield calculated Q values of 3415, 3173, 3112, 3113, 3109, 3176, 3230, and 3289, respectively. The calculated Q value increases by an increase in absolute value of FPEAK. FPEAK value of -0.5 results in a slight change in source profiles. FPEAK value of zero where the change in Q is slight is accepted as the optimum value.

### 3. RESULTS AND DISCUSSIONS

Utilization of PMF2 with robust mode, FPEAK value of 0.0 and five factors yielded the optimum solution that explains variation in the VOC data generated at residential site in Ankara during summer campaign. Factors identified by PMF2 are interpreted qualitatively by evaluating source profiles (i.e., factor loadings), time variations in source contributions (i.e., factor scores) and explained variations (EV) that are generated by the model. The source profiles are used as the final criteria for source identification. The EV profiles are provided for reference only. Source profiles generated by the model for summer data set are shown in Figure 1.

Source profile computed for Factor 1 (F1 in Figure 1) shows that isoprene and n-pentane are the most abundant species in this factor. This factor explains greater than 80% of the variance in isoprene concentration. Isoprene is a very well known marker for biogenic emissions (Watson et al., 2001). Source contribution for this factor is the highest during daytime and the lowest during nighttime sessions. This pattern indicates the relationship between sunlight and emission from Factor 1, which agrees with the sunlight dependence of emissions from plants. Factor 1 is identified as biogenic emission source.

Source profile computed for Factor 2 (F2 in Figure 1) indicates that toluene is the most abundant compound in this factor. Greater than 30% of variances in chloroform, toluene and cyclohexane concentrations are explained by Factor 2. Source contribution for Factor 2 is the highest during morning session and negligible during night session. Source profile for Factor 2 is compared with the profiles available in the literature to identify the source accurately. Among various profiles used in comparison, Factor 2 shows the best fit ( $R^2 > 0.90$ ) with architectural coating profiles (i.e., Na et al., 2004; Scheff et al., 1989). Contributions of toluene, m&p-xylene and o-xylene are 62%, 6% and 2% in Factor 2, respectively. High abundance of aromatic compounds in Factor 2 indicates that Factor 2 is solvent-based architectural coating source rather than water-based coating.

Source profile for F3 depicted in Figure 1 shows that toluene, m&p-xylene, benzene and 2-methylpentane are among the most abundant compounds in Factor 3. This factor explains variation in most of the light hydrocarbons. Abundance of BTEX compounds and the source profile pattern indicate that this profile could be associated with motor vehicle emissions.

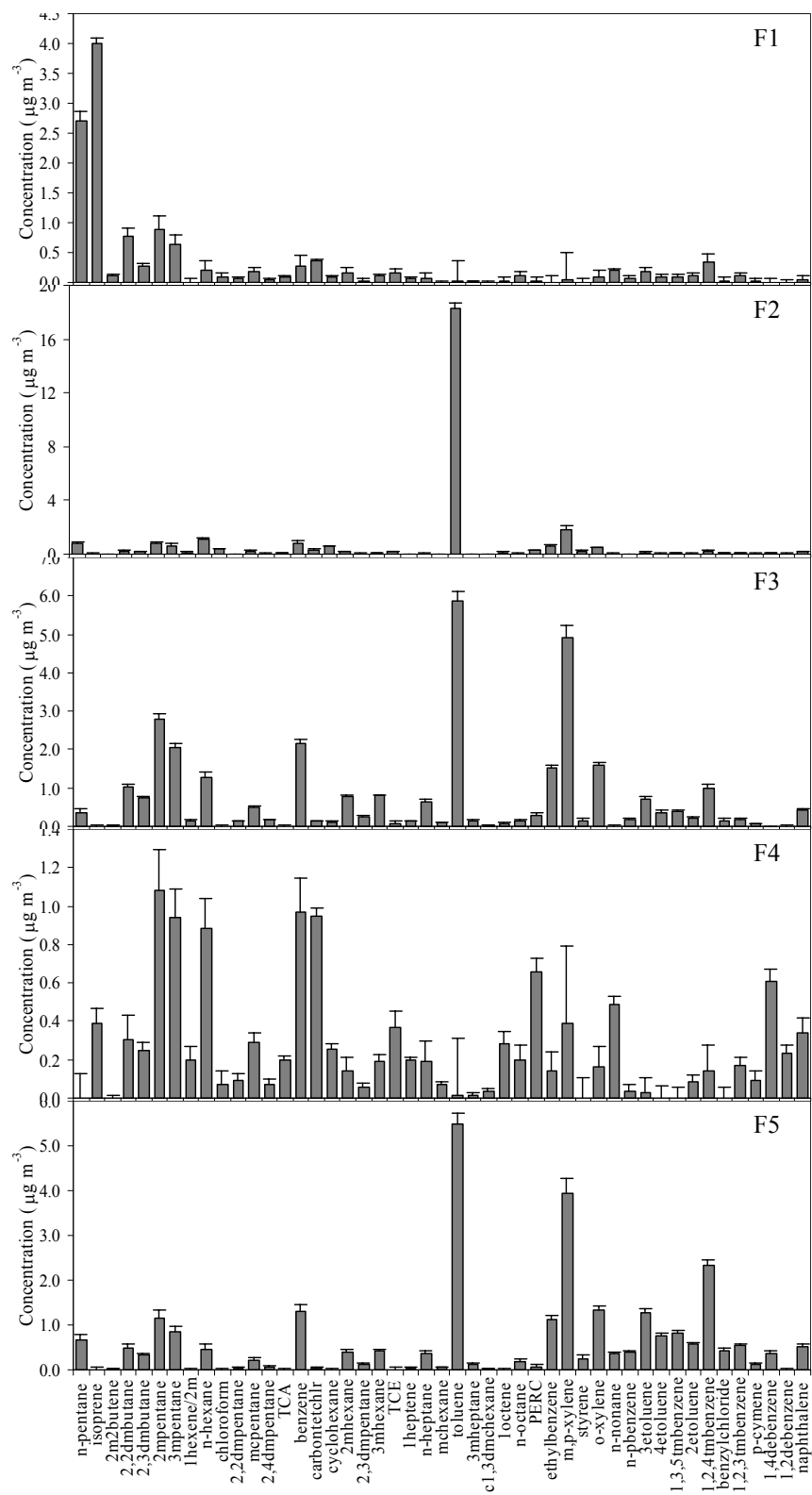


Figure 1. Source profiles computed by PMF2 for summer data set.

Source contribution averaged over sampling sessions for Factor 3 show a well-defined diurnal variation, which is typical for traffic emissions with high contributions during morning and evening rush hours and low contribution during noon session. The source profile is compared with the running vehicle exhaust and fuel profiles generated as part of this research and presented by Kuntasal et al. (2005). Good correlation ( $R^2 > 0.90$ ) is observed with exhaust profiles. Correlation with gasoline fuel profile is better than diesel fuel profile. Thus, Factor 3 is identified as motor vehicle exhaust associated mostly with gasoline-derived vehicles.

Factor 4 source profile (F4 in Figure 1) shows that carbon tetrachloride, benzene, perchloroethylene (PERC) are among the most abundant compounds. PERC is a very well known marker for dry cleaning (Scheff *et al.*, 1989; Watson *et al.*, 2001). Factor 4 explains most of the variation in carbon tetrachloride, 1,1,1-trichloroethane (TCA) and trichloroethylene (TCE) that are commonly used in industrial processes and household cleaners and polishes (Nazaroff and Weschler, 2004). The source contribution does not present a significant diurnal pattern for this factor. Factor 4 is identified as a solvent-use source from various applications.

Toluene, m&p-xylene, 1,2,4-trimethylbenzene and benzene are the most abundant compounds in the source profile for Factor 5 (F5 in Figure 1). This factor explains variation in most of the heavy hydrocarbons. Source contribution indicate a diurnal variation that is very similar to that observed for Factor 3 that is higher source contributions during morning and evening rush hours than noon hours are observed. Factor 5 is compared with the running vehicle exhaust and fuel profiles generated as part of this research. Good correlations ( $R^2 > 0.90$ ) are observed with exhaust profiles. As this profile explains most of the variation in heavy hydrocarbons, Factor 5 is identified as running vehicle exhaust strongly influenced by diesel emissions.

The source contribution estimates (SCE) are calculated for each identified factor using linear regression method that is suggested by Andersen et al. (2001). The results are provided in Table 1. Gasoline vehicle exhaust (42%) and diesel vehicle exhaust (30%) sources contribute most of the VOC emissions observed at the residential station during summer campaign. Architectural coating is the third most abundant source with 12% contribution. Biogenic emissions and solvent use account for 9% and 8% of the total VOC concentration, respectively.

Table 1. Summary of VOC sources and computed SCEs at residential station.

VOC Source	SCE (%)
Gasoline vehicle exhaust	41.77
Diesel vehicle exhaust	29.71
Architectural coating (solvent based)	12.11
Biogenic emissions	8.75
Solvent use	7.66

Model performance is also evaluated. Linear regression between the modeled (i.e., predicted) and measured (i.e., observed) VOC data is performed. The model results reveal a very good fit with the measured VOC data (see Figure 2) with  $R^2$  value of 0.99 and intercept of 1.05. Ratio of the modeled to measured total VOC concentration is 0.95.

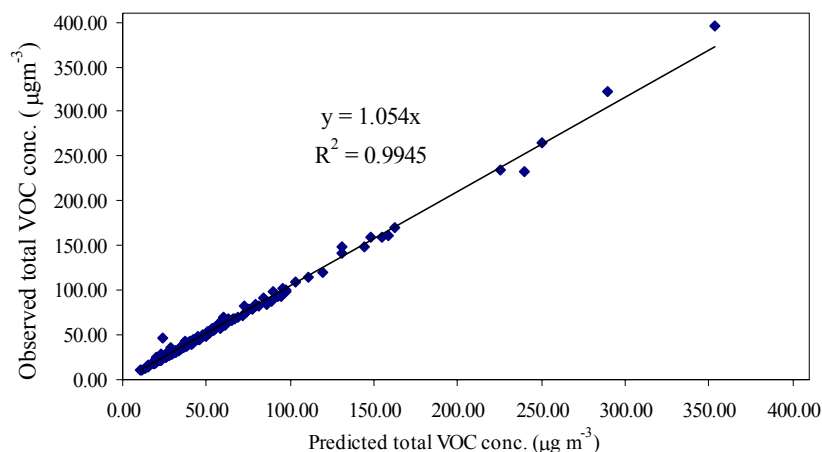


Figure 2. Observed vs. Predicted VOC concentration

Scaled residual errors are also inspected to investigate performance of the model. Most of the scaled residuals are between -2.0 and 2.0 with a random distribution of positive and negative values. The frequency distributions of scaled residual errors for isoprene, TCA, n-octane and 2-ethyltoluene are shown in Figure 3 as example.

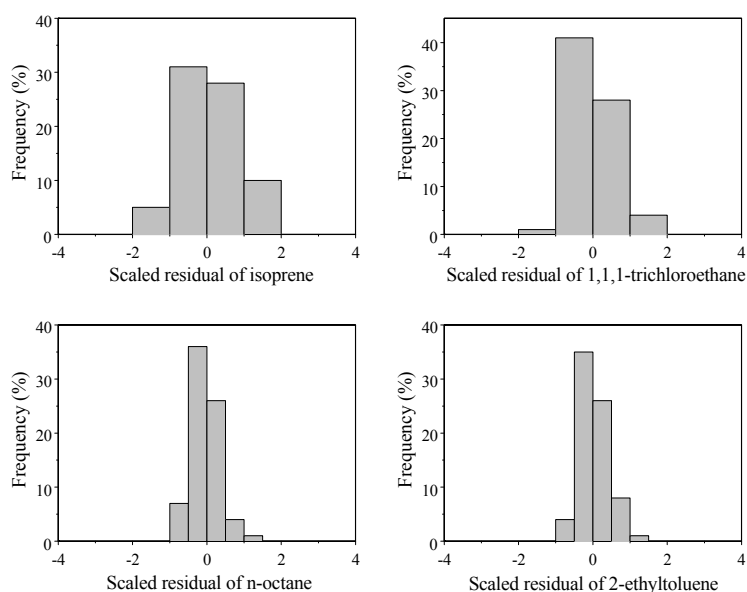


Figure 3. Frequency distribution plot for scaled residual errors.

#### 4. CONCLUSIONS

Receptor modeling technique is applied for the first time in Turkey to VOC data collected in an urban atmosphere. Positive Matrix Factorization model is used to apportion sources of VOCs at a residential site in Ankara during summer season. On the average, motor vehicle related sources contribute to 72% of the total VOC concentration. The solvent related sources including solvent use and the architectural coating result in about 20% contribution. Motor vehicles are the major source of VOCs measured at the residential station in Ankara during summer campaign. PMF runs also successfully resolved biogenic emission source that is effective during summer campaign.

#### 5. ACKNOWLEDGEMENTS

Authors thank to Dr. Daniel Wang from Environment Canada and Prof. Stuart Batterman from University of Michigan for their helps during development of analytical methodology via personal communications; and Gürkan Kuntasal, Ozan Aktan and Kutay Erbayat for their enormous help during fieldwork.

#### REFERENCES

- Anderson, M.J., Miller, S.L. and Milford, J.B., 2001. "Source apportionment of exposure to toxic volatile organic compounds using positive matrix factorization", *Journal of Exposure Analysis and Environmental Epidemiology*, Vol. 11, pp. 295-307.
- Hopke, P.K., 2003. "Recent developments in receptor modeling", *Journal of Chemometrics*, Vol. 17, pp. 255-265.
- Kuntasal, Ö.O., Karman, D., Tuncel, S.G. and Tuncel, G., 2004. "Determination of Volatile Organic Compounds in Ambient Air by Multibed Adsorption/Short Path Thermal Desorption-GC/MS", *Proceedings of the 4<sup>th</sup> Aegean Analytical Chemistry Days*, September 29 - October 3, Kuşadası, Turkey, pp. 339-341
- Kuntasal, Ö.O., Tuncel, G. and Karman, D., 2005. "Motorlu Taşıtlardan Kaynaklanan Toksik Hava Kirleticilerinin Emisyon Profillerinin Tayini", *Proceedings of the IX. Otomotiv ve Yan Sanayi Sempozyumu*, organized by Turkish Chamber of Mechanical Engineers, May 27-28, Bursa, Turkey, pp. 111-114. (in Turkish)
- Na, K., Kim, Y.P., Moon, I. and Moon, K.C., 2004. "Chemical composition of major VOC emission sources in the Seoul atmosphere", *Chemosphere*, Vol. 55, pp. 585-594.
- Nazaroff, W.W. and Weschler, C.J., 2004. "Cleaning products and air fresheners: exposure to primary and secondary air pollutants", *Atmospheric Environment*, Vol. 38, pp. 2841-2865.
- Paatero, P. and Tapper, U., 1994. "Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values", *Environmetrics*, Vol. 5, pp. 111-126.

- Paatero, P., Hopke, P.K., Song, X.H. and Ramadan, Z., 2002. "Understanding and controlling rotations in factor analytic models", *Chemom. And Intelligent Laboratory Systems*, Vol. 60, pp. 253- 264.
- Paterson, K.G., Sagady, J.L., Hooper, D.L., Bertman, S.B., Carroll, M.A. and Shepson, P.B., 1999. "Analysis of air quality data using positive matrix factorization", *Environmental Science and Technology*, Vol. 33, pp. 635-641.
- Polissar, A.V., Hopke, P.K., Malm, W.C. and Sisler, J.F., 1998. "Atmospheric aerosol over Alaska: 2. Elemental composition and sources", *J. of Geophysical Research*, Vol. 103, pp. 19045-19057.
- Scheff, P.A., Wadden, R.S., Bates, B.A. and Aronian, P.F., 1989. "Source fingerprints for receptor modeling of volatile organics", *JAPCA*, Vol. 39, pp. 469-478.
- Watson, J.G., Chow, J.C. and Fujita, E.M., 2001. "Review of volatile organic compounds source apportionment by chemical mass balance", *Atmospheric Environment*, Vol. 35, pp. 1567–1584.
- Zhao, W., Hopke, P.K. and Karl, T., 2004. "Source identification of volatile organic compounds in Houston, Texas", *Environmental Science and Technology*, Vol. 38, Iss. 5, pp. 1338-1347.