Source apportionment of VOC in 3 French sites by CMB and PMF models and critical analysis

F. Troussier, N. Locoge & J. C. Galloo
LCSQA-Department Chemistry and Environment,
Ecole des Mines de Douai, France

Abstract

Hourly concentrations of 31 ambient volatile organic compounds (VOC) were analysed between 2001 and 2003 at three sites located close to three large French cities, in order to calculate contributions of each VOC source. Winter and summer VOC datasets were analysed using chemical mass balance (CMB) and positive matrix factorisation (PMF). For each site, some common sources (vehicle exhaust, gasoline evaporation, leaks of gas, biogenic and domestic heating) were found. Whatever the site of measurement, the automobile exhaust source is most important (35.4 to 61.4% of the total contributions according to the season and the site) in terms of contribution of source. First PMF and CMB results are quite different because PMF finds factor relating to distant sources. New CMB modelling using distant sources profile were made and so CMB and PMF exhibit nearly identical solutions.
Keywords: volatile organic compound, urban-periurban-rural sites, data analysis, source apportionment, receptor modelling.

1   Introduction

The volatile organic compounds (VOC) are important trace species emitted into the atmosphere from anthropogenic and biogenic sources [1], which can influence its chemistry in many ways [2].

In France, VOC measurements are carried out by air quality monitoring networks. Continuous hourly measurements of 31 VOC from C2 to C9 have been performed since 2001 in three regional networks (Marseille, Strasbourg and Grenoble) and so a big VOC concentration database is now available. Each site presents a specific typology: urban site for Marseille, periurban site for

Strasbourg and rural site under industrial influence for Grenoble. The sources likely to influence the ambient concentrations of VOC on each site were systematically indexed and three databases were made up starting from the VOC concentrations data.

In recent years, many studies have been conducted on VOC ambient [3–7]. In the same way, many studies explain CMB [8–10] or PMF [11, 12] results on VOC datasets. However, among these studies a restricted number relates to a comparison between PMF and CMB results [3, 13]. According to our exceptional dataset (more than 1.200.000 data), source apportionment was performed and receptor models were applied to VOC data divided into seasonal periods (summer and winter).

So, in the present study, the objectives are:
- to use receptor models to calculate relative contributions of each source at the three sites
- to compare PMF and CMB results
- to optimise CMB modelling according to PMF results

2   Field studies and sampling sites

We systematically listed the whole of the VOC sources likely to influence the three receptor sites located near 3 large French cities. This localization is necessary. Indeed, without a good knowledge of the sources likely to influence the ambient concentrations, it will be impossible to insert good profiles sources in the CMB and to explain the results of the different modelling especially for PMF modelling.

2.1 Marseille Site (43°16' N, 5°23' E)

For this site, the VOC measurement is led at the Prado station which also corresponds to the central station of Marseille air quality monitoring network (AIRMARAIX) located in the southern half of the city on one of the busiest arteries (27,000 vehicles/day) of Marseille. It is an urban station. In a radius of 1 km around the measurement site, there are seven laundries, four little printing works or painting companies and more than twenty garages or petrol stations. In addition to the specific sources of emission which have just been located, the sources of emission likely to influence the site measurement are the automobile traffic of the Prado Avenue, the adjacent streets and one partly underground highway (A50). Lastly, a chemical factory specialized in the manufacture of an amino acid (approximately 10 km in the east of the site, 358 t.y^{-1} of VOC) has been located.

2.2 Strasbourg Site (48°36' N, 7°42' E)

For Strasbourg, the VOC measurement is led on the Strasbourg central station of air quality monitoring network (ASPA). This sampling site is in Schiltigheim, town of approximately 31,000 inhabitants situated next to Strasbourg. It is a periurban site. The presence in the vicinity (approximately 50 m) of the receptor

site of a small gasoline depot is able to influence VOC concentrations. Taking into account the localization of the site, only a few local sources can be at the origin of ambient VOC concentrations. Consequently, some industrial sites relatively strong VOC emitters (refinery, production of elastomers, printing works) but quite far away from the receptor site (more than 2 kilometers) are able to influence the measured concentrations.

2.3 Grenoble Site (45°06' N, 5°43' E)

For the zone covered by the Grenoble network (ASCOPARG), the VOC measurement is led on a rural station under industrial influence: the Champagnier station. The sampling site can be described as slightly influenced. Indeed, it does not have activities recognized like strongly emitting of VOC (laundry, garage, petrol station,) except the relatively limited automobile traffic (proximity of a secondary road). Consequently, the station could mainly be subjected to the action of a near factory site which mainly produces polychloroprene rubber. The site could also be influenced by the emissions of more distant industries such as the chemical pole of Jarrie (production of synthetic gems and basic chemicals) or the emissions of Pont de Claix industries (manufacture of chlorinated compounds, inks production and printing works).

3   Receptor models

For this study, we have used the Chemical Mass Balance (CMB) [14–16] and Positive Matrix Factorization (PMF) [17, 18].

CMB predicts the contribution of different sources to measured receptor concentrations using an inverse variance weighted least-squares linear regression. Mathematically the system is written:

$$C_{ik} = \sum_{j=1}^{p} F_{ij} \cdot S_{jk} \quad i = 1,...,m ;\ k = 1,...,n \tag{1}$$

where C_{ik} is the concentration of compound i for the k observation, F_{ij} is mass fraction of compound i from source j (weighting percentage), S_{jk} is the contribution of source j for the k observation ($\mu g/m^3$).

In contrast to other receptor models (PMF and UNMIX for example), which extract source compositions from the data, CMB requires the user supply source profiles. Also in contrast to the other models, CMB is applied separately to each observation, rather than operating on the data set as a whole. Assumptions made in using CMB and more details on mathematical resolutions have been discussed elsewhere [14]. CMB is largely used for environmental data and generally produce good results for VOC data [13, 19].

PMF is a multivariate receptor model, which analyze a data series of n observations simultaneously in an attempt to determine the number of sources, p; their chemical composition, F_{ij} ; their contributions to each observation S_{jk} and the residual error, ε_{ik} such that:

$$C_{ik} = \sum_{j=1}^{p} F_{ij} \cdot S_{jk} + \varepsilon_{ik} \qquad i = 1,...,m\,; \; k = 1,...,n \qquad (2)$$

The resolution of the equation is done by minimization of Q.

$$Q = \sum_{i=1}^{m} \sum_{j=1}^{n} \left(\frac{\varepsilon_{ij}}{\sigma_{ij}} \right)^{2} \qquad (3)$$

where σ_{ik} is the variation of the data point.

In PMF, the user chooses the number of factors p. Two parameters can help to choose between different sources numbers in order to obtain the optimal number of sources :

❶ IM : "the maximum individual column mean"

❷ IS : "the maximum individual column standard deviation"

More details are given by Lee et al. [20].

4 Results

For this study, the dataset of each site has been divided into summer (June–August) and winter (December–February) periods.

4.1 CMB results

For CMB modelling, the whole of the profiles relating to the VOC sources highlighted during the sources indexing (see section 2) were tested. First modelling will be done with the whole of the sources profiles likely to influence the concentrations on the receptor sites. Then, this modelling is repeated by removing the profiles of the sources giving negative contributions or standards errors larger than the contribution value [8]. Indeed, although present on the study area, some sources cannot contribute significantly to the ambient concentrations. For the sources which influence the receptor site, denominations are given in table 1. Whatever the season and the site, very few sources seem able to influence ambient concentrations (table 2). There are 5 common sources (vehicle exhaust, gasoline evaporation, leaks of gas, biogenic and domestic heating) to which we must add:

- rubber factory source for Grenoble
- printing works and solvent sources for Strasbourg
- printing works, solvent use and amino-acid factory sources for Marseille

Whatever the site of measurement, the automobile exhaust source contribution is most important (35.4 to 61.4% of the total contributions according to the season and the site). Despite the fact that Marseille is the most urban site, the share of the vehicle exhaust source is the weakest of the three sites. Consequently, relative percentage of vehicle exhaust source depends not only on urban traffic intensity but also on the number and the intensity of the other sources and we must be careful not to give hasty conclusions.

Table 1:   Summary of the sources influencing the three receptor sites.

Name	Description
EXHAUST	Traffic exhaust
EVAP	Gasoline evaporation
LEAK	Leak of gas
BIOGENIQUE	Biogenic sources
HEATING	Domestic heating
SOLVENT	Solvent use
PRINTING	Printing works
RUBBER	Rubber factory
AA FACTORY	Amino acid factory

Table 2:   Relative contribution of the different sources on the three receptor sites according to the season.

Source	Marseille		Strasbourg		Grenoble	
	summer (%)	winter (%)	summer (%)	winter (%)	summer (%)	winter (%)
EXHAUST	35.4	41.1	46.0	53.8	58.9	61.4
EVAP	25.6	21.3	22.2	18.6	16.4	9.5
LEAK	9.2	14.6	16.6	18.7	15.9	17.4
BIOGENIQUE	0.7	-	0.8	-	2.4	-
HEATING	-	0.3	-	3.3	-	9.5
SOLVENT	20.0	9.3	13.1	5.2	-	-
PRINTING	2.1	3.8	1.4	0.5	-	-
RUBBER	-	-	-	-	6.4	2.1
AA FACTORY	6.9	9.6	-	-	-	-

On each sampling site, the gasoline evaporation source is overall the second more important source. Thus, the automobile traffic through the evaporation and exhaust sources is very mainly responsible for the measured VOC concentrations (between 61 and 75.3% of the total contents). This result is in agreement with other CMB results obtained elsewhere [9, 21]. In addition, the contribution of gasoline evaporation source is stronger in summer when the higher temperature allows a stronger hydrocarbons evaporation. In agreement with the urban location, the lowest latitude and the increased sunlight of the Marseille site, the contribution of the source gasoline evaporation is the strongest compared to the other sites. Besides, the clear difference between the percentages of this same contribution for the site of Strasbourg (22.2% the summer and 18.6% the winter) and Grenoble (16.4% the summer and 9.5% the winter) can be explained by the presence of a fuel depot near the Strasbourg site.

For the three receptor sites, the leak of gas source is one of the major sources. However, bibliography shows that the share of this source, in the case of urban studies, is not negligible [9, 22, 23]. Taking into account the life times of the two major compounds of the profile of this source (τ(ethane) = 23 days and τ(propane) = 5 days), it is sometimes difficult to distinguish the share taken by

the contribution of old air masses brought back by air transport towards the sampling site. Thus, Hellén et al. [10, 24] integrates the concept of ''remote sources'' in her work with a profile mainly associated with alkanes (20 to 24% in mass for ethane and 15 to 18% for propane depending on the season).

Biogenic source appears only during the summer period when sunlight and temperature are higher [25]. The results show particularly an important percentage (2.4%) for this source for Grenoble site mainly due to forests in the vicinity.

Generally domestic heating source misses for studies using CMB. In our case, among the four profiles tested, only the profile resulting from GENEMIS database could give satisfactory results. Considering the exclusively winter character of this source, the contribution of the source ''domestic heating'' is quantifiable only during this season. On Grenoble site, taking into account the lower urban density (limitation of the number of sources influencing the site) and the average winter temperature very low (between 0 and 2°C), the domestic heating source has one of the strongest contributions.

Solvent use source could be quantified only on Marseille and Strasbourg sites. Indeed, the site of Grenoble is distant from any activity being able to use solvents.

The three other sources which were listed are of industrial origins. Their contributions can present an important percentage (up to 9.6%) and a seasonal variation.

4.2 PMF results and discussion

PMF modelling were made initially on the Grenoble data. The optimal numbers of sources p were chosen according to the two parameters described in section 3. Best results were obtained with 7 factors for winter and summer dataset. In order to compare PMF and CMB results, factors modelling exhaust due to diesel and gasoline were combined in one profile (figure 1).
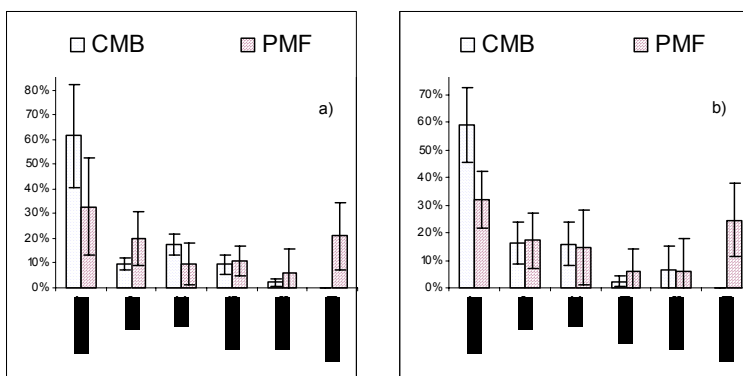

Figure 1:   Comparison between CMB and PMF results for Grenoble data. a) winter period. b) summer period.

   The results of figure 1 show a clear difference (about 30%) for the percentage of the exhaust source contribution calculated by the 2 models. For the other sources, contributions are similar taking into account uncertainties (standard deviation of calculated contribution). However it seems, for winter modelling, that the PMF over-estimates the contribution of the gasoline evaporation source (same contribution winter and summer) and under-estimates the contribution of the leak of gas source. In addition, an unknown source (about 20% of total contribution) was found by PMF for summer and winter dataset.

   To complete this work, we have tried to identify the unknown source. The latter is principally composed of alkanes (ethane 34.1%; butane 15.5%). In recent articles, Héllen et al. [24] and Gouw et al. [26] describe distant sources emissions which during their transport are oxidised remaining thus the less reactive species such as alkanes. Consequently, our unknown source could be a distant source. Starting from the Grenoble dataset, we tried to find observations made in high wind conditions, at night and preferably at week ends. Thus, the influence of the local sources (especially traffic exhaust) will be minimized, and we will be able to reach an approximation of the concentrations profile due to the distant sources (extreme conditions VOC profile). Figure 2 shows the comparison between PMF unknown factor and extreme conditions VOC profile for Grenoble summer data. The two profiles are nearly identical and it seems that PMF was able to distinguish the distant sources in the ambient concentrations.
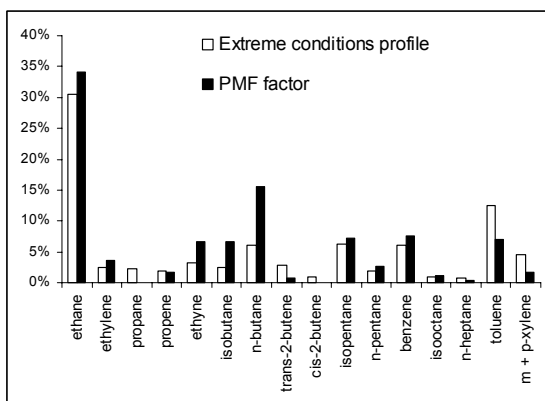

Figure 2:   Comparison between PMF unknown factor and extreme conditions VOC profile for Grenoble site and summer data.

4.3 New CMB modelling

Considering the previously results, we made new CMB modelling on Grenoble summer dataset with the profile (created above) corresponding to distant sources. Thus, the whole of the sources influencing the site of Grenoble the summer will be (this time) entered in CMB model. The results show an average contribution of about 40% for exhaust sources and 20% for distant sources. For the other

sources relative contributions are globally unchanged. Finally, CMB and PMF exhibit, now, nearly identical solutions. Thus, CMB users finding strong contribution for the exhaust source (without using profiles relating to the remote sources) must pay attention. Indeed exhaust source contribution can be over-estimated. Future tests on the Strasbourg and Marseille data will be carried out in order to know the global influence of the distant sources.

5   Conclusions

Hourly concentrations data of 31 VOC measured on three sites located close to three large French cities were analysed using chemical mass balance (CMB) and positive matrix factorisation (PMF). For each site, five common sources (vehicle exhaust, gasoline evaporation, leaks of gas, biogenic and domestic heating) were found. Whatever the site, the automobile exhaust source is most important (35.4 to 61.4% of the total contributions according to the season and the site) in terms of contribution of source. An increase of evaporative source contribution (gasoline evaporation and solvent use) was observed during the passage from the winter months to the summer ones. Two sources appear only during one season: the biogenic source presents a quantifiable share only during the summer months when sunlight and temperature are higher and the heating source is lost during the passage from the winter months to the summer ones. The latter has a strong contribution on Grenoble site taking into account the low number of sources influencing the site and the very low average winter temperature. The industrial sources can have important percentage (up to 9.6%). First PMF and CMB results on Grenoble dataset are quite different because (i) PMF finds unknown factor which contribute to about 20% (ii) an important difference is observed for the exhaust source contribution. Study of bibliography can permit to conclude that the unknown source could be associated to a distant source. New CMB modelling on Grenoble summer dataset with the profile created with extreme conditions corresponding to distant sources was made. So CMB and PMF nearly exhibit identical solutions and CMB users studying rural site must pay attention not to over-estimate contribution of exhaust sources.

Acknowledgements

References

[1]   Sawyer R. F., Harley R. A., Cadle S. H., Norbeck J. M., Slott R. & Bravo H. A., Mobile sources critical review: 1998 NARSTO assessment, *Atmospheric Environment,* **34**(12-14), pp. 2161-2181, 2000.
[2]   Atkinson R. & Arey J, Atmospheric Degradation of Volatile Organic Compounds, *Chemical Reviews*, **103**(12), pp. 4605-4638, 2003.

[3] Latella A., Stani G., Cobelli L., Duane M., Junninen H., Astorga C. & Larsen B. R., Semicontinuous GC analysis and receptor modelling for source apportionment of ozone precursor hydrocarbons in Bresso, Milan, 2003, *Journal of Chromatography A*, **1071**(1-2), pp. 29-39, 2005.

[4] Lin T.-Y., Sree U., Tseng S.-H., Chiu K. H., Wu C.-H. & Lo J.-G., Volatile organic compound concentrations in ambient air of Kaohsiung petroleum refinery in Taiwan, *Atmospheric Environment*, **38**(25), pp. 4111-4122, 2004.

[5] Zhao W., Hopke P. K. & Karl T., Source identification of volatile organic compounds in Houston, Texas, *Environmental Science and Technology*, **38**, pp. 1338-1347, 2004.

[6] Hsieh C.-C. & Tsai J.-H., VOC concentration characteristics in Southern Taiwan, *Chemosphere*, **50**(4), pp. 545-556, 2003.

[7] Na K., Kim Y.P., Moon K., Moon I. & Fung K., Concentrations of volatile organic compounds in an industrial area of Korea, *Atmospheric Environment*, **35**(15), pp. 2747-2756, 2001.

[8] Fujita E.M, Watson J.G. & Chow J.C., Receptor model and emissions inventory source apportionments of nonmethane organic gases in California's San Joaquin valley and San Francisco bay area, *Atmospheric Environment*, **29**(21), pp. 3019-3035, 1995.

[9] Vega E., Mugica V., Carmona R. & Valencia E., Hydrocarbon source apportionment in Mexico City using chemical mass balance receptor model, *Atmospheric Environment*, **34**(24), pp. 4121-4129, 2000.

[10] Hellén H., Hakola H. & Laurila T., Determination of source contributions of NMHCs in Helsinki (60°N, 25°E) using chemical mass balance and the Unmix multivariate receptor models, *Atmospheric Environment*, **37**(11), pp. 1413-1424, 2003.

[11] Jorquera H. & Rappenglück B., Receptor modelling of ambient VOC at Santiago, Chile, *Atmospheric Environment*, **38**(25), pp. 4243-4263, 2004.

[12] Buzcu B. & Fraser M. P., Source identification and apportionment of volatile organic compounds in Houston, TX, *Atmospheric Environment*, **40**(13), 2385-2400, 2006.

[13] Miller S.L., Anderson M.J., Daly E.P. & Milford J.B., Source apportionment of exposures to volatile organic compounds. I. Evaluation of receptor models using simulated exposure data, *Atmospheric Environment,* **36**(22), pp. 3629-3641, 2002.

[14] Watson J.G., Cooper J.A. & Huntzicker J.J., The effective variance weighting for least squares calculations applied to the mass balance receptor model, *Atmospheric Environment,* **18**(7), pp. 1347-1355, 1984.

[15] Watson J.G., Robinson N.F, Fujita E.M., Chow J.G., T.G., Lewis C. & Coulter T., *CMB8 Applications and validation protocol for PM$_{2.5}$ and VOCs*, Desert Research Institute Document N°1808.2D1, 1998.

[16] Watson J.G., Chow J.C., Fujita E.M., Review of volatile organic compound source apportionment by chemical mass balance, *Atmospheric Environment,* **35**(9), pp. 1567-1584, 2001.

[17] Paatero P. & Tapper U., Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data values, *Environmetrics*, **5**, pp. 111-126, 1994.

[18] Paatero P. & Hopke P. K., Discarding or downweighting high-noise variables in factor analytic models, *Analytica Chimica Acta*, **490**(1-2), pp. 277-289, 2003.

[19] Vega E., Mugica V., Carmona R. & Valencia E., Hydrocarbon source apportionment in Mexico City using the chemical mass balance receptor model, *Atmospheric Environment*, **34**(24), pp. 4121-4129, 2000.

[20] Lee E., Chan C.H. & Paatero P., Application of positive matrix factorization in Source apportionment of particulate pollutants in Hong Kong, *Atmospheric Environment*, **33**(19), pp. 3201-3212, 1999.

[21] Srivastava A., Source apportionment of ambient VOCS in Mumbai city, *Atmospheric Environment*, **38**(39), pp. 6829-6843, 2004.

[22] Badol C., Caractérisation des composés organiques volatils dans une atmosphère urbaine sous influence industrielle: de l'identification à la contribution des sources, *thesis*, Université des Sciences et Technologies de Lille, pp. 375, 2005.

[23] Borbon A., Locoge N., Veillerot M., Galloo J. C. & Guillermo R., Characterisation of NMHCs in a French urban atmosphere: overview of the main sources, *The Science of The Total Environment*, **292**(3), 177-191, 2002.

[24] Hellén H., Hakola H., Pirjola L., Laurila T. & Pystynen K.-H., Ambient Air Concentrations, Source Profiles, and Source Apportionment of 71 Different C2-C10 Volatile Organic Compounds in Urban and Residential Areas of Finland, *Environmental Science and Technology*, **40**(1), pp. 103-108, 2006.

[25] Guenther A.B., Zimmerman P.R., Harley P.C., Monson R.K. & Fall R., Isoprene and monoterpene emission rate variability: Model evaluation and sensitivity analyses, Journal of Geophysical Research, **98**(D7), pp. 12 609-12 671, 1993.

[26] Gouw J. A., Middlebrook A. M., Warneke C., Goldan P. D., Kuster W. C., Roberts J. M., Fehsenfeld F. C., Worsnop D. R., Canagaratna M. R., Pszenny A. A. P., Keene W. C., Marchewka M., Bertman S. B. & Bates T. S., Budget of organic carbon in a polluted atmosphere: Results from the New England Air Quality Study in 2002, *Journal of Geophysical Research D: Atmospheres*, **110**(16), pp. 1-22, 2005.