Receptor modeling of ambient VOC at Santiago, Chile

Héctor Jorquera[a],*, Bernhard Rappenglück[b]

[a] *Departamento de Ingeniería Química y Bioprocesos, Pontificia Universidad Católica de Chile, Vicuña Mackenna 4860, Santiago 6904411, Chile*
[b] *Forschungszentrum Karlsruhe GMBH, Institut für Meteorologie und Klimaforschung (IMK-IFU), Bereich Atmosphärische Umweltforschung, Kreuzeckbahnstr. 19, D-82467 Garmisch-Partenkirchen, Germany*

Abstract

Ambient VOC measured at a 1996 spring campaign at Santiago, Chile, have been analyzed using the receptor models UNMIX and Positive Matrix Factorization (PMF). The ambient campaign took place at two sites: a downtown site, close to major traffic emissions, and a residential site, downwind of major industrial sources and highways. At the downtown site the following source apportionments estimates were obtained: fuel evaporation, $29.7 \pm 5.6\%$; gasoline exhaust, $22.0 \pm 3.4\%$; diesel exhaust, $18.1 \pm 2.9\%$; biogenic, LPG and evaporative emissions, $18.0 \pm 3.4\%$, unexplained: 12%. At the residential site, the following source apportionment was obtained: transported gasoline exhaust, $31.2 \pm 4.1\%$; local gasoline exhaust, $25.5 \pm 4.0\%$; evaporative losses, $11.7 \pm 2.8\%$; LPG losses, $11.0 \pm 2.5\%$; biogenic emissions, $7.7 \pm 1.7\%$; diesel exhaust, $6.2 \pm 1.5\%$; unexplained, 7.7%. Thus, near 70% of ambient VOC impacts at both sites are due to mobile sources. The receptor analyses produced source profiles that had distinctive, dominant compounds; in addition, source contributions exhibited diurnal profiles that were consistent with ambient temperature and wind speed data, and the expected activity patterns within the city. Typical errors in the source contributions vary between 15% for the larger sources—like gasoline exhaust—and 25% for the smaller sources—like biogenic emissions. It was found that the number of factors given by the UNMIX model was a good starting point to refine the solution using PMF. Both models showed good performance at discriminating between source profiles that had similar compositions in subsets of common species, but PMF was able to find better, cleaner source profiles that did UNMIX. At both monitoring sites LPG losses appear mixed in with other source profiles, and this feature could not be solved by adding more source profiles in the analyses; this was likely due to a lack of C_3 measurements needed to better resolve an LPG source profile.

1. Introduction

Volatile organic compounds (VOC) are abundant in urban atmospheres, because they come from mobile sources and its associated fuel distribution chain, and from the intensive use of solvents in non-combustion processes. VOC contribute in the complex photochemical cycle leading to high tropospheric amounts of ozone, peroxyacyl nitrates (PANs) and other oxidants in urban areas (Seinfeld and Pandis, 1998). In order to establish ozone abatement policies in a given region, authorities need to assess quantitative estimates of photochemical impacts under different emission control scenarios.

VOC emission inventories are difficult to assemble, because significant emissions come from fugitive (evaporative) sources during the fuel distribution cycle (gasoline, liquefied petroleum gas—LPG—), that are

*Corresponding author. Tel.: +562-686-4421; fax: +562-686-5803.

E-mail addresses: jorquera@ing.puc.cl (H. Jorquera), bernhard.rappenglueck@imk.fzk.de (B. Rappenglück).

hard to quantify (Kourtidis et al., 1999). A quantitative assessment of the contributions from different sources to ambient VOC levels can be obtained by applying receptor models, so this process provides an independent check of the quality of the VOC emission inventories.

2. Receptor models

Receptor models are mathematical procedures for identifying and quantifying the sources of ambient air pollution and their effects at a site (receptor), primarily on the basis of concentration measurements at the receptor site and generally, without need of emission inventories and meteorological data (Willis, 2000). In mathematical terms, the general receptor modeling problem can be stated in terms of the contributions from p independent sources to all chemical species in a given sample as follows (Hopke, 1985):

$$x_{ij} = \sum_{k=1}^{p} g_{ik} f_{kj} + e_{ij}, \qquad (1)$$

where x_{ij} is the jth species concentration measured in the ith sample, g_{ik} is the VOC mass concentration from the kth source contributing to the ith sample, f_{kj} is the jth species mass fraction from the kth source, e_{ij} is a residual associated with the jth species concentration measured in the ith sample, and p is the total number of independent sources. The corresponding matrix equation is

$$X = GF + E, \qquad (2)$$

where X is a $n \times m$ data matrix with n measurements and m number of chemical species; E is a $n \times m$ matrix of residuals; G is a $n \times p$ source contribution matrix with p sources; F is a $p \times m$ source profile matrix. As pointed out by Henry (1987), there are an infinite number of possible solutions to the factor analysis problem (rotations of G and F matrices).

One commonly used multivariate receptor model is the Principal Component Analysis (PCA) (Jollife, 1986). Although PCA has been applied as a tool for source apportionement of ambient VOC (Borbon et al., 2001, 2003; Ho et al., 2002), it suffers from several drawbacks. The factors of PCA are rarely physically explainable without further transformations (rotations), and no fully satisfactory rotation techniques have yet been found (Paatero, 1997). Besides, PCA cannot properly handle missing and below-detection-limit (BDL) data, both common in environmental data.

Another receptor model widely used is the Chemical Mass Balance method (Watson et al., 1990), based in inverse variance weighted least-squares regression. Inputs to the model are chemical composition profiles (known beforehand) for likely sources and the chemical compositions of receptor concentrations. This model has

been applied for ambient VOC source apportionment in Berlin (Thijsse et al., 1999), Cairo (Abu-Allaban et al., 2002), Helsinki, Finland (Hellén et al., 2003), Mexico City (Vega et al., 2000), North Carolina (Lawrimore and Aneja, 1997), and southern California (Fujita et al., 2003) among other locations; a recent review has been published by Watson et al. (2001). However, for Santiago, Chile, no source profiles have been reported so far, so we cannot apply this technique in our case study.

In this work, two advanced receptor models have been applied: UNMIX and Positive Matrix Factorization (PMF); both circumvent the traditional limitations of PCA, by adding constraints into the mathematical analysis. Below we describe those two methodologies.

2.1. Positive matrix factorization (PMF)

In this approach constraints are integrated into the computational process (Paatero and Tapper, 1993, 1994). PMF can be expressed as the following minimization of the goodness of fit function Q:

$$\min_{F,G} Q(X, \sigma, G, F) = \left\| \frac{(X - GF)}{\sigma} \right\|^2 = \sum_i \sum_j \left(\frac{e_{ij}}{\sigma_{ij}} \right)^2,$$

$$e_{ij} = x_{ij} - \sum_{k=1}^{p} g_{ik} f_{kj}; \ g_{ik} \geqslant 0; \ f_{kj} \geqslant 0,$$

$$(3)$$

where σ is the known matrix of error estimates of X. PMF uses a unique algorithm described by Paatero (1997) in which matrices G and F are varied simultaneously on each least-squares step. The solution given by PMF has a degree of freedom corresponding to a scaling coefficient s_k:

$$x_{ij} = \sum_{k=1}^{p} f_{ik} g_{kj} = \sum_{k=1}^{p} f_{ik} \frac{s_k}{s_k} g_{kj} = \sum_{k=1}^{p} f'_{ik} g'_{kj}. \qquad (4)$$

Assuming that all of the sources contributing mass to the VOC samples have been identified, the sum of the mass contributions should add to the measured VOC concentration. Thus, a regression of the measured mass concentration of VOC against the source contribution values permits the estimation of those scaling factors. In the applications reported here, we have chosen the option of normalizing the source profiles f_{ik}, hence the source contributions g_{kj} are directly computed, without applying the regression scheme, that is, without assuming that all possible sources have been identified.

PMF has been recently applied to the source identification of aerosol data from Thailand (Chueinta et al., 2000), Phoenix, AZ (Ramadan et al., 2000) and from Atlanta, GA (Kim et al., 2003). For ambient VOC analysis, PMF has been applied by Miller et al. (2002) for simulated exposure data, and by Anderson et al. (2001, 2002) for actual VOC exposure data.

2.2. The unmix receptor model

UNMIX is a type of factor analysis, but it is geometrically constrained to generate source contributions and profiles with the physically meaningful attribute of non-negativity (Henry, 1997). UNMIX is based on two key algorithms. A first procedure finds data "edges" (more generally, hyperplanes) in a $(k-1)$ dimensional space; since UNMIX implicitly assumes that there are some cases in which only one source is dominant, the edges are supposed to represent those dominant, single source profiles (Henry, 1997, 2003). Once the edges are found, the number of sources is estimated by a Monte Carlo resampling technique in which random subsets of the data are succesively fit with UNMIX; this algorithm, NUMFACT (Henry et al., 1999), estimates the number of sources (k) that can be resolved relative to the given ambient data set.

UNMIX has been used in aerosol characterization studies in Phoenix, AZ (Lewis et al., 2003), at a remote Vermont site (Poirot et al., 2001) and near a phosphorous production facility (Willis et al., 2001). For ambient VOC analysis, UNMIX has been applied at an industrial site in the US (Henry, 2000), at Helsinki, Finland (Hellén et al., 2003) and for simulated exposure data (Miller et al., 2002).

3. Data description

Santiago, Chile, located at $33°27'$S and $70°40'$W, has a population of 5.6 million people, covers an area of roughly $30 \times 35 \, km^2$ and has a dry spring and summer season. The data come from the photochemical campaign conducted in November–December 1996 by Rappenglück et al. (2000). In that work, no analysis of the data by receptor models was attempted; emphasis was placed upon evaluating photochemical products (ozone, PAN) and estimating the level of photochemical activity in the greater metro area. During the campaign, strong radiation and clear skies were frequent, leading to high temperatures between $25°C$ and $33°C$.

Measurements were taken at two locations (separated by approximately 18 km) within the Santiago area. One set of instruments was installed at a facility of the air quality network operated by the Health Ministry (SESMA, 2004) close to a side entrance of Parque O'Higgins (henceforth denoted as POH) located almost in the center of Santiago (540 m MSL). This site is about 300 m from a major highway that crosses the city, and about 150 m from a busy freeway. The other instrumentation was set up in the suburban area of Las Condes (henceforth denoted as LAC) on the northeast side of Santiago RM (802 m MSL). This site is located in the small Mapocho valley leading to the Andes range, far from industrial sources. At both sites, SESMA

routinely measures ozone, CO, NO_x, and meteorological parameters with commercial instruments.

3.1. Data measurement protocol

For the photochemical campaign, the University of Munich equipped both measurement sites with gas chromatographic (GC) systems (Siemens AG RGC 402) for quasi-continuous online NMHC measurements in the range of C_4–C_{12} with a temporal resolution of 30 min. This system has been documented in previous publications and has been widely used in urban air quality studies, e.g. in Athens/Greece (Rappenglück et al., 1998, 1999; Kourtidis et al., 1999), in Munich/Germany (Rappenglück and Fabian, 1999a, b), and in Santiago de Chile (Rappenglück et al., 2000). It has been tested against canister measurements and other GC systems, and cross checked with different calibration gas sources, e.g. calibration gas cylinders, permeation tubes (Rappenglück et al., 1998). Within the German Tropospheric Research Focus (TFS) the GC instruments were subject to comprehensive independent quality assurance procedures (TFS) (Volz-Thomas et al., 2002). In test series with certified synthetic standards in the ppbv range our systems agreed to the instruments of the best participants within $\pm 20\%$ for most compounds. The results suggest that these instruments are very reliable in urban air conditions.

The GC-systems were calibrated, both at the home lab and at the sites in Santiago de Chile, with a certified 27-component hydrocarbon mixture calibration gas cylinder purchased from the National Physical Laboratory (NPL), United Kingdom. This mixture contains aromatics, alkenes, alkanes in the ppbv range. For calibrating other hydrocarbons, the response factor obtained by the most similar hydrocarbon component in respect of both its hydrocarbon class (alkene, alkane, aromatic) and its number of C-atoms was used. Detection limit is 0.01–0.02 ppbv, for C_4 hydrocarbons detection limits are higher (0.03–0.05 ppbv). The detection limits were taken as three times the standard deviation of the corresponding blank values. The estimated overall accuracy was about $\pm 20\%$ and the precision ± 10–15%, depending on the specific hydrocarbon compound.

Several hundreds of 30-min NMHC values were obtained at both sites in Santiago de Chile in 1996. Table 1 displays summary statistics at both measurement sites; the convention taken here was that the BDL cases were assigned half the minimum detection level; thus the missing values are only due to technical problems in the measurement protocol.

3.2. Details on PMF implementation

The protocol of analyses begins by selecting the following error model for the sampled data points, in

Table 1
Statistical summary of measurement campaign at Santiago, Spring 1996 (data in $\mu g\,m^{-3}$)

Compound	Results for downtown, POH site					Results for residencial, LAC site				
	1st Q	Median	3rd Q	BDL[a]	Missing	1st Q	Median	3rd Q	BDL[a]	Missing
Benzene	1.26	2.20	6.73		20	2.36	3.975	6.15		68
Toluene	4.21	7.83	19.42		23	8.35	11.780	16.41		53
Ethylbenzene	0.48	1.00	2.60	15	41	1.17	1.690	2.49		55
p-Xylene	0.48	1.00	2.72	12	41	1.26	1.860	2.64	1	66
m-Xylene	1.21	2.34	6.85		34	2.94	4.315	6.55		54
Cumene	0.02	0.02	0.02	584	22	0.02	0.100	0.20	555	54
o-Xylene	1.08	1.82	4.55		27	1.99	2.860	4.08		52
Limonene	0.03	0.44	0.95	7	23	0.03	0.030	0.33	710	52
n-Propylbenzene/dodecane	0.02	0.02	0.54	345	23	0.25	0.390	0.59	28	52
p/m-Ethyltoluene	0.39	0.93	2.46	44	23	1.23	1.820	2.65	3	52
i-Butylbenzene	0.03	0.03	0.03	585	23	0.03	0.030	0.16	798	52
Mesitylene	0.02	0.33	0.78	178	23	0.54	0.790	1.18		52
Styrene	0.02	0.30	0.72	235	23	0.30	0.470	0.64	11	52
o-Ethyltoluene	0.02	0.20	0.74	294	23	0.34	0.540	0.83	25	52
p-Cymene	0.03	0.03	0.03	640	23	0.02	0.020	0.02	1016	52
1,2,4-Trimethylbenzene	0.64	1.18	3.09	10	23	1.47	2.110	3.04		52
n-Butane/1-butene	2.03	2.80	5.66	3	23	5.20	9.225	14.07	13	50
t-2-Butene	0.01	0.01	0.01	512	16					
c-2-Butene	0.01	0.01	0.01	580	16	1.53	2.270	3.18	28	50
i-Pentane	1.44	2.68	7.82	7	16	3.97	6.585	10.56	14	50
1-Pentene	0.01	0.37	1.05	268	16					
n-Pentane/2-methyl-1-butene	1.05	1.91	6.07	7	16	1.97	3.500	5.62	15	50
t-2-Pentene/isoprene/*c*-2-pentene	5.62	9.98	16.66	8	16	0.99	1.640	2.64	34	50
Cyclopentane/2,3-dimethylbutane/2-methylpentane	0.66	1.78	5.40	85	16					
3-Methylpentane	0.35	0.74	2.52	132	16	0.77	1.370	2.18	36	50
1-Hexene	0.02	0.02	0.57	436	16	0.02	0.170	0.45	530	50
n-Hexane	1.20	2.15	4.96	7	16	1.09	1.760	2.53	13	50
t-2-Hexene/2-methyl-2-pentene	0.96	1.79	4.11	27	16	0.34	0.580	0.89	76	50
c-2-Hexene	0.02	0.02	0.02	537	16	0.02	0.020	0.02	1110	50
Methylcyclopentane/*t*-3-methyl-2-pentene	0.82	2.34	7.32	68	16	1.41	2.525	3.89	21	50
Cyclohexane	0.02	0.58	1.72	198	16	0.02	0.140	0.24	403	50
2-Methylhexane	0.02	0.57	1.84	193	16	0.70	1.100	1.64	15	50
2,3-Dimethylpentane	0.02	0.02	0.02	540	16	0.41	0.820	1.76	55	50
1-Heptene	0.02	0.64	1.60	238	16	0.02	0.200	0.44	454	50
n-Heptane	0.61	1.10	2.80	51	16	0.82	1.310	1.92	16	50
Methylcyclohexane	0.56	1.12	3.39	70	16	1.00	1.680	2.49	19	50
2,4/2,5-Dimethylhexane	0.02	0.02	0.56	429	16	0.19	0.420	0.70	120	50
2,3,4-trimethylheptane	0.03	0.03	0.03	518	16	0.03	0.170	0.35	394	50
2-Methylheptane	0.02	0.37	1.26	293	16	0.42	0.750	1.21	18	50
3-Methylheptane	0.02	0.02	0.79	346	16	0.37	0.610	0.98	20	50
1-Octene	0.02	0.02	0.55	369	16	0.23	0.370	0.55	86	50
n-Octane	0.61	0.89	1.96	28	16	0.79	1.170	1.73	15	50
Ethylcyclohexane	0.02	0.02	0.64	360	16	0.37	0.600	0.87	37	50
Nonene	0.03	0.03	0.72	378	16	0.46	0.670	0.98	22	50
Nonane	1.10	1.68	3.87	6	16	1.73	2.410	3.40	17	50
n-Decane	0.70	1.28	3.02	19	16	0.99	1.570	2.32	32	50
Undecane	0.83	1.34	2.55	55	16	0.19	0.510	1.02	254	50

[a] Number of measurements below detection limit.

ppbv

$$\sigma_{ij} = 0.01 + 0.15 \times \max\{x_{ij},\, y_{ij}\}, \qquad (5)$$

that is, there is a fixed contribution associated to the detection limit of 0.01 ppbv plus a percentage of the observed data (15%) associated with the uncertainty in the analytical determination by the measurement technique (cf. Section 3.1). For a large value (a possible contamination-type outlier) a large standard deviation is obtained, thus a contaminated value never gets a large

weight. For small observed values the choice given by Eq. (5) takes the larger of y_{ij} (fitted value) and x_{ij} (observed value) as the basis for standard deviation estimates. In addition the robust mode was chosen (Huber, 1981), so that extreme values (and outliers) could be handled to reduce their influence on the objective function Q in Eq. (3).

There is rotational ambiguity in the results of all two-way factor analytic programs, including PMF. One means for choosing between different possible solutions is "FPEAK", a "peaking parameter" that can be selected in the optional parameters for PMF runs. By setting a positive value to FPEAK (0.1, 0.2, …, 1.0, say) one forces the routine to search for such solutions that have extreme values among the F factor values, so there are "peaks" on the F side, that is, source profiles with "sparse" compositions. Likewise, the choice FPEAK <0.0 generates "peaks" on the G side (Paatero et al., 2002), that is, cases when few sources dominate the ambient mass. A large value (positive or negative) of FPEAK leads to worsening of the fit: additional rotations are then only possible if the components in the factors change (in order to avoid violating the non-negativity constraints) and this makes the fit worse. Hence Q increases and this can be checked to accept or reject a given FPEAK value (Paatero, 2003).

3.3. Details on UNMIX implementation

The UNMIX version used for the present work was a stand-alone executable (EPA Unmix 2.3). The data file is constructed so that the BDL values were set at half the

analytic detection limit for each species—the same was done for the PMF input files. The most important decision made by the user is the selection of species to be used in the model (Henry, 2001). The results given in this work were derived by the approach that consists of (1) generating X–Y scatter plots of all the major species against total VOC mass to identify those compounds with "good edges", that is, candidates that fulfill UNMIX assumptions; (2) using the latter species in UNMIX to generate a "minimal" solution whose diagnostic indicators are acceptable; and (3) adding additional species to explore whether any of the additional species can lead to a better solution with a larger number of sources. Fig. 1 shows a panel of the scatterplots referred to in step (1). These can be generated within UNMIX itself. Those species exhibiting well-defined upper edges include benzene, i-pentane, methylcyclopentane, nonane and n-decane, among others. These species were used to generate the initial solution, that was subsequently refined to add in as many sources and chemical species as possible. The model requires that the sources that enter into the solution must explain at least 80% of the variance of each species in the source profiles, and that the signal to noise ratio is larger than 2.0 for each source contribution.

4. Results for the downtown site POH

The campaign was conducted between 25 November and 11 December 1996 (17 days). Meterology was
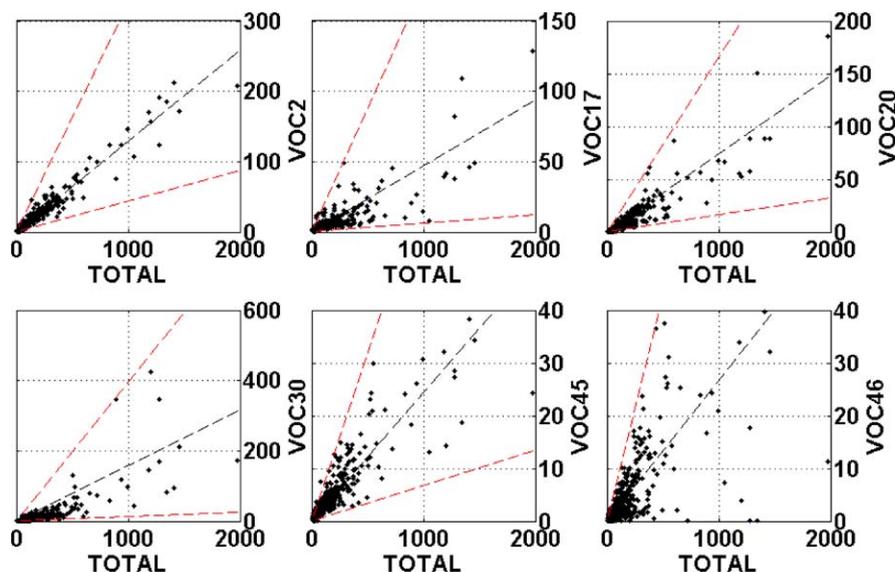

Fig. 1. Scatterplots of selected VOC species versus total VOC measured at POH site (all in µg m^{-3}). Compounds: VOC1: benzene; VOC17: n-butane/1-butene, VOC20: i-pentane, VOC30: methylcyclopentane/t-3-methyl-2-pentene, VOC45: nonane and VOC46: n-decane.

typical of springtime conditions in Santiago, mostly warm and dry. Some of the samples could not be determined, yielding a total of 679 data containing 47 VOC; cumene, *i*-butylbenzene, and *p*-cymene were discarded because they presented so many BDL values and small concentrations—see Table 1—so they fell into the "weak variable" category as defined by Paatero and Hopke (2003); hence the final data matrix analyzed had dimension 679×44.

Fig. 2 shows box and whisker plots of the diurnal profiles of selected species and the total mass of VOC sampled at the POH site. It can be seen that the total mass of VOC has a strong morning peak, indicating that mobile sources dominate impacts at that site. Likewise, many of the measured species display that behavior; in particular toluene (VOC2), *n*-butane/1-butene (VOC17), *i*-pentane (VOC20), and *n*-decane (VOC46), among others. Even isoprene (VOC23) shows a morning rise—ascribed to contributions from *t*-2-pentene and

Fig. 2. Box and whisker plots of selected species and total VOC measured at POH site (all in $\mu g\,m^{-3}$). The compounds displayed are total VOC, toluene (VOC2), *n*-butane/1-butene (VOC17), *i*-pentane (VOC20), isoprene/*t*-2-pentene/*c*-2-pentene (VOC23), and *n*-decane (VOC46) . The black and white box denotes the first and third quartile of the data (lower and upper edges, respectively), the inner boundary between black and white regions denotes the median; square brackets stand for median value plus or minus 1.5 times the box length (inter quartile range); outliers are denoted by single lines.

c-2-pentene in traffic emissions—and then displays a distinctive afternoon peak that all other VOC species (except limonene) do not exhibit at all.

4.1. UNMIX results

UNMIX found a four-source solution including 22 VOC, 626 valid samples, a minimum R^2 of 0.84 for the species in the solution and a minimum Signal to Noise ratio of 4.1 for the sources resolved. Several attempts to include more sources or compounds into the UNMIX solution failed to obtain feasible solutions.

Table 2 shows the VOC source profiles resolved by UNMIX; notice that all negative compositions are not significant ($p > 0.05$). Table 2 also includes the average mass contribution associated to each source; notice that UNMIX produces source contributions that add up to 100% of the observed concentrations. The first source profile (F1) contains benzene, toluene, xylenes, *i*-pentane, *n*-pentane, cyclopentane, methylcyclopentane, *n*-heptane and methylcyclohexane, and its contribution to total VOC mass is 18%. Fig. 3 shows the diurnal profile for the corresponding source contribution (denoted as G1); a strong peak associated to the early morning traffic rush is clear for this factor—the correlation coefficient with ambient CO measurements is $r = 0.18$—thus we identify the source as gasoline exhaust emissions. The second source profile contains significant amounts of isoprene. However, there are also contributions of *n*-butane, *n*-pentane and *i*-pentane in this profile; these are likely coming from evaporative sources like LPG and gasoline vapor leaks that also increase as ambient temperature rises; hence we identify this source as a mix of biogenic and evaporative emissions. The total contribution to ambient VOC mass of this mixed source is about 17%. Fig. 3 shows the diurnal profile of the corresponding source contribution (denoted as G2). The temperature modulation is clear on the profile, so its dependence with ambient temperature must be strong, and this is consistent with what is known about biogenic emissions. In fact, the isoprene diurnal profile (see Fig. 2) is similar to the diurnal profile of G2. The third source profile in Table 2 contributes approximately 32% to VOC mass and it is enriched with *i*-pentane, *n*-butane, *n*-pentane, benzene, etc. that indicate this source corresponds to fuel evaporation. Fig. 3 shows a diurnal plot for the source contribution of this factor (G3), so there is a strong indication that this source corresponds to traffic evaporative emisssions (running losses, cold start VOC emissions) mixed in with gasoline vapor—the correlation coefficient between G3 and ambient CO is high, $r = 0.226$. The low values for G3 in the afternoon are ascribed to better ventilation conditions and lower traffic flows, the *i*-pentane diurnal profile (see Fig. 2) is very similar to the one for G3. The fourth source profile is the one that contains most
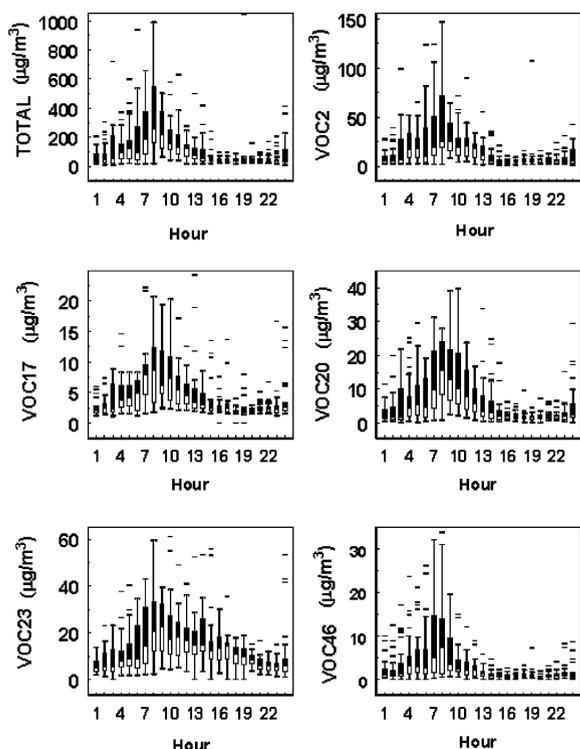
Table 2
Results of UNMIX[a] and PMF (FPEAK = 0.0) for POH site[b]

Compound	UNMIX				PMF			
	F1	F2	F3	F4	F1	F2	F3	F4
Benzene	0.025	(0.006)	0.061	0.045	0.071	0.023	0.049	0.051
Toluene	0.118	0.111	0.109	0.161	0.084	0.088	0.183	0.195
Ethylbenzene	0.010	0.009	0.014	0.024	0.017	0.007	0.015	0.033
p-Xylene	0.011	0.009	0.017	0.026	0.018	0.008	0.014	0.037
m-Xylene	0.032	(0.017)	0.043	0.067	0.045	0.012	0.037	0.109
o-Xylene	0.020	0.030	0.022	0.036	0.023	0.028	0.019	0.059
Limonene					(0.00)	0.017	0.001	(0.00)
n-Propylbenzene/dodecane					0.008	(0.00)	0.000	0.005
p/m-Ethyltoluene					0.032	(0.00)	0.004	0.046
Mesitylene					0.009	(0.00)	0.002	0.012
Styrene					0.010	(0.00)	0.001	0.009
o-Ethyltoluene					(0.00)	0.002	0.002	0.008
1,2,4-Trimethylbenzene	0.015	0.014	0.017	0.034	0.029	0.005	0.011	0.056
n-Butane/1-butene	0.028	0.116	0.044	(0.004)	(0.00)	0.080	0.052	0.042
t-2-Butene					0.001	(0.00)	(0.00)	0.001
c-2-Butene					(0.00)	0.000	0.000	0.000
i-Pentane	0.040	0.069	0.107	0.027	0.055	0.052	0.076	0.016
1-Pentene					(0.00)	0.010	(0.00)	0.000
n-Pentane/2-methyl-1-butene	0.027	0.041	0.068	0.035	0.059	0.033	0.049	0.022
t-2-Pentene/isoprene/c-2-pentene	(0.036)	0.325	0.067	0.075	0.137	0.447	(0.00)	(0.001)
Cyclopentane/2,3-dimethylbutane/2-methylpentane	0.058	(−0.010)	0.053	0.024	0.067	0.001	0.065	(0.00)
3-Methylpentane	0.012	(−0.001)	0.026	0.017	0.033	(0.00)	0.027	0.003
1-Hexene					0.010	(0.00)	(0.00)	(0.00)
n-Hexane	0.018	0.031	0.045	0.022	0.034	0.047	0.055	0.003
t-2-Hexene/2-methyl-2-pentene					0.006	0.045	0.050	(0.00)
c-2-Hexene					(0.00)	0.001	(0.00)	(0.00)
Methylcyclopentane/t-3-methyl-2-pentene	0.314	(−0.023)	0.036	0.024	0.003	0.001	0.134	0.004
Cyclohexane					0.018	(0.00)	0.021	(0.00)
2-Methylhexane	0.010	(−0.005)	0.026	0.011	0.032	(0.00)	0.016	(0.00)
2,3-Dimethylpentane					0.001	(0.00)	(0.00)	0.001
1-Heptene					0.021	0.001	0.010	(0.00)
n-Heptane	0.013	(0.005)	0.029	0.015	0.027	0.018	0.029	(0.001)
Methylcyclohexane	0.017	(−0.001)	0.032	0.024	0.032	0.002	0.036	0.020
2,4/2,5-Dimethylhexane					0.013	(0.00)	0.001	(0.00)
2,3,4-Trimethylheptane					0.006	(0.00)	(0.00)	(0.00)
2-Methylheptane					0.027	(0.00)	0.005	0.003
3-Methylheptane					0.020	(0.00)	0.003	0.001
1-Octene					0.004	0.005	(0.00)	(0.00)
n-Octane					0.016	0.018	0.011	0.018
Ethylcyclohexane					0.009	(0.00)	0.003	0.004
Nonene					0.007	0.002	(0.00)	0.003
Nonane	0.020	0.032	(0.007)	0.049	0.011	0.020	0.019	0.079
n-Decane	0.015	0.035	(−0.006)	0.063	0.005	0.002	(0.00)	0.103
Undecane	(0.007)	0.053	(−0.006)	0.029	(0.00)	0.025	0.001	0.056
% of mass contribution (standard error)	18(3.2)	17(3.5)	32(4.7)	32(5.1)	22(3.4)	18(3.4)	30(5.6)	18(2.9)
Correlation coefficient with ambient CO[c]	0.167	0.059	0.226	0.530	0.400	0.139	0.163	0.568

[a] Empty entries mean that the species did not enter into UNMIX's computations.
[b] Composition estimates that are not significant ($p > 0.05$) are enclosed in parenthesis.
[c] The ambient CO is monitored following standard EPA protocols by SESMA (2004).

toluene, xylenes, 1,2,4-trimethylbenzene, nonane, n-decane and undecane, contributing to 32% of the ambient VOC mass. Fig. 3 displays a diurnal profile for

the source contributions (G4); because of the shape of this source contribution, the very high correlation coefficient with ambient CO ($r = 0.53$) and the relative
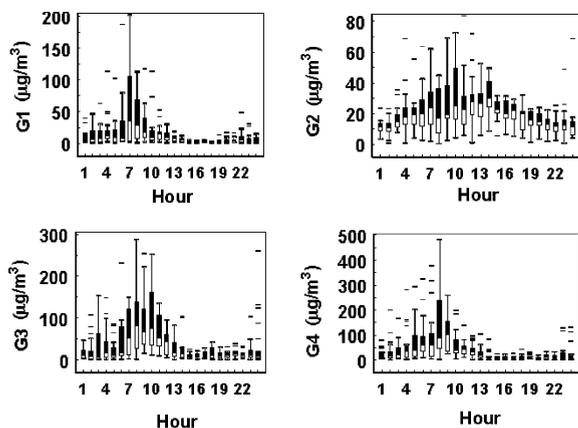
Fig. 3. Box and whisker plots of diurnal profiles for the four source contributions (in μg m⁻³) obtained using UNMIX for the POH data set. The plotting conventions are the same as in Fig. 2.

Fig. 4. Box and whisker plots of diurnal profiles for the four source contributions (in μg m⁻³) obtained by PMF (FPEAK = 0.0) for the POH data set. The plotting conventions are the same as in Fig. 2.

enrichment in high molecular weight compounds, we identify this source as diesel exhaust. For example, the *n*-decane diurnal profile (see Fig. 2) is similar to the G4 profile. The estimated errors in these source contributions vary from 15% for gasoline exhaust emissions through 21% for the evaporative emissions, so the latter ones have more variability, as expected for this kind of diffuse source.

In summary, source contributions G1, G3 and G4 display similar diurnal patterns, as many of the VOC that contribute to those three sources, which are all coming essentially from mobile sources. By contrast, source G2 is dominated by biogenic emissions that peak in the early afternoon. The sharp fall of ambient concentrations in the afternoon is associated with better ventilation conditions.

4.2. PMF results

Table 2 also shows PMF results for the POH data set, and for the parameter FPEAK set to 0.0, when the number of factors is chosen as four—this particular choice of FPEAK is justified below. In addition, the source profiles are listed in the same order as in the UNMIX solution, to facilitate comparisons.

The first source profile is rich on benzene, toluene, cyclopentane, 3-methylpentane, cyclohexane, 2-methyl-hexane, 1-heptene, methylcyclohexane, and 2-methyl-heptane, among other VOC. This profile corresponds to gasoline exhaust emissions and its contribution to ambient VOC mass is of 22%. The diurnal profile of this source contribution plotted in Fig. 4 is quite similar to the first source profile in UNMIX solution (see Fig. 3), but its correlation coefficient with ambient CO is higher ($r = 0.40$). The second source profile contains limonene and isoprene and also contributions from toluene, *n*-
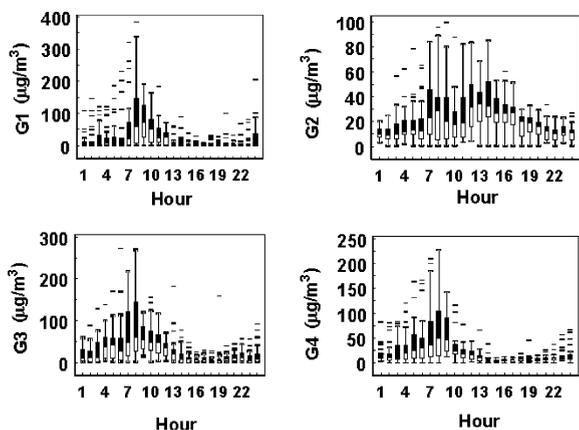
butane, *i*-pentane, *n*-pentane, *n*-hexane, that is, evapora-tive emissions likely to come from anthropogenic sources but, because of their high depedendence upon ambient temperature, appear mixed in with the biogenic emissions. The diurnal profile of this source contribu-tion, plotted in Fig. 4, looks quite similar to the second UNMIX source profile—its correlation coefficient with ambient CO is the lowest ($r = 0.14$) among the four sources, and its average contribution to VOC mass is 18%. The third source profile is rich on toluene, *n*-butane, *i*-pentane, *n*-pentane, cyclopentane and *n*-hexane, among other species. Thus, it corresponds to evaporative emissions from the fuel distribution cycle, and it contributes aproximately 30% to ambient mass. The diurnal profile, plotted also in Fig. 4, looks similar to the corresponding evaporative factor in UNMIX's solution—the correlation coefficient between this source contribution and ambient CO is $r = 0.16$. Finally, the fourth source profile is rich on toluene, xylenes, mesitylene, styrene, 1,2,4 trimethylbenzene, nonane, *n*-decane and undecane. Because of the shape of the source contribution profile and its very high correlation with ambient CO ($r = 0.57$), we identify it as diesel exhaust, that is, similar to the fourth source profile in UNMIX's solution. This source contributes approximately 18% to the total ambient mass. Hence, the four sources found in the PMF solution with FPEAK = 0.0 are identified to correspond to the same four sources resolved by UNMIX. However, the total contributions to ambient VOC mass differ between those two methods. UNMIX estimates mass contributions by multiple linear regres-sion of the identified, acceptable source profiles against total VOC mass, so the sum of source constributions is 100%; by contrast, PMF does not make that assumption (see Section 2.1) and the total percentage of explained mass is 88%, leaving 12% of unidentified mass.

Regarding source contribution uncertainties, these do not change significantly when the FPEAK parameter is modified in PMF, and the values range between 15% for the gasoline exhaust source through 20% for the mixed biogenic and evaporative source. A similar range of uncertainty is obtained in UNMIX solution (see Table 2).

4.3. Comparison of PMF and UNMIX solutions for POH data set

To explore the range of possible PMF solutions, the parameter FPEAK was varied and the corresponding Q values of the fit analyzed, along with the mass apportionment and source profiles obtained for each solution. Inspection of the Q versus FPEAK values indicates that there is a minimum near FPEAK = 0.0 and that there is little change in the Q objective function between FPEAK = −0.15 and 0.4. It was also found that for FPEAK ⩾ 0.5 the Q values start to increase rapidly and so those solutions are not likely to be good candidates. It was not possible to obtain physically meaningful solutions for FPEAK values < −0.2.

In Fig. 5 the mass apportionment is presented for UNMIX and all PMF solutions computed for the POH data. There is a continuous change in source contributions when the FPEAK parameter is modified, and the major change is when FPEAK crosses the zero value, when gasoline exhaust contributions decrease and "biogenic" contributions increase, while diesel exhaust and evaporative contributions do not change significantly.

In order to better understand the trend in the solutions, in Fig. 6 we have plotted the mass fractions of selected species, for the four factors analyzed at POH site. The top panel in Fig. 6 displays some species in the "biogenic" and evaporative emissions. The isoprene mass fraction in the biogenic source profile is quite

steady near 0.45 for all values of FPEAK. The mass fraction of *i*-pentane in evaporative emissions is also fairly constant near 0.07 for most of the PMF solutions; this same, almost constant behavior is observed for limonene in the biogenic profile and *n*-butane in the evaporation profile. UNMIX results for toluene and *i*-pentane are, respectively, lower and higher than the values found in PMF solutions; this is due to UNMIX's solution mass normalization with only 22 VOC species.

The middle panel in Fig. 6 displays selected species in the gasoline exhaust profile. Most of UNMIX and PMF solutions with negative FPEAK values have species compositions on the lower level of reported values, while PMF solutions with FPEAK ⩾ 0 have similar compositions, that are closer to average source profiles reported in the literature (Mugica et al., 1998; Vega et al., 2000; Watson et al., 2001). The lower panel in Fig. 6 shows selected species in the diesel exhaust source profile. UNMIX and negative FPEAK PMF solutions display lower compositions than the solutions with FPEAK ⩾ 0; it can be seen that UNMIX and PMF results with negative FPEAK values predict too high isoprene fractions in the diesel exhaust profile; this unfeasible result is corrected when FPEAK increases from zero to higher values.

The combination of these results suggest that the profiles resolved by PMF with values of FPEAK ⩾ 0 produce the most credible profiles, and that they exhibit better source profiles than the ones produced by the standard UNMIX solution; in order to pick one of those solutions, we have chosen the solution with the lowest Q values—that is, the better relative fit—and this explains the choice of FPEAK = 0.0 in Fig. 4 and Table 2. For instance, the PMF solution for the solvent source profile does not contain significant isoprene in that source profile, while UNMIX's solution does. In addition, PMF can incorporate VOC with low ambient concentrations or many BDL samples (like styrene, limonene, etc.) and these compounds allow us to better identify the source profiles. Nonetheless, UNMIX is good at picking the number of factors to be used in the analysis, a useful feature in receptor analysis.

We attempted to include more sources in the analysis, so we run PMF with five, six and seven sources and varying parameter FPEAK. However, we could not find additional credible solutions. We conclude that the ambient VOC at that site is so dominated by mobile sources that it is quite difficult to identify other sources, except biogenic emissions that have a distinctive daily emission profile.
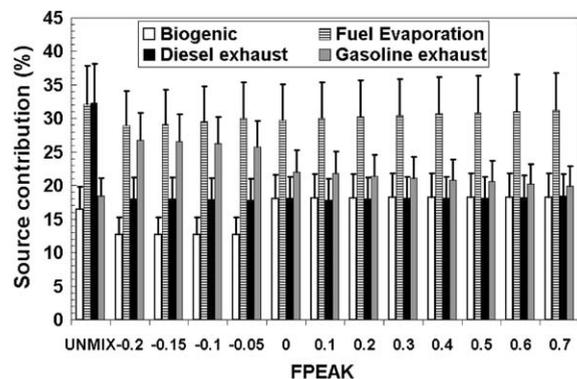

Fig. 5. Comparison of source contributions at POH, for the four source solutions given by UNMIX and PMF (several FPEAK values); the standard errors in each source contribution are also displayed.

5. Results for the residential site LAC

The campaign was conducted between 13 November and 11 December 1996 (29 days). Some of the samples
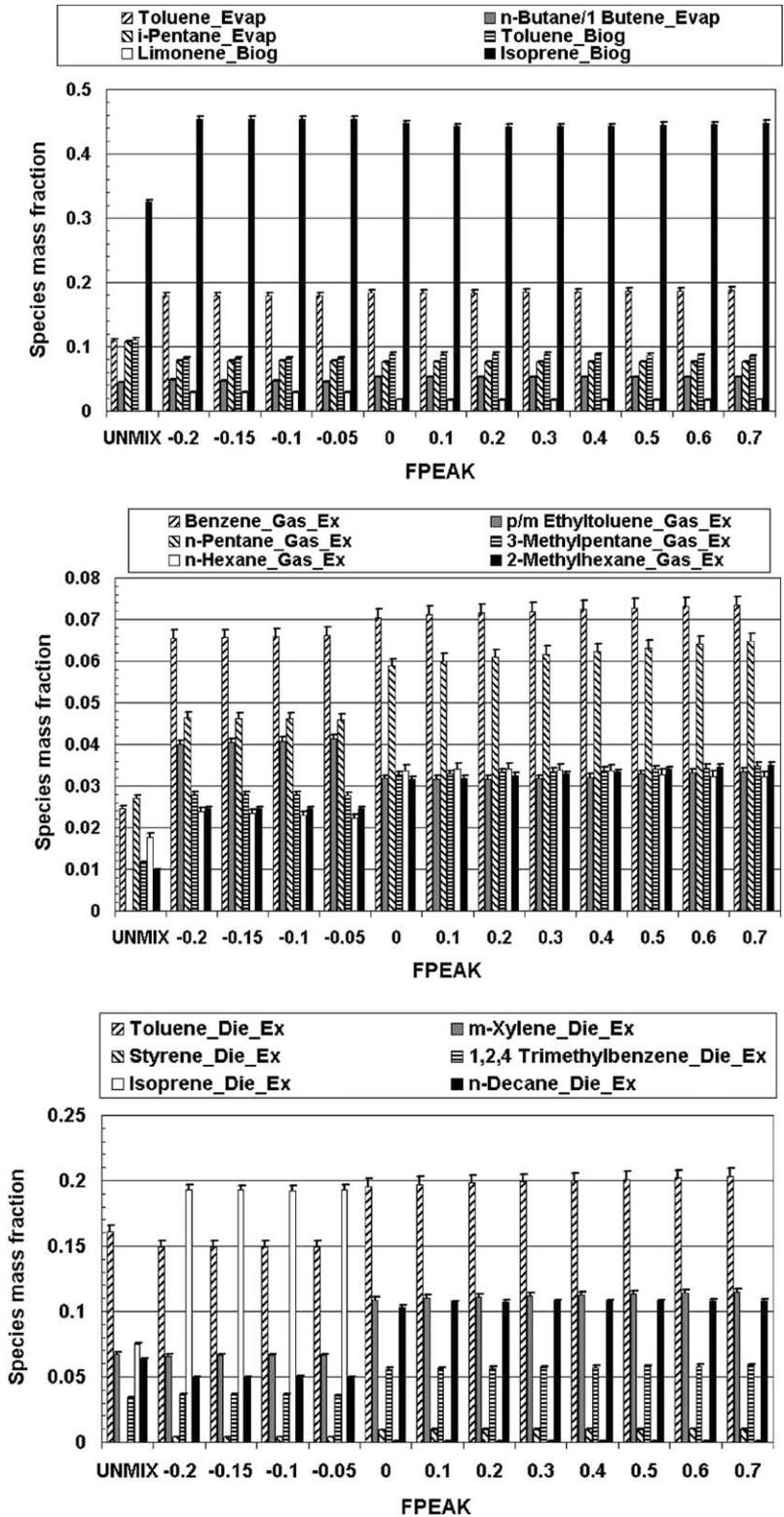
Fig. 6. Species mass fractions for the evaporative and biogenic sources in the four source solutions given by UNMIX and PMF (several FPEAK values). Top panel: evaporative and biogenic profiles; middle panel: gasoline exhaust profile; bottom panel: diesel exhaust profile. The standard errors in each species composition are also shown.

could not be determined, yielding a total of 1198 data containing 44 VOC; cumene, *i*-butylbenzene, *p*-cymene and *c*-2-hexene were discarded because they presented too many BDL values and small concentrations (see Table 1), and they fell into the "weak variable" category as defined by Paatero and Hopke (2003); hence the final data matrix analyzed had dimension 1198 × 40.

Fig. 7 shows box and whisker plots of the diurnal profiles of selected species and the total mass of VOC sampled at the LAC site. The shape of these plots indicate that there is an initial peak of traffic impacts near 7–8 a.m. and a major peak near noon, that is, this site is a receptor site and the plume of the city takes a few hours to reach that monitor site—notice that upwind emissions must be quite larger than the local ones to produce a similar peak at noon, when ventilation conditions are much better than in the morning. In addition, evening emissions linger on that site because of mountain valley circulation, so the concentrations do not decrease overnight as at the POH site (see Fig. 2). There is greater variability among the species diurnal
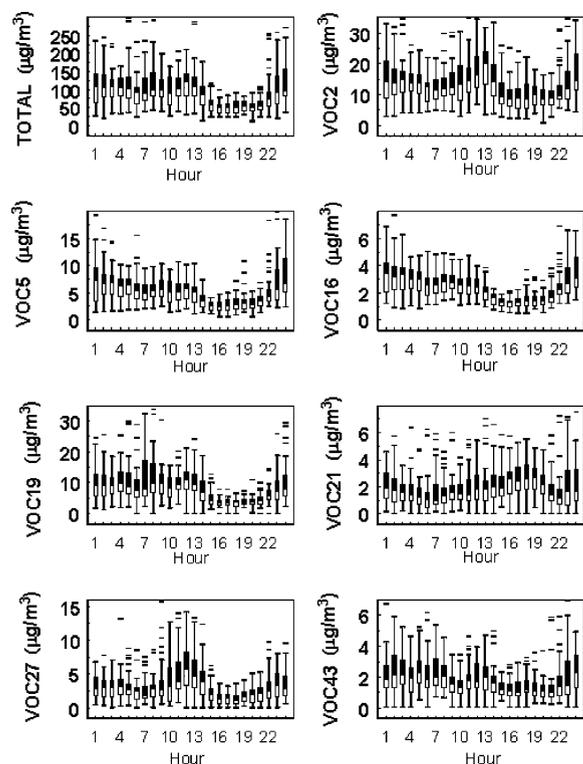

Fig. 7. Box and whisker plots of selected species and total VOC measured at LAC site (all in $\mu g\,m^{-3}$). The compounds displayed are total VOC, toluene (VOC2), *m*-xylene (VOC5), 1,2,4 trimethylbenzene (VOC16), *i*-pentane (VOC19), isoprene/*t*-2-pentene/*c*-2-pentene (VOC21), methylcyclopentane/*t*-3-methyl-2-pentene (VOC27) and *n*-decane (VOC43). The plotting conventions are the same as in Fig. 2.

profiles: toluene (VOC2) and methylcyclopentane (VOC27) have a distinctive peak at noon; isoprene (VOC21) has a late afternoon peak; *m*-xylene (VOC5) and 1,2,4 trimethylbenzene (VOC16) suggest a rather steady emission profile, perhaps increasing in the afternoon; *i*-pentane (VOC19) and *n*-decane (VOC43) display profiles that suggest more than one kind of source explaining their diurnal variation. All species have a midnight accumulation maximum related to mountain to valley downslope flow that brings back pollution to the eastern side of the city.

5.1. UNMIX results

UNMIX found a solution with 26 VOC, 1112 valid samples, a minimum $R^2 = 0.80$ for the species entering the final solution and a minimum signal to noise ratio of 3.45 for the five sources resolved. Table 3 summarizes the source profiles obtained and Fig. 8 shows box and whisker plots of the corresponding source contributions. The first source profile is rich on isoprene and on fuel evaporation species like toluene, *m*-xylene, *n*-butane and *c*-2-butene. Thus it corresponds to a mixture of biogenic emissions and fugitive emissions, that accounts for 15% of the observed mass; the corresponding diurnal profile has a clear maximum in the afternoon, and the correlation coefficient of this source contribution with ambient CO is negligible ($r = 0.015$). The second source profile is enriched with benzene, toluene, xylenes, *i*-pentane, *n*-pentane, methylcyclopentane, 3-methylpentane and *n*-hexane, among other VOC. This source profile corresponds to gasoline vehicles, and it contributes with 24% of the observed VOC mass. This second source peaks near noon and so it corresponds to motor vehicle emissions being transported from upwind sources in the city—in springtime the dominant diurnal winds are southwesterly (Rappenglück et al., 2000)—the correlation coefficient of this "carry-over" source contribution with ambient CO is $r = 0.147$. The third source profile is enriched in toluene, xylenes, 1,2,4 trimethylbenzene, nonane, and *n*-decane. The diurnal profile of the corresponding source contribution indicates a strong correlation with ambient CO measurements ($r = 0.292$). Hence, this source is identified as diesel exhaust emissions, and it contributes about 9% of the total VOC at LAC site. The fourth source profile is highly enriched with *i*-pentane, *n*-pentane, *n*-butane, so it corresponds to fuel evaporation sources coming mostly from LPG losses mixed with gasoline vapor. The shape of the diurnal profile indicates a source that has a peak in the morning—coincident with cooking activities in the city—a secondary peak near noon—carry over emissions from upwind in the city—and an accumulation late in the evening, with a weaker correlation with ambient CO ($r = 0.096$). This source explains 16% of the total ambient concentrations at the

Table 3
Results of UNMIX[a] for LAC site[b]

Compound	F1	F2	F3	F4	F5
Benzene	0.035	0.063	(0.017)	0.018	0.066
Toluene	0.161	0.194	0.117	(−0.036)	0.166
Ethylbenzene	0.022	0.023	0.033	(−0.008)	0.028
p-Xylene	0.027	0.023	0.036	(−0.010)	0.031
m-Xylene	0.056	0.047	0.110	(−0.019)	0.079
o-Xylene	0.041	0.030	0.065	(−0.015)	0.047
Limonene					
n-Propylbenzene/dodecane					
p/m-Ethyltoluene	0.020	0.021	0.031	(−0.008)	0.034
Mesitylene	0.010	0.006	0.021	(−0.004)	0.016
Styrene	0.005	0.005	0.010	(−0.002)	0.008
o-Ethyltoluene					
1,2,4-Trimethylbenzene	0.027	0.019	0.042	(−0.011)	0.039
n-Butane/1-butene/t-2-butene	0.122	0.076	(0.030)	0.359	0.093
c-2-Butene	0.109	(0.009)	(0.000)	0.050	0.011
i-Pentane	(−0.027)	0.056	(0.034)	0.435	0.051
n-Pentane/2-methyl-1-butene	0.033	0.032	(0.015)	0.179	0.031
t-2-Pentene/isoprene/c-2-pentene	0.111	(−0.001)	(0.001)	(0.014)	0.013
3-Methylpentane	(0.007)	0.035	(0.000)	0.029	0.014
1-Hexene					
n-Hexane	0.019	0.028	(0.006)	0.022	0.021
t-2-Hexene/2-methyl-2-pentene					
Methylcyclopentane/t-3-methyl-2-pentene	(0.005)	0.110	(0.001)	0.018	0.010
Cyclohexane					
2-Methylhexane					
2,3-Dimethylpentane					
1-Heptene					
n-Heptane	0.016	0.022	(0.009)	(−0.001)	0.020
Methylcyclohexane	0.016	0.027	0.019	(−0.002)	0.027
2,4/2,5-Dimethylhexane	0.001	0.008	0.005	(0.001)	0.008
2,3,4-Trimethylheptane					
2-Methylheptane					
3-Methylheptane					
1-Octene					
n-Octane	0.014	0.017	0.016	(−0.005)	0.021
Ethylcyclohexane					
Nonene					
Nonane	0.037	0.026	0.065	−0.015	0.035
n-Decane	0.022	0.002	0.125	−0.010	0.017
Undecane	(0.000)	(−0.001)	0.092	(−0.001)	(0.000)
% of mass contribution (standard error)	14.5 (3.9)	23.7 (4.3)	9.1 (3.0)	16.3 (2.8)	36.4 (3.4)
Correlation coefficient with ambient CO[c]	0.015	0.147	0.292	0.096	0.453

[a] Empty entries mean that the species did not enter into UNMIX's computations.

[b] Composition estimates that are not significant ($p > 0.05$) are enclosed in parenthesis.

[c] The ambient CO is monitored following standard EPA protocols by SESMA (2004).

residential LAC site. Finally, the fifth source profile is the most strongly correlated with ambient CO ($r = 0.453$) and it has several species compositions similar to the ones in the second source profile— compare compositions of benzene, ethlybenzene, i-pentane, n-pentane, n-heptane, and methylcyclohexane in Table 3. This source accounts for 36% of the ambient VOC mass, hence we identify it as local gasoline exhaust emissions. The estimated uncertainties in source contributions range from 14% in the gasoline exhaust sources through 24% for the diesel exhaust, that is, the dominant sources are estimated more accurately than the smaller sources.

In summary, diurnal profiles of the five identified VOC sources reflect activity patterns that coincide with the identification made based upon their chemical
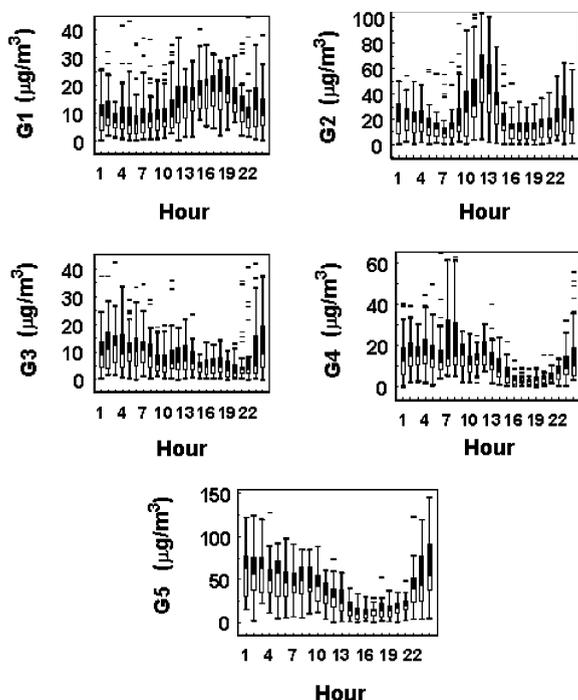
Fig. 8. Box and whisker plots of diurnal profiles for the five source contributions ($\mu g\,m^{-3}$) resolved by UNMIX for the LAC data set. The plotting conventions are the same as in Fig. 2.

composition. However, due to transport phenomena these emission profiles are not necessarily restricted to local activities, and here the UNMIX model was able to discriminate between local and upwind gasoline exhaust emissions impacting this residential site, because they have different temporal patterns of source contributions.

Therefore, the identified source contributions at LAC also show an area heavily dominated by mobile sources. The typical errors in the source contributions estimates vary from 9% for the local gasoline exhaust source through 33% for diesel exhaust emissions, that is, the smaller source contributions are known with higher error and vice versa.

5.2. PMF results

We first tried running PMF with a total of five sources, and we obtained similar results as the UNMIX solution, both for source profiles and source contributions. In order to try to obtain less mixed source profiles, we increased the number of sources and run PMF with different FPEAK values. Table 4 below displays the six source profiles obtained by setting FPEAK = 0.0, and Fig. 9 displays box and whisker plots of the six source contributions daily profiles. The first source profile

contains isoprene and limonene, and significant amounts of toluene, xylenes and nonane that come from fugitive (evaporative) emissions. Nonetheless, neither n-butane, i-pentane nor n-pentane contribute at all to this source profile, indicating a "cleaner" profile resolved by the PMF method. This factor accounts for 7.7% of the ambient VOC concentration. Notice also in Fig. 9 that the source contribution has a distinctive shape with a clear afternoon peak, no signal of morning emissions and clearly this source shuts down overnight. The second source profile contains benzene, toluene, n-butane, i-pentane, n-pentane, 3-methylpentane and n-hexane, with no enrichment in compounds like n-decane, and highly correlated with ambient CO ($r = 0.284$) so it is identified as transported gasoline exhaust emissions, and this source contributes with 31% of the mass of ambient VOC. The third source profile is rich in xylenes, toluene, nonane, n-decane and undecane that also has a high correlation with ambient CO ($r = 0.238$), so we identify it as diesel exhaust. Its daily source contribution profile shows a small peak near noon, coming from traffic sources upwind in the city, and then a midnight accumulation likely coming from local sources being recirculated by the wind. This third source accounts for 6% of the ambient VOC concentrations. The fourth source profile is enriched with benzene, i-pentane, n-pentane, 3-methylpentane, isoprene and n-butane, so it is identified as evaporative emissions dominated by gasoline vapor; this source has a weaker correlation with ambient CO ($r = 0.146$). The daily profile of the corresponding source contribution has a clear morning contribution followed by a late evening accumulation peak. This source accounts for 11.7% of the ambient VOC mass. The fifth source profile is the one with the highest correlation with ambient CO ($r = 0.393$), and its source profile is similar to the second source profile identified by PMF—see for instance benzene, ethylbenzene, p-xylene, i-pentane, n-pentane, 2-methylheptane, and 3-methylheptane compositions in Table 4—hence it is identified as local gasoline exhaust emissions that account for 25% of the VOC mass. Finally, the sixth source profile is enriched in n-butane, c-2-butene, i-pentane, n-pentane and so it corresponds to LPG losses mixed with fuel evaporation; this source accounts for 11% of the total VOC. Note in Fig. 9 that this source has a clear peak near at noon, so its major contributions come from sources upwind in the city; this source tends to shut down overnight, because it does not have a midnight peak; the correlation coefficient with ambient CO is the lowest among this six source solution ($r = 0.043$). The errors in the estimated source contributions vary from 14% for gasoline exhaust emissions through 23% and 24% for the "biogenic" and diesel exhaust emissions, respectively. That is, dominant sources are resolved with higher uncertainty than smaller strength sources. These error estimates do not

Table 4
Results of PMF (FPEAK = 0.0) for LAC site[a]

Compound	F1	F2	F3	F4	F5	F6
Benzene	0.022	0.064	(0.004)	0.052	0.056	0.051
Toluene	0.256	0.194	0.184	0.038	0.129	0.041
Ethylbenzene	0.028	0.028	0.033	0.005	0.025	(0.000)
p-Xylene	0.038	0.026	0.031	0.005	0.029	(0.000)
m-Xylene	0.053	0.056	0.089	0.014	0.090	(0.000)
o-Xylene	0.055	0.033	0.061	0.004	0.054	0.006
Limonene	0.021	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
n-Propylbenzene/dodecane	0.009	0.006	0.008	(0.000)	0.005	0.001
p/m-Ethyltoluene	0.024	0.022	0.029	0.002	0.041	0.004
Mesitylene	0.009	0.006	0.019	0.001	0.022	(0.000)
Styrene	0.006	0.006	0.008	(0.000)	0.009	0.002
o-Ethyltoluene	0.006	0.004	0.004	0.001	0.017	(0.000)
1,2,4-Trimethylbenzene	0.035	0.019	0.039	0.004	0.053	(0.000)
n-Butane/1-butene/t-2-butene	(0.000)	0.087	(0.000)	0.257	0.137	0.232
c-2-Butene	(0.000)	(0.000)	(0.000)	(0.000)	0.024	0.196
i-Pentane	(0.005)	0.074	0.006	0.236	0.070	0.185
n-Pentane/2-methyl-1-butene	(0.000)	0.041	(0.000)	0.134	0.038	0.081
t-2-Pentene/isoprene/c-2-pentene	0.139	(0.000)	(0.000)	0.079	0.002	(0.000)
3-Methylpentane	(0.000)	0.031	(0.000)	0.037	0.001	0.036
1-Hexene	0.001	(0.000)	(0.000)	0.021	(0.000)	(0.000)
n-Hexane	0.016	0.029	0.009	0.027	0.016	0.029
t-2-Hexene/2-methyl-2-pentene	0.014	0.004	(0.000)	(0.000)	(0.000)	0.043
Methylcyclopentane/t-3-methyl-2-pentene	0.003	0.069	(0.000)	(0.000)	(0.000)	0.062
Cyclohexane	(0.000)	(0.000)	(0.000)	0.006	0.003	0.002
2-Methylhexane	0.013	0.022	0.013	0.004	0.010	0.008
2,3-Dimethylpentane	(0.000)	0.022	0.015	0.021	0.003	(0.000)
1-Heptene	0.004	0.002	0.002	0.010	(0.000)	(0.000)
n-Heptane	0.025	0.024	0.009	0.011	0.015	0.003
Methylcyclohexane	0.017	0.030	0.015	0.011	0.025	0.002
2,4/2,5-Dimethylhexane	(0.000)	0.009	0.005	0.005	0.005	(0.000)
2,3,4-Trimethylheptane	(0.000)	0.004	0.003	0.002	0.003	(0.000)
2-Methylheptane	0.006	0.014	0.007	0.005	0.012	(0.000)
3-Methylheptane	0.004	0.011	0.005	0.004	0.010	0.001
1-Octene	0.028	0.004	0.003	0.001	0.001	(0.000)
n-Octane	0.022	0.017	0.015	0.004	0.021	0.002
Ethylcyclohexane	0.010	0.009	0.008	0.001	0.009	(0.000)
Nonene	0.030	0.009	0.016	(0.000)	0.006	0.002
Nonane	0.070	0.026	0.056	(0.000)	0.040	0.007
n-Decane	0.029	0.000	0.179	(0.000)	0.021	(0.000)
Undecane	(0.000)	(0.000)	0.126	(0.000)	(0.000)	(0.000)
% of mass contribution (standard error)	7.7 (1.7)	31.2 (4.1)	6.2 (1.5)	11.7 (2.8)	25.5 (4.0)	11.0 (2.5)
Correlation coefficient with ambient CO[b]	−0.238	0.284	0.238	0.146	0.393	0.043

[a] Composition estimates that are not significant ($p > 0.05$) are enclosed in parenthesis.
[b] The ambient CO is monitored following standard EPA protocols by SESMA (2004).

change significantly when the FPEAK parameter is changed.

When we tried to increase the number of sources to seven or eight, PMF was not able to extract more meaningful source profiles, so we will discuss only the six source solution below.

Fig. 10 shows the VOC mass apportionment for the PMF results; it can be seen that biogenic emissions decrease steadily their contributions as FPEAK increases from −0.1 to 0.7, lowering the biogenic contribution from 8% to 5%; in this parameter scan, local and transported gasoline exhaust emissions present a minimum and a maximum, respectively, when FPEAK is near 0.2–0.3 while LPG losses do not change significantly and stay near 11%. By contrast, evaporative emissions are more important contributions when

FPEAK is positive, reaching 20% of the ambient mass, while they contribute only 12% when FPEAK is negative.
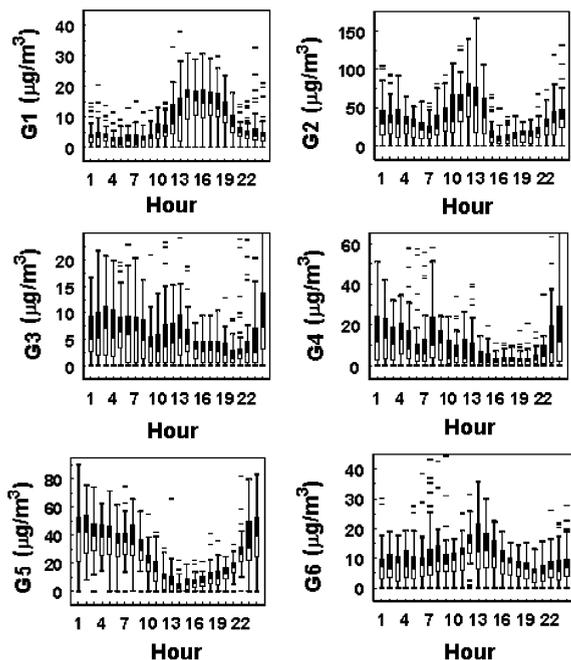

Fig. 9. Box and whisker plots of diurnal profiles for the six source contributions ($\mu g\, m^{-3}$) resolved by PMF (FPEAK = 0.0) for the LAC data set. The plotting conventions are the same as in Fig. 2.

To further explore the nature of the six source solutions with PMF, we have plotted selected species in the source profiles identified by the model. Fig. 11 (top panel) shows that as FPEAK increases, limonene and isoprene slowly increase their compositions in the biogenic source profile; at the same time, species such as xylenes and 1,2,4 trimethylbenzene reduce their compositions in this profile. However, toluene composition is not changed significantly as parameter FPEAK is varied, indicating that some fraction of the toluene emissions cannot be discriminated from the biogenic source profile, indicating a strongly correlated toluene and biogenic emission.

Fig. 11 (middle panel) displays the source compositions of several species in the evaporative emissions, and we have included also the butane and butene concentrations in the sixth source profile obtained with PMF, that is, the one identified with LPG losses. The compositions of benzene, i-pentane and n-pentane in the evaporative source profile are fairly constant across the range of FPEAK values, whereas as FPEAK increases, the toluene mass fraction decreases and the butane and butene concentration increases, but this latter value is always lower than the butane and butene concentrations in the LPG source profile. Thus, we ascribe that increase in concentrations to an artifact of the solution and not to that evaporative profile becoming like another LPG fugitive emission.

Fig. 11 (bottom panel) displays concentration of several species in the transported gasoline vehicles source profile. It can be seen that the values do not
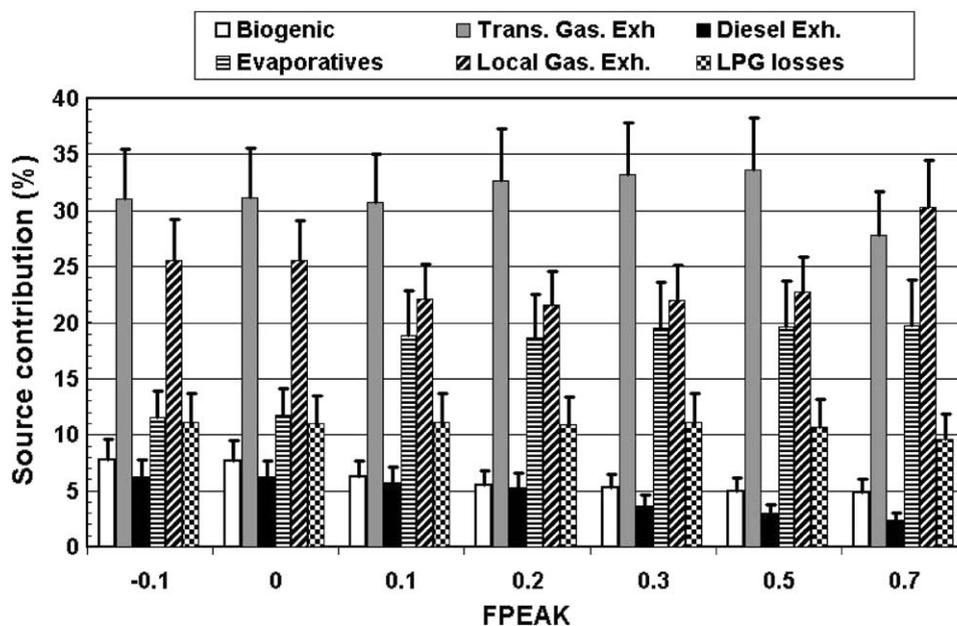

Fig. 10. Comparison of source contributions at LAC, for the six source solution given by PMF and several FPEAK values; the standard errors in each source contribution are also displayed.
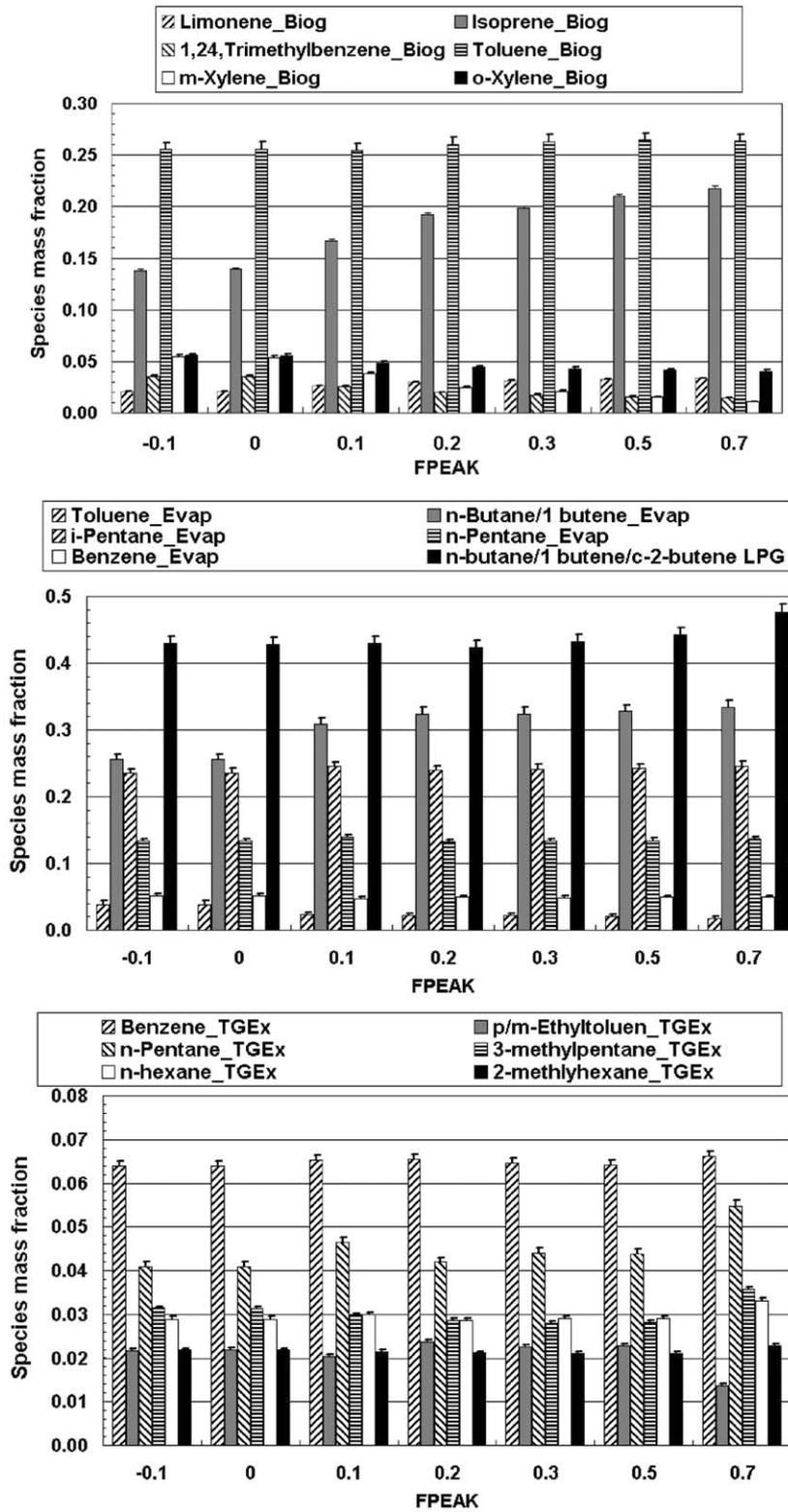
Fig. 11. Species mass fractions for the biogenic source profile in the six source solution given by PMF for several FPEAK values. Top panel: biogenic profile; middle panel: evaporative profile; bottom panel: transported gasoline exhaust profile. The standard errors in each species composition are also shown.

change significantly as the FPEAK parameter is changed—except when FPEAK reaches the value 0.7—that is, this is a very well defined source profile.

Fig. 12 (upper panel) shows the species compositons in the local gasoline exhaust source profile. It can be seen that compounds like toluene, m-xylene, styrene and 1,2,4 trimethylbenzene show compositions that do not change when parameter FPEAK is modified, thus this is a very stable source profile. In the case of the diesel exhaust profile (Fig. 12, lower panel), nonane compositions do not change much when FPEAK is changed, but

n-decane shows a steady grow as FPEAK increases, reaching 45% of mass when FPEAK = 0.7, an unlikely result that we ascribe to an artifact of the PMF solution when FPEAK is increased too much. A similar behavior is displayed by n-undecane in this source profile.

From the above results we conclude that the six source solution with FPEAK = 0.0 is a reasonable choice because it is characterized by a smaller value of the Q function—see Eq. (3) above—and the source profile compositions are not significantly different from close solutions with positive values of FPEAK.
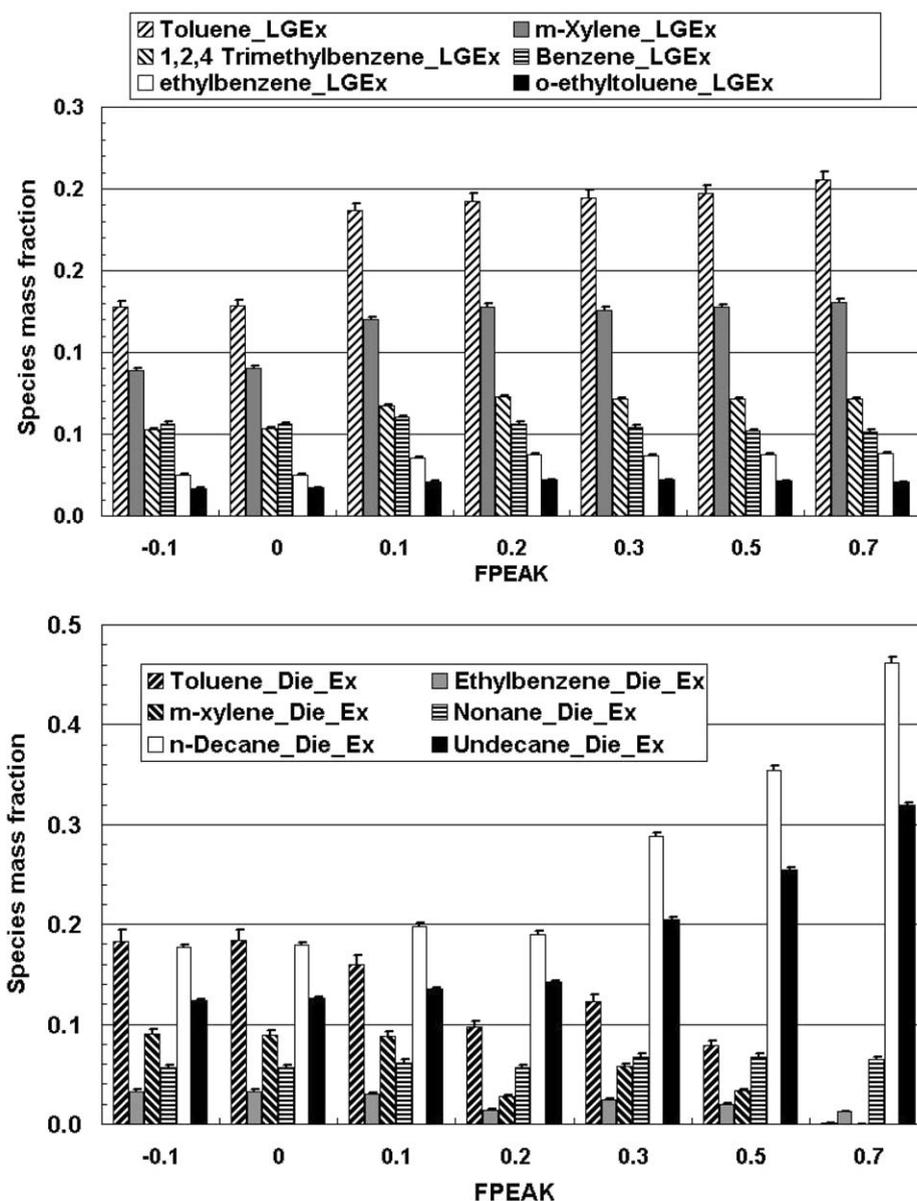

Fig. 12. Species mass fractions for the in the six source solution given by PMF for several FPEAK values. Upper panel: local gasoline exhaust; lower panel: diesel exhaust source profiles. The standard errors in each species composition are also shown.

5.3. Comparison of model solutions

Both models were able to extract more sources at LAC than at the POH site. This was possible because more data points were available at LAC. However, PMF was able to extract more information from the data than did UNMIX for this site, producing a six source solution with three different source profiles for biogenic, evaporative and LPG emissions, that are more credible than the five source solution given by UNMIX, where biogenic, evaporative and LPG source profiles are clearly mixed.

One possible explanation for this difference in performance is that UNMIX implicitly assumes that there are times when only one source is dominanting ambient concentrations. In this case of VOC emissions, there are several sources that do not ever shut down, as clearly shown in the daily profiles in Figs. 8 and 9.

6. Discussion

The application of PMF and UNMIX models produced results that were similar at the downtown (POH) site but that differed at the residential (LAC) site. From the composition of selected species within the resolved source profiles we conclude that the PMF solution—obtained with FPEAK = 0.0 in both cases—was more credible than the solution obtained with UNMIX. For instance, UNMIX produced source profiles that were a mixture of two or three source profiles, like the "biogenic" profile, making difficult to estimate the different source contributions in quantitative terms. In addition, UNMIX produces normalized source contributions that add up to 100% of the observed concentrations, something that was avoided in the application of PMF, the latter results showing an estimate of unresolved VOC sources.

Below we discuss some issues that might have had an impact on the analyses and conclusions reported here.

6.1. Absence of C_2 and C_3 compounds

In the 1996 photochemical campaing at Santiago, Chile, no C_2–C_3 compounds were measured due to instrumental limitations of the online technique. Species that could not be determined are for instance ethylene, acetylene and propane. There is no way of knowing how much of a bias is produced by lack of those light compounds in the VOC receptor model analyses. For instance, it is likely that inclusion of propane measurements could have led us to better resolve the LPG fugitive emission source, something that was partially achieved at the LAC site but was not feasible at POH site. In the case of motor vehicles, inclusion of ethylene and acetylene would have certainly improved the out-

come of the receptor modeling at both sites. Nonetheless, the higher correlation coefficients between carbon monoxide ambient measurements and the resolved source profiles attributed to mobile sources lends support to the results reported here.

6.2. Changing the number of sources in the analysis

The solutions provided by UNMIX had a fixed number of sources, a feature that is automatically controlled by UNMIX on the basis of signal-to-noise ratios. PMF allows the user to select any number of sources, so we performed an additional experiment increasing the number of sources for each data set. When PMF was run with additional sources at POH, we could not obtain sensible results for five, six or higher number of sources. The reason is that the POH site is so dominated by traffic sources, that increasing the number of sources in the analysis only leads to artifacts like splitting the diesel exhaust source profile into low- and high-molecular weight source profiles that cannot be further identified. When the number of sources is increased to six in PMF analysis of LAC data set, a better solution is obtained: the LPG, evaporative and biogenic emission profiles are now better distinguished, and quantitative estimates of the corresponding source contributions can be made with the new, cleaner profiles. Further increases in the number of sources do not add more information, so we select the six source solution from PMF as the final result for the analysis of data at LAC site.

6.3. The issue of VOC reactivities

There is a great range of VOC reactivities among all VOC species measured in the 1996 photochemical campaign at Santiago. Hence, it is likely that most reactive VOC tend to be underestimated from the source profiles resolved by the receptor models, so that the source profiles are enriched in the less reactive VOC. In other words, the source profiles obtained here consist of "aged" profiles that differ from the initially released "fresh" profiles, and the magnitude of this discrepancy would depend upon each specific VOC emitted, the time required for the species to reach the receptor site and the specific meteorological conditions (temperature, radiation, etc.). There is no simple way of estimating that effect within the receptor model framework, only the use of *state-of-the-science* air quality models could help in resolving how much that aging effect distorts the original, fresh emission profiles. Therefore, care must be taken when the source profiles reported here are compared with other source profiles reported in the literature, especially those used in chemical mass balance receptor modeling.

7. Conclusions

We have applied two different receptor models to data from an ambient VOC campaign taken at Santiago, Chile in 1996 at two different sites: a downtown site (POH) and a residential site (LAC). The following outcomes were obtained:

(a) At both monitoring sites mobile sources dominate ambient VOC measurements. Gasoline exhaust and evaporative emissions are the larger sources, followed by LPG losses and biogenic emissions. At the downtown site the following source apportionments estimates were obtained: fuel evaporation, $29.7\pm5.6\%$; gasoline exhaust, $22.0\pm3.4\%$; diesel exhaust, $18.1\pm2.9\%$; biogenic, LPG and evaporative emissions, $18.0\pm3.4\%$, unexplained: 12%. At the residential site, the following source apportionment was obtained: transported gasoline exhaust, $31.2\pm4.1\%$; local gasoline exhaust, $25.5\pm4.0\%$; fuel evaporation, $11.7\pm2.8\%$; LPG losses, $11.0\pm2.5\%$; biogenic emissions, $7.7\pm1.7\%$; diesel exhaust, $6.2\pm1.5\%$; unexplained, 7.7%. Thus, near 70% of ambient VOC impacts at both sites are due to mobile sources.

(b) When these results are compared with a metropolitan area with a similar climate like southern California (Fujita et al., 2003), we find that at Santiago there are higher contributions from evaporative sources—that is, gasoline vapor—and lower contributions from gasoline exhaust than in California. A plausible explanation is that in 1996 no regulation on controling gasoline vapor leaks from pump stations had been enacted at Santiago, whereas California has long been into tighter regulation for VOC fugitive emissions in the fuel distribution system.

(c) UNMIX is good at selecting a suitable number of factors, based on signal to noise ratios in environmental data sets. PMF was good at completing the source profiles by adding all measured species—except the ones having too many BDL samples—and this feature allowed us to better identify the resulting factors, because key species such as limonene, styrene, *n*-propylbenzene, among others, were incorporated in PMF solutions. At both sites the solution given by PMF model was more credible than the solution produced by the UNMIX model, and at the residential (LAC) site PMF was able to extract an additional source (six sources versus five sources in UNMIX solution) that yielded more "cleaner" (less mixed) source profiles than in UNMIX solution. Hence the source apportionment results quoted above are for the PMF solutions at both measurement sites.

(d) The above results for VOC source apportionment can be used to assess the quality of urban VOC emission inventories, by providing quantitative bounds for the emission strengths of the different source categories resolved by the analyses reported here.

Acknowledgements

References

Abu-Allaban, M., Gertler, A.W., Lowenthal, D.H., 2002. A preliminary apportionment of the sources of ambient PM10, PM2.5 and VOCs in Cairo. Atmospheric Environment 36, 5549–5557.

Anderson, M.J., Miller, S.L., Milford, J.B., 2001. Source apportionment of exposure to toxic volatile organic compounds using positive matrix factorization. Journal of Exposure Analysis and Environmental Epidemiology 11, 295–307.

Anderson, M.J., Daly, E.P., Miller, S.L., Milford, J.B., 2002. Source apportionment of exposures to volatile organic compounds: II. Application of receptor models to team study data. Atmospheric Environment 36, 3643–3658.

Borbon, A., Fontaine, H., Veillerot, M., Locoge, N., Galloo, J.C., Guillermo, R., 2001. An investigation into the traffic-related fraction of isoprene at an urban location. Atmospheric Environment 35, 3749–3760.

Borbon, A., Fontaine, H., Locoge, N., Veillerot, M., Galloo, J.C., 2003. Developing receptor-oriented methods for non-methane hydrocarbon characterisation in urban air—Part I: Source identification. Atmospheric Environment 37, 4051–4064.

Chueinta, W., Hopke, P.K., Paatero, P., 2000. Investigation of sources of atmospheric aerosol at urban and suburban residential area in Thailand by positive matrix factorization. Atmospheric Environment 34, 3319–3329.

Fujita, E.M., Campbell, D.E., Zielinska, B., Sagebiel, J.C., Bowen, J.L., Goliff, W.S., Stockwell, W.S., Lawson, D.R., 2003. Diurnal and weekday variations in the source contributions of ozone precursors in California's South Coast Air Basin. Journal of the Air and Waste Management Association 53, 844–863.

Hellén, H., Hakola, H., Laurila, T., 2003. 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, 1413–1424.

Henry, R.C., 1987. Current factor analysis models are ill-posed. Atmospheric Environment 21, 1815–1820.

Henry, R.C., 1997. History and fundamentals of multivariate air quality receptor models. Chemometrics and Intelligent Laboratory Systems 37, 37–42.

Henry, R.C., 2000. Reconciliation of receptor modeling and emission inventories of VOCs in a heavy industrial area. Paper #871, Proceedings from the 93rd Annual Meeting and Exhibition. Air and Waste Management Association, Salt Lake City, Utah, 18–22 June.

Henry, R.C., 2001. UNMIX Version 2.3 Manual; available with UNMIX software (rhenry@usc.edu).

Henry, R.C., 2003. Multivariate receptor modeling by N-dimensional edge detection. Chemometrics and Intelligent Laboratory Systems 65, 179–189.

Henry, R.C., Park, E.S., Spiegelman, C.H., 1999. Comparing a new algorithm with the classic methods for estimating the number of factors. Chemometrics and Intelligent Laboratory Systems 48, 91–97.

Ho, K.F., Lee, S.C., Chiu, G., 2002. Characterization of selected volatile organic compounds, polycyclic aromatic hydrocarbons and carbonyl compounds at a roadside monitoring station. Atmospheric Environment 36, 57–65.

Hopke, P.K., 1985. Receptor Modeling in Environmental Chemistry. Wiley, New York.

Huber, P.J., 1981. Robust Statistic. Wiley, New York.

Jollife, I.T., 1986. Principal Component Analysis. Springer, New York.

Kim, E., Hopke, P.K., Edgerton, E.S., 2003. Source identification of Atlanta aerosol by positive matriz factorization. Journal of the Air and Waste Management Association 53, 731–739.

Kourtidis, K.A., Ziomas, I.C., Rappenglück, B., Proyou, A., Balis, B., 1999. An estimation of evaporative emissions, traffic CO and speciated hydrocarbon emissions from the city of Athens. Atmospheric Environment 33, 3831–3842.

Lawrimore, J.H., Aneja, P., 1997. A chemical mass balance analysis of nonmethane hydrocarbon emissions in North Carolina. Chemosphere 35, 2751–2765.

Lewis, C.W., Norris, G.A., Conner, T.L., Henry, R.C., 2003. Source apportionment of phoenix PM2.5 aerosol with the UNMIX receptor model. Journal of the Air and Waste Management Association 53, 325–338.

Miller, S.M., Anderson, M.J., Daly, E.P., Milford, J.B., 2002. Source apportionment of exposures to volatile organic compounds. I. Evaluation of receptor models using simulated exposure data. Atmospheric Environment 36, 3629–3641.

Mugica, V., Vega, E., Arriaga, J.L., Ruiz, M.E., 1998. Determination of motor vehicle profiles for non-methane organic compounds in the Mexico City metropolitan area. Journal of the Air and Waste Management Association 48, 1060–1068.

Paatero, P., 1997. Least square formulation of robust nonnegative factor analysis. Chemometrics and Intelligent Laboratory Systems 37, 23–35.

Paatero, P., 2003. User's Guide for Positive Matrix Factorization Programs PMF2 and PMF3, Part 1: Tutorial. ftp://rock.helsinki.fi/pub/misc/pmf (accessed September 2003).

Paatero, P., Tapper, U., 1993. Analysis of different modes of factor analysis as least squares fit problems. Chemometrics and Intelligent Laboratory Systems 18, 183–194.

Paatero, P., Tapper, U., 1994. Positive Matrix Factorization: a nonnegative factor model with optimal utilization of error estimates of data values. Environmetrics 5, 111–126.

Paatero, P., Hopke, P.K., 2003. Discarding or downweighting high-noise variables in factor analytic models. Analytica Chimica Acta 490, 277–289.

Paatero, P., Hopke, P.K., Song, X.H., Ramadan, Z., 2002. Understanding and controlling rotations in factor analytic models. Chemometrics and Intelligent Laboratory Systems 60, 253–264.

Poirot, R.L., Wishinski, P.R., Hopke, P.K., Polissar, A.V., 2001. Comparative application of multiple receptor methods to identify aerosol sources in Northern Vermont. Environmental Science and Technology 35, 4622–4636.

Ramadan, Z., Song, X.-H., Hopke, P.K., 2000. Identification of sources of phoenix aerosol by positive matrix factorization. Journal of the Air and Waste Management Association 50, 1308–1320.

Rappenglück, B., Fabian, P., Kalabokas, P., Viras, L.G., Ziomas, I.C., 1998. Quasi-continuous measurements of non methane hydrocarbons (NMHC) in the Greater Athens Area during MEDCAPHOT-TRACE. Atmospheric Environment 32, 2103–2121.

Rappenglück, B., Fabian, P., 1999a. An analysis of simultaneous GC-measurements of BTEX-aromatics at three selected sites in the Greater Munich area. Journal of Applied Meteorology 38, 1448–1462.

Rappenglück, B., Fabian, P., 1999b. Non methane hydrocarbons (NMHC) in the Greater Munich Area/Germany. Atmospheric Environment 33, 3843–3857.

Rappenglück, B., Kourtidis, K., Melas, D., Fabian, P., 1999. Observations of biogenic and anthropogenic NMHC in Athens during the PAUR campaign. Physics and Chemistry of the Earth (B) 24, 717–724.

Rappenglück, B., Oyola, P., Olaeta, I., Fabian, P., 2000. The evolution of photochemical smog in the metropolitan area of Santiago de Chile. Journal of Applied Meteorology 39, 275–290.

Seinfeld, J.H., Pandis, S.N., 1998. Atmospheric Chemistry and Physics 2nd Edition. Wiley, New York.

Thijsse, T.M., van Oss, R.F., Lenschow, P., 1999. Determination of source contributions to ambient volatile organic compound concentrations in Berlin. Journal of the Air and Waste Management Association 49, 1394–1404.

Vega, E., Mugica, V., Carmona, R., Valencia, E., 2000. Hydrocarbon source apportionment in Mexico City using chemical mass balance receptor model. Atmospheric Environment 34, 4121–4129.

Volz-Thomas, A., Slemr, J., Konrad, S., Schmitz, T., Apel, A.C., Mohnen, V.A., 2002. Quality assurance of hydrocarbon measurements for the German Tropospheric Research Focus (TFS). Journal of Atmospheric Chemistry 42, 255–279.

Watson, J.G., Robinson, N.F., Chow, J.C., Henry, R.C., Kim, B.M., Nguyen, Q., Meyer, E.L., Pace, T.G., 1990. Receptor Model Technical Series, Vol. III, CMB7 User's Manual. US Environmental Protection Agency, Report No. EPA-450/4-90-004.

Watson, J.G., Chow, J.C., Fujita, E.M., 2001. Review of volatile organic compound source apportionment by chemical mass balance. Atmospheric Environment 35, 1567–1584.

Willis, R.D., 2000. Workshop on UNMIX and PMF as applied to PM2.5. US Environmental Protection Agency, Report No. EPA/600/A-00/048, Research Triangle Park, NC, 2000, 26pp. Available for download at http://www.epa.gov/ttn/amtic/unmixmtg.html (accesed October 2003).

Willis, R.D., Ellenson, W.D., Conner, T.L., 2001. Monitoring and source apportionment of particulate matter near a large phosphorous production facility. Journal of the Air and Waste Management Association 51, 1142–1166.