

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/10735889>

# Source Apportionment of Polycyclic Aromatic Hydrocarbons in the Urban Atmosphere: A Comparison of Three Methods

ARTICLE *in* ENVIRONMENTAL SCIENCE AND TECHNOLOGY · JUNE 2003

Impact Factor: 5.33 · DOI: 10.1021/es0206184 · Source: PubMed

---

CITATIONS

421

---

READS

276

## 2 AUTHORS:



**Randolph Kenneth Larsen**

St. Mary's College of Maryland

12 PUBLICATIONS 594 CITATIONS

SEE PROFILE



**Joel E Baker**

University of Washington Tacoma

121 PUBLICATIONS 5,351 CITATIONS

SEE PROFILE

## Article

## Source Apportionment of Polycyclic Aromatic Hydrocarbons in the Urban Atmosphere: A Comparison of Three Methods

Randolph K. Larsen, and Joel E. Baker

*Environ. Sci. Technol.*, **2003**, 37 (9), 1873-1881 • DOI: 10.1021/es0206184 • Publication Date (Web): 22 March 2003

Downloaded from <http://pubs.acs.org> on May 18, 2009

### More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 19 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

[View the Full Text HTML](#)



**ACS Publications**  
High quality. High impact.

# Source Apportionment of Polycyclic Aromatic Hydrocarbons in the Urban Atmosphere: A Comparison of Three Methods

RANDOLPH K. LARSEN, III AND  
JOEL E. BAKER\*

*Chesapeake Biological Laboratory, University of Maryland  
Center for Environmental Science, 1 Williams Street,  
Solomons, Maryland 20688*

Polycyclic aromatic hydrocarbons (PAHs) are ubiquitous pollutants in urban atmospheres. Several PAHs are known carcinogens or are the precursors to carcinogenic daughter compounds. Understanding the contributions of the various emission sources is critical to appropriately managing PAH levels in the environment. The sources of PAHs to ambient air in Baltimore, MD, were determined by using three source apportionment methods, principal component analysis with multiple linear regression, UNMIX, and positive matrix factorization. Determining the source apportionment through multiple techniques mitigates weaknesses in individual methods and strengthens the overlapping conclusions. Overall source contributions compare well among methods. Vehicles, both diesel and gasoline, contribute on average 16–26%, coal 28–36%, oil 15–23%, and wood/other having the greatest disparity of 23–35% of the total (gas- plus particle-phase) PAHs. Seasonal trends were found for both coal and oil. Coal was the dominate PAH source during the summer while oil dominated during the winter. Positive matrix factorization was the only method to segregate diesel from gasoline sources. These methods indicate the number and relative strength of PAH sources to the ambient urban atmosphere. As with all source apportionment techniques, these methods require the user to objectively interpret the resulting source profiles.

## Introduction

Polycyclic aromatic hydrocarbons (PAHs) are ubiquitous pollutants in urban atmospheres. Several PAHs are known carcinogens or are the precursors to carcinogenic daughter compounds. Understanding the contributions of the various sources is critical to appropriately manage PAH levels in the environment. This study compares the results of three factor-analysis-based source apportionment methods to determine the sources of ambient atmospheric particulate and gas-phase PAHs in Baltimore, MD. Agreements among the methodologies reinforce the findings, while divergent solutions indicate areas of uncertainty or weakness in the models. The three models employed in this study are principal component analysis followed by multiple linear regression (1), UNMIX (2), and positive matrix factorization (3). The chemical mass balance model is the U.S. Environmental

Protection Agency's principal receptor model for inorganic particulate matter. However, less consensus has been reached on the appropriate method for organic source apportionment. There are two significant concerns regarding the use of PAHs in source apportionment studies. First, the tendency of various PAHs to partition between gas and particulate phases complicates both the sampling methodology and characteristic source signatures. Few, if any, source apportionment studies have used only the gas-phase compounds. In this study, the measured gas- and particulate-phase concentrations of 32 individual PAHs were combined to avoid problems associated with partitioning. The second concern is the loss of the source signature by destruction of PAHs by photochemical processes. One method of assessing the degree of photochemical destruction is to evaluate the ratio of benzo[*e*]pyrene (BeP) to benzo[*a*]pyrene (BaP). BaP is highly susceptible to photochemical decay, whereas BeP is much more stable in the atmosphere (4). Similar ratios of BeP to BaP emissions are reported for a variety of sources (4). Therefore, changes in source apportionment should have only a minor effect on the ambient BeP/BaP ratio relative to the loss of BaP from photochemistry. A high BeP-to-BaP ratio indicates loss of the more reactive BaP (4). Samples in this study were grouped by season. A Student *t*-test of the summer samples (June, July, and August) versus the winter samples (December, January, and February) found no significant differences in the mean values at the 95% confidence interval. This indicates that the degree of photochemical losses of PAHs does not significantly change seasonally. Although the absolute amount of photochemical loss is unquantified, these results indicate that it is relatively constant and would therefore not impact the resulting source contribution, although it will create a systematic error in source composition.

**Use of PAH Source Markers.** Vehicles are a dominant source of PAHs in several cities around the world (1, 5–7). Several studies have sought to determine the percent contribution of PAHs from mobile versus nonmobile sources (5, 6, 8, 9). Some studies have gone further to separate mobile sources as diesel- or gas-powered vehicles. Nonmobile sources include domestic oil heating (10), meat grilling (11), wood combustion (9), coal and coking processes (12), fumigation of creosote logs (13), and volatilization from contaminated surface waters (14). The following is a summary of studies to identify individual compounds or specific ratios of compounds that are indicative of various PAH sources.

**Determining the Traffic Contribution of the Total PAH Load.** Correlating PAH levels with other pollutants is one method for determining source contributions. One study found that the majority of individual particulate-phase PAHs were significantly positively correlated to carbon monoxide and NO<sub>x</sub> (6). Assuming that CO and NO<sub>x</sub> were solely due to traffic, then linear regression of these compounds with PAHs allows for estimates of traffic contribution. The estimated traffic contribution of particle-phase four- to seven-ring PAHs was between 59–87% and 47–81% based on the respective CO/PAH and NO<sub>x</sub>/PAH regressions.

Several tunnel and dynamometer studies have been used to identify compounds that are specific to mobile sources. Coronene (Cor) and benzo[*g,h,i*]perylene (B[*g,h,i*]P) were found to be enriched in traffic tunnels along with BaP (15). It was later shown that B[*g,h,i*]P concentrations from traffic studies were independent of the diesel population (16). Because B[*g,h,i*]P has been linked to vehicle emissions (17, 18) and B[*g,h,i*]P has a positive correlation with total PAHs on PM 2.5, it was used to attribute vehicles as a primary

\* Corresponding author phone: (410) 326-7205; fax: (410) 326-7341; e-mail: baker@cbl.umces.edu.

source of PAHs on PM 2.5 in Hong Kong (7). Particulate-phase Cor in Copenhagen has been hypothesized to be exclusively traffic related. The ratio of Cor to BeP representative of a traffic source was  $1.54 \pm 0.16$  in Copenhagen (5). With this ratio, the excess of BeP to Cor above 1.54 was assumed to be unrelated to traffic. The same method was applied in Birmingham, U.K., where the Cor/BeP ratio attributable to traffic was measured at 1.80 (6). The ratio of B[g,h,i]P to BeP for traffic (2.37) and nontraffic sources (0.8) were combined with Cor/BeP ratios for traffic (1.8) and nontraffic (0.3) sources to form two equations that were solved simultaneously. The mean of the two solutions for the fractional traffic contribution was comparable to the Cor/BeP calculation alone. These techniques are adequate when location-specific ratios of source markers are available. However, these two examples show a 20% difference between location for the Cor/BeP ratio, which brings into question the applicability of using them at other locations.

**Segregating Gasoline- versus Diesel-Fueled Vehicles.** Several additional PAHs and PAH ratios have been used to segregate diesel from gasoline exhaust. Ratios of fluoranthene to fluoranthene plus pyrene on the order of 0.40–0.45 have been reported in exhausts of gasoline-fueled vehicles (19). Cyclopenta[cd]pyrene was reported to be a unique tracer for gasoline, while benzo[b]naphtho[2,1-d]thiophene was reported to be a unique tracer for diesel emissions (18). Diesel exhaust is believed to be enriched in fluoranthene, chrysene, and pyrene relative to gasoline-fuel exhaust (20). Increased concentrations of benzo[k]fluoranthene relative to other PAHs (21) and to carbon monoxide (16) have implied increased proportion of diesel vehicles in the traffic fleet. Different ratios of indeno[1,2,3-cd]pyrene to indeno[1,2,3-cd]pyrene plus benzo[g,h,i]perylene have been reported for diesel vehicles (22) and gasoline-fueled vehicles (23). The ratio of the sum of methylated phenanthrenes (4,5-MePh, 3-MePh, 9-MePh, 1-MePh, 2-MePh) to phenanthrene is different for diesel versus gasoline engine exhaust. A  $\Sigma\text{MePh}/\text{Ph}_{\text{gasoline}}$  ratio equal to 0.7 and  $\Sigma\text{MePh}/\text{Ph}_{\text{diesel}}$  ratios equal to 5.5 and 3.89 have been reported (5, 6).

**Segregating Motor Vehicles and Wood Combustion.** Motor vehicle emissions were segregated from residential wood combustion using the comparison of 1,7-dimethylphenanthrene (1,7-DMP) to 2,6-dimethylphenanthrene (2,6-DMP) (9). 1,7-DMP is primarily emitted by the burning of soft wood, whereas 2,6-DMP is emitted by both residential wood combustion and motor vehicles (9). Retene was suggested as a marker for soft wood combustion (12). It forms, along with 1,7-DMP, from the abietic and pimaric acids contained in wood during the combustion process (24).

**Chemical Mass Balances.** Chemical mass balances (CMBs) are routinely used in atmospheric source apportionment of inorganic compounds. Li and Kamens (1993) found that normalized source signatures of PAHs from various sources were better able to discriminate between sources than nonnormalized PAH profiles (25). Photochemical degradation rates for PAHs have been reported in the literature and can be applied to the CMBs to adjust for PAH losses (26, 27).

Duval and Friedlander used a CMB with first-order decay to resolve PAH sources to the Los Angeles atmosphere (10). In this study, fingerprints of coal combustion (anthracene, phenanthrene, fluoranthene, pyrene, BaA, and chrysene), coke production (anthracene, phenanthrene, BaP, B[g,h,i]P), incineration (phenanthrene, fluoranthene, and especially pyrene), wood combustion (anthracene, phenanthrene, fluoranthene, and pyrene), oil burning (fluoranthene and pyrene), gas-powered vehicles (fluoranthene, pyrene, B[g,h,i]P, and coronene), and diesel-powered vehicles (BbF, BkF, and thiophene compounds) were used. However, this technique was not suitable for the current study for the following reasons: (a) the fingerprints used in the CMBs were for the

particle phase only, (b) a priori knowledge of the number of PAH sources for the Baltimore region are not known, (c) location-specific profiles for the Baltimore PAH sources are not known, and (d) the error resulting from using source profiles from different regions to represent local sources are not known.

**PCA with Multiple Linear Regression Analysis.** Harrison et al. (1996) combined gas- and particle-phase PAHs with inorganic particulate measurements into a principal component analysis (PCA) using Varimax rotation, retaining principal components whose eigenvalues were greater than 1 (1). The relationship between the principal component and the chemical compound is indicated by the factor loadings and relates to the source emission composition. Multiple linear regression analysis (MLRA) of the factor scores is used to quantify the source contribution. Using the absolute factor scores as independent variables, MLRA was run with the chemical concentrations as the dependent variables. The regression coefficients obtained by MLRA were used to convert the absolute daily factor scores to estimates of source mass contributions (1). Six sources were identified: road dust, oil, coal, vehicles, incineration, and road salt. The majority of the PAH mass was associated with road dust, from which the authors conclude that road traffic is responsible for the major source of PAHs in the Birmingham, U.K., atmosphere.

Similar factor and regression analyses were performed by Simcik et al. (1999) on ambient gas- and particulate-phase PAHs in Chicago, IL (12). However, Simcik et al. (1999) did not use inorganic compounds in the PCA; therefore, identification of sources was based on an evaluation of the calculated source profiles with referenced information on the emission characteristics of several local source types (12). Their results indicated that coal combustion accounted for almost half of the PAH concentration in the Chicago atmosphere, followed by natural gas combustion, coke ovens, and vehicle emissions. The authors supported their results with estimates of local fossil fuel consumption.

## Experimental Section

**Sampling.** Atmospheric sampling of gas- and particle-phase PAHs occurred approximately every ninth day between March 1997 and December 1998. Twenty-four-hour samples were collected using a modified high-volume air sampler. Each high volume air sample consisted of a 0.4- $\mu\text{m}$  glass fiber filter (GFF) to collect particle-phase PAHs, followed by a polyurethane foam filter (PUFF) for gas-phase PAHs. The sampling location was on the roof of the Maryland Science Center in downtown Baltimore, Maryland.

**Extraction and Analysis.** The GFFs and PUFFs were extracted in Soxhlets for 24 h with dichloromethane and petroleum ether, respectively. The filters were spiked with four perdeuterated PAH surrogates prior to extraction. The extracts were subsequently reduced by rotoevaporation, transferred to test tubes, and further reduced to approximately 1 mL using dry nitrogen. The extracts were then spiked with an additional four perdeuterated PAHs that were used as internal standards. A HP 5890 gas chromatograph and 5972 mass-selective detector were operated in select ion mode to quantify PAH concentrations. The mean surrogate recoveries for the gas-phase PAHs were 64% for d<sub>8</sub>-naphthalene, 81% for d<sub>10</sub>-fluorene, 83% for d<sub>10</sub>-fluoranthene, and 76% for d<sub>12</sub>-perylene. The mean surrogate recoveries for the particle-phase PAHs were 62% for d<sub>8</sub>-naphthalene, 81% for d<sub>10</sub>-fluorene, 75% for d<sub>10</sub>-fluoranthene, and 82% for d<sub>12</sub>-perylene. Details of the sampling, analytical method, quality assurance and measured concentrations are described in ref 28.



**TABLE 1. Rotated Component Matrix of Gas- and Particle-Phase PAHs from Baltimore, Maryland<sup>a</sup>**

principal components	1	2	3	4
naphthalene	0.23	-0.01	0.90	-0.05
2-methylnaphthalene	0.20	-0.04	0.91	-0.04
1-methylnaphthalene	0.32	-0.04	0.91	-0.01
acenaphthylene	0.52	0.05	0.62	-0.02
acenaphthene	0.45	0.44	0.20	-0.01
fluorene	0.48	0.29	0.67	0.04
phenanthrene	0.24	0.67	-0.02	0.35
anthracene	0.64	0.61	0.11	0.21
1-methylfluorene	0.59	0.18	0.19	0.23
4,5-methylfluorene	-0.004	0.90	0.11	0.21
2-methylphenanthrene	0.16	0.16	0.40	0.70
2-methylantracene	0.55	0.30	-0.06	0.63
1-methylantracene	0.16	0.14	-0.09	0.90
1-methylphenanthrene	0.34	0.35	-0.12	0.80
fluoranthene	0.17	0.93	-0.13	0.14
pyrene	0.16	0.95	0.02	0.13
3,6-dimethylphenanthrene	0.06	0.06	0.15	0.03
9,10-dimethylphenanthrene	0.29	-0.04	0.38	0.15
benzo[a]fluorene	0.71	0.61	0.10	0.14
benzo[b]fluorene	0.82	0.48	0.14	0.04
benzo[a]anthracene	0.95	0.15	0.21	0.09
chrysene + triphenylene	0.90	0.33	0.13	0.10
retene	-0.10	0.33	-0.06	0.58
benzo[b]fluoranthene	0.95	0.12	0.20	0.11
benzo[k]fluoranthene	0.91	0.13	0.28	0.03
benzo[e]pyrene	0.94	0.10	0.23	0.13
benzo[a]pyrene	0.95	0.06	0.22	0.12
perylene	0.91	0.18	0.23	0.15
indeno[1,2,3-cd]pyrene	0.90	0.12	0.25	0.13
benzo[g,h,i]perylene	0.95	0.05	0.17	0.16
dibenz[ah,ac]anthracene	0.93	0.01	0.18	0.03
coronene	0.91	0.07	0.15	0.21
estimated source	vehicle	coal	oil	wood
variance (%)	51.5	14.9	7.7	6.0

<sup>a</sup> Rotation method: Varimax with Kaiser normalization. Only significant factors are listed.

## Results

**Source Estimates From PCA.** Principal component analysis (PCA) followed by multiple linear regression of the data was performed using SPSS software. The purpose of PCA is to represent the total variability of the original PAH data in a minimum number of factors. Each factor is orthogonal to all others, which results in the smallest possible covariance. The first factor represents the weighted (factor loadings) linear combination of the original variables (i.e., individual PAHs) that account for the greatest variability. Each subsequent factor accounts for less variability than the previous (Table 1). By critically evaluating the factor loadings, an estimate of the chemical source responsible for each factor can be made. The number of significant factors was determined during the stepwise multiple linear regression, which identified those factors that significantly improved the regression between the factors and the measured total PAH concentration.

**Factor 1.** The first factor is responsible for 51% of the total variance. This factor is heavily weighted by coronene, perylene, chrysene + triphenylene, indeno[1,2,3-cd]pyrene and all of the measured benzo derivatives of fluorene, fluoranthene, anthracene, pyrene, and perylene. According to the literature, this source appears vehicular (gas and diesel) in nature. Benzo[g,h,i]perylene and coronene have been identified as tracers of auto emissions (1, 25, 26). Elevated levels of benzo[k]fluoranthene relative to other PAHs have been suggested to indicate diesel vehicles (21). Indeno[1,2,3-cd]pyrene was also found in both diesel and gas engine emissions (22). Acenaphthylene, which is moderately loaded on this factor, has also been used as a tracer for vehicles (12).

**Factor 2.** The second factor is responsible for 15% of the total variance. This factor is predominately weighted in

pyrene, fluoranthene, and 4,5-methylfluoranthene. It also has the largest weighting of phenanthrene and anthracene. This profile is indicative of coal combustion. Several authors report fluoranthene, pyrene, phenanthrene, and anthracene as predominant coal combustion profiles (1, 12, 20, 29).

**Factor 3.** The third factor is responsible for 8% of the total variance. This factor is predominately composed of naphthalene, 2-methylnaphthalene, and 1-methylnaphthalene. These three components are the three lightest PAHs quantified. Initially, this was believed to be indicative of volatilization from creosote or coal tar, of which naphthalene is a characteristic marker. This hypothesis was tested by rerunning the PCA with the inclusion of daily maximum temperature, wind speed, wind direction, and precipitation as initial variables. The results continued to show factor 3 with the same grouping, but it also included a substantial inverse weighting with daily maximum temperature. This finding negates volatilization as the potential source of this factor. Oil combustion is another potential source of this factor. Harrison et al. (1996) reported a similar bimodal pattern, heavily weighted in the more volatile PAH species with moderate loadings of higher-molecular-weight compounds (1). Furthermore, Masclet et al. (1986) found oil-fired power-generation stations characterized by two- and three-ring PAHs, specifically methylnaphthalenes and phenanthrene (20). These references, along with the pattern of increasing importance with decreasing maximum temperatures, indicate that this factor is an oil combustion source.

**Factor 4.** The fourth factor is responsible for 6% of the total variance. This factor is weighted in phenanthrene, methylated phenanthrenes and anthracenes, and retene. The use of retene as a marker for wood combustion has been reported by several authors (9, 30–33). According to these references, factor 4 indicates a wood combustion source.

**Multiple Linear Regression.** The ultimate goal of performing PCA/MLR is to determine the percent contribution of different PAH sources for a given sample of ambient air. The basic equation of a multiple linear model is

$$y = \sum m_i X_i + b \quad (1)$$

This model is only valid if there is no collinearity in  $X_i$ . Noncollinearity, by definition, is ensured by selecting the PCA factor scores as the independent variables,  $X_i$ . The dependent variable,  $y$ , is  $\Sigma$ PAHs. The influence of each dependent variable on the independent variable can be directly compared by the regression coefficients, if the independent and dependent variables are "normally standardized". The equation for normal standardization is

$$\text{standardized normal deviate of } x = (x - \bar{x})/\sigma_x \quad (2)$$

where  $x$  is any variable,  $\bar{x}$  is the mean of  $x$ , and  $\sigma_x$  is the standard deviation of  $x$ . This new variable represents the distance, in standard deviation units, of a given sample from the mean. When the variables of eq 1 are normalized, the regression coefficients are represented as  $B_i$ , and the intercept ( $b$ ) is 0

$$z = \sum B_i X_i \quad (3)$$

$z$  becomes the standardized normal deviates of  $\Sigma$ PAHs. By definition, the factor scores have a mean of 0 and a standard deviation of 1; therefore,  $X_i$  after application of eq 2 remains unchanged. The partial regression coefficient  $B_i$  is also the partial correlation coefficient,  $r_{ip}$ , such that the squared multiple correlation coefficient,  $R^2$ , can be expressed as

$$R^2 = \sum B_i^2 \quad (4)$$

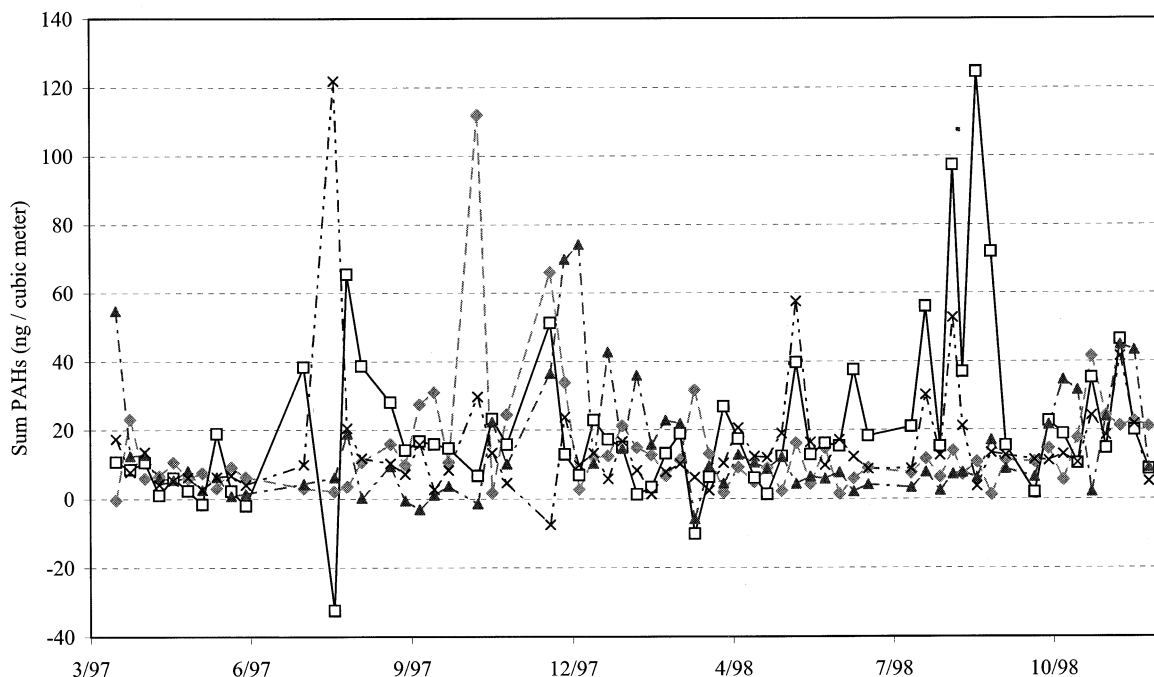


FIGURE 1. Time series of daily source contributions based on principal component analysis/multiple linear regression model. Symbols represent: ◆, vehicle; □, coal; ▲, oil; and ×, wood.

The multiple linear regression was performed stepwise using SPSS software. The stepwise procedure allows for the addition of each independent variable into the regression, from highest to lowest, in order of their individual simple correlation with the dependent variable. Each independent variable will only enter into the regression equation if it can significantly increase the correlation. A default level of significance of 0.05 was used. As each new variable enters into the regression equation, the previously entered variables lose some of their explanatory power. If the significance of a variable already in the regression equation falls below 0.10, it is removed from the equation.

Factor scores 1–4 representing vehicle, coal, oil, and wood sources, respectively, were regressed against the standard normalized deviate ( $z$ ) of the sum of 32 gas- plus particle-phase PAHs. The resulting equation is

$$\hat{Z} = 0.438FS_1 + 0.589FS_2 + 0.429FS_3 + 0.458FS_4 \quad (R^2 = 0.929) \quad (5)$$

By expanding  $\hat{Z}$  and rearranging terms, the multiple linear regression equation becomes

$$\sum_{32}(\text{gas} + \text{particle PAHs}) = 0.438\sigma_{\text{PAH}}FS_1 + 0.589\sigma_{\text{PAH}}FS_2 + 0.429\sigma_{\text{PAH}}FS_3 + 0.458\sigma_{\text{PAH}}FS_4 + \text{mean} \left[ \sum_{32}(\text{gas} + \text{particle PAHs}) \right] \quad (6)$$

where  $\sigma_{\text{PAH}}$  is  $26.2 \text{ ng}^{-3}$ , and the mean  $\sum_{32}(\text{gas plus particle PAHs})$  is  $43.6 \text{ ng}^{-3}$ . The calculation of the mean percent contribution becomes

$$\text{mean contribution of source } i (\%) = 100 \times (B_i / \sum B_i) \quad (7)$$

Thus, the mean percent contribution is 26% for the vehicular source, 28% for the coal source, 22% for the oil source, and 24% for the wood source. From these mean contributions,

the daily contribution is calculated as

$$\text{daily contribution of source } i (\text{ng PAH}^{-3}) = \text{mean} \sum_{32}(\text{gas} + \text{particle PAHs}) \times (B_i / \sum B_i) + B_i \sigma_{\text{PAH}} FS_i \quad (8)$$

where  $B_i / \sum B_i$  is the ratio of the regression coefficient for factor  $i$  to the sum of all the regression coefficients and  $FS_i$  is the factor score for factor  $i$ . Figure 1 shows the estimated daily contributions for each of the proposed sources. The relative contributions between sources vary substantially from the overall mean estimates. During the winter months, there are increased contributions from the oil combustion source. The coal combustion source appears to have the greatest range as a percentage of total and absolute concentrations. Approximately 10% of the individual samples contain apparent negative source contributions, which is physically impossible. PCA's ability to generate negative source contributions is a known concern. Examination of the source contributions shows that the largest negative source contribution occurred on July 31, 1997. The large negative contribution of the coal source is offset by a large positive contribution of the wood source. In fact, the wood source does not exhibit even half of the July 31, 1997, daily source contribution at any other time during the study. Examination of factor loadings reveals that 1-methylantracene (1-MeA) has the highest loading on the wood source. 1-MeA also has the highest concentration during the study on July 31, 1997. These facts indicate that PCA created a factor to primarily account for the July 31, 1997, outlying 1-MeA concentration. This is an example of PCA/MLRA's inability to effectively model extreme data. This concept is an expression of the "nonrobustness" in the PCA/MLRA method. When the PCA/MLRA analysis was rerun without the July 31, 1997, datum, a five-source solution was created. The factor loadings of the first four sources were similar to the original vehicle, coal, oil, and wood profiles. The fifth source was unidentifiable because of the lack of prominent chemical loadings. This truncated PCA is provided in the Supporting Information. The average source contributions of this truncated dataset were 28% vehicular, 34% coal, 25% oil, and 12% total from the combination wood and unidenti-

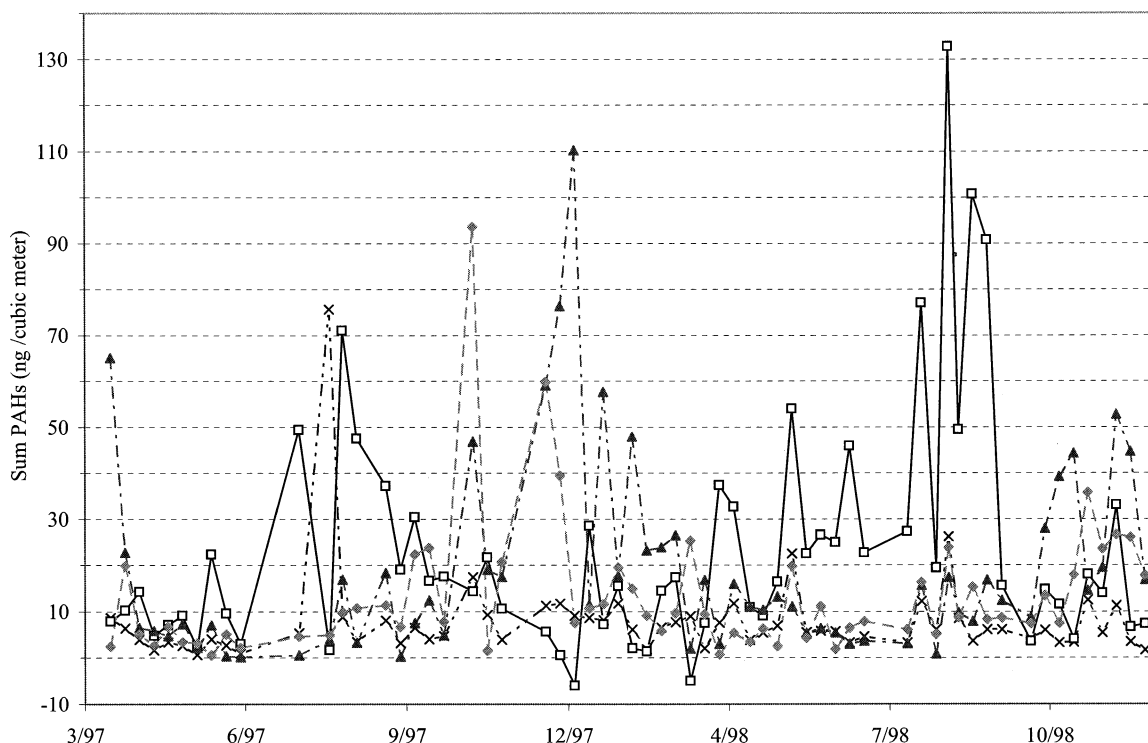


FIGURE 2. Time series of daily source contributions based on the UNMIX model. Symbols represent: ♦, vehicle; □, coal; ▲, oil; and ×, other.

TABLE 2. UNMIX-Determined Partial Source Compositions of Ambient PAH Levels in Baltimore (ng/m<sup>3</sup>)

source	1	2	3	4
naphthalene	0.119	1.853	0.003	0.178
2-methylnaphthalene	0.257	2.830	0.145	0.209
1-methylnaphthalene	0.027	0.853	0.000	0.192
phenanthrene	8.006	0.995	1.165	4.067
1-methylanthracene	0.480	0.086	1.591	0.000
fluoranthene	2.003	0.051	0.000	0.919
benz[a]anthracene	0.010	0.021	0.004	0.093
chrysene + triphenylene	0.061	0.036	0.010	0.164
benzo[b]fluoranthene	0.025	0.040	0.011	0.156
benzo[e]pyrene	0.012	0.025	0.007	0.094
benzo[a]pyrene	0.003	0.021	0.004	0.105
perylene	0.002	0.003	0.000	0.015
indeno[1,2,3-cd]pyrene	0.016	0.039	0.006	0.153
benzo[g,h,i]perylene	0.012	0.037	0.013	0.212
coronene	0.014	0.018	0.011	0.101
estimated sources	coal	oil	other	vehicle

fied sources. Efforts to develop other statistically driven source-receptor models that do not require a priori knowledge of source profiles and are "robust" have been published. Two such models, UNMIX and positive matrix factorization, have been applied to the same Baltimore dataset to elucidate differences in their output.

**UNMIX Model.** UNMIX was developed to address the shortcomings of both PCA and CMBs (2). Results from the UNMIX model are constrained to positive values, which address the most significant concern of PCA. Unlike CMBs, UNMIX does not require a priori knowledge of the number of sources or their compositions. Given a data matrix of  $n$  samples with  $m$  chemical components, the model performs a singular value decomposition of the  $n \times m$  matrix after it has been normalized such that all species have a mean of 1. This step reduces the dimensionality of the data space to the number of sources. UNMIX further reduces the normalized source composition by projecting the data to a plane perpendicular to the first axis of  $N$ -dimensional space. The

boundaries or edges of the projected data represent the samples that characterize the sources. Henry et al. (1994) provides further details of the UNMIX model (2). Unlike PCA and other forms of source modeling, UNMIX repeats the model calculations for each possible combinations of  $m$  chemicals and retains only those chemicals that contribute to improving the model's signal-to-noise ratio.

From the original matrix of Baltimore PAHs, four "seed" compounds were chosen to start the "overnight" mode of UNMIX, and 15 remained at the completion of the analysis. Each compound had a correlation coefficient of 0.8 for a particular source. As with the PCA/MLRA method, UNMIX determined four sources of PAHs in the data (Table 2). In a manner similar to the PCA/MLRA analysis, it is clear that one source has a vehicular signature with high levels of Cor, I[1,2,3]P, and B[g,h,i]P. One source is similar to the oil signature with high levels of Nap, 2-MeN, and 1-MeN. Another source similar to the coal signature had the highest fraction of fluoranthene and phenanthrene. The last source contained the highest levels of 1-MeA as well as substantial levels of phenanthrene, which is similar to the wood source found in the PCA/MLRA. Unfortunately, retene did not correlate to the 0.8 minimum in the wood-like source. Therefore, this source is assigned as "other combustion". The daily contributions of each source are presented in Figure 2. The same cyclic patterns for coal and oil as shown in the PCA/MLRA analysis are present. However, with UNMIX, sources no longer exhibit extreme negative contributions.

**Positive Matrix Factorization Model.** Positive matrix factorization (PMF) is a generic factorization methodology that has been used to determine source compositions and contributions of aerosols (34, 35), precipitation (3), general air-quality parameters (36), water-quality parameters (37), and molecular line spectra (38). PMF is a factor-analysis-based method that assumes that the initial data matrix can be expressed as

$$\mathbf{X} = \mathbf{Y} + \mathbf{E} \quad (9)$$

TABLE 3. Source Profiles Obtained from Positive Matrix Factorization (ng/m<sup>3</sup>)<sup>a</sup>

source	1	2	3	4	5	6
naphthalene	<b>0.208</b>	0.020	<b>0.035</b>	0.001	0.000	0.010
2-methylnaphthalene	<b>0.323</b>	0.023	<b>0.073</b>	0.000	<b>0.066</b>	0.016
1-methylnaphthalene	<b>0.116</b>	0.009	<b>0.019</b>	<b>0.014</b>	0.000	0.004
acenaphthylene	<b>0.080</b>	0.000	0.000	<b>0.013</b>	0.000	0.000
acenaphthene	0.000	<b>0.018</b>	<b>0.012</b>	<b>0.011</b>	<b>0.011</b>	0.000
fluorene	<b>0.158</b>	0.021	<b>0.024</b>	<b>0.079</b>	<b>0.054</b>	0.031
phenanthrene	0.038	<b>0.542</b>	<b>0.320</b>	<b>0.418</b>	<b>0.324</b>	<b>0.325</b>
anthracene	0.000	0.000	0.000	<b>0.039</b>	0.010	<b>0.032</b>
1-methylfluorene	0.001	<b>0.039</b>	<b>0.048</b>	0.000	0.000	<b>0.012</b>
4,5-methylenepheneanthrene	0.000	0.000	<b>0.005</b>	0.000	0.007	<b>0.053</b>
2-methylphenanthrene	<b>0.057</b>	<b>0.089</b>	0.000	0.000	0.029	<b>0.117</b>
2-methylanthracene	0.000	0.000	0.001	<b>0.005</b>	0.000	<b>0.005</b>
1-methylanthracene	<b>0.003</b>	<b>0.062</b>	0.002	0.000	<b>0.069</b>	<b>0.078</b>
1-methylphenanthrene	<b>0.006</b>	<b>0.030</b>	0.000	0.000	0.000	<b>0.044</b>
fluoranthene	0.000	<b>0.090</b>	<b>0.091</b>	<b>0.055</b>	<b>0.046</b>	<b>0.105</b>
pyrene	0.004	<b>0.041</b>	<b>0.111</b>	0.000	<b>0.045</b>	<b>0.119</b>
3,6-dimethylphenanthracene	0.000	<b>0.001</b>	0.002	<b>0.003</b>	<b>0.323</b>	0.000
9,10-dimethylanthracene	0.000	0.000	<b>0.006</b>	0.000	0.000	0.000
benzo[a]fluorene	<b>0.001</b>	<b>0.003</b>	<b>0.008</b>	<b>0.011</b>	<b>0.005</b>	0.004
benzo[b]fluorene	0.000	<b>0.002</b>	0.001	<b>0.009</b>	0.002	0.003
benz[a]anthracene	0.000	0.000	<b>0.015</b>	<b>0.025</b>	0.000	0.001
chrysene + triphenylene	0.000	<b>0.004</b>	<b>0.033</b>	<b>0.038</b>	0.003	0.005
retene	0.000	0.003	<b>0.027</b>	0.004	0.000	<b>0.030</b>
benzo[b]fluoranthene	<b>0.001</b>	0.001	<b>0.038</b>	<b>0.044</b>	0.001	0.001
benzo[k]fluoranthene	0.000	0.000	0.000	<b>0.044</b>	0.001	0.000
benzo[e]pyrene	0.000	0.001	<b>0.022</b>	<b>0.029</b>	0.001	0.000
benzo[a]pyrene	0.000	<b>0.001</b>	<b>0.015</b>	<b>0.029</b>	0.000	0.000
perylene	0.000	<b>0.000</b>	0.000	<b>0.006</b>	0.000	0.000
indeno[1,2,3-cd]pyrene	<b>0.001</b>	0.000	<b>0.030</b>	<b>0.048</b>	0.002	0.000
benzo[g,h,i]perylene	<b>0.002</b>	0.000	<b>0.038</b>	<b>0.051</b>	0.001	0.002
dibenz[ah,ac]anthracene	<b>0.000</b>	0.000	<b>0.002</b>	<b>0.002</b>	0.000	0.000
coronene	<b>0.001</b>	0.000	<b>0.023</b>	<b>0.022</b>	0.000	0.002
estimated source	oil	coal	gasoline	diesel	other	wood

<sup>a</sup> Values in boldface are greater than twice the estimated standard deviation.

where **X** is the initial data matrix ( $n \times m$ ), **Y** is the calculated matrix that can be either two or three dimensions, and **E** is the residual matrix. A two-dimensional solution was chosen in this study to provide a direct comparison of the PMF results with UNMIX and PCA/MLR. In a two-dimensional solution, **Y** is expressed as

$$\mathbf{Y} = \mathbf{GF} \quad (10)$$

where **G** is the source contribution matrix ( $n \times p$ ) and **F** is the source profile ( $p \times m$ ). Elements of the residual matrix are defined as

$$e_{ij} = x_{ij} - \sum_{k=1}^p f_{ik} g_{kj} \quad (11)$$

where  $i = 1, \dots, m$  elements;  $j = 1, \dots, n$  samples; and  $k = 1, \dots, p$  sources. PMF uses the residual matrix elements,  $e_{ij}$ , along with an uncertainty estimate,  $s_{ij}$ , provided by the user to calculate a minimum  $Q$  value defined as

$$Q = \sum_{i=1}^m \sum_{j=1}^n \left( \frac{e_{ij}}{s_{ij}} \right)^2 \quad (12)$$

The algorithm for  $Q$  uses an iterative least-squares methodology that simultaneously varies the **G** and **F** matrices. The iterations halt when the difference in successive values of  $Q$  change by less than a predefined limit. The default value for this limit is set at 0.01. During this analysis, the limit was reduced to 0.0001. By reducing this limit, we ensure that the iterations for  $Q$  have, in fact, converged. The method has several other additional features such as a penalty function that prevents negative solutions for either **G** or **F**, error estimates for both the **G** and **F** matrices and a robust mode that reweights an individual datum to prevent extreme values

from significantly influencing the solution (39). This robust mode was used in this analysis.

The **X** matrix for this study was  $61 \times 32$ ; therefore, the theoretically optimum value of  $Q$  equals 1952. The values in the **X** matrix consisted of the gas-phase PAH concentration plus the particle-phase PAH concentration. If either the gas or particle value was less than detectable, then one-half the limit of detection (LOD) was used as a proxy. The limit of detection was expressed as 3 times the mean blank concentration for the respective phase (gas or particle). The **S** matrix consisted of the individual error estimates,  $s_{ij}$ , which consisted of the propagated error estimates of both the gas- and particle-phase PAH concentrations. This propagated error was the square root of the sum of the squared gas-phase and squared particle-phase PAH error estimates. These gas- and particle-phase PAH concentration estimates were calculated as the recovery standard deviation of the surrogate representative of compound  $i$  multiplied by  $x_{ij}$  plus  $1/2$  LOD. The standard deviations for the gas-phase surrogates were 22% for  $d_8$ -naphthalene, 22% for  $d_{10}$ -fluorene, 22% for  $d_{10}$ -fluoranthene, and 19% for  $d_{12}$ -perylene. The standard deviations for the particle-phase surrogates were 19% for  $d_8$ -naphthalene, 16% for  $d_{10}$ -fluorene, 15% for  $d_{10}$ -fluoranthene, and 16% for  $d_{12}$ -perylene. The program PMF3 was used to solve the source apportionment in a two-dimensional fashion. This required the third dimension of the output to be forced to 1. The number of sources was varied from 2 to 10, with 10 repeat solutions for each source.

The optimum number of sources appears to be six, as after 10 repeated solutions, the source profiles are consistently similar to two different patterns. The calculated global minimum  $Q$  value was 2989; local minimum  $Q$  values ranged from 2989 to 3668. Increasing or decreasing the number of sources resulted in increasing numbers of nonunique potential solutions. Table 3 outlines the typical source compositions based on a six-source solution. The decision



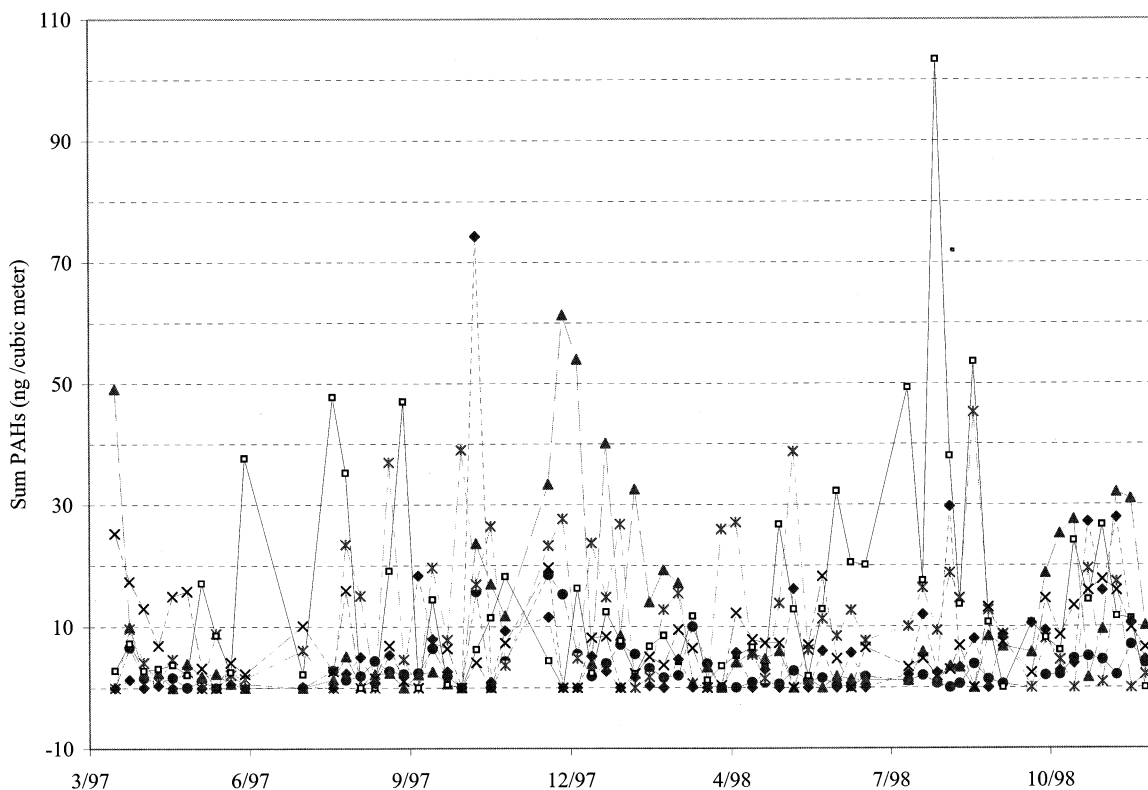


FIGURE 3. Time series of daily source contributions based on the positive matrix factorization model. Symbols represent:  $\blacklozenge$ , gasoline;  $\bullet$ , diesel;  $\square$ , coal;  $\blacktriangle$ , oil;  $*$ , other; and  $\times$ , wood.

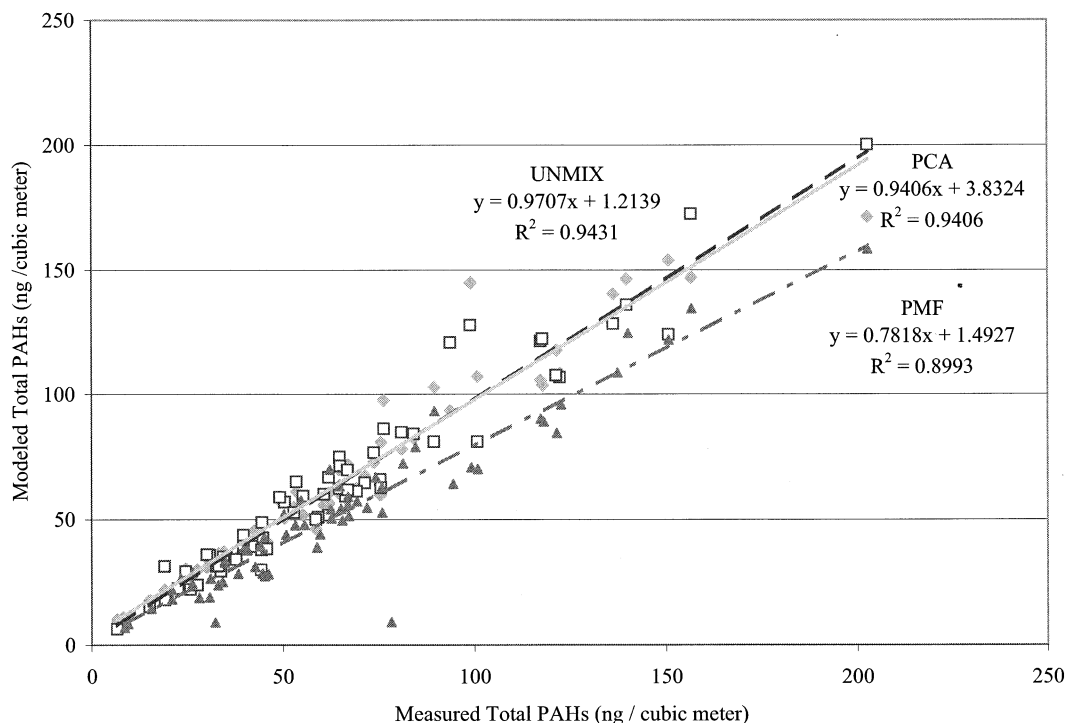


FIGURE 4. Correlation of modeled results with original measurements. Symbols represent:  $\blacklozenge$ , PCA/MLR;  $\square$ , UNMIX; and  $\blacktriangle$ , PMF.

of which solution to present was slightly subjective on the part of the researcher, on the basis of whether the profiles can be explained by known source patterns. Several of the factors are greater than twice their uncertainty, which indicates that there is little rotational ambiguity in their possible solution.

The factor loadings from the PCA/MLRA cannot quantitatively be compared to the source profiles calculated by

PMF because they are not of the same units. However, qualitatively the two outputs are similar in several areas. Likewise the relative order of sources presented in PMF and UNMIX are arbitrary and do not relate to their significance. The PMF pattern for source 1 is similar to the PCA oil profile with significant levels of naphthalene, 2-MeN, 1-MeN, acenaphthylene, fluorene, 2-MeP, BaA, BkFl, BeP, and I[1,2,3]P. Therefore, this source is also identified as oil

**TABLE 4. Overall Contributions of Sources to the Total Ambient PAH Concentration**

source	contribution (%)		
	PCA/MLRA	UNMIX	PMF
vehicle <sup>a</sup>	26	23	16
coal	28	29	36
oil	22	23	15
other/wood <sup>b</sup>	24	23	35

<sup>a</sup> PMF value for vehicles is the combination of 7% diesel-powered and 8% gasoline-powered vehicles. <sup>b</sup> PMF value for other/wood is the combination of 21% wood combustion and 14% other categories.

combustion. The PMF pattern for source 5 is similar to the PCA coal profile with significant levels of phenanthrene, 1-MeP, 4,5-MeF, fluoranthene, pyrene, BaF, BbF, and chrysene + triphenylene. Therefore, this source is also identified as coal combustion. PMF sources 1 and 6 are both heavily weighted in the vehicular-like compounds coronene, B[g,h,i]P, and I[1,2,3]P; however, source 1 is enriched in B[k]F relative to source 6. This indicates that source 1 might be diesel-like, whereas source 6 is more similar to a gasoline-powered vehicle profile. The remaining PMF sources 4 and 2 are more difficult to identify. Source 4 is most like the wood combustion pattern presented in the PCA, because it has retene and substantial loadings of phenanthrene, methylated phenanthrenes, and anthracenes. Because of the lack of any particular marker in source 2, this source is labeled "other" and could be either one or several other PAH sources.

The daily contributions of each source to the total ambient PAH burden are presented in Figure 3. As previously seen in the other methods, annual cyclic patterns for coal and oil are present, along with the addition of vehicle sources, now broken down into diesel and gasoline sources. As designed, the fit between the modeled total PAH concentrations and the measured concentrations presented in Figure 4 is strong; however, the UNMIX model had the closest 1:1 regression. This was expected because the UNMIX model selectively removes variables from the initial dataset to maximize the final correlation. PMF consistently underestimated the modeled output, but this can be explained by the way PMF operates. PMF requires an uncertainty estimate for each variable. Part of the uncertainty estimate is proportional to the total concentration; therefore, higher concentrations have larger uncertainties. In addition, PMF was run in robust mode, which downweights outlying variables, i.e., those values beyond 4 times the mean concentration. Therefore, PMF purposely underestimates the modeled concentrations of the highest measured values. Furthermore, PMF allows for additional dimensions affecting the measured concentrations not explained by the sources alone, such as weather, additional transient sources, or sampling artifacts, but this feature was not tested.

The overall source contributions presented in Table 4 compare well among the three methods. Vehicles, both diesel and gasoline, contribute on average 16–23% of the PAHs in ambient Baltimore air, coal 28–36%, oil 15–29%, and wood/other having the greatest disparity of 23–35%. By comparison, vehicles (both gas and diesel) accounted for 9 ± 4% while coal accounted for 48 ± 5% of the PAHs in Chicago (12).

Any of these methods can be used without a priori knowledge to determine the number of PAH sources and their contributions. However, the greater difficulty lies in proper source identification. Future research and monitoring efforts should strive to include key PAH source markers that will improve the ability to identify the outputs from these models. For example, the addition of 1,7-DMP and 2,6-DMP to the initial suite of 32 PAHs could have substantially improved our ability to distinguish wood combustion.

Likewise, the addition of cyclopenta[cd]pyrene might have improved our basis for separating diesel from gasoline with minimal effort. Although it would require greater analytical effort, combining the PAH data with other characteristic organic and inorganic source markers would further resolve source identification and contributions.

## Acknowledgments

We thank Dr. Pentti Paatero (Department of Physical Sciences, University of Helsinki) for considerable assistance with PMF and Dr. Ron Henry for providing help with UNMIX. This is the University of Maryland Center of Environmental Science Contribution No. 3651.

## Supporting Information Available

Table containing results of principal component analysis using truncated PAH data. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## Literature Cited

- (1) Harrison, R. M.; Smith, D. J. T.; Luhana, L. *Environ. Sci. Technol.* **1996**, *30*, 825–832.
- (2) Henry, R.; Lewis, C.; Collins, J. *Environ. Sci. Technol.* **1994**, *28*, 823–832.
- (3) Junntto, S.; Paatero, P. *Environmetrics* **1994**, *5*, 127–144.
- (4) Nielsen, T. *Atmos. Environ.* **1988**, *22*, 2249–2254.
- (5) Nielsen, T. *Atmos. Environ.* **1996**, *30*, 3481–3490.
- (6) Lim, L. H.; Harrison, R. M.; Harrad, S. *Environ. Sci. Technol.* **1999**, *33*, 3538–3542.
- (7) Zheng, G. J.; Lam, M. H. W.; Lam, P. K. S.; Richardson, B. J.; Man, B. K. W.; Li, A. M. Y. *Mar. Pollut. Bull.* **2000**, *40*, 1210–1214.
- (8) Nielsen, T.; Jorgensen, H. E.; Larsen, J. C.; Poulsen, M. *Sci. Total Environ.* **1996**, *190*, 41–49.
- (9) Benner, B. A.; Wise, S. A.; Currie, L. A.; Klouda, G. A.; Klinedinst, D. B.; Zweidinger, R. B.; Stevens, R. K.; Lewis, C. W. *Environ. Sci. Technol.* **1995**, *29*, 2382–2389.
- (10) Duval, M.; Friedlander, S. Source Resolution of Polycyclic Aromatic Hydrocarbons in the Los Angeles Atmospheres—Application of a CMB with First-Order Decay, U.S. Environmental Protection Agency (EPA), U.S. Government Printing Office: Washington, DC, 1981.
- (11) Dyremark, A.; Westerholm, R.; Overvik, E.; Gustavsson, J. A. *Atmos. Environ.* **1995**, *29*, 1553–1558.
- (12) Simcik, M. F.; Eisenreich, S. J.; Liroy, P. J. *Atmos. Environ.* **1999**, *33*, 5071–5079.
- (13) Gevao, B.; Jones, K. C. *Environ. Sci. Technol.* **1998**, *32*, 640–646.
- (14) Bamford, H. A.; Offenberg, J. H.; Larsen, R. K.; Ko, F. C.; Baker, J. E. *Environ. Sci. Technol.* **1999**, *33*, 2138–2144.
- (15) Sawicki, E. *J. Am. Ind. Hyg. Assoc.* **1962**, *23*, 137–142.
- (16) Hering, S.; Miguel, A.; Dod, R. *Sci. Total Environ.* **1984**, *36*, 39–45.
- (17) Baek, S.; Field, R.; Goldstone, M.; Kirk, P.; Lester, J.; Perry, R. *Water, Air, Soil Pollut.* **1991**, *60*, 279–300.
- (18) Daisey, J.; Cheney, J.; Liroy, P. *J. Air Pollut. Control Assoc.* **1986**, *36*, 17–33.
- (19) Aceves, M.; Grimalt, J. O. *Environ. Sci. Technol.* **1993**, *27*, 2896–2908.
- (20) Masclet, P.; Mouvier, G.; Nikolaou, K. *Atmos. Environ.* **1986**, *20*, 439–446.
- (21) Venkataraman, C.; Lyons, J. M.; Friedlander, S. K. *Environ. Sci. Technol.* **1994**, *28*, 555–562.
- (22) May, W.; Wise, S. *Anal. Chem.* **1984**, *56*, 225–232.
- (23) Stenberg, V.; Alsberg, A.; Westerholm, R. *Environ. Health Perspect.* **1983**, *47*, 53–56.
- (24) Ramdahl, T. *Nature* **1983**, *306*, 580–582.
- (25) Li, C. K.; Kamens, R. M. *Atmos. Environ. A* **1993**, *27*, 523–532.
- (26) Miguel, A.; Pereira, P. *Aerosol Sci. Technol.* **1989**, *10*, 292–295.
- (27) Sexton, K.; Liu, K.; Hayward, S.; Spengler, J. *Atmos. Environ.* **1985**, *19*, 1225–1236.
- (28) Larsen, R. K. Ph.D. Dissertation, University of Maryland, College Park, MD, 2002.

- (29) Mastral, A. M.; Callen, M.; Murillo, R. *Fuel* **1996**, 75, 1533–1536.
- (30) Schauer, J. J.; Rogge, W. F.; Hildemann, L. M.; Mazurek, M. A.; Cass, G. R. *Atmos. Environ.* **1996**, 30, 3837–3855.
- (31) Simcik, M. F.; Eisenreich, S. J.; Golden, K. A.; Liu, S. P.; Lipiatou, E.; Swackhamer, D. L.; Long, D. T. *Environ. Sci. Technol.* **1996**, 30, 3039–3046.
- (32) Sheffield, A. E.; Gordon, G. E.; Currie, L. A.; Riederer, G. E. *Atmos. Environ.* **1994**, 28, 1371–1384.
- (33) Halsall, C. J.; Barrie, L. A.; Fellin, P.; Muir, D. C. G.; Billeck, B. N.; Lockhart, L.; Rovinsky, F. Y.; Kononov, E. Y.; Pastukhov, B. *Environ. Sci. Technol.* **1997**, 31, 3593–3599.
- (34) Huang, S.; Rahn, K. A.; Arimoto, R. *Atmos. Environ.* **1999**, 33, 2169–2185.
- (35) Sirois, A.; Barrie, L. A. *J. Geophys. Res.* **1999**, 104, 11, 599–511, 618.
- (36) Paterson, K. G.; Sagady, J. L.; Hooper, D. L. *Environ. Sci. Technol.* **1999**, 33, 635–641.
- (37) Reinikainen, S.; Laine, P.; Minkinen, P.; Paatero, P. *Fresenius' J. Anal. Chem* **2001**, 369, 727–732.
- (38) Juvela, M.; Lehtinen, K.; Paatero, P. *Mon. Not. R. Astron. Soc.* **1996**, 280, 616–626.
- (39) Paatero, P. *Chemom. Intell. Lab. Syst.* **1997**, 37, 23–35.

*Received for review February 26, 2002. Revised manuscript received January 24, 2003. Accepted February 3, 2003.*

ES0206184