Source Apportionment of Sediment PAHs in Lake Calumet, Chicago: Application of Factor Analysis with Nonnegative Constraints

PHILIP A. BZDUSEK AND
ERIK R. CHRISTENSEN*
Department of Civil Engineering and Mechanics and Center

(*19*). In short, the cumulative percent variance, coefficient of determination (*20*), Exner function (*21*) and convergence of the nonnegative rotations were all considered for determineTJT*12 1 Tft6l,2 1 T(deter-)-t(

samples (*15*). As a result of difficulties associated with GC separation of BbFlA and BkFlA the sum of the two isomers was used as the abbreviation Bb+kFlA (*15*). Also, samples with concentrations less than 2 μg/g were not considered (*15*). Thus, the resulting data matrix was 15 × 49. FA was applied for a two-, three-, four-, five-, and six-source solution; however, the FA model only yielded convergent solutions for the two- and six-source solutions.

In addition to the 15 × 49 data set, a 15 × 43 data set was modeled. This data set eliminated core K, where all six samples had significantly higher IP concentrations than the average IP concentration for the data set. Two- and five-source FA solutions were obtained; three-, four-, and six-source applications did not yield convergent solutions. PAH sources were determined, qualitatively, by visually comparing patterns between literature source profiles and factor loadings, and quantitatively, by the sum of squares difference between normalized literature PAH source profiles, and normalized FA model factor loadings.

Monte Carlo Simulation. To determine the uncertainty of the FA model results Monte Carlo simulation was utilized. The governing equation follows

where D_{ij} is the generated PAH concentration from PAH i and sample j, A_{ij} is the starting concentration of PAH i from sample j, C_{ij} is the coefficient of variation of PAH i from sample j, erf^{-1} is the inverse Gaussian error function and R_{ij} is a random number between 0 and 1.

Using Monte Carlo simulation nine artificial data matrices **D** were generated and modeled by FA to yield nine factor-loading and nine factor-score matrices. The standard deviation of the mean for each entry in the factor-loading and factor-score matrices was calculated to be the uncertainty.

A_{ij} was taken to be the modified 15 × 49 data set. The coefficient of variation for each sample was obtained from the standard deviations of replicate experimental measurements (*15*). If the coefficient of variation for a given sample was less than 20% the value was used in the Monte Carlo simulation, otherwise 20% was used as the coefficient of variation value. If the coefficient of variations were much larger than 20% the FA model would converge for less than 50% of the model runs, demonstrating the limits of the

TABLE 1. Results of Diagnostic Tools Application for the Determination of the Number of Significant Factors for Lake Calumet (IL)

PAHs	coefficient of determination factors 1	2	3	4	5	6
acenaphthylene (AcNP)	0.31	0.45	0.76	0.95	0.99	1.00
acenaphthene (AcN)	0.35	0.37	0.48	0.70	0.99	0.99
fluorene (Fl)	0.30	0.61	0.62	0.66	0.78	0.79
phenanthrene (PhA)	0.79	0.83	0.83	0.90	0.93	0.95
anthracene (An)	0.65	0.92	0.93	0.93	0.93	0.94
fluoranthene (FlA)	0.77	0.87	0.88	0.96	0.97	0.97
pyrene (Py)	0.66	0.85	0.85	0.98	0.98	0.98
benz[*a*]anthracene (BaA)	0.91	0.95	0.95	0.97	0.97	0.97
chrysene (Chy)	0.96	0.96	0.96	0.97	0.98	0.98
benzo[*b*]+[*k*]fluoranthene (Bb+kFlA)	0.93	0.93	0.94	0.94	0.95	0.96
benzo[*e*]pyrene (BeP)	0.92	0.93	0.96	0.96	0.97	0.98
benzo[*a*]pyrene (BaP)	0.79	0.82	0.90	0.93	0.96	0.97
indeno[123-*cd*]pyrene (IP)	0.63	0.86	0.87	0.87	0.89	1.00
dibenz[*a,h*]anthracene (DBahA)	0.31	0.76	0.95	0.95	0.99	1.00
benzo[*ghi*]perylene (BghiP)	0.79	0.88	0.94	0.95	0.96	0.98
cumulative % variance	88.4	93.3	95.7	97.3	98.6	99.2
Exner function	0.13	0.07	0.06	0.04	0.03	0.02

TABLE 2. Sum of Squares for All PAH Compounds[a]

factor loadings (Figures 3 and 4)	literature PAH profiles (Figure 2) power plant	coal residential	coke oven	wood burning I	gasoline engine	diesel engine	traffic tunnel
1 of 2[b]	0.061	0.130	**0.007**	0.093	0.046	0.107	0.013
2 of 2	0.025	0.089	0.019	0.028	0.021	0.053	**0.015**
1 of 6	0.087	0.167	**0.016**	0.119	0.063	0.143	0.024
2 of 6[c]	0.033	0.092	0.063	0.018	**0.018**	0.042	0.038
3 of 6	0.077	0.123	0.036	0.090	0.051	0.103	**0.027**
4 of 6	0.128	0.220	**0.024**	0.168	0.100	0.202	0.055
5 of 6[d]	0.042	0.045	0.085	**0.049**	0.057	0.037	0.062

[a] Bold faced type indicates probable PAH source profile as discussed in the text. [b] IP concentration reduced to 30% of the value modeled. [c] BghiP concentration changed to the amount present in the literature gasoline engine profile. [d] PhA and FlA were not considered in the analysis because of high uncertainties in the literature source profiles for wood burning.

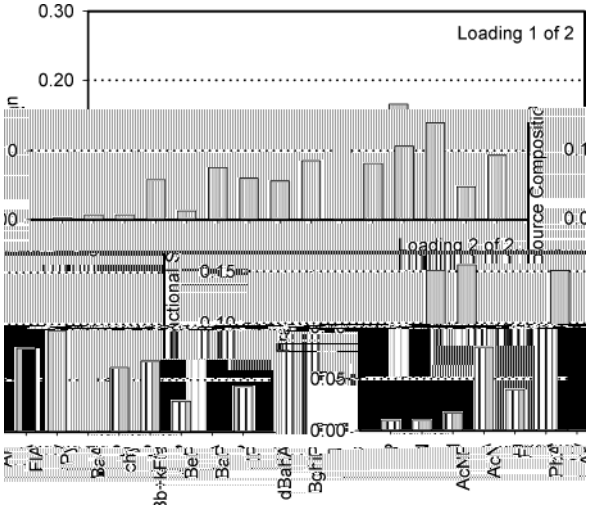

FIGURE 3. Factor loading plots for Lake Calumet two-source factor analysis solutions.

are overestimated, while benzo[*ghi*]perylene (BghiP) is underestimated. The percent contributions from the coke oven emissions and traffic tunnel were calculated to be 45% and 55%, respectively, based on the average of the four individual percent contributions from cores D, E, I, and K. This agrees well with the overall results of 52% from coal-related sources and 48% from traffic tunnel obtained from CMB modeling Operation #5 (*15*) and the results of 51% from coke oven and 42% from both gasoline and diesel engine emissions by CMB Operation #2 (*15*). Operations #2 and #5 were particularly significant statistically.

Note that the source profiles derived from two-factor solution of FA (Figure 3) match coke oven and traffic tunnel sources, not the "Coal Average" and "Traffic Average" shown in Figure 2 of ref *15*. In fact, these two averages show considerable similarities, which resulted in poor CMB modeling results (high percentage of source contribution estimates were "inestimable"). Averaging source profiles across categories diminishes the difference, resulting in unsuccessful model runs.

The agreements between the literature (Figure 2) and the model-derived source profiles and between the relative source contributions resulting from the CMB and FA models are remarkable. The CMB model relies on the availability and the adequacy of the source profiles of all major sources, from which the source contributions were computed using statistical techniques, such as, the effective variance weighted solution used in EPA's CMB8.2. In contrast, no *a priori* knowledge about the source emission is needed to run the FA model. Compared with previous applications of factor analysis for atmospheric apportionment (*9, 24*), our FA modeling with nonnegative constrains has the advantage of detailed comparison with literature source profiles and with results of CMB modeling.

The six-source factor-loading solution is presented in Figure 4. By comparing with the literature source profiles shown in Figure 2, and sum of squares calculations in Table 2, the six sources include two coke oven sources, a gasoline engine source, a traffic tunnel source, a wood burning/ residential coal source, and a loading dominated by IP.

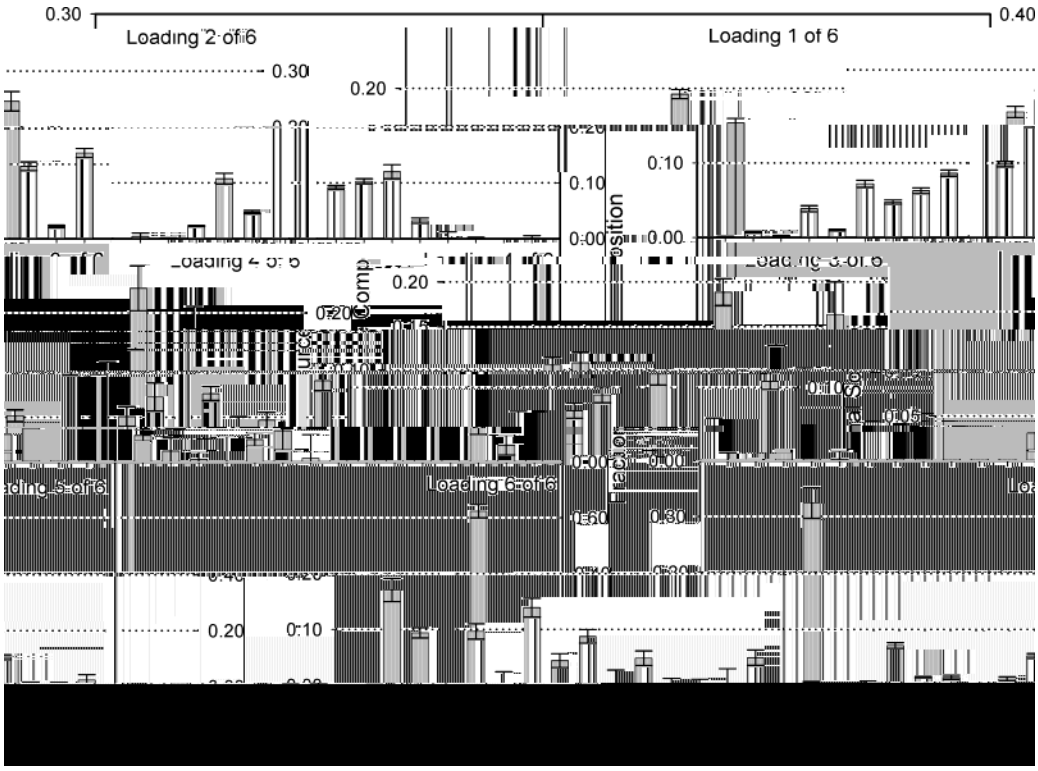

FIGURE 4. Factor loading plot for Lake Calumet six-source factor analysis solution with error bars representing the standard deviation of the mean for nine FA model runs using data sets created by Monte Carlo simulation.

Loading 1 of 6 is a coke oven profile. The model is not able to reproduce BaP and IP really well, i.e., the modeled profiles do not match the literature profiles. The model reproduces most other PAHs accurately, as supported by the sum of squares calculation (Table 2) for coke oven, (loading 1 of 6) which is the lowest.

Loading 2 of 6 represents a gasoline engine profile. A distinct pattern is observed between the literature and modeled profiles for the PAHs FlA, Py, BaA, Chy, Bb+kFlA, BeP, and BaP. This pattern is not present in other literature source profiles, such as wood burning and power plant. The model has difficulties reproducing BaP and BghiP to the fractional source compositions from the literature. Uncertainty analysis does indicate BaP and BghiP are present in some simulations. For the sum of squares calculation the concentration of BghiP was increased to that of the gasoline engine literature profile to determine if the remaining PAHs were modeled well. Loading 3 of 6 is a traffic tunnel profile. Low molecular weight PAHs are reproduced fairly accurately by the model; however, some higher molecular weight PAHs such as BaP, IP, and dibenz[*a*,*h*]anthracene (dBahA) vary from the literature profiles. The variance of IP and dBahA could result from the poor separations between IP and dBahA (*15*), and also the uncertainty for dBahA is the largest observed. The uncertainty for BaP is also significant. Sum of squares calculations (Table 2) indicate traffic tunnel is the best fit to model data.

An interesting point from the above discussion is that there are two traffic sources generated by the model. Considering the primary traffic source to be I-94, located just west of Lake Calumet (Figure 1), one may wonder how this is possible. We believe the separate sources are a result of varying inputs from traffic sources. For example, I-94 may have primary gasoline engine traffic during daytime hours and more of a mixture of traffic tunnel sources during the nighttime. The split in traffic sources could produce two sources that are distinguished by the FA model.

Loading 4 of 6 is another coke oven profile. This profile is low in PhA and Py; however, it is enriched with BaP and An. Uncertainty analysis does indicate PhA and Py are present and quite uncertain. This profile could result from a separate coke oven facility, possibly in Gary, Indiana. This would indicate the first coke oven source is from the coke plant located two miles north of site J. Another explanation for the second coke oven source is that it represents a degraded coke oven profile, with losses of PhA and Py observed. It is also possible that the same coke oven emits two different signals based on fuel source, installation of controls, and so forth, although the prevalence of the secondary profile (Loading 4 of 6) in core I and the primary profile (Loading 1 of 6) in core D (Figure 5) suggest two different coke oven sources.

Loading 5 of 6 has properties of a wood burning profile or a mix of a wood burning and residential coal profile. Most of the literature sources for wood burning do not measure acenaphthene (AcN); however, it is present in loading 5 of 6 and is much more abundant than indicated by the literature sources. Two outlier values of AcN in the data set, J-3 and I-5, may be influencing the FA model. PhA is very abundant in loading 5 of 6, possibly a result of combined residential coal and wood burning source or an artifact of the model. FlA is abundant in wood burning profiles but only seen in the uncertainty analysis for loading 5 of 6. The wood burning literature profiles are uncertain as seen with Wood Burning I and Wood Burning II (Figure 2), thus variance is expected between loading 5 of 6 and a literature wood burning profile. The sum of squares calculation (Table 2) for loading 5 of 6 eliminates PhA and FlA because of high uncertainties in the literature source profiles for wood burning. As a result of the small percent contribution (2.3%), from loading 5 of 6, it is possible that other PAH sources such as power plant are influencing the source profile. Source contributions from loading 5 of 6 will be discussed below and provide evidence for a wood burning or combined wood burning - residential coal source. Loading 6 of 6 is dominated by IP and believed to be a result of the high IP concentrations in core K.

To examine if high IP values in core K would account for a factor of six, a FA model run was made where core K was

With the exception of 1997, traffic profiles are fairly constant in Figure 5; however, a small peak in 1988 is observed. Wood burning and coal residential have small contributions in both models. However, in Figure 5 there appears to be a minimum between 1982 and 1992, indicating that a wood burning source may be more likely than a coal residential source (*18*). The variance between Figure 6 and Figure 5 may be attributed to several outlier IP values in core K, with Figure 5 probably being more accurate since IP here is represented by a separate factor.

Overall, combined coke oven (47%) and combined traffic (45%) have the largest PAH contribution to Lake Calumet. Traffic contributions increase significantly after the 1960s, and there seems to be a consistent maximum contribution in the late 1980s and early 1990s. Coke oven contributions are highest in core K, which is located near the coke oven facility. Wood burning is a minor PAH source. Loading 5 of 6 (Figure 5) is taken as the mix of residential coal combustions and wood burning rather than only residential coal combustion. The reason is that this source contribution is observed to continue after the 1950s, while coal was not used as domestic heating fuel much after the 1950s, and cores D and I have minimum wood burning values in the 1950s and 1960s. From the literature (*18*, *25*, *26*), a wood burning minimum in U.S. consumption is seen in the 1960s.

Compared with previous FA models for atmospheric PAHs, we show here that coke oven emissions is the most important coal-related PAH source, and we provide uncertainties of the estimated source profiles. In addition, time records of wood consumption are used to demonstrate that the low-molecular weight dominated PAH source profile is likely to be from wood burning rather than residential coal.

Acknowledgments

This work was supported by the U.S. National Science Foundation grant number BES-0107402. Student support is also provided by the Pilot Grant Program of the Illinois Education and Research Center, National Institute of Occupational Safety & Health (NIOSH) Award #: T42/CCT510424-09.

Supporting Information Available

Additional references, a summary of the wood type and phase measured (Table S1), and a summary of literature PAH source compositions normalized to benzo[*e*]pyrene (Table S2) for wood burning I and II. This material is available free of charge via the Internet at http://pubs.acs.org.

Literature Cited

(1) Neff, J. M. *Polycyclic Aromatic Hydrocarbons in the Aquatic Environment: Sources, Fates, and Biological Effects*; Applied Science Publishers: London, 1979.

(2) Harrison, R. M.; Smith, D. J. T.; Luhana, L. *Environ. Sci. Technol.* **1996**, *30*, 825 - 832.

(3) Baek, S. O.; Field, R. A.; Goldstone, M. E.; Kirk, P. W.; Lester, J. N.; Perry, R. *Water, Air, Soil Pollut.* **1991**, *60*, 279 - 300.

(4) Christensen, E. R.; Li, A.; Ab Razak, I. A.; Rachdawong, P.; Karls,