Electronic Supplementary Material (ESI) for Environmental Science: Processes & Impacts.

Supporting Information

Poor Regulation Implications in a Low and Middle Income Country Based on PAHs Source Apportionment and Cancer Risk Assessment

Wiaam Jaafar[§], *Vera Zaherddine*[§], *Fatima Hussein, Najat Aoun Saliba, and Nathalie Hayeck**[&]

Department of Chemistry, Faculty of Arts and Sciences, American University of Beirut, Beirut, Lebanon

[&]Department of Natural Sciences, Lebanese American University, Chouran, Beirut 1102-2801, Lebanon

Corresponding Author Information

*Corresponding Author:

Nathalie Hayeck, Tel: +961 1 786456/3412. E-mail: nathalie.hayeck@lau.edu.lb

[§] Contributed equally

Source Apportionment using Positive Matrix Factorization

This model works by solving the chemical mass balance (CMB) equation:

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

where, X_{ij} refers to the concentration for the j_{th} specie in the i_{th} sample; p represents the number of factors; g_{ik} is the relative contribution of factor k to i_{th} sample; f_{kj} refers to the profile factor of each source for the j_{th} specie; and e_{ij} is the PMF residual error for the i_{th} sample and the j_{th} species not fitted by the model.[25]

An optimal source apportionment analysis is accomplished when the minimal value of Q_{robust} (Eq.2) for a given p is found. Q_{robust} is the goodness of fit parameter calculated excluding outliers. However, Q_{true} is the goodness of fit parameter calculated including all collected data. For a stable and reliable solution, Q_{robust} and Q_{true} should be comparable with a ratio of Q_{true}/Q_{robust} less than 1.5, meaning that the outliers are not affecting the modeling of the system.[23]

$$Q = \sum_{i=1}^{n} \sum_{j=0}^{m} \left(\frac{e_{ij}}{u_{ij}} \right)^2 \qquad (Eq.2)$$

Here, n refers to the number of samples and m to the number of species; and u_{ij} refers to the uncertainty of the measured concentration for the i_{th} sample and the j_{th} species.[27]

To evaluate the robustness solution for the system, additional steps are done such as signal-to-noise ratio, base model, bootstrap (BS), displacement (DISP), and bootstrap-displacement (BS-DISP) that are offered by the model, further explanation about these steps are found in the supporting information.

In this study, 55 samples of 16 PAHs were placed in the EPA PMF 5.0 model with their uncertainties to estimate the different source contributions to the measured PAHs.

Measurement uncertainty (u_{ij}) is very critical to determine the optimal solution to the analytical equation by minimizing the function Q. In the absence of samples replicates, the analytical uncertainties cannot be calculated, however uncertainties can be estimated based on uncertainty equations.

Four of the most commonly used equations (Table 1) were tested to identify the best uncertainties to our study.

Table S1: Methods for calculating uncertainties for PMF analyses

PMF Uncertainty (u_{ij})	References
$\sigma ij = 0.05 \times Xij + Dij$	28
$U_{ij} = 0.1N_{ij} + \frac{D_{ij}}{3}$	29
$\sqrt{(Error\ Fraction \times Conc.)^2 + MDL^2}$	30
$\sigma ij = C1(Nij + \bar{N}j)$ then, $sij = \sigma ij + C3Nij$	31, 32

Where: X_{ij} and N_{ij} are the concentrations of the species; D_{ij} and MDL are the detection limits; C_1 and C_3 are constants between 0.01-0.05 and 0.1-0.5 respectively; $\bar{N}j$ is the average concentration of species.

Each equation was tested based on the five essential steps, explained in the supporting information, to determine which uncertainty calculation compliments the concentrations. $\sigma ij = C_1(N_{ij} + \bar{N}j)$ then,

$s_{ij}=\sigma_{ij}+C_3N_{ij}$ showed the best fit for our data, with C_1 and C_3 being 0.02 and 0.2 respectively, which were determined by trial-and-error tests.

To evaluate the solution for the system, five essential steps are required: Signal-to-noise ratio, base model, bootstrap, displacement, and bootstrap-displacement. The signal-to-noise ratio indicates if the measurements are above or within the noise (detection limit) of the data. The model categorizes the signals as "strong" if $S/N > 2$, "weak" if $0.2 \leq S/N \leq 2$, and "bad" if $S/N < 0.2$.[1] If a species is classified as weak, then it is down weighted by the PMF model, and if a species is classified as bad, it is excluded from the PMF analysis.

The base model run determines the Q_{true} and Q_{robust}. The ratio of Q_{true}/Q_{robust} should be smaller than 1.5 to indicate that the outliers are not affecting the modeling of the system.[2] Another output of the base model run is the residual analysis, which is the degree of adjustment of each species by the model. An ideal result is when the residual analysis is normally distributed between ±3 for each species.[3]

The bootstrap (BS) calculation validates the base model by predicting the results using the developed model. A model represents a good fit if the bootstrap analysis shows values of 70-75 and higher for each factor.[2]

The displacement (DISP) runs the solution infinite times to give the maximum variation in Q_{robust} (dQmax). A model is considered a good fit if the displacement analysis returns a dQmax less than 1% of the initial Q_{robust}.[3]

The last step to be discussed is the bootstrap-displacement (BS-DISP) run. A combination of the bootstrap and displacement analysis evaluates the effect of random errors and returns with a decrease in Q_{robust}. If the decrease in Q_{robust} is less than 0.5% then the model is a good fit and the analyses can be continued.[3]

Number of Factors from PMF model

Using Eq. 4 from Table 1 with C_1 and C_3 determined as 0.02 and 0.2, respectively, 2 to 6 factors were tested to identify the number of sources that contributed to PAHs using PMF version 5.0. The optimal number of three factors was chosen based on the signal-to-noise ratio, Q_{true}/Q_{robust} ratio of 1, scaled residuals normally distributed between ±3, BS of 96%, 96%, and 92%, DISP and BS-DISP decrease in dQmax by less than 1% and 0.5%, respectively.

Cancer Risk Calculation

Table S2: Toxicity Equivalence Factor (TEF) for each PAH

PAHs	TEF[4]	**Carcinogenic Group (IARC 2018)**
Naphthalene (Nap)	0.001	2B
Acenaphthylene (Acy)	0.001	-
Acenaphtene (Ace)	0.001	3
Fluorene (Flu)	0.001	3
Phenanthrene (Phe)	0.001	3
Anthracene (Ant)	0.01	3
Fluroanthene (Flt)	0.001	3
Pyrene (Pyr)	0.001	3
Benzo[a]anthracene (BaA)	0.1	2B
Chrysene (Chr)	0.01	2B
Benzo[k]fluoranthrene (BkF)	0.1	2B
Benzo[b]fluoranthrene (BbF)	0.1	2B
Benzo[a]pyrene (BaP)	1	1
Benzo[g,h,i]perylene (BghiP)	0.01	3
Dibenz[a,h]anthracene (DBahA)	1	2A
Indeno[1,2,3-c,d]pyrene (IP)	0.1	2B

PAHs concentrations and Weather Data

Table S3: Total PAHs with the recorded temperature, wind speed and direction for each sampling day.

Sampling Date	Total PPAH (ng/m^3)	Temperature (°C)	Wind Speed (m/s)	Wind Direction
11/16/2016	15.4	23.0	6.8	S/SW
11/18/2016	11.4	22.0	2.8	N/NE
11/24/2016	12.3	23.0	0.5	E/SE
11/30/2016	11.0	22.0	7.2	S/SE
12/5/2016	14.4	21.0	2.0	E/NE
12/11/2016	14.6	19.0	2.6	S
12/17/2016	16.0	16.0	1.3	E/NE
12/28/2016	14.3	16.0	3.9	S
1/2/2017	15.2	14.0	7.9	S/SE
1/8/2017	12.7	16.0	10.6	S
1/14/2017	11.1	16.0	0.7	SE
1/24/2017	14.9	17.0	0.9	E/SE
1/30/2017	15.8	13.0	3.1	E
2/6/2017	16.3	17.0	1.9	E/SE
2/12/2017	9.6	16.0	1.9	E/SE
2/20/2017	17.8	16.0	2.4	E/NE
2/26/2017	15.6	13.0	2.7	E/NE
3/5/2017	9.5	18.0	4.4	N/NE
3/11/2017	11.3	19.0	10.0	S/SW
3/17/2017	10.2	17.0	7.3	S
3/23/2017	10.1	21.0	3.6	W
3/29/2017	8.5	21.0	3.6	N/NE
4/4/2017	10.6	21.0	4.9	N/NE

4/10/2017	7.8	21.0	5.4	N
4/16/2017	8.3	20.0	4.8	SW
4/22/2017	8.1	26.0	1.2	E
4/26/2017	9.7	25.0	5.0	N
5/2/2017	8.4	22.0	4.7	SW
5/8/2017	10.0	25.0	5.4	SW
5/14/2017	9.6	25.0	5.2	W/SW
5/20/2017	8.1	24.0	7.6	S/SW
6/11/2017	6.8	27.0	5.2	SW
6/17/2017	6.5	28.0	5.5	W/SW
6/23/2017	6.6	28.0	4.2	W/SW
6/29/2017	7.7	29.0	4.8	W/SW
7/6/2017	5.1	29.0	4.7	W/SW
7/12/2017	6.9	31.0	4.6	W/SW
7/18/2017	5.3	31.0	4.3	W
7/24/2017	5.5	31.0	4.2	W/SW
8/8/2017	6.3	30.0	4.2	N/NW
8/14/2017	6.2	30.0	4.5	W/SW
8/21/2017	6.1	30.0	3.9	W
8/27/2017	6.6	32.0	5.2	S/SW
9/2/2017	8.3	28.0	3.8	SW
9/8/2017	9.1	28.0	4.2	N
9/14/2017	9.3	30.0	4.0	N
9/20/2017	9.3	29.0	4.8	SW
10/10/2017	10.5	23.0	4.1	SW

10/16/2017	9.2	23.0	2.8	N
10/19/2017	10.2	28.0	1.5	E/SE
10/25/2017	7.5	25.0	2.6	SW
11/10/2017	13.3	22.0	1.7	E/NE
11/14/2017	11.4	22.0	3.3	S/SW
11/18/2017	15.6	23.0	1.8	E/SE
11/22/2017	13.3	18.0	2.6	W/NW

Figure S1 shows the negative correlation between total PAHs and wind speed, where during high wind speed, the total sum of PAHs captured is low and vice versa. The wind speed and direction were obtained from Earth NullSchool (https://earth.nullschool.net/). Samples were collected on non-rainy days. This is in good accordance with data recorded by other studies.[5-7]

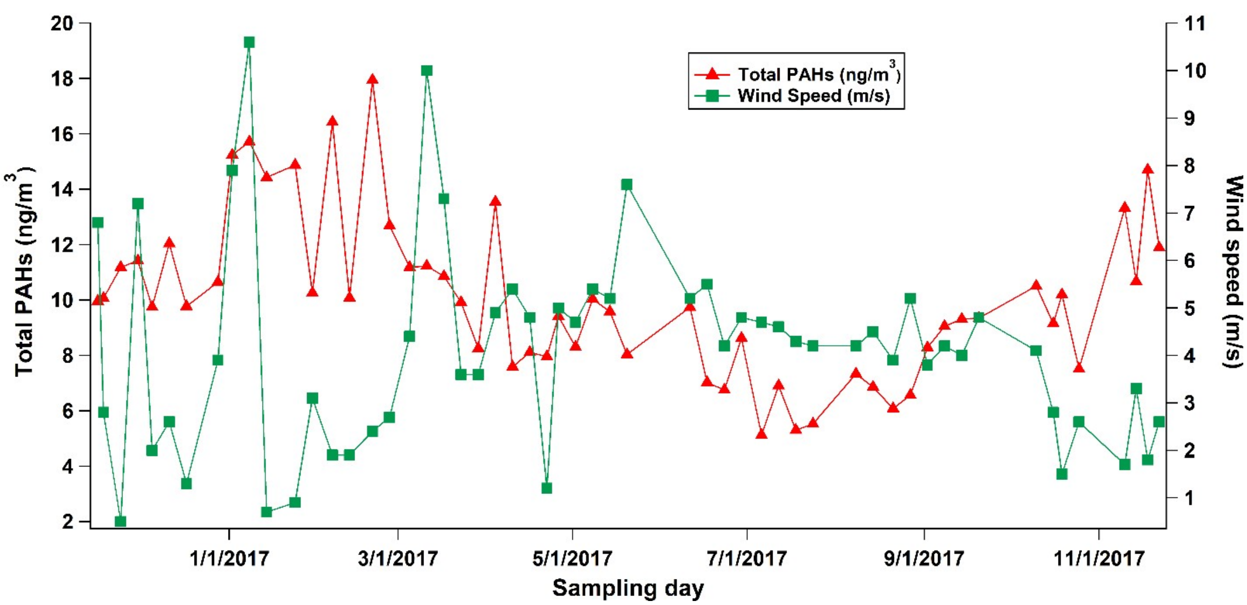

Figure S1: Variation of concentration of total PAHs with respect to wind speed over the period extended between November 2016 and November 2017

References

1. P. Paatero and P. K. Hopke, *Analytica Chimica Acta*, 2003, **490**, 277-289.
2. C. Belis, O. Favez, M. Mircea, E. Diapouli, M.-I. Manousakas, S. Vratolis, S. Gilardoni, M. Paglione, G. Močnik, D. Mooibroek, S. Takahama, R. Vecchi, P. Paatero, P. Salvador and S. Decesari, *European guide on air pollution source apportionment with receptor models - Revised version 2019*, 2019.
3. G. Norris and R. Duval, EPA Positive Matrix Factorization (PMF) 5.0 Fundamentals and User Guide, https://cfpub.epa.gov/si/si_public_file_download.cfm?p_download_id=523318&Lab=NERL).
4. I. C. Nisbet and P. K. Lagoy, *Regulatory toxicology and pharmacology*, 1992, **16**, 290-300.
5. A. C. Amarillo and H. Carreras, *Atmospheric Pollution Research*, 2016, **7**, 597-602.
6. M. Amodio, M. Caselli, G. de Gennaro and M. Tutino, *Environmental Research*, 2009, **109**, 812-820.
7. A. Nadali, M. Leili, A. Bahrami, M. Karami and A. Afkhami, *Ecotoxicology and Environmental Safety*, 2021, **209**, 111807.