Data Supplement

Title:
Ambient $PM_{2.5}$ temporal variation and source apportionment in Mbarara, Uganda

Authors:
Silver Onyango[1]*, Crystal M. North[2-4]*, Hatem A. Ellaithy[2], Paul Tumwesigye[5], Choong-Min Kang[3], Vasileios Matthaios,[3,6] Martin Mukama[5], Nuriat Nambogo[5], J. Mikhail Wolfson[3], Stephen Ferguson[3], Data Santorino[1,5], David C. Christiani[2-4], Petros Koutrakis[3]
*Co-first author

Affiliations:
[1]Mbarara University of Science and Technology, Mbarara, Uganda
[2]Massachusetts General Hospital, Boston, MA
[3] Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA
[4]Harvard Medical School, Boston, MA
[5]Consortium for Affordable Medical Technologies, Uganda
[6]School of Geography Earth and Environmental Science, University of Birmingham, UK

Methods – Source Apportionment

The mathematical background of PMF analysis has been comprehensively described.[1] Briefly, this statistical method uses a mass balance equation, which in the receptor model is expressed as:

$$X_{ij} = \sum_{k=1}^{p} G_{ik} F_{kj} + E_{ij} \quad (1),$$

Here, X_{ij} is the concentration of j species measured in sample i and G_{ik} is the species contribution of the k source to sample i. F_{kj} (frequently reported as "source profiles") is the fraction of j species from the k source while E_{ij} is residual associated with the j species concentration measured in the i sample. Finally, p denotes the total number of sources. The goal of the model is to reproduce x_{ij} matrix by finding values for G_{ik} and F_{kj} matrices for a given P. The values of G_{ik} and F_{kj} matrices are adjusted until a minimum Q "the loss function" for a given P is found.[2] PMF solves the receptor modelling problem by minimizing the loss function Q based on the uncertainty of each observation by the following equation:

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

where σ_{ij} is an estimate of the uncertainty for the j^{th} species in the i sample, n is the number of samples and m is the number of species.

To evaluate the reproducibility of the PMF solution and the adequate number of PMF factors three validation methods were used: bootstrap (BS), displacement (DISP) and bootstrap-displacement (BS-DISP). BS is a re-sampling method that explores random errors in data values and identifies observations that may affect the solution. A base bootstrap model method was carried out, executing 100 iterations, using a random start and a minimum Pearson correlation coefficient (R-value) of 0.65 as suggested in the literature.[3–5] All the modelled factors were well reproduced over at least 90% of runs, indicating good model results. The remaining 10% was distributed amongst the existing factors, while it should be noted that no runs were unmapped (unmapped is considered a factor when the bootstrap factor is not correlated to any of the base factors). For all the factors, 90% of the species (of the base run) were within the interquartile range (25th to 75th percentile) of the bootstrap runs. DISP explores the rotational ambiguity by assessing the largest range of source profile values. In DISP validation no factor swaps were observed. BS-DISP is a combination of the former two validation methods and examines random errors in conjunction with rotational ambiguity of the factors. In the BS-DISP, 94% of the PMF solutions were accepted and 3 (traffic related), 4 (fine soil) and 2 (secondary aerosol & biomass burning and heavy oil & fuel combustion) factor swaps were observed. These swaps are attributed to the fact that these sources have common element tracers that are overlapping. For example earth crustal elements such as Al, Si Mg, Ca can be present in both traffic related and fine soil dust sources,[6] while S, K, Pb have been reported in secondary aerosol & biomass burning and heavy oil & fuel combustion sources.[7] Further, solution validation metrics included 1) the symmetric distribution of scaled residuals (±3σ), 2) the investigation of species $Q/Q_{expected}$ values ≤ 2 (Supplemental Figure 1), and 3) inter-relationship investigations between the predicted and observed volume mixing ratios. Supplemental Figure 2 shows the correlation between total $PM_{2.5}$ reconstructed concentrations from all six factors with observed $PM_{2.5}$ concentrations which had an $R^2 = 0.924$ indicating good agreement between the receptor model and the observations.
and the observations.

Table S1. Ambient $PM_{2.5}$ concentrations[a] by day of week

Characteristic	Monday N = 29	Tuesday N = 29	Wednesday N = 29	Thursday N = 25	Friday N = 28	Saturday N = 26	Sunday N = 20	p-value[b]
Mbarara								
$PM_{2.5}$								0.74
Median (IQR)	26.7 (17.8, 37.7)	28.9 (23.6, 32.1)	27.1 (22.8, 38.4)	27.5 (22.2, 36.6)	25.7 (22.8, 31.6)	26.7 (23.1, 35.8)	23.6 (17.7, 31.7)	
Range	7.4, 47.7	9.9, 43.3	15.7, 62.7	11.5, 173.1	10.7, 45.6	15.2, 56.4	11.8, 44.6	
AQI Category								
Good	4 (14%)	1 (3.4%)	0 (0%)	1 (4.0%)	1 (3.6%)	1 (3.8%)	3 (15%)	
Moderate	20 (69%)	26 (90%)	23 (79%)	18 (72%)	25 (89%)	24 (92%)	14 (70%)	
Unhealthy for Sensitive Groups	5 (17%)	2 (6.9%)	6 (21%)	4 (16%)	2 (7.1%)	1 (3.8%)	3 (15%)	
Unhealthy	0 (0%)	0 (0%)	0 (0%)	1 (4.0%)	0 (0%)	0 (0%)	0 (0%)	
Very Unhealthy	0 (0%)	0 (0%)	0 (0%)	1 (4.0%)	0 (0%)	0 (0%)	0 (0%)	
Above 2006 WHO Guideline	16 (55%)	19 (66%)	17 (59%)	16 (64%)	16 (57%)	16 (62%)	9 (45%)	0.85
Kampala								
$PM_{2.5}$								0.54
Median (IQR)	56.3 (50.0, 65.4)	57.2 (50.9, 66.8)	64.7 (53.3, 75.8)	55.7 (47.6, 70.2)	60.0 (47.7, 68.7)	58.7 (54.2, 62.5)	60.2 (50.4, 69.2)	
Range	37.3, 106.1	32.6, 86.3	35.2, 84.1	23.8, 90.0	24.5, 105.9	25.4, 97.8	25.7, 90.9	
AQI Category								
Good	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	
Moderate	0 (0%)	1 (3.0%)	0 (0%)	2 (5.9%)	2 (5.9%)	2 (6.2%)	1 (3.0%)	
Unhealthy for Sensitive Groups	27 (82%)	27 (82%)	18 (56%)	24 (71%)	26 (76%)	25 (78%)	24 (73%)	
Unhealthy	6 (18%)	5 (15%)	14 (44%)	8 (24%)	6 (18%)	5 (16%)	8 (24%)	
Very Unhealthy	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	
Above 2006 WHO Guideline	34 (100%)	36 (100%)	33 (100%)	34 (97%)	33 (97%)	32 (100%)	33 (100%)	0.87

[a]n (%), unless otherwise specified

[b]Kruskal-Wallis rank sum test or Pearson's Chi-squared test, as appropriate

$PM_{2.5}$: particulate matter less than 2.5 microns in diameter; IQR: interquartile range; AQI: air quality index; WHO: world health organization

Table S2. Mean concentration of species (ug/m^3) at the Mbarara sampling site.

Element or Parameter	Mean	Standard Deviation	Minimum	Maximum	Median
BC	3.47999	2.05442	1.49653	19.65143	3.11959
K	1.17434	0.72752	0.3997	6.3032	1.0026
Si	0.79256	0.64382	0.0584	5.5214	0.6769
Al	0.68153	0.4306	0.0661	3.5688	0.5963
S	0.59717	0.51713	0.1802	4.4334	0.4882
Fe	0.47059	0.3097	0.0331	2.5504	0.4296
Cl	0.25281	0.22597	0.0335	1.1761	0.176
Na	0.12779	0.07695	0.0243	0.583	0.1145
Ca	0.11121	0.13018	0.0087	1.0993	0.0852
Mg	0.06208	0.07601	0	0.6286	0.0459
Ti	0.03392	0.02284	0.0026	0.1941	0.0309
Br	0.01584	0.00807	0.0054	0.0607	0.0142
Zn	0.01367	0.02074	0.0016	0.1748	0.0099
Pb	0.0084	0.00733	0.0005	0.0457	0.007
Mn	0.00765	0.0056	0	0.0464	0.0072
Cu	0.00574	0.00962	0.0013	0.0804	0.0039
Sr	0.00309	0.00216	0.0005	0.0143	0.0028
V	0.00159	0.00195	0	0.0164	0.0013
Ni	0.00152	0.00183	0	0.0104	0.001
Co	0.00127	0.00136	0	0.0077	0.0011

Figure S1. Comparison of ambient $PM_{2.5}$ concentrations by season of measurement, weekday versus weekend, and day of the week

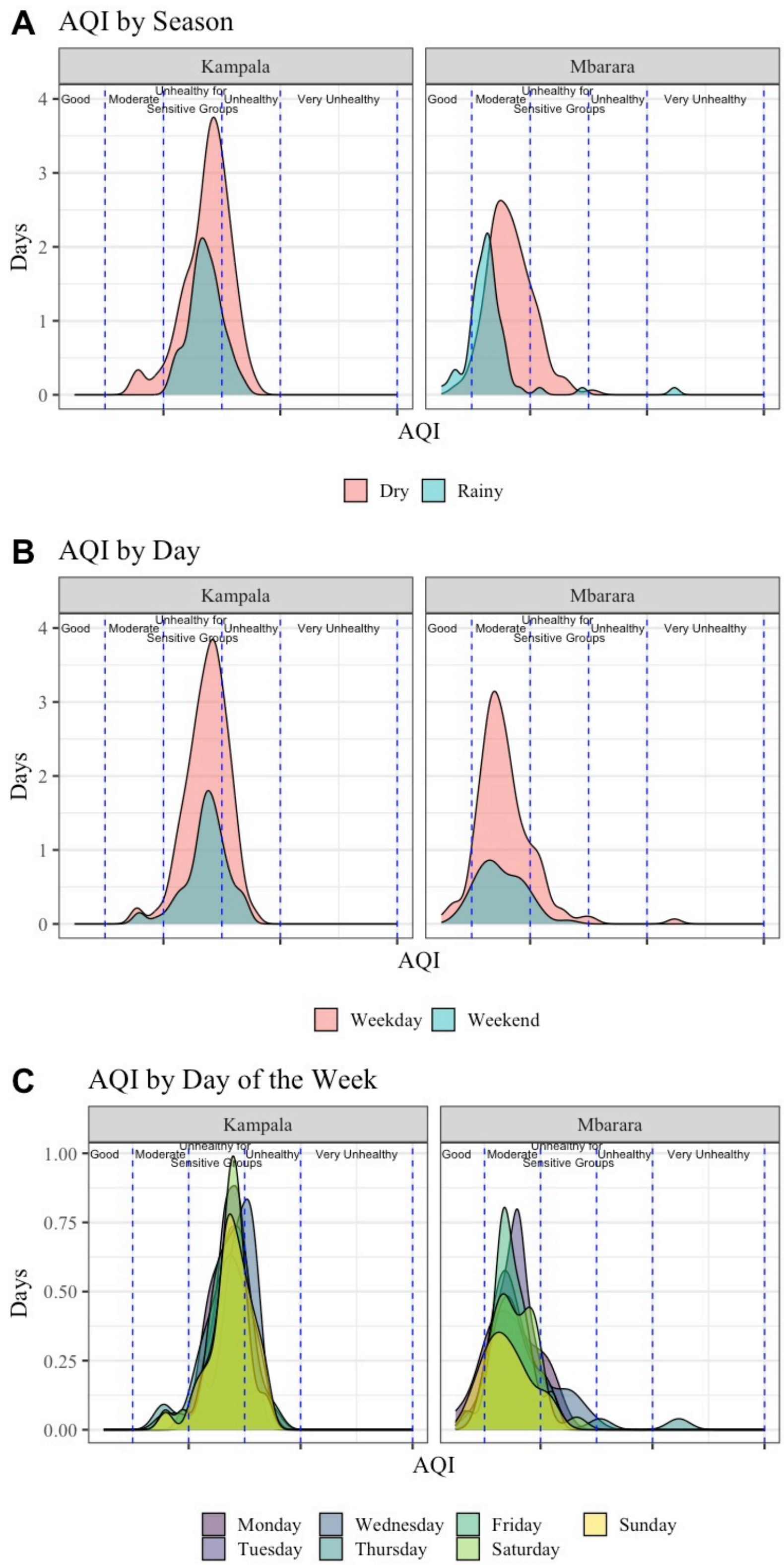


Figure S2. Species specific Q/Qexp values

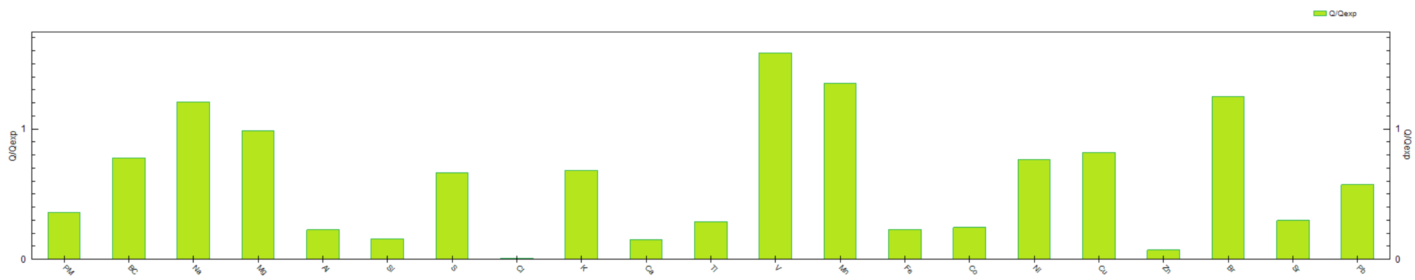
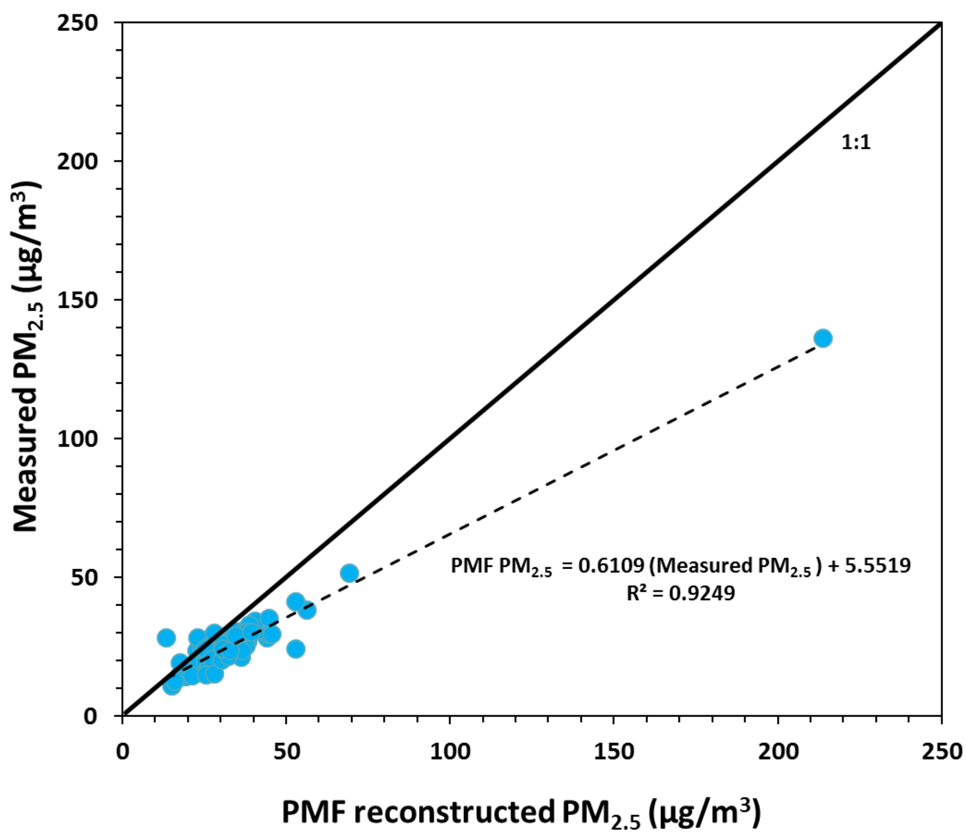


Figure S3. Observed and predicted $PM_{2.5}$ values from the six factor solution.



References

1. Paatero, P. & Tapper, U. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics* **5**, 111–126 (1994).
2. Paatero, P., Hopke, P., Song, X.-H. & Ramadan, Z. Understanding and Controlling Rotations in Factor Analytic Models. *Chemometrics and Intelligent Laboratory Systems* **60**, 253–264 (2002).
3. Bressi, M. *et al.* Sources and geographical origins of fine aerosols in Paris (France). *Atmospheric Chemistry and Physics* **14**, 8813–8839 (2014).
4. Baudic, A. *et al.* Seasonal variability and source apportionment of volatile organic compounds (VOCs) in the Paris megacity (France). *Atmospheric Chemistry and Physics* **16**, 11961–11989 (2016).
5. Matthaios, V. N. *et al.* Quantifying factors affecting contributions of roadway exhaust and non-exhaust emissions to ambient PM10–2.5 and PM2.5–0.2 particles. *Science of The Total Environment* **835**, 155368 (2022).
6. Matthaios, V. N. *et al.* Sources of indoor PM2.5 gross α and β activities measured in 340 homes. *Environmental Research* **197**, 111114 (2021).
7. Liu, J. *et al.* Emission control priority of PM2.5-bound heavy metals in different seasons: A comprehensive analysis from health risk perspective. *Sci Total Environ* **644**, 20–30 (2018).