Elemental characterization of personal filter samples from the Household Air Pollution Intervention Network (HAPIN) trial for Source Apportionment

Erick E. Mollinedo

2024-04-19

## 1 Abstract

Household air pollution is one the most prominent risks in low- and middle-income countries. Fine particulate matter (PM2.5) is a major air pollutant from household air pollution. The chemical and physical composition of PM2.5 depends heavily on its sources. Source apportionment is a methodology that aims to identify sources of pollution to reconstruct their impacts, using methods such as positive matrix factorization (PMF). The current project aims to conduct the initial analysis before conducting PMF modeling, to describe the most important chemical species in personal exposure filter samples from rural Guatemala. Six-hundred and twenty-nine PTFE filter samples from the HAPIN trial, a randomized fuel intervention controlled trial in Guatemala, were analyzed using X-Ray fluorescence to determine the presence of inorganic elements. Transmissometry was used to determine the black carbon (BC) concentrations and gravimetry to determine the PM2.5 concentrations. Multivariate linear regression models and gamma GLM models were used to predict which elements can be used to determine total PM2.5 mass concentrations and which questionnaire based exposures can infer BC mass concentrations. It was determined that the concentrations were significantly lower in samples from users that cooked using liquefied petroleum gas (LPG) fuel, compared to users that cooked with biomass fuels. The linear regression model was the most appropriate to determine the apportionment of each chemical species to total PM2.5 concentrations, and BC, magnesium and potassium seemed the most important predictors. The gamma GLM model was the most fit to predict the impact of categorical exposures on BC concentrations, where type of fuel, exposure to kerosene and a neighboring biomass stove were the most significantly associated exposures. However, caution should be taken into using questionnaire exposures to infer predictions. In addition, the ground for future source apportionment analyses was founded through these statistical analyses. The inclusion of additional chemical species could better inform the total composition of PM2.5 so PMF can be used for source apportionment.

## 2 Introduction

Household air pollution (HAP), due to the combustion of solid fuels for cooking and heating, is one of the most prominent environmental health issues in low- and middle-income countries (LMICs) [1]. HAP has been listed as the second environmental health risk factor worldwide, just after outdoor air pollution. Exposure to HAP has been linked to more than 2.31 million deaths in 2019 and it is associated with adverse cardiovascular and respiratory health outcomes [2, 3]. Particulate matter with an aerodynamic diameter less than 2.5 micrometers (PM2.5) is one of the major air pollutants chemically and physically composed of crustal material, black carbon (BC), organic compounds, metals and others. Toxicity from PM2.5 depends, among many factors, on the chemical composition of the particles which heavily depend on the sources of emissions [4, 5].

The identification of point sources of pollution is important for the development of strategies to improve air quality. Source apportionment (SA) is a methodology to reconstruct the impact of emissions from different sources, a method applied to different pollutants such as particulate matter [6]. The most used SA methods includes receptor models, an example of this is the positive matrix factorization (PMF). One advantage of PMF is that it does not require the source profiles prior to analysis and there is no limit on the number of sources, however, it is required the knowledge of potential source profiles. Said requirement to know the potential sources is associated with the profiles of chemical species, given that the concentration of certain organic and inorganic species alongside their uncertainties can be a good estimate of the point sources [7]. Another important piece before conducting source apportionment is to correctly process the data (which includes concentrations and uncertainties), and to explore associations to try to identify the potential sources. Just a few studies about SA in LMICs have been published, therefore it is important to characterize chemical composition and potential sources of air pollutants in these settings [8–10].

The Household Air Pollution Intervention Network (HAPIN) trial was a randomized controlled trial conducted in four different LMICs, where a liquefied petroleum gas (LPG) stove was delivered as an intervention to half of the pregnant women participants, meanwhile the other half remained cooking using their biomass stove. The intervention was evaluated in terms of reduction of exposures and specific health outcomes by comparing the two study groups (Control and Intervention). Baseline measurements were taken to the participants, and then followed-up for about 18 months, conducting exposure measurements three times during pregnancy and three times after the birth of the child. A description of the HAPIN trial and the methods for exposure sampling can be found elsewhere [4, 11]. A source apportionment pilot study from HAPIN was conducted in 2022. This study was conducted in 64 and 59 personal exposure filter samples from the pilot phase in Guatemala and Rwanda, respectively. Twenty-two chemical species were analyzed in all samples using X-ray fluorescence (XRF), a non-destructive spectrometry analytical method, but only 12 species were detected at both study sites in comparable levels. Based on the chemical species detected, four potential sources of PM2.5 were identified in both sites. One study caveat is the small sample size and limited assessment of chemical composition and sources between study arms, so further studies to complement the findings are needed [12].

The following study aims to produce the clean dataset from the HAPIN trial in Guatemala, needed for PMF analysis. This study also aims to conduct main and exploratory analysis of this chemical species dataset to answer some questions before the PMF modelling: 1.) What chemical species are correlated among each other from these filter samples?, 2.) What are the differences in concentrations of the chemical species if the samples are categorized by type of fuel and by study arm? 3.) What chemical species are associated with the variations in concentrations of total PM2.5? and 4.) What exposures can be modeled as predictors of BC concentration?

## 3 Methods

### 3.1 Data collection

The main exposure data comes from filter samples of PM2.5 personal exposure measurements in pregnant women participants from the HAPIN trial, in Guatemala. The Enhanced Children’s MicroPEM (ECM) was the instrument used for the measurement of PM2.5 exposures. In summary, the instrument works as an air pump with a standard flow of 0.3 liters per minute, so the particles get into an inlet and get collected into the filter. In total, six hundred and forty eight sample filters were available, where eighteen correspond to field blanks. From the HAPIN trial study design, in total, six exposure measurements (visits) were conducted during pregnancy in around 800 participants, but for this scenario, only 648 filter samples were randomly selected for chemical species characterization. Other exposures where collected as categorical variables using standardized questionnaires administered in each visit. These exposures capture some of the sources of exposure such as tobacco smoke, trash burning, use of incense or coil, kerosene, exposure to other biomass stoves and others.

### 3.2 Concentration estimates

The filters were pre-weighed (before sampling) in a Sartorius cubis microbalance, and then post-weighed in the same instrument after sampling. The difference in mass deposition was calculated as the post-pre weigh difference. The average 24-hour PM2.5 concentrations were calculated using the following equation:

Where, M is the mass deposition of particles in the filter and V is the volume of air the sampler used to collect the particles.

Black carbon was estimated measuring transmissometry via a Magee Sootscan instrument. The black carbon 24-hour concentrations were calculated using the following equation:

Where, If is the post-attenuation and I0 is the pre-attenuation, A is the area of the filter, V is the volume of air the sampler used to collect the particles and is the mass absorption coefficient (a constant).

The concentration of 22 chemical species (elements) on the filters were determined using a non-destructive spectrometer technique: X-Ray Fluorescence (XRF), and the uncertainties were estimated also based on the XRF instrument.

### 3.3 Concentration and Uncertainties data processing

All data cleaning and processing were performed on R (version 4.3.2). First, all chemical species where more than 50% of the samples are below the detection limit (LOD) were filtered out, leaving a total of 10 chemical species. For these remaining species, the concentrations were adjusted based on the blank filters, and the values below LOD were replaced as their corresponding LOD divided by the square root of 2. Finally, the concentrations were transformed to micrograms per cubic meter (ug/m3). Also, the uncertainties were transformed to ug/m3 to match the units of the concentrations. In addition, the black carbon concentrations were added to add another chemical species to the analysis.

The uncertainties were processed by, first removing all the chemical species that had more than 50% of their samples below the detection limit, as seen above. Then, the uncertainties were adjusted based on equations derived from the law of propagation of uncertainty. In brief, the equations consider the individual uncertainties from each one of the variables, in this case the XRF instrument uncertainty, the area of the filter, volume, and the attenuation (in case of the Black carbon measurements).

### 3.4 Statistical analyses

All statistical analyses were performed on R (version 4.3.2). An initial exploration of the data was performed, to determining missing values for all possible predictors and outcomes and trends for all chemical species. These explorations served in order to skim down which variables to use as predictors and outcomes and to compute their summary statistics.

The correlation (r2) between chemical species were determined using paired correlations in a correlation plot matrix. To assess differences in concentrations between samples by type of fuel and by study arm, a Welch two-sample T-test was computed. Multivariate regression models were computed to assess which elements are likely to explain the variations in concentrations of total PM2.5 and to determine which exposures are most associated with the BC concentrations. For each one of the pollutants as outcomes (PM2.5 and BC), two different models were computed and assessed. Model 1 was defined as a linear multivariate regression model, meanwhile Model 2 was a Generalized linear model with gamma distribution, due to the approximation of the distributions observed for the concentrations of each chemical species. The PM2.5 model used all the chemical species to be included for PMF as predictors. For the BC model, the predictors were all the categorical exposures with two levels (Yes or No) registered from the questionnaires. The final model for each pollutant was reported and selected based on certain parameters such as the metrics, uncertainties and residuals.

## 4 Results

Six hundred and twenty-nine filter samples were considered in the final analyses, due to one filter sample that was damaged and the concentration was not available. Among the 22 chemical species, only 11 were included to be used for future PMF modeling, which means only 11 species had more than 50% of their observations above the limit of detection. Black carbon was also included into the chemical species, due to be a representative of a good portion of total PM2.5. Table 1 and Table 2 show a summary of the descriptive statistics for each chemical species by type of fuel. All the statistics are listed in ug/m3 units. As a note, all the observations of the baseline measurements were categorized as biomass fuel type, since at this point, the LPG intervention was not yet allocated. This explains the larger number of samples belonging to the biomass fuel type.

Table 1: **Summary statistics: Biomass fuel samples**

| Species | N | Mean | SD | Median | IQR | Min | Max |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Al | 375 | 0.68 | 0.46 | 0.59 | 0.43 | 0.01 | 2.63 |
| BC | 375 | 12.05 | 7.02 | 11.49 | 6.65 | 1.73 | 95.61 |
| Ca | 375 | 0.40 | 0.33 | 0.32 | 0.35 | 0.03 | 2.95 |
| Fe | 375 | 0.32 | 0.27 | 0.25 | 0.33 | 0.01 | 2.05 |
| K | 375 | 1.49 | 1.30 | 1.10 | 1.23 | 0.18 | 9.72 |
| Mg | 375 | 0.14 | 0.09 | 0.12 | 0.10 | 0.01 | 0.53 |
| Mn | 375 | 0.02 | 0.02 | 0.02 | 0.02 | 0.00 | 0.15 |
| S | 375 | 0.44 | 0.29 | 0.37 | 0.28 | 0.06 | 2.01 |
| Si | 375 | 0.91 | 0.92 | 0.56 | 1.11 | 0.01 | 4.43 |
| Ti | 375 | 0.04 | 0.03 | 0.03 | 0.04 | 0.01 | 0.28 |
| Zn | 375 | 0.01 | 0.02 | 0.01 | 0.01 | 0.00 | 0.18 |

Table 1: **Summary statistics: LPG fuel samples**

| Species | N | Mean | SD | Median | IQR | Min | Max |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Al | 254 | 0.65 | 0.40 | 0.53 | 0.42 | 0.01 | 2.02 |
| BC | 254 | 4.82 | 4.39 | 2.90 | 3.15 | 0.95 | 39.85 |
| Ca | 254 | 0.22 | 0.18 | 0.18 | 0.18 | 0.01 | 1.20 |
| Fe | 254 | 0.29 | 0.26 | 0.22 | 0.28 | 0.01 | 1.55 |
| K | 254 | 0.41 | 0.28 | 0.34 | 0.30 | 0.00 | 1.96 |
| Mg | 254 | 0.10 | 0.07 | 0.08 | 0.07 | 0.01 | 0.44 |
| Mn | 254 | 0.02 | 0.02 | 0.01 | 0.02 | 0.00 | 0.10 |
| S | 254 | 0.32 | 0.25 | 0.24 | 0.21 | 0.00 | 1.45 |
| Si | 254 | 0.76 | 0.75 | 0.57 | 1.02 | 0.01 | 3.77 |
| Ti | 254 | 0.04 | 0.03 | 0.03 | 0.04 | 0.01 | 0.19 |
| Zn | 254 | 0.01 | 0.01 | 0.00 | 0.01 | 0.00 | 0.06 |

### 4.1 Chemical Species correlations

In terms of correlations, Spearman correlation was estimated between all the chemical species and [Figure 1](#fig-correlations) shows the associations highlighting the R2 value. It is observed that the highest correlation is observed between Si and Fe, Ti-Fe, and Al-Si. These values could indicate that these species come from the same point sources, something to account when conducting the source apportionment analysis using PMF modeling.

|  |
| --- |
| Figure 1: Correlation plot for chemical species |

### 4.2 Chemical species concentrations

The concentration of the chemical species are distributed in different ranges depending on their type. For example, the highest concentrations were determined to be for Black carbon (BC), with values from 1 to 96 ug/m3. The remaining chemical species have lower concentrations, ranging from less than 0.1 ug/m3 to 10 ug/m3. Other noticeable patterns is that some of the species were found in higher concentrations for personal samples from participants that used biomass stoves, compared to LPG stoves ([Figure 2](#fig-concentrations)). When conducting the two-sided T-test, it was determined that all of the chemical species, except Fe, Al and Zn were significantly different between type of fuel. The differences in concentrations were also analyzed by study arm, however, given that all baseline measurements belonged to biomass fueltype observations, these results were not presented but can be observed in the Supplementary material (Figure 1).

|  |
| --- |
| Figure 2: Chemical species concentrations by type of fuel |

### 4.3 Chemical species modeling

#### 4.3.1 Modeling PM2.5

The PM2.5 model included aluminum, calcium, iron, potassium, magnesium, manganese, sulfur, silicon, titanium, zinc and black carbon as predictors in the final model. Between both computed models, the linear regression model was selected as the most appropriate. Table 3 summarizes three metrics used for the comparisons between both models: RMSE, R-squared and MAE. As seen from these metrics, there was a drastic change when comparing both types of models, showing an improvement in the linear model. This was confirmed due to the under-prediction of certain values as seen in the observed-predicted plots and residuals (Supplementary material, Figures 5 and 6) for the GLM gamma model.

Based on the coefficients and the statistical significance (p-value), it was determined that only magnesium, potassium and black carbon had a significant effect on total mass concentration of PM2.5. Black carbon and potassium had an increasing effect, meanwhile magnesium had a decreasing effect for predicting total PM2.5 mass concentration.

Table 1: Table 3

**Metrics summary: PM2.5 Models**

| Metric | Linear | GLM |
| --- | --- | --- |
| RMSE | 53.5700562 | 231.0584395 |
| R-squared | 0.7327571 | 0.3587276 |
| MAE | 32.1838830 | 54.1840008 |

#### 4.3.2 Modeling Black Carbon

For the model to predict black carbon concentrations, the predictors considered in the final model were: smoking, use of mosquito coils, trash burning, use of kerosene, cooking smoked foods, crop burning, exposed to a biomass stove (or a neighboring biomass stove) and type of fuel. Use of stove, incense and a generator for corn grinding were not included due to the small variability of answers (most or all of them coded as Yes or No). When comparing the linear and GLM regression models for BC, both seem to perform similarly based on the metrics (Table 4) and the residuals and observed vs predicted values plots (Supplementary: Figures 7 and 8).

The gamma GLM model was selected as the most appropriate, given its ability to more accurately discriminate total mass BC concentrations from the mentioned categorical variables. Based on the coefficients and the p-values it was determined that type of fuel, exposure to a neighboring biomass stove and use of kerosene could be predictors of BC concentrations. The use of LPG as a source of fuel has a significant reduction of BC concentrations, meanwhile the use of kerosene and exposure to a neighboring biomass stove increases BC total mass concentration.

Table 1: Table 4

**Metrics summary: BC Models**

| Metric | Linear | GLM |
| --- | --- | --- |
| RMSE | 4.535811 | 4.6019779 |
| R-squared | 0.380425 | 0.3638607 |
| MAE | 3.346001 | 3.3907889 |

## 5 Discussion

Although the scope of this study was not to conduct the source apportionment methodology, the findings reported here are helpful for the early stages of the overall positive matrix factorization (PMF) modeling of the HAPIN Guatemala personal exposure samples. Based on the limit of detection, only 10 of the 22 chemical species (elements) were deemed important to be included in the PMF model. The model is sensitive to the mass concentrations of all species used, that is why the remaining 11 species were left out of the analyses. Some important species such as lead, copper, chromium, iodine, arsenic and others could provide a better profile of the chemical composition of total PM2.5. However, since these were in trace concentrations, maybe they are not as important as others, in terms of determining the sources of PM2.5 in this location. For the chemicals that were incorporated in the final analyis, it was observed that BC had overall the highest concentrations. This could indicate that BC is a chemical component that comes from many different sources besides just biomass burning, which is one of the most expected exposures/sources that contribute to these samples. The following element in highest concentration appeared to be potassium, which again, based on literature [13, 14], it is an indicator of biomass burning.

When comparing the mass concentrations between samples from users that cooked using biomass fuels against cooking with LPG, it was observed that this provided a more accurate picture of the differences by the LPG intervention only. As seen in the analyses, dividing the samples in control and intervention arms produced similar results that when dividing by type of fuel. However there were higher concentration values observed for some species in the intervention samples, which is explained by the inclusion of all baseline measurements into this division. As explained in the HAPIN study description [4, 11], the intervention was allocated after baseline, and during baseline it was observed that the concentrations of the main pollutants was virtually the same between both arms. In conclusion, it is considered that if looking at differences, categorizing by type of fuel, provides a clearer picture of the exposure contrasts. The correlations explored between the selected chemical species provide a glimpse of what elements could be more associated in terms of a common source. The more notables are iron, silicon and aluminum, that, based on the literature, are mostly associated by a crustal source of pollution [8]. Crustal means ground or dust generated particles. Other patterns are not so easily visualized, but there is a possible association between magnesium and calcium, which could be an indicator of burning of some sort of plant-based material [8, 13, 14]. Black carbon was moderately associated with potassium, which is a possible indicator of a biomass burning source, but this moderate to low correlation indicates that it could come from many other sources.

In terms of model selection for predicting PM2.5 and BC concentrations, multivariate models were deemed the most appropriate due to the presence of many predictors. In the BC prediction model, the gamma GLM model seemed more appropriate compared to the linear regression model, however this may not be the most accurate and other types of model could be a better fit. It would be interesting to fit a LASSO regression model, in order to better discriminate which of the categorical exposures could be used to explain the variations in concentrations of BC. Another explanation why this model under-performed could be to the data collection process, there were just a few records of certain exposures such as trash burning, smoking, incense, such that this could have driven the model to the wrong conclusions. This is why it is believed that modelling exposures is more accurate when considering actual exposure measurements (pollutants or chemicals), instead of relying on questionnaire based exposures that could provide biased answers. On another hand, by considering only kerosene use, exposure to a neighboring biomass stove and type of fuel as accurate predictors of BC, we could infer that the model under-performed. However, it is also important to note that these are good representations of use of biomass fuel source, which again, could be the most significant source of pollution, instead of considering other such as vehicle emissions, crustal or others.

For the PM2.5 models, the multivariate linear regression model seems the most appropriate as seen on the evaluation criteria (Table 3 and Supplementary material). However, it is important to point out that the moderate R-squared value implies that not all of the chemical species were included in the model to explain the total mass concentrations of PM2.5. Besides BC, no other organic compounds were analyzed and included in the model, and it is known that some compounds like ammonia, elemental carbon, levoglucosan and many others [15] have very important roles in particulate matter composition. It was surprising that the GLM model did not perform as good as expected, usually under-predicting extreme values. One possible way to overcome this could have been to test a multivariate linear model log-transforming the outcomes and predictors, this is due to them not following a normal distribution. In terms of the most significant species, the significant contribution of potassium and black carbon only add to the evidence that biomass burning could be the most important contributor of particulate matter in this location.

One of the limitations for this analyses is that, as pointed before, some chemical species were not included in the model. This is partly because of the source of the chemical species data, in this case it was performed only transmissometry, gravimetry and x-ray fluorescence to determine some of the species. It would be impossible to determine all the chemical species that conform the PM2.5 filter samples, however, adding a chromatography analysis could improve the resolution of the model by detecting additional compounds. Another goal accomplished by the analyses was to compute the final concentrations and uncertainties for all the chemical species and filter samples that will be analyzed using PMF. Currently, the PMF algorithm is licensed by the Environmental Protection Agency, under the EPA PMF 5.0 software [16]. There was an attempt to compute the whole PMF analysis for source apportionment in R, however, this algorithm is not available publicly due to the patent. However, there is one specific packages ‘pmfr’ found online in a Github repository (https://github.com/skgrange/pmfr) that allows working with the outputs of the EPA PMF 5.0 software to do additional analysis and plotting.

## 6 Conclusions

Twelve chemical species from the filter samples in rural Guatemala were found important for PMF modeling: Mg, Mn, Al, Si, S, Ca, K, Zn, Ti, Fe and BC. Of these species, all of them except Al, Zn and Fe had significantly smaller mass concentrations in LPG fuel samples compared to biomass fuel samples. BC and K are two of the most available chemical species found in the samples, which can be associated to the heavy influence of biomass burning in this location. When modelling which exposures were associated to BC concentrations, kerosene use, exposure to a neighboring biomass stove and type of fuel were significant predictors according to a GLM gamma model. However, the low diversity in answers for these categorical variables could have biased the results. On the other hand, a linear regression model was found more accurate to define total PM2.5 mass concentrations with all the chemical species. BC, potassium and magnesium were the most significant predictors in the PM2.5 model. The current study also sets the final concentration and uncertainty values needed to perform source apportionment using the PMF algorithm developed by the Environmental Protection Agency.

## 7 References

1. Gordon SB, Bruce NG, Grigg J, et al (2014) Respiratory risks from household air pollution in low and middle income countries. The Lancet Respiratory Medicine 2(10):823–860. <https://doi.org/10.1016/S2213-2600(14)70168-7>

2. Murray CJL, Aravkin AY, Zheng P, et al (2020) Global burden of 87 risk factors in 204 countries and territories, 1990–2019: A systematic analysis for the Global Burden of Disease Study 2019. The Lancet 396(10258):1223–1249. <https://doi.org/10.1016/S0140-6736(20)30752-2>

3. Pope CA, Dockery DW (2006) Health Effects of Fine Particulate Air Pollution: Lines that Connect. Journal of the Air & Waste Management Association 56(6):709–742. <https://doi.org/10.1080/10473289.2006.10464485>

4. Clasen T, Checkley W, Peel JL, et al (2020) Design and Rationale of the HAPIN Study: A Multicountry Randomized Controlled Trial to Assess the Effect of Liquefied Petroleum Gas Stove and Continuous Fuel Distribution. Environmental Health Perspectives 128(4):047008. <https://doi.org/10.1289/EHP6407>

5. Mohammed MOA, Wei SW, Li MW, et al Potential Toxicological and Cardiopulmonary Effects of PM2.5 Exposure and Related Mortality: Findings of Recent Studies Published during 2003-2013. Biomed Environ Sci

6. Viana M, Kuhlbusch TAJ, Querol X, et al (2008) Source apportionment of particulate matter in Europe: A review of methods and results. Journal of Aerosol Science 39(10):827–849. <https://doi.org/10.1016/j.jaerosci.2008.05.007>

7. Reff A, Eberly SI, Bhave PV (2007) Receptor Modeling of Ambient Particulate Matter Data Using Positive Matrix Factorization: Review of Existing Methods. Journal of the Air & Waste Management Association 57(2):146–154. <https://doi.org/10.1080/10473289.2007.10465319>

8. Zhou Z, Dionisio KL, Verissimo TG, et al (2014) Chemical Characterization and Source Apportionment of Household Fine Particulate Matter in Rural, Peri-urban, and Urban West Africa. Environmental Science & Technology 48(2):1343–1351. <https://doi.org/10.1021/es404185m>

9. Sharma SK, Mandal TK, Jain S, Saraswati, Sharma A, Saxena M (2016) Source Apportionment of PM2.5 in Delhi, India Using PMF Model. Bulletin of Environmental Contamination and Toxicology 97(2):286–293. <https://doi.org/10.1007/s00128-016-1836-1>

10. Martínez-Cinco M, Santos-Guzmán J, Mejía-Velázquez G (2016) Source apportionment of PM for supporting control strategies in the Monterrey Metropolitan Area, Mexico. Journal of the Air & Waste Management Association 66(6):631–642. <https://doi.org/10.1080/10962247.2016.1159259>

11. Johnson MA, Steenland K, Piedrahita R, et al (2020) Air Pollutant Exposure and Stove Use Assessment Methods for the Household Air Pollution Intervention Network (HAPIN) Trial. Environmental Health Perspectives 128(4):047009. <https://doi.org/10.1289/EHP6422>

12. Kearns KA EXPOSURES TO HOUSEHOLD AIR POLLUTION IN WOMEN ACROSS FOUR DIVERSE COUNTRIES AS PART OF A MULTI-COUNTRY COOKSTOVE INTERVENTION TRIAL

13. Rai P, Furger M, El Haddad I, et al (2020) Real-time measurement and source apportionment of elements in Delhi’s atmosphere. Science of The Total Environment 742:140332. <https://doi.org/10.1016/j.scitotenv.2020.140332>

14. Pervez S, Dubey N, Watson JG, Chow J, Pervez Y (2012) Impact of Different Household Fuel Use on Source Apportionment Results of House-Indoor RPM in Central India. Aerosol and Air Quality Research 12(1):49–60. <https://doi.org/10.4209/aaqr.2011.08.0124>

15. Gadi R, Shivani, Sharma SK, Mandal TK (2019) Source apportionment and health risk assessment of organic constituents in fine ambient aerosols (PM2.5): A complete year study over National Capital Region of India. Chemosphere 221:583–596. <https://doi.org/10.1016/j.chemosphere.2019.01.067>

16. Norris G, Duvall R, Brown S, Bai S Positive Matrix Factorization (PMF) 5.0 Fundamentals and User Guide