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

Project: Part 4

Erick E. Mollinedo

2024-04-05

## 1 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 and to conduct the exploratory and main data analysis 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? and 3.) What exposures are associated with the variations in concentrations of the most significant chemical species?

## 2 Methods

### 2.1 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 are 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 on incense or coil, kerosene and others.

### 2.2 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 X-Ray Fluorescence (XRF), and the uncertainties were estimated also based on the XRF instrument.

### 2.3 Concentration and Uncertainties data processing

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).

### 2.4 Statistical analysis

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, a Welch two-sample T-test was computed. Additional plots were presented to visualize those differences between samples.

Multivariate regression models were computed for each one of the key chemical species to assess which exposures are likely to explain the variations in their concentrations. Given that there is low variability in the answers for some of the categorical exposures, in addition to computing a multivariate model with all the predictors, another model was evaluated. This second model considers other stoves and fueltype as predictors, alongside a new combined variable called ‘micro exposures’ that englobes exposures to kerosene, cigarette, coils and incense and the variable ‘macro exposures’ which englobes exposures to trash burning and crop residues. Given that the concentrations can be approximated to a gamma distribution, the generalized linear model with gamma distribution was used to compute the models. The root mean square error (RMSE) and R-squared were the metric parameters to evaluate which one of both models for all chemical species fits best.

## 3 Results

Six hundred and twenty-nine filter samples were considered in the final analyses. [Figure 1](#fig-stats1) presents a summary of the concentrations from the biomass fueltype samples. Meanwhile [Figure 2](#fig-stats2) shows the summary of concentrations from the LPG observations. 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.

|  |
| --- |
| Figure 1: Summary statistics of the concentrations in the biomass group |

|  |
| --- |
| Figure 2: Summary statistics of the concentrations in the LPG group |

### 3.1 Chemical Species correlations

In terms of correlations, Spearman correlation was estimated between all the chemical species and [Figure 3](#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 sources, so it will probably be easier to identify using the PMF software after the conclusion of this project.

|  |
| --- |
| Figure 3: Correlation plot for chemical species |

### 3.2 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 100 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 4](#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 4: Chemical species concentrations by type of fuel |

### 3.3 Chemical species models

Model 1, for each one of the chemical species, was computed to determine which of the exposures could better explain the variations in the concentrations (Supplementary Material). Given that none of the observations recorded to have used a generator machine for corn mill grinding, this categorical variable was left out. When computing this model, none of the predictors were significant to predict the concentrations of Aluminim (Al), Iron (Fe) and Zinc (Zn). Kerosene was found a significant predictor of Black carbon (BC) and Magnesium (Mg) concentrations (p < 0.05). Exposure to smoke emitted from a biomass stove other than the one in their homes was also found as a significant predictor (p < 0.05) for BC and potassium (K). Also, exposure to cigarette smoke in the homes was associated to K concentration. Meanwhile, the strongest predictor was found to be type of fuel (p < 0.01), where exposure to biomass fuel was associated to an increase of the concentration of BC, Ca, K, Mg, Mn, S, Si and Ti. These last results are consistent to what was observed in terms of significant differences among samples by type of fuel ([Figure 4](#fig-concentrations)).

Model 2 was tested as another way to evaluate which predictors could explain the variations in concentrations by grouping certain exposures. Similar results were observed as from Model 1. Micro exposures, which is a category that includes smoking, incense use, kerosene and mosquito coils explained the variations in BC concentrations (p < 0.05). Exposure to other biomass stoves was a predictor for BC and K (p < 0.05). Fueltype was still the strongest predictor (p < 0.01) for the variations in concentrations for BC, Ca, K, Mg, Mn, S, Si and Ti.

In terms of comparison between models, both models seem to under-perform in terms of predictions (Table 1, Supplementary Material). The lowest RMSE was observed in Model 2 for all the chemical species, meanwhile the highest R-square was observed in Model 1. In this case, Model 2 with grouping of variables could be a better model to explain the effect of certain exposures to the concentrations of the selected chemical species.

## 4 Discussion

## 5 Conclusions

## 6 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