

Towards retrieving aerosol chemical composition from temporal variations of total PM mass concentrations: Theoretical approach, insights, and the promise of machine learning techniques

Danish Ansari¹, Prateek Verma², Tabish Ansari³

¹Independent Researcher, NGEF layout, Sadananda Nagar, Bangalore

²Department of Chemical Engineering, University of Arkansas, Fayetteville, United States

³Campus Fryslan, University of Groningen, Leeuwarden, the Netherlands

Correspondence to: t.u.ansari@rug.nl

Abstract: Fine and ultrafine ambient particulate matter (PM) has major health and climate impacts. Chemical composition of PM is required for better estimation of these impacts but is considerably expensive to measure as compared to the total PM mass concentrations. We explore the indirect estimation of PM chemical composition by analysing the temporal variation of, relatively inexpensively measured, total PM.

Keywords: Air pollution, PM₁, PM_{2.5}, Aerosol, Machine Learning, megacities

Introduction

Ambient particulate matter (PM) or aerosols with diameter smaller than 2.5 microns (PM_{2.5}) and smaller than 1 micron (PM₁) have deleterious effects on human health including mortality and morbidity¹⁻⁶. Ambient PM levels consistently exceed daily and yearly safe standards set by the World Health Organization (WHO) in various parts of the world^{7,8}. Health impacts are often estimated using response curves that relate reduced mortality to total PM mass⁹. However, better estimates of health effects require chemical makeup of PM, in addition to its total mass^{10,11}. Additionally, knowledge of PM chemical composition aids in understanding of cloud formation and other climate effects^{12,13}. PM includes chemical components such as organic carbon (ORG), black carbon (BC), ammonium (NH₄), nitrate (NO₃), sulphate (SO₄), chloride (CHL), mineral dust and other heavy metals²⁴.

While it is relatively inexpensive to measure total PM mass concentrations, it is much more expensive and challenging to observe PM chemical composition. *Offline* sampling techniques of estimating aerosol chemical composition involve retrospective lab-based chemical analysis of filter papers from PM measuring

instruments ¹⁴⁻¹⁶. Besides requiring human intervention and expensive laboratory analysis ^{17,18}, they yield longer time-averaged (typically daily- or weekly-averaged) chemical composition information and are thus better suited for limited time-span field campaigns than for real-time data streaming service to citizens and stakeholders, for instance. *Online* techniques involve direct measurement of aerosol chemical components using mass spectrometers ¹⁹ such as the Aerosol Mass Spectrometer (AMS) or Aerosol Chemical Speciation Monitor (ACSM) ^{20,21} which can be prohibitively expensive (approximately, AMS: 500,000\$, ACSM: 150,000-200,000\$; personal communication), incur expensive maintenance, and require trained personnel to operate. Due to these limitations, chemical composition of aerosols remains unobserved over large parts of the globe including major population centres such as the Indo Gangetic Plain²⁹. On the contrary, total PM can be measured using reference monitors that cost around (based on personal communication) 20,000\$-30,000\$ or low cost sensors that cost between 200\$-2000\$ (e.g., Purple Air; <https://www2.purpleair.com/>). Unsurprisingly, observations of total PM are more abundant than chemical compositions. Chemical composition observations over time are available for multiple sites in United States³⁰, Europe³¹ and China^{22,23}, but with several gaps³¹, thereby hindering scientists from developing a consistent global picture of aerosol chemical composition at a high temporal resolution.

Total PM concentrations show distinct temporal variations at the diurnal as well as seasonal scales ^{24,25}. This temporal signature depends on a number of local and regional factors such as meteorology (wind speed and direction, relative humidity, boundary layer mixing), chemistry (chemical rates of formation and transformation of

aerosols) and emissions (their strength and temporal variations). Moreover, temporal signature of total PM concentration can be viewed as the sum of the temporal signatures of each of its chemical components. Hourly data from various AMS/ACSM measurements, particularly for urban locations, indicate that all chemical components do not exhibit identical temporal variations²⁴⁻²⁷. We notice that, generally, primary aerosols (BC and primary ORG) show a more spikey signal, highlighting that they are freshly emitted and preserve some characteristics of the temporal profiles of their emission sources. Secondary aerosols (NH₄, NO₃, SO₄) show a smoother temporal signal which is evidence of the timescales of chemical conversion of precursor gases to aerosols. Concentrations of precursor gases (e.g., NO₂ and SO₂) might retain the temporal signal of emission sources but since they are lost during chemical transformation, are not preserved in NO₃ and SO₄ aerosol observations.

Superimposed upon the emissions/chemistry signal is the meteorological signal but regional meteorology affects all chemical components simultaneously, rather than individually. So, all chemical components will show a simultaneous decline if the wind direction changes to one which brings clean air or if wind speed increases. Turbulent mixing in the atmospheric boundary layer, however, has a somewhat different effect on primary and secondary aerosols²⁹. Primary aerosols have a stronger vertical gradient than secondary aerosols because all primary emission sources are situated at the surface while secondary aerosols have a weaker vertical gradient within the boundary layer due to the spatial scales of chemical production. Therefore a reduction in boundary layer height leads to much steeper increases in

primary aerosol concentrations as compared to relatively modest increases in secondary aerosol concentrations, see, for example, Ansari *et al.*, 2019²⁷.

Therefore, we posit that, under similar meteorological conditions, a highly spikey and fluctuating temporal profile of the total PM indicates a dominant fraction of primary aerosols while a smoother temporal signal of total PM suggests a larger proportion of secondary aerosols. These distinct temporal responses of different chemical components hold the key for their indirect estimation. This approach is valid, especially for urban locations where aerosols are *externally-mixed*; freshly emitted and contain strong temporal signals of their sources (e.g., traffic, industry, households). However, observations made at remote locations with far-away emission sources have weaker source-signals and the aerosols are often in an *internally-mixed* state; different chemical components stick to each other into a single particle, such that, the chemical composition of each particle is the same as the overall composition of the aerosol mixture²⁴, which, in turn, will likely hinder the ability to make good estimations of compositions based on the temporal profile of total PM alone.

Apart from temporal variations, it is important to study the relationship between chemical component fractions and total PM mass. It is crucial to understand if the fractions persist over a given range of total PM mass concentrations or if they change more randomly across the total PM range. Some chemical components are formed at the expense of others. For example, SO_4 is preferentially formed before NO_3 in presence of excess NH_3 ²⁴. Such co-dependence must be analysed vis-a-vis total PM concentration and aids in estimating chemical fractions in the absence of

their actual measurements. To illustrate this approach, we used the publicly available PM₁ chemical components data over Delhi from Gani *et al.*, 2019²⁸.

Figure 1 shows the variation of six chemical fractions against 25 µg m⁻³ bins of total PM₁. The median fractions, across all bins, range from 0.42 to 0.65 for ORG, 0.05 to 0.13 for NH₄, 0.03 to 0.22 for SO₄, 0.04 to 0.11 for NO₃, 0.01 to 0.28 for CHL and 0.05 to 0.13 for BC. The ORG fraction shows a general increase from bottom to top suggesting that the higher PM₁ episodes are caused by additional organic aerosol contributions. However, this increase is not monotonic. ORG fraction remains somewhat constant between 25-175 µg m⁻³ PM₁ range, showing a slight reduction in the 225-250 µg m⁻³ bin, followed by an increasing trend until 400 µg m⁻³. However, the median ORG fraction shows an abrupt decrease between the 400-500 µg m⁻³ bins. Interestingly, the decline in ORG fraction across these bins is compensated by a corresponding increase in the CHL fraction and, to a lesser extent, NH₄ fraction. Other fractions remain nearly unchanged across these bins.

NH₄ and SO₄ fractions show a decreasing trend from bottom (cleaner) to top (more polluted conditions). NO₃ shows an increasing trend until 225 µg m⁻³ after which it declines. CHL shows an overall increasing trend beginning at 0.01 at the lowest bin to 0.09 at the highest bin with an abrupt increase in the 425-500 µg m⁻³ range as noted earlier. BC fraction shows a modest decreasing trend from 0.13 in the lowest bin to 0.05 at 450 µg m⁻³ after which it steadily increases to 0.13 in the highest bin. The interquartile range (IQR) for NH₄, SO₄, NO₃ and BC were generally low (within 0.1), while those for ORG and CHL were the largest, up to 0.3, and generally increased as we moved up the bins. For SO₄, the IQR was largest for cleaner

periods (0-75 $\mu\text{g m}^{-3}$) after which it diminished consistently. The consistent decreasing trend in SO_4 fraction shows the physical limit of SO_4 formation in Delhi given the regional precursor emissions.

We further investigated the competition effect between CHL and ORG fractions as noted from Figure 1 and found a significant anticorrelation ($r=-0.74$) between the two, for cases where $\text{CHL}>0.1$ (Figure 2). An even tighter anticorrelation was found between NH_4 and CHL fractions ($r=-0.87$) while CHL and NH_4 fractions showed a positive correlation ($r=0.61$). Such relationships between different chemical fractions, especially if true across multiple sites, can potentially be exploited by machine learning algorithms to establish more generalizable relationships among chemical fractions and/or relate them to the total PM mass.

Conclusion:

In this study we have laid out the theoretical foundation for potentially retrieving hourly aerosol chemical composition information without directly measuring them by utilising merely the temporal variations of total PM mass concentrations. We have discussed the peculiarities of temporal signals of different aerosol components based on their physico-chemical properties and their differing response to meteorology. We have also analysed a publicly available dataset of aerosol chemical composition in a novel way to establish the relationships between total PM_{10} and individual chemical fractions. We believe these distinct properties of aerosol chemical components provide for an excellent use-case for machine learning techniques to uncover valuable climate- and health-relevant information in a

cost-effective way. We will present our results from the ML model in a subsequent manuscript.

References:

1. Cohen, A. J., Brauer, M., Burnett, R., et al., Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015. *Lancet*, 2017, **389**, 1907–1918.
2. HEI, H. E. I. and IHME, H. M. and E., a Special Report on Global Exposure To Air Pollution and Its Disease Burden. 2018.
3. Pope, C. A. 3rd and Dockery, D. W., Health effects of fine particulate air pollution: lines that connect. *J. Air Waste Manag. Assoc.*, 2006, **56**, 709–742.
4. Anenberg, S. C., Horowitz, L. W., Tong, D. Q., and West, J. J., An Estimate of the Global Burden of Anthropogenic Ozone and Fine Particulate Matter on Premature Human Mortality Using Atmospheric Modeling. *Environ. Health Perspect.*, 2010, **118**, 1189–1195.
5. Lelieveld, J., Evans, J. S., Fnais, M., Giannadaki, D., and Pozzer, A., The contribution of outdoor air pollution sources to premature mortality on a global scale. *Nature*, 2015, **525**, 367–71.
6. Maji, S., Ghosh, S., and Ahmed, S., Association of air quality with respiratory and cardiovascular morbidity rate in Delhi, India. *Int. J. Environ. Health Res.*, 2018, **28**, 471–490.

7. Kanawade, V. P., Srivastava, A. K., Ram, K., et al., What caused severe air pollution episode of November 2016 in New Delhi? *Atmos. Environ.*, 2020, **222**, 117125.
8. Morawska, L., Zhu, T., Liu, N., et al., The state of science on severe air pollution episodes: Quantitative and qualitative analysis. *Environ. Int.*, 2021, **156**, 106732.
9. Lim, S. S., Vos, T., Flaxman, A. D., et al., A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet (London, England)*, 2012, **380**, 2224–2260.
10. Molina, C., Toro A., R., Manzano, C. A., Canepari, S., Massimi, L., and Leiva-Guzmán, M. A., Airborne Aerosols and Human Health: Leapfrogging from Mass Concentration to Oxidative Potential. *Atmosphere (Basel)*., 2020, **11**.
11. Leni, Z., Cassagnes, L. E., Daellenbach, K. R., et al., Oxidative stress-induced inflammation in susceptible airways by anthropogenic aerosol. *PLoS One*, 2020, **15**, e0233425.
12. Charlson, R. J., Schwartz, S. E., Hales, J. M., et al., Climate Forcing by Anthropogenic Aerosols. *Science (80-.)*., 1992, **255**, 423–430.
13. Almeida, G. P., Brito, J., Morales, C. A., Andrade, M. F., and Artaxo, P., Measured and modelled cloud condensation nuclei (CCN) concentration in São Paulo, Brazil: the importance of aerosol size-resolved chemical

composition on CCN concentration prediction. *Atmos. Chem. Phys.*, 2014, **14**, 7559–7572.

14. Holzinger, R., Kasper-Giebl, A., Staudinger, M., Schauer, G., and Röckmann, T., Analysis of the chemical composition of organic aerosol at the Mt. Sonnblick observatory using a novel high mass resolution thermal-desorption proton-transfer-reaction mass-spectrometer (hr-TD-PTR-MS). *Atmos. Chem. Phys.*, 2010, **10**, 10111–10128.
15. Timkovsky, J., Chan, A. W. H., Dorst, T., Goldstein, A. H., Oyama, B., and Holzinger, R., Comparison of advanced offline and in situ techniques of organic aerosol composition measurement during the CalNex campaign. *Atmos. Meas. Tech.*, 2015, **8**, 5177–5187.
16. Malaguti, A., Mircea, M., Torretta, T. M. G. La, et al., Comparison of Online and Offline Methods for Measuring Fine Secondary Inorganic Ions and Carbonaceous Aerosols in the Central Mediterranean Area. *Aerosol Air Qual. Res.*, 2015, **15**, 2641–2653.
17. Deng, Y., Fujinari, H., Yai, H., et al., Offline analysis of the chemical composition and hygroscopicity of sub-micrometer aerosol at an Asian outflow receptor site and comparison with online measurements. *Atmos. Chem. Phys. Discuss.*, 2021, **2021**, 1–32.
18. Rai, P., Slowik, J. G., Furger, M., et al., Highly time-resolved measurements of element concentrations in PM₁₀ and PM_{2.5}: comparison of Delhi, Beijing, London, and Krakow. *Atmos. Chem. Phys.*, 2021, **21**, 717–730.

19. Nash, D. G., Baer, T., and Johnston, M. V, Aerosol mass spectrometry: An introductory review. *Int. J. Mass Spectrom.*, 2006, **258**, 2–12.
20. Ng, N. L., Herndon, S. C., Trimborn, A., et al., An Aerosol Chemical Speciation Monitor (ACSM) for Routine Monitoring of the Composition and Mass Concentrations of Ambient Aerosol. *Aerosol Sci. Technol.*, 2011, **45**, 780–794.
21. Fröhlich, R., Cubison, M. J., Slowik, J. G., et al., The ToF-ACSM: a portable aerosol chemical speciation monitor with TOFMS detection. *Atmos. Meas. Tech.*, 2013, **6**, 3225–3241.
22. Dao, X., Lin, Y.-C., Cao, F., et al., Introduction to the National Aerosol Chemical Composition Monitoring Network of China: Objectives, Current Status, and Outlook. *Bull. Am. Meteorol. Soc.*, 2019, **100**, ES337–ES351.
23. Li, Y. J., Sun, Y., Zhang, Q., et al., Real-time chemical characterization of atmospheric particulate matter in China: A review. *Atmos. Environ.*, 2017, **158**, 270–304.
24. Seinfeld, J. H. and Pandis, S. N., *Atmospheric Chemistry and Physics : From Air Pollution to Climate Change* John Wiley & Sons, Incorporated, New York, United States, 2016.
25. Sun, X., Luo, X.-S., Xu, J., et al., Spatio-temporal variations and factors of a provincial PM_{2.5} pollution in eastern China during 2013–2017 by geostatistics. *Sci. Rep.*, 2019, **9**, 3613.

26. Sun, Y., Wang, Z., Dong, H., et al., Characterization of summer organic and inorganic aerosols in Beijing, China with an Aerosol Chemical Speciation Monitor. *Atmos. Environ.*, 2012, **51**, 250–259.
27. Ansari, T. U., Wild, O., Li, J., et al., Effectiveness of short-term air quality emission controls: a high-resolution model study of Beijing during the Asia-Pacific Economic Cooperation (APEC) summit period. *Atmos. Chem. Phys.*, 2019, **19**, 8651–8668.
28. Gani, S., Bhandari, S., Seraj, S., et al., Submicron aerosol composition in the world's most polluted megacity: the Delhi Aerosol Supersite study. *Atmos. Chem. Phys.*, 2019, **19**, 6843–6859.
29. Brooks, J., Allan, J. D., Williams, P. I., et al., Vertical and horizontal distribution of submicron aerosol chemical composition and physical characteristics across northern India during pre-monsoon and monsoon seasons. *Atmos. Chem. Phys.*, 2019, **19**, 5615–5634.
30. Zhang, Q. and Handschy, A., Aerosol Mass Spectrometry (AMS) Global Database, <https://sites.google.com/site/amsglobaldatabase/>
31. EMEP ACSM Station Data Coverage. 2019.,
<https://www.psi.ch/en/acsm-stations/overview-full-period>

Figure captions:

Figure 1: Panelplot showing median values and interquartile ranges of each chemical fraction against total PM_{10} bins of range $25 \mu g m^{-3}$. The first column shows the frequency distribution of total PM_{10} for the entire period. Other columns show the corresponding chemical fractions. Median values are shown as black dots while lower and upper quartiles are shown in grey crossmarks.

Figure 2: Selected scatterplots between chemical component fractions with high correlations.



