

A Comparison of Chemical Mass Balance and Positive Matrix Factorisation Methods for Quantification of Fine Particle Air Pollution Sources in Sydney, Australia.

David D Cohen¹, Stuart Hargreaves², Ed Stelcer¹, Olga Hawas¹, Adam Sarbutt¹,
David Button¹

¹Australian Nuclear Science and Technology Organisation,
PMB1, Menai, NSW, 2234, Australia.

²Department of Computing and Telecommunications, University of Wollongong,
Wollongong, NSW, 2522, Australia.

1. Introduction

Australia is currently considering new fine particle (particulate matter with aerodynamic diameters less than 2.5 μm , PM_{2.5}) air pollution standards or goals for protection of public health. These will be quite stringent compared with other PM_{2.5} standards around the globe. It is therefore important for environmental pollution agencies to better understand and quantify sources of fine particle air pollution in urban environments. The simultaneous application of PIXE, PIGE, RBS and PESA ion beam analysis (IBA) techniques is ideally suited to producing multi-elemental fingerprints of such pollution and this has been reported extensively elsewhere [1-6].

Recently two methods have come to the fore for quantitative source apportionment or air pollution – the Chemical Mass Balance (CMB) method and the most recent powerful Positive Matrix Factorisation (PMF) method [8-10]. This paper discusses and compares data analysed by these two source receptor methods, at the urban site of Mascot in Sydney during 2003 and 2004.

2. Sampling and Analysis Methods

To perform the comparison of these two source apportionment methods we have used the daily data obtained at the Mascot site in Sydney over the period January 2003 to December 2004. Samples were collected on Teflon filters over a 24 hour period from midnight to midnight every Sunday and Wednesday. Typically each filter contained fine particles from 32m³ of sampled air. Fig. 1 shows the monthly average of the daily gravimetric fine mass/m³ of air sampled over the 2 year study period. This monthly gravimetric mass varies between 5 $\mu\text{g}/\text{m}^3$ and 12 $\mu\text{g}/\text{m}^3$ which is equivalent to between 70 μg and 170 μg total particulate mass on each filter.

Four simultaneous IBA techniques of PIXE, PIGE, PESA and RBS were used to analyse each of these Teflon filters for the 22 elements, H, C, F, Na, Al, Si, P, S, Cl, K, Ca, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, Br and Pb. Minimum detection limits (MDLs) were typically 1-10 ng/m³ of air sampled. HeNe laser techniques were also used to measure the total black carbon (BC) or soot content on each filter [2,5-6]. The precise measurement of so many different elemental components was the key to obtaining good 'mass closure'. This is important if one is to achieve reliable source fingerprints and good source apportionment estimates. 'Mass closure' can be defined as the ratio of the reconstructed mass (RCM), obtained by adding together all the analysed components, and the gravimetric mass. Generally for

this type of work, by assuming various elements occur in their oxide or common chemical forms, we are able to obtain percentage reconstructed masses (%RCM) varying between 80% and 120% of the measured gravimetric mass [7]. The missing mass is usually nitrate compounds (not measured here ~5-10%) and water vapour (depending on the humidity ~5-15%). Table 1 below shows the major component estimates of the PM_{2.5} fine fraction for the Mascot site covering the period January 2003 to December 2004. Salt was assumed to be sea salt and composed mainly of NaCl, soil was estimated from the oxides of Al, Si, Ti, Ca and Fe, the sulfate was assumed to be fully neutralised and occurring as (NH₄)₂SO₄ and organics were estimated from the H, C and O content in the standard way [1,2,7]. As the table shows this gave good 'mass closure' with two year

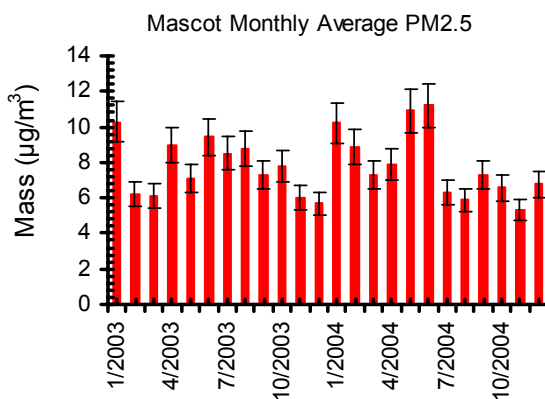


Fig. 1. Average monthly fine mass at Mascot

%RCM=(86±14)%. This is an excellent result for this type of work and makes for reliable, consistent source fingerprinting and apportionment estimates as shown below.

PM _{2.5} µg/m ³	Gravimetric Mass	Salt (NaCl)	BC (soot)	Soil	Ammonium. Sulfate	Organics	%RCM
Average	7.68±4.7	1.00±1.3	1.55±1.2	0.43±0.36	1.69±1.2	1.87±2.4	86±14
Median	6.35	0.70	1.19	0.27	1.42	1.27	86
Min.	1.56	0.00	0.18	0.06	0.33	0	-
Max.	32.85	8.81	6.89	2.02	6.09	13.46	-

Table 1. The average of all daily data for the Mascot site for PM_{2.5} fine particles for 2003-04 study period.

The numbers in Table 1 are good estimates of the fine particle composition but they do not provide much explicit information on the actual source contributions to the measured fine particle mass. In order to achieve this we need to apply Chemical Mass Balance of Positive Matrix Factorisation methods to these IBA analyses.

3. Chemical Mass Balance (CMB) Method

The CMB method is a two step process, first you need to determine the elemental fingerprints then fit these (using least squared fitting techniques) to the gravimetric mass. Generally Principal Component Analysis (PCA) methods are used to identify elements that are associated within any given fingerprint. Then the gradient of the correlation plots are used to estimate the fractional contributions of each element to a fingerprint. This has been discussed in detail elsewhere [2,4-6].

Five generic fingerprints were needed to fit the Mascot data during 2003-04 using the US EPA CMB7 code [9]. These were Autos, Smoke, Soil, Sea spray and Industry and are shown in Figs 2(a) to (e) below. The smoke fingerprint did not occur for all months of the study period but was prevalent mainly during the cooler winter months, during controlled bush burning periods or during actual bushfire episodes. The average percentage contributions for 2003-04 at the Mascot (Sydney site) were, 36±20%, 3±12%, 4±1%, 14±10% and 43±15% for the Autos, Smoke, Soil, Sea spray and Industry fingerprints respectively. The CMB7 codes provide three parameters as measures of the goodness of fit, namely, R², χ² and Mass%. R² is similar to a correlation coefficient and estimates how well the fingerprints used fit the measured gravimetric masses. R² between 0.8 and 1.0 are acceptable. χ² generally lies between 0-2 for a good fit, 2-4 for an acceptable fit and >4 for a poor fit, but this is a function of the input errors used.

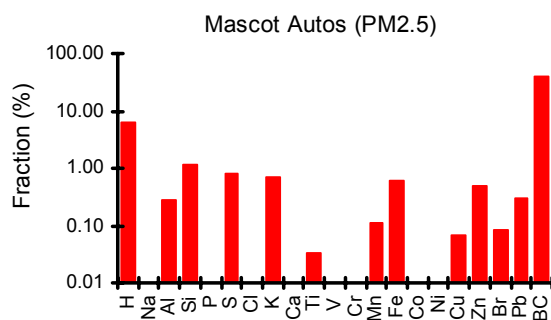


Fig. 2a Auto fingerprint

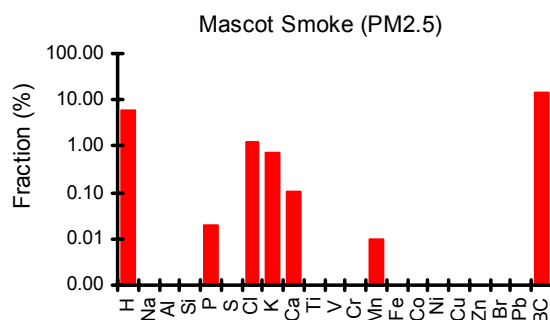


Fig. 2b Smoke fingerprint

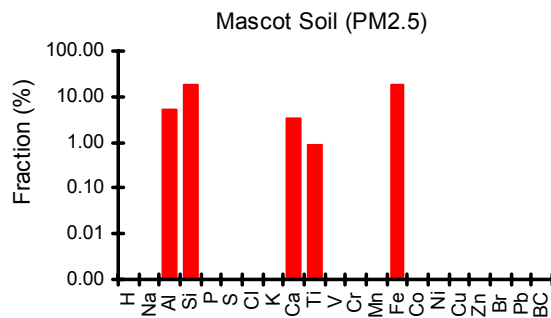


Fig. 2c Soil fingerprint

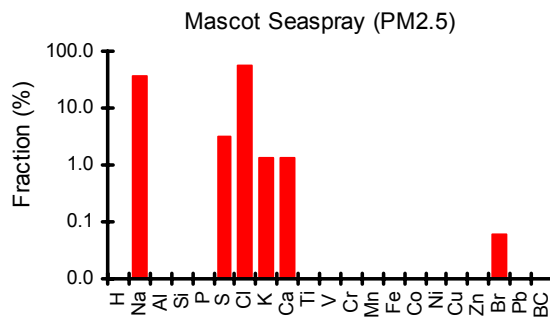


Fig. 2d Sea spray fingerprint

It estimates how well the calculated fingerprint concentrations fit the measured elemental concentrations. The Mass% parameter represent the percentage of the total mass that can be explained by the least squares summing of the individual fingerprints. Values of Mass% between 80-120% are acceptable depending again on the sizes of the measurement errors. For example, at Mascot the CMB fingerprint estimates and their fits to the total mass were excellent with average $R^2=0.96\pm 0.02$, $\chi^2=0.59\pm 0.14$ and Mass%=99 \pm 5% over the 2003-04 sampling period.

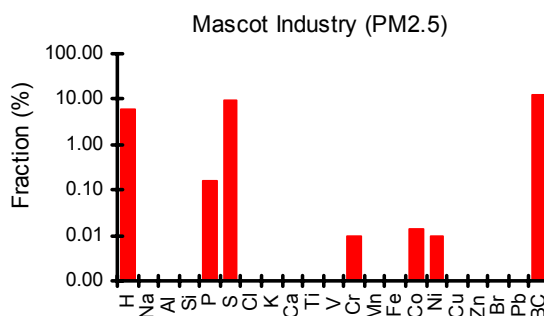
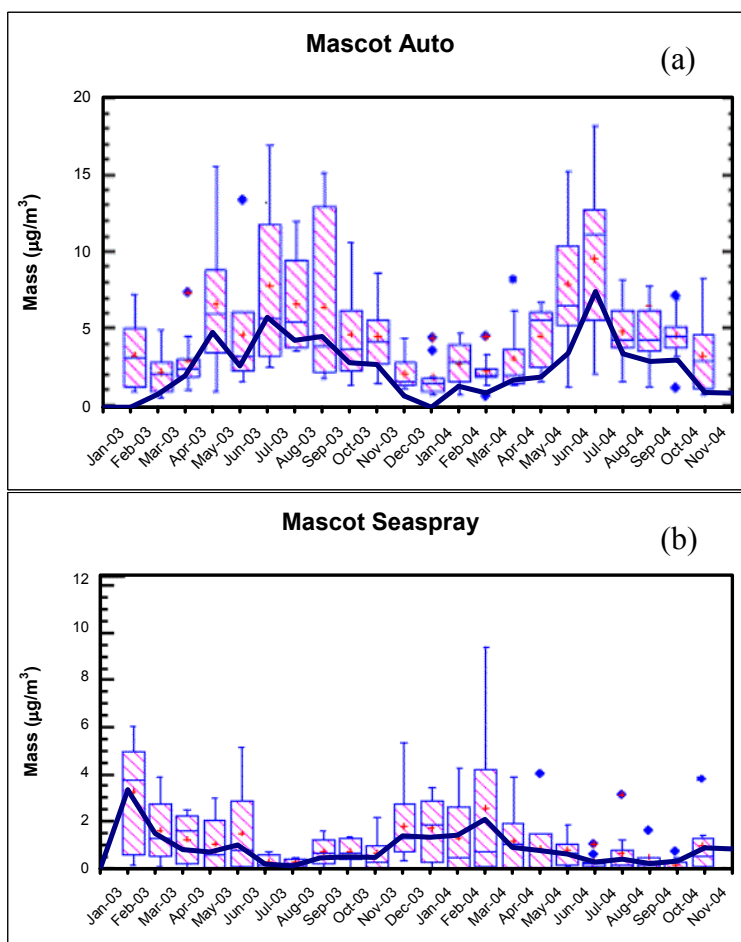


Fig. 2e Industry fingerprint

4. Comparisons with Positive Matrix Factorisation Methods (PMF)

PMF is a one step process it does not require predetermined fingerprints. These are estimated automatically from the input data together with their contribution estimates [10,11]. The user provides the number of required sources (or factors), a matrix of the elemental concentrations, the measurement errors and the minimum detection limits for each element and measurement on each day. Generally PMF produces 2 to 3 more primary source fingerprints than required by CMB methods. For example PMF analysis of the same data as used in the CMB calculations above on Mascot data for 2003-04 produced optimal fits for 7 individual source fingerprints. These were, as expected very similar to the ones used in the CMB calculations, however, the Sea spray fingerprint split into two fingerprints both with high Na as the primary identifying element. The first sea spray fingerprint had the expected Na to Cl fractional ratios for sea salt while the second had reduced Cl and high S indicating Cl loss in high sulfate environments. Interestingly, the PMF Auto fingerprint also split into two fingerprints, representing automobiles with high Pb and Br and little organics (petrol based) and automobiles with high organics and sulfur and low Pb and Br (diesel vehicles). This was a good demonstration of the increased power of PMF methods over CMB methods.

In order to directly compare the 5 average monthly CMB source fingerprint contributions with the daily 7 PMF source fingerprints, the PMF data was reduced to monthly averages and the two Sea spray and two Autos fingerprints combined. The results are plotted in Figs 3(a) to (d) as a solid curve for the CMB data and a box and whisker plot for the PMF data. The (+) and the horizontal bars in the box PMF data are the mean and median monthly values respectively. The correlations between the results of two methods are good for all four fingerprints plotted. The seasonal summer/ winter trends are faithfully reproduced by both methods. The CMB (solid curve) tends to follow the PMF median values (horizontal bars) better than the average values (+) within each box. Unlike the CMB data the PMF results generally do not go to zero which for a source like Autos at an urban site like Mascot, is a lot more logical. Also the PMF data generally over predict



the CMB data for the same month. This is because the PMF individual source fingerprints generally contain more element species than the manually generated CMB ones and hence have a higher mass content.

The average percentage source contributions for 2003-04 at Mascot were, $45 \pm 25\%$, $10 \pm 6\%$, $7 \pm 7\%$, $21 \pm 21\%$ and $22 \pm 14\%$ for the PMF Autos, Smoke, Soil, Sea spray and Industry fingerprints respectively. These values are similar but higher than the CMB estimates for reasons already given. Also the seasonal variations in the PMF data are more pronounced than the CMB data giving larger standard deviations in the average source contribution results.

Data for the smoke source contributions was not provided because of space limitations and its occurrence basically only in winter months.

5. Summary

The CMB and PMF methods produced consistent results, identifying similar source fingerprints and contributions. The PMF method generally produced more factors (sources) with more elemental

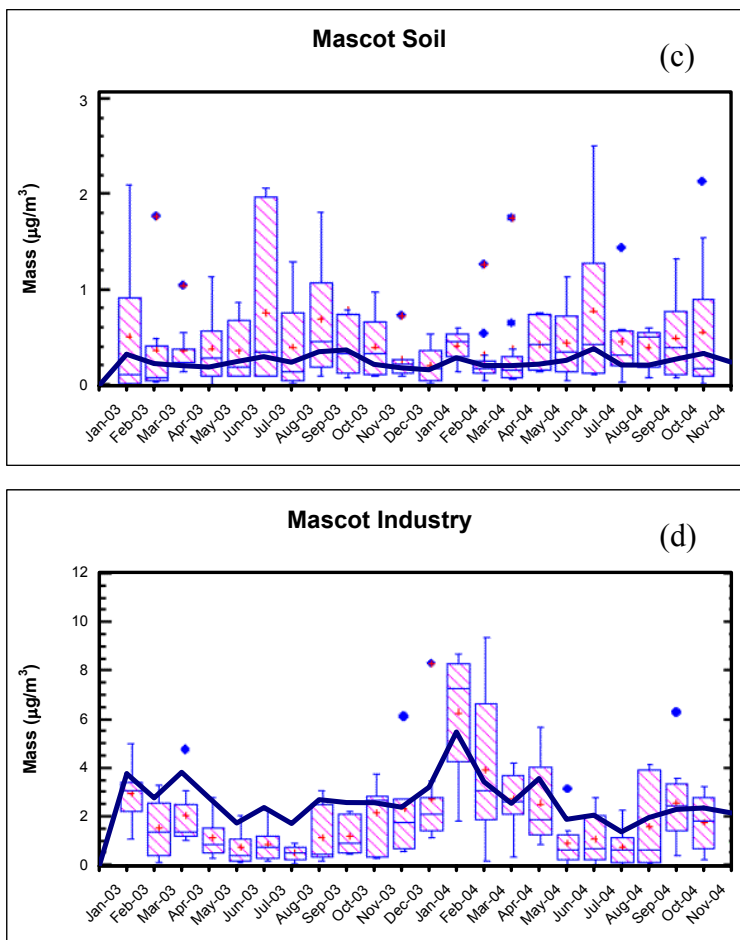


Fig. 3. (a)-(d) CMB (solid) and PMF (boxes) source comparisons detail in individual source fingerprints. PMF is a one step process and hence more efficient and less time consuming than the CMB method. Both methods rely heavily on a good knowledge of all the errors (and MDLs) contributing to the data set. Multi-elemental analytical IBA techniques like, PIXE, PIGE, PESA and RBS are ideally suited to both these source fingerprinting and apportionment methods because they provide datasets with excellent mass closure and can analyse hundreds of samples in a very short time, essential prerequisites.

6. Acknowledgements

The authors would like to thank all ANSTO accelerator staff for assistance with IBA analyses and Botany Council for financial support to operate the Mascot sampling site.

7. References

- [1]. D. D. Cohen, G.M. Bailey, R. Kondepudi, Nucl. Instr. and Methods, 109 (1996) 218-226.
- [2]. D. D. Cohen, Nucl. Instr. and Methods, 136B (1998) 14-22.
- [3]. D. D. Cohen, Ed Stelcer, David Garton, Nuclear Instr. and Methods, B190 (2002) 466-470
- [4]. D. D. Cohen, D. Garton, E. Stelcer, O. Hawas, T. Wang, S. Poon, J. Kim, B. Choi, S. Oh, H. Shin, M. Young, K. Uematsu, Journal of Geophysical Research, 109 (2004) D19S12, doi:10.1029/2003JD003569.
- [5]. D. D. Cohen, E. Stelcer, O. Hawas, D. Garton, Nucl. Instr. and Methods B219-220 (2004) 145-152.
- [6]. D. D. Cohen, D. Garton, E. Stelcer, O. Hawas, Radiation Physics and Chemistry, 71 (2004) 759-767.
- [7]. W.C. Malm, J.F. Sisler, D. Huffman, R.A. Eldred, T.A. Cahill, J. Geophys. Research, 99 (1994) 1347.
- [8]. P.K Hopke, Receptor Modeling for Air Quality Management, Elsevier Science, Amsterdam, (1991).
- [9]. J.G. Watson, N.F. Robinson, J.C. Chow, et al Receptor model technical series, volume III, CMB7 User's manual, US Environmental Protection Agency, Report EPA-450/4-90-004, January, 1990.
- [10]. P. Paatero and U Tapper, Environmetrics, 5 (1994) 111.
- [11]. P. Paatero, Chemometrics and Intelligent Laboratory Systems 37 (1997) 23.