

Carbonaceous aerosols contributed by traffic and solid fuel burning at a polluted rural site in the Northwestern England - Supplementary material

The purpose of this document is to describe the data processing methods used in the main article and the criteria that were employed to derive a given solution set. The procedure followed is similar to that of Allan et al. (2010), in that the number of factors were first constrained, followed by the correct value of 'fpeak'. The PMF Evaluation Toolkit (Ulbrich et al., 2009) was used, employing version 4.2 of the PMF2 executable (Paatero, 1997).

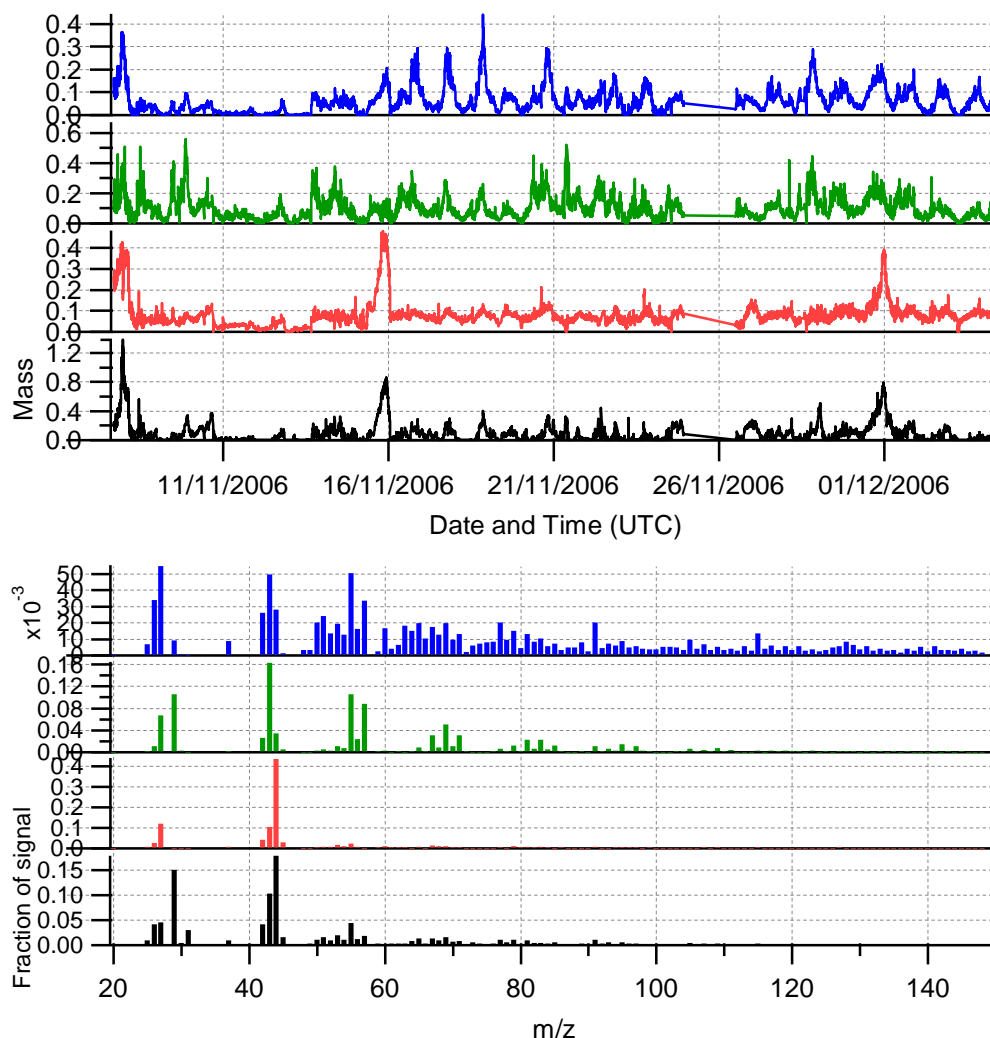
1. Data pretreatment

As this was an early deployment of the Manchester C-TOF-AMS, the operational methodologies were still in development and some issues were noted in certain channels relating to digitiser operation. Specifically these related to shifts in the electronic baseline, a fraction of single ion signals being lost to thresholding (Drewnick et al., 2005) and larger signals saturating. These resulted in slight nonlinearities in reported signals. These were most evident in some data products having nonzero values during routine filter tests and periods of low signal intensity. The channels affected were those with very large gas-phase or background signals or those that use subtractions based on these large signals (principally $m/z=28$) in the fragmentation table procedure (Allan et al., 2004), specifically $m/z=15, 16, 29$ and 44 . Without correcting for these, it was found that PMF would produce factors that were more indicative of these artefacts rather than real atmospheric phenomena. This was seen by the presence of step changes in the PMF time series corresponding to the shifts in instrument characteristics (corresponding to calibrations, power downs, etc.). An effort was made to remove these artefacts by applying time-dependent modifications to the fragmentation table, based on the filter tests and changes in the instrument performance (diagnosed by changes in indicators such as the ratios of air signals at $m/z=28$ and 32). While this approach was somewhat arbitrary, it did result in PMF generating more contiguous outputs, so it was deemed to be largely successful. However, given the nonlinear nature of the problem, it is possible that some artefacts remain in the data, though none are observable. Note that any remaining effects would diminish for the total mass concentrations reported in the main article, as large fractions of the summed signals are from channels in the mass spectra that are unaffected by this.

In order to avoid duplication of the $m/z=44$ signal, the contributions below $m/z=20$ (which are predominantly based on this signal) were removed for subsequent re-insertion after the data had been processed. There was also deemed to be little usable signal in the channels above $m/z=150$, so these were also removed to reduce the influence from 'weak' variables (Paatero and Hopke, 2003).

2. Selection of number of factors

A 4-factor solution was disregarded, as it was found to introduce 'splitting' between factors (Ulbrich et al., 2009), specifically the OOA1 (LV-OOA) factors (solution for $f_{\text{peak}}=0$ shown).



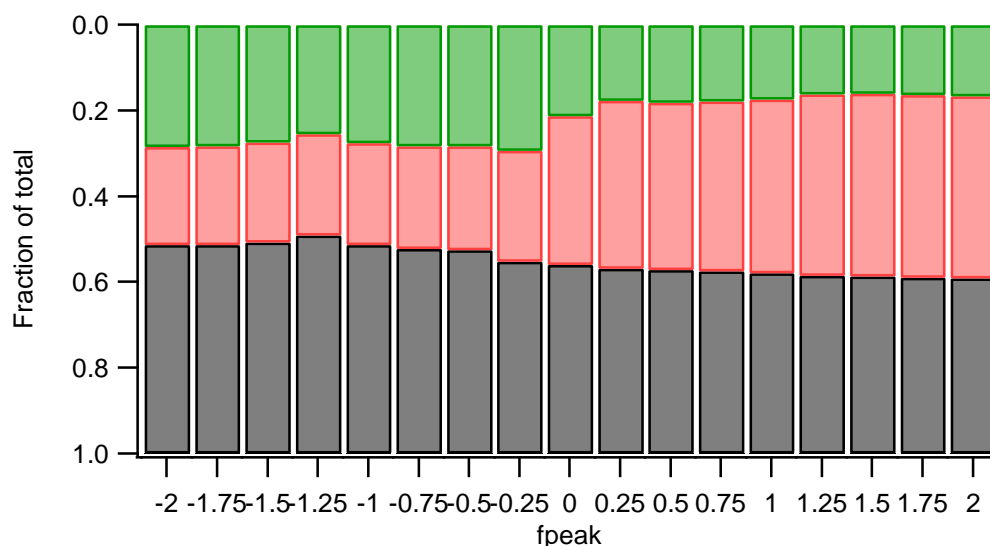
Note the similarity in the time series and profiles of factors 1 and 2 (the bottom two factors). While it is possible that this represents real chemical variability, it could also be further manifestations of the aforementioned instrument artefacts, especially given that the principal peak ($m/z=44$) was one of those affected. Given that no suitable external comparisons were available that could be used to verify whether this was the case or not, this solution was not considered reliable. Increasing the number of factors further resulted in more splitting still.

A 3-factor solution was found to consistently give unique solutions, which are the ones used in the main article). The bootstrapping diagnostics were not as favourable as the factors used in Allan et al. (2010) with the 3 time series reporting SD_{TS}/TS values of 13.5, 10.2 and 16.7 % respectively (the Allan et al. values of the solutions used were all less than 10 %). Also, the maximum profile standard deviations were 0.7, 1.1 and 2.1 % respectively (the Allan et al. values were all less than 1 %). However, there was no dependency on starting seed noted, so the solutions were still deemed acceptable.

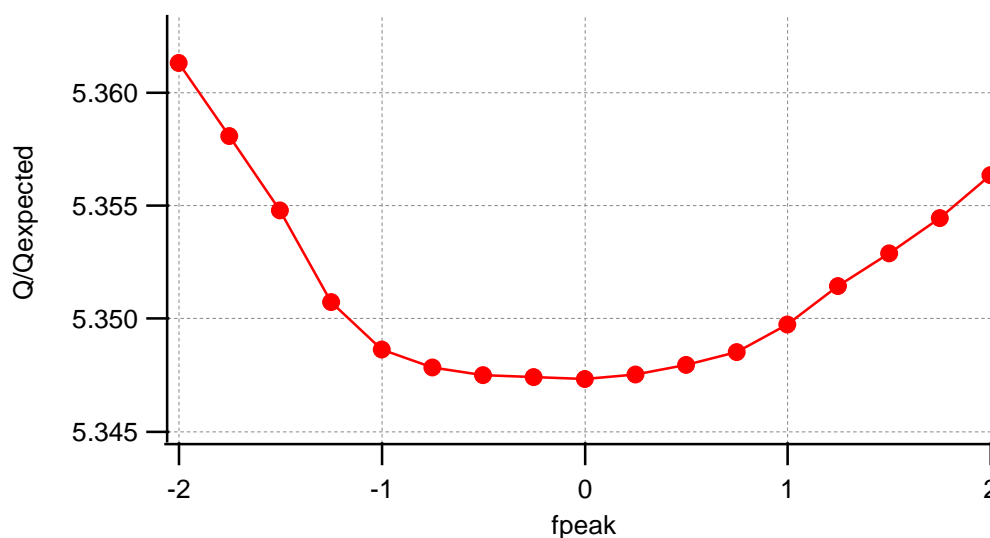
3. Selection of f_{peak}

As with most multivariate analysis, the outputs of PMF carry a certain amount of ambiguity associated with rotations within the solution space. The PMF2 algorithm comes with the facility to vary a parameter known as 'fpeak' to facilitate exploring this ambiguity. However, it should be noted that as this is only a single parameter being used to explore a multidimensional space, it will not necessarily produce the most definitive solution for a given dataset, so any chosen solution can only be seen as a 'best estimate' within this context. It may be possible to improve the solution by applying further constraints to the dataset such as target factors (e.g. Lanz et al., 2008), but this involves applying a priori knowledge of the behaviour of the particulates and risks biasing the outcome.

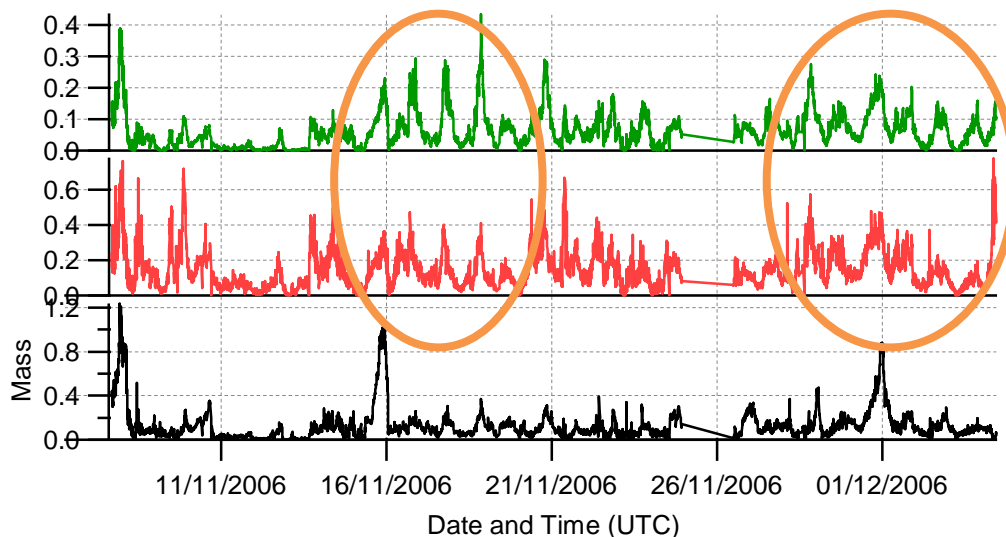
For the 3-factor solution, the behaviour of the fits seemed to occupy one of two states, depending on whether fpeak was positive or negative:



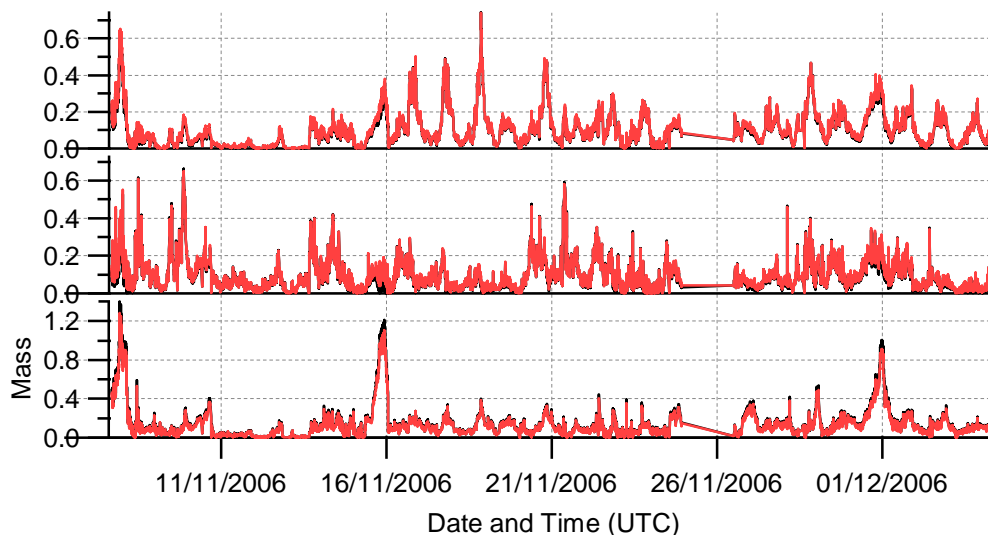
The optimum values of Q seemed to form a trough between -1 and 0.75:



For non-negative values, two factors began to show too much temporal similarity (the top two traces in the fpeak=0.25 example shown below, similarities highlighted for clarity). This would indicate that an unacceptable amount of 'mixing' is taking place (Ulbrich et al., 2009):



This mixing was seen to be less of a problem with the solutions using negative values of f_{peak} . The changes within the acceptable range of negative values were very subtle, as shown in the following comparison of -1 and -0.25:



The main differences relate to the events on 7/11 and 15/11. However, these differences are very slight overall. Therefore, the solution for -0.25 was chosen as this had the smallest Q/Q_{exp} value (5.347).

Allan, J. D., Coe, H., Bower, K. N., Alfarra, M. R., Delia, A. E., Jimenez, J. L., Middlebrook, A. M., Drewnick, F., Onasch, T. B., Canagaratna, M. R., Jayne, J. T., and Worsnop, D. R.: A generalised method for the extraction of chemically resolved mass spectra from aerodyne aerosol mass spectrometer data, *J. Aerosol. Sci.*, 35, 909-922, 2004.

Allan, J. D., Williams, P. I., Morgan, W. T., Martin, C. L., Flynn, M. J., Lee, J., Nemitz, E., Phillips, G. J., Gallagher, M. W., and Coe, H.: Contributions from transport, solid fuel burning and cooking to primary organic aerosols in two uk cities, *Atmos. Chem. Phys.*, 10, 647-668, 2010.

Drewnick, F., Hings, S. S., DeCarlo, P., Jayne, J. T., Gonin, M., Fuhrer, K., Weimer, S., Jimenez, J. L., Demerjian, K. L., Borrmann, S., and Worsnop, D. R.: A new time-of-flight aerosol mass spectrometer (tof-ams)-instrument description and first field deployment, *Aerosol Sci. Technol.*, 39, 637-658, 2005.

Lanz, V. A., Alfarra, M. R., Baltensperger, U., Buchmann, B., Hueglin, C., Szidat, S., Wehrli, M. N., Wacker, L., Weimer, S., Caseiro, A., Puxbaum, H., and Prevot, A. S. H.: Source attribution of submicron organic aerosols during wintertime inversions by advanced factor analysis of aerosol mass spectra, *Environ. Sci. Technol.*, 42, 214-220, Doi 10.1021/Es0707207, 2008.

Paatero, P.: Least squares formulation of robust non-negative factor analysis, *Chemometr Intell Lab*, 37, 23-35, 1997.

Paatero, P., and Hopke, P. K.: Discarding or downweighting high-noise variables in factor analytic models, *Anal Chim Acta*, 490, 277-289, Doi 10.1016/S0003-2670(02)01643-4, 2003.

Ulbrich, I. M., Canagaratna, M. R., Zhang, Q., Worsnop, D. R., and Jimenez, J. L.: Interpretation of organic components from positive matrix factorization of aerosol mass spectrometric data, *Atmos. Chem. Phys.*, 9, 2891-2918, 2009.