

Simultaneous factor analysis of organic particle and gas mass spectra: AMS and PTR-MS measurements at an urban site

J. G. Slowik^{1,3}, A. Vlasenko^{1,3}, M. McGuire^{2,3}, G. J. Evans^{2,3}, and J. P. D. Abbatt^{1,3}

¹Department of Chemistry, University of Toronto, Toronto, Canada

²Department of Chemical Engineering and Applied Chemistry, University of Toronto, Toronto, Canada

³Southern Ontario Centre for Atmospheric Aerosol Research, University of Toronto, Toronto, Canada

Received: 19 February 2009 – Published in Atmos. Chem. Phys. Discuss.: 11 March 2009

Revised: 7 January 2010 – Accepted: 11 January 2010 – Published: 19 February 2010

Abstract. During the winter component of the SPORT (Seasonal Particle Observations in the Region of Toronto) field campaign, particulate non-refractory chemical composition and concentration of selected volatile organic compounds (VOCs) were measured by an Aerodyne time-of-flight aerosol mass spectrometer (AMS) and a proton transfer reaction-mass spectrometer (PTR-MS), respectively. Sampling was performed in downtown Toronto ~15 m from a major road. The mass spectra from the AMS and PTR-MS were combined into a unified dataset, which was analysed using positive matrix factorization (PMF). The two instruments were given balanced weight in the PMF analysis by the application of a scaling factor to the uncertainties of each instrument. A residual based metric, $\Delta\bar{e}_{sc}$, was used to evaluate the instrument relative weight within each solution. The PMF analysis yielded a 6-factor solution that included factors characteristic of regional transport, local traffic emissions, charbroiling and oxidative processing. The unified dataset provides information on emission sources (particle and VOC) and atmospheric processing that cannot be obtained from the datasets of the individual instruments: (1) apportionment of oxygenated VOCs to either direct emission sources or secondary reaction products; (2) improved correlation of oxygenated aerosol factors with photochemical age; and (3) increased detail regarding the composition of oxygenated organic aerosol factors. This analysis represents the first application of PMF to a unified AMS/PTR-MS dataset.

1 Introduction

Air pollutants have important effects on ecosystems (Schindler, 1988; Driscoll et al., 2003), human health (Dockery and Pope, 1994; Pope and Dockery, 2006), atmospheric visibility (Watson, 2002) and climate change (Jacobson, 2001; Ramanathan et al., 2001). Organic pollutants exist in both the gas and particle phases and vary in terms of their composition and source. Both particulate organic species and volatile organic compounds (VOCs) may enter the atmosphere either as a result of primary emissions such as fossil fuel combustion or through secondary processes such as gas-phase or heterogeneous chemical reactions. A quantitative understanding of VOC and particulate organic sources and atmospheric processing is necessary to reduce uncertainties in global climate models and for the development of pollution mitigation strategies to improve air quality (Kanakidou et al., 2005).

One approach to estimating the effects of source contributions and atmospheric processing to particle and VOC composition and concentration is through the use of receptor modelling techniques such as positive matrix factorization (PMF) (Paatero and Tapper, 1994; Paatero, 1997) and UNMIX (Lewis et al., 2003). Multivariate statistical techniques are used to deconvolve a time series of simultaneous measurements into a set of factors and their time-dependent concentrations. These factors may then be related to emission sources, chemical composition and/or atmospheric processing, depending on their specific chemical and temporal characteristics. Because receptor models require no a priori knowledge of meteorological conditions or emission inventories, they are ideal for use in locations where emission inventories are poorly characterised or highly complicated (e.g. urban areas), or where atmospheric processing plays a major role.



Correspondence to: J. P. D. Abbatt
(jabbatt@chem.utoronto.ca)

Factor analysis techniques have been previously applied to a range of VOC measurements (Buzcu and Frazier, 2006; Holzinger et al., 2007; Lanz et al., 2008b), yielding factors related to atmospheric processing and sources such as traffic and biogenic emissions. Although PMF has previously been applied to particle measurements (Ramadan et al., 2000; Polissar et al., 2001; Lee et al., 2003; Owega et al., 2004), a detailed treatment of the organic component has only recently been attempted. Lanz et al. (2007) applied PMF to organic aerosol mass spectra obtained from an aerosol mass spectrometer (AMS), obtaining six distinct factors relating to aerosol composition, volatility range and specific sources such as charbroiling and wood burning emissions. Zhang et al. (2005) developed a technique for deconvolving AMS mass spectra into oxygenated organic aerosol (OOA) and hydrocarbon-like organic aerosol (HOA) using m/z 44 (CO_2^+) and m/z 57 (C_4H_9^+ , $\text{C}_3\text{H}_5\text{O}^+$) as OOA and HOA tracers. Other studies have typically included selected AMS mass spectral fragments in receptor modelling (typically restricted to inorganic species, m/z 44 and m/z 57) (Buset et al., 2006; Quinn et al., 2006), classified organics based on their thermal properties (Zhao and Hopke, 2006), or treated the organics as a single species for analysis.

Recent studies indicate that the traditional binary treatments of atmospheric organics as either gases or particles may be inadequate (Robinson et al., 2007). A proposed alternative is the treatment of organic species through the use of a volatility basis set (Donahue et al., 2006), in which the partitioning behaviour of organics are considered over a range of volatilities. Such issues highlight the need for analytical approaches capable of simultaneous, cohesive analysis of gas and particle data. One such approach is presented here, through the application of the PMF receptor modelling technique to coupled gas and particle data.

In this experiment, simultaneous measurements of the mass spectra of particulate organics and VOCs were obtained using an Aerodyne aerosol mass spectrometer (AMS) and a proton transfer reaction mass spectrometer (PTR-MS). The measurements from these two instruments were combined into a single dataset and analysed using PMF. This analysis yielded factors related to emission sources and chemical composition, specifically the degree of oxygenation. These factors were compared to the results obtained from PMF analysis conducted separately on the individual AMS and PTR-MS datasets. This is the first application of PMF analysis to a unified AMS/PTR-MS dataset. In the present study, benefits of the unified analysis are evident in the apportionment of gas and particle constituents to primary emission and secondary reaction processes, correlation of oxygenated aerosol factors with photochemical age and detailed composition of oxygenated organic aerosol factors.

2 Materials and methods

2.1 Sampling and instrumentation

During the winter component of the SPORT (Seasonal Particle Observations in the Region of Toronto) field campaign (22 January 2007 to 5 February 2007), a time-of-flight aerosol mass spectrometer (C-ToF-AMS) (Aerodyne Research, Inc., Billerica, MA, USA) and a proton transfer reaction-mass spectrometer (PTR-MS) (Ionicon Analytik, Innsbruck, Austria) were deployed in downtown Toronto (Wallberg Building, University of Toronto). The sampling inlet consisted of a 10 cm diameter circular duct located ~ 5 m above ground and ~ 15 m north of College Street. College Street has a weekday traffic volume of approximately 33 000 vehicles per day, similar to other major roadways in Toronto (Godri et al., 2009). The site is situated in a mixed commercial/residential area. Known local particle emissions sources include automobile traffic, street food vendors and restaurants. Ambient air was sampled continuously at a rate of 300 L/min through a 10.2 cm outer diameter duct. The AMS sampling line was ~ 7 m long and constructed from ~ 6 m stainless steel and ~ 1 m conductive silicone tubing (TSI, inc., Shoreview, MN, USA). The PTR-MS utilized a Teflon sampling line with a length of ~ 2.5 m.

The literature provides detailed descriptions of the AMS (Drewnick et al., 2005; Canagaratna et al., 2007) and PTR-MS (Hansel et al., 1995; Lindinger et al., 1998). The AMS provides the size-resolved, non-refractory composition of submicron particles, while the PTR-MS provides the concentrations of VOCs with a proton affinity greater than that of water. The AMS and PTR-MS recorded data on 1 min and 30 s time intervals, respectively; data from both instruments were re-averaged into 15 min time intervals. For the analysis of both individual and unified datasets, time periods containing mass spectra from only one instrument were excluded, yielding a total of 1148 analysed mass spectra.

The longer 15 min interval was selected for analysis because of (1) signal-to-noise considerations (e.g. for PTR-MS aromatics, the longer averaging periods increased signal-to-noise by approximately a factor of 5.5, from 1.0–1.5 to 6.0–9.0), and (2) an indication from preliminary PMF analysis of the unified AMS/PTR-MS dataset indicating that the shorter averaging times led to solutions in which the resolved factors contained data from either the AMS or PTR-MS, but not both. It is speculated that this second issue is caused by the different residence times in the instrument sampling lines, causing imperfect synchronization of the instrument sampling intervals. The importance of this effect is reduced by a longer averaging interval (because the non-synchronized averaging time is a smaller fraction of the total interval). This is supported by the observation that longer averaging intervals (not shown) provide consistent results with the 15 min dataset. Comparison of the effect of averaging time with other datasets (e.g. rural locations where the

particle/gas composition is less affected by rapidly changing point sources) will provide further insight.

AMS data analysis was performed using the ToF-AMS Analysis Toolkit v.1.44 (D. Sueper, University of Colorado-Boulder, Boulder, CO, USA) for the Igor Pro software package (Wavemetrics, Inc., Portland, OR, USA). The organic components of $m/z \leq 300$ were included in the PMF analysis. Mass fragments containing no organic signal were excluded, resulting in 270 analysed m/z . At m/z that contain signals from both inorganic and organic ions, the organic contribution was determined through a fragmentation pattern-based analysis routine (Allan et al., 2004). The procedure for calculating AMS uncertainties is described in detail in the literature (Allan et al., 2003) and summarized briefly as follows. The distribution of ion signals recorded for a given ensemble are represented as a Poisson distribution and convolved with a detector-dependent Gaussian distribution representing the variation in signal obtained for a single ion. During operation, the particle beam is alternately blocked (yielding a background measurement) and unblocked. Uncertainties are calculated independently for each mode and summed in quadrature, yielding the expression $\Delta I_d = \alpha \frac{\sqrt{I_o + I_b}}{\sqrt{t_s}}$. Here I_o and I_b are the ion signals in the unblocked and blocked (background) positions, t_s is the sampling time and α is a factor accounting for the width of the Gaussian ion signal distribution.

Due to signal-to-noise constraints imposed by the 30 s sampling intervals, the PTR-MS was not used to scan the entire mass spectrum and instead was set to measure specific masses. Ions at m/z 31 (formaldehyde), 43 (alkyl fragments, propylene, acetic acid, acetone, peroxyacetyl nitrate (PAN)), 45 (acetaldehyde), 59 (acetone, propanal, glyoxal), 61 (acetic acid), 73 (methyl ethyl ketone (MEK), methylglyoxal, butanal), 79 (benzene), 93 (toluene), 107 (xylenes, ethyl benzene, benzaldehyde), and 121 (trimethyl benzene, ethyl toluene, propyl benzene) were included in the PMF analysis (de Gouw and Warneke, 2007), while m/z 33 (methanol), 37 (water dimer), 42 (acetonitrile), 69 (isoprene) and 129 (naphthalene) were measured but excluded from the PMF due to poor signal-to-noise (m/z 42, 69, 129), signal exclusively due to the water dimer ion (m/z 37) or problems with the measurement dynamic range due to persistent local sources (spikes of m/z 33 (methanol) from windshield washer fluid). Uncertainties for the PTR-MS were calculated from background levels and Poisson ion counting statistics as described in the literature and summarized below (de Gouw et al., 2003). Typical uncertainty values were in the range of 2 to 18% of signal, depending on the m/z . Background levels were obtained by sampling through a charcoal cartridge (Supelco, Bellefonte, PA, USA). The overall uncertainty is given by $\Delta(I - I_c) = \sqrt{\frac{I}{\tau} + \frac{I_c}{\tau_c}}$, where I is the ion signal, τ is the averaging time and the “c” subscript denotes background measurements. PTR-MS calibration was performed using a custom-made standard (Apel-Riemer Environmental

Inc., Broomfield, CO, USA), yielding species-dependent calibration factors and detection limits as described elsewhere (Vlasenko et al., 2009). Because of the large number of species fragmenting to m/z 43, the calibration factor at this m/z was estimated as bounded by those of the oxygenated species and aromatics. For factors dominated by one or the other, this could yield uncertainties of up to $\sim \pm 50\%$ in the mixing ratio at this m/z . However, m/z 43 signals are sufficiently low that even this worst case would not greatly influence the reported factor mixing ratios.

2.2 Positive Matrix Factorization (PMF)

The AMS and PTR-MS mass spectral time series and uncertainties obtained, as described above, were analysed using the PMF2 software package version 4.2 (P. Paatero, U. of Helsinki, Finland), together with a modified version of the CU AMS PMF Tool (Ulbrich et al., 2009a). Two methods of analysis were employed. In the first method, PMF was separately applied to the AMS and PTR-MS data. In the second method, the data from the two instruments were combined into a single dataset and PMF was applied to this unified dataset.

The PMF model is described in detail in the literature (Paatero and Tapper, 1994; Paatero, 1997). Here, we provide a brief summary and discuss the special considerations required to apply PMF to the unified dataset. PMF operates on the input data matrix \mathbf{X} and the corresponding uncertainty matrix \mathbf{S} . In the present study, \mathbf{X} is the time series of mass spectra collected by the AMS and/or PTR-MS. The matrix \mathbf{S} , therefore, contains the uncertainty in the measurement of the signal of each m/z at every point in time. The PMF model is described by the matrix equation:

$$\mathbf{X} = \mathbf{G}\mathbf{F} + \mathbf{E} \quad (1)$$

Here the columns of the \mathbf{G} matrix contain the factor time series and the rows of the \mathbf{F} matrix contain the factor mass spectra. The number of factors in a solution is user-determined through criteria discussed later. The \mathbf{E} matrix contains the residuals and is defined by Eq. (1). The PMF model solves Eq. (1) by using a weighted least-squares algorithm to minimize the sum of squares, Q , defined as:

$$Q = \sum_i^n \sum_j^m (e_{ij}/s_{ij})^2 \quad (2)$$

Here e_{ij} are the elements of the residual matrix \mathbf{E} and s_{ij} are the elements of the uncertainty matrix \mathbf{S} (i and j are the time and m/z indices, respectively, while n and m denote the number of time points and number of m/z). The theoretical value of Q , denoted Q_{expected} , can be estimated as:

$$Q_{\text{expected}} = \text{NumElements}(\mathbf{X}) - (\text{NumElements}(\mathbf{G}) + \text{NumElements}(\mathbf{F})) \quad (3)$$

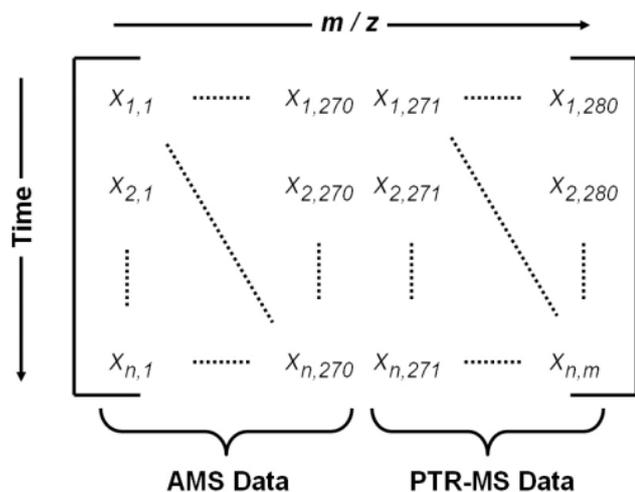


Fig. 1. Schematic of the unified AMS/PTR-MS data matrix (\mathbf{X}).

Here, the NumElements operation denotes the number of elements in the indicated matrix. In practice, Q is expected to be somewhat larger than Q_{expected} for ambient data because the data cannot be perfectly represented by a finite number of factors.

The unified AMS/PTR-MS data matrix, \mathbf{X}_{UN} , is shown in Fig. 1. The associated uncertainty matrix, \mathbf{S}_{UN} , is constructed similarly. In examining the solutions to the unified dataset, an important consideration is the fit quality of the PMF model to the data from each instrument. When no instrument weighting is applied to the unified dataset, the AMS component of the dataset is well-represented in the solution, while the PTR-MS component is poorly represented. This is due to (1) the large size of the AMS dataset (270 AMS m/z vs. 10 PTR-MS m/z), (2) co-variance between m/z of a particular instrument (e.g. AMS m/z 43, 57, 71, 85, etc. are somewhat correlated because they all contain contributions from alkane fragments), and (3) the signal-to-noise ratio within the instrument datasets (m/z with higher signal-to-noise typically have fractionally more signal apportioned to factors (instead of the residuals) than do low signal-to-noise m/z). In the present case, the m/z -to- m/z variations in signal-to-noise for a given instrument are larger than systematic differences between the instruments; however, instrument differences may still exert some influence). Therefore, it is necessary to increase the weight of the PTR-MS component so that the PMF solution provides a balanced representation of the data from both instruments. Here, the instruments are balanced by the application of a weighting factor to the PTR-MS uncertainties and the instrument's relative weight is evaluated utilizing a scaled residual-based metric, as discussed in Sect. 2.2.2. While other weighting methods and evaluation metrics could potentially be devised (e.g. an alternate weighting method is briefly discussed in Sect. 3.3.3), the evaluation of the instrument weight is essential to ensure that

Table 1. Input parameters for PMF analysis.

	Dataset		
	AMS	PTR-MS	Unified
Matrix rows	1148	1148	1148
Matrix columns	270	10	280
Range of p analysed	1 to 10	1 to 8	1 to 10
Range of fPeak analysed	-1.5 to 1.5	-1.5 to 1.5	-1.5 to 1.5
Number of random starts	100	100	100
Outlier limit (α)	4	4	4
C3 error coefficient	0	0	0

the PMF algorithm does not discard data from an included instrument. A consequence of the selected uncertainty weighting method is that the PMF2 robust operating mode (in which outlying data points are iteratively downweighted) cannot be used. Therefore, an alternate method of outlier downweighting is developed (pseudo-robust method) and discussed in Sect. 2.2.1. For consistency, the pseudo-robust method was utilized for both the individual and unified datasets.

Input parameters for the three datasets are summarized in Table 1; the α parameter in this table is defined below in Eq. (4). Note that matrix rotations were explored through the fPeak parameter and the identification of a global minimum solution was supported by initiating PMF from 100 random starting points (seed parameter). For the individual AMS and PTR-MS datasets, both robust and pseudo-robust methods were used, yielding nearly identical \mathbf{F} and \mathbf{G} matrices. In most cases, convergence was obtained when 5 solution steps yielded $\Delta Q < 0.1$. The exceptions are solutions to the AMS dataset at fPeak < -0.75 , where ΔQ values from 0.2 to 5 were required.

The uncertainty matrix \mathbf{S} was in all cases calculated from instrument operating principles and an error coefficient C3=0 value was selected. This choice sometimes resulted in Q -values being significantly higher than expected. The selection of C3=0 and its implications are discussed further in Sect. 3.3.4.

2.2.1 Pseudo-robust outlier treatment

An outlier is defined as a data point which satisfies:

$$|e_{ij}/s_{ij}| > \alpha \quad (4)$$

In the robust mode, PMF2 iteratively downweights outliers, preventing them from dominating the model fit. As discussed in the next section, the uncertainty weighting method uses a modified set of uncertainties, denoted $s_{\text{inst},ij}$, where the $s_{\text{inst},ij}$ of one instrument are scaled in relation to those of the other. Therefore, the robust mode cannot be used because Eq. (1) outliers cannot be correctly identified and Eq. (2) for strongly weighted $s_{\text{inst},ij}$, the robust mode counteracts the

uncertainty weighting, thereby preventing a balanced solution from being reached.

In the present dataset, outliers mostly occur during periods of high particle and/or gas concentrations and at m/z with consistently high signal-to-noise. Under such conditions, e_{ij}/s_{ij} may become large while e_{ij}/x_{ij} remains small. This is a result of issues such as (1) minor variations in source profiles with time and (2) the general approximation inherent in PMF that ambient data may be represented through a finite number of static factors. As a result, it is desirable to retain information from these periods, but to prevent them from unduly pulling the model fit. Therefore, we treat outliers with a downweighting procedure, rather than excluding data altogether. It is desirable to obtain a solution in which the relationship between scaled residuals is preserved, but the outliers do not dominate the fit.

The pseudo-robust method introduced here is modelled on the robust PMF analysis (Paatero, 1997). We describe the method in terms of a generic uncertainty matrix \mathbf{S} (which would be \mathbf{S}_{inst} in the pseudo-robust analysis of the unified dataset, see Sect. 2.2.2). In robust PMF, the PMF task is defined as:

$$\underset{\mathbf{G}, \mathbf{F}}{\text{argmin}} = \sum_{j=1}^m \sum_{i=1}^n \frac{(\mathbf{X} - \mathbf{Y})^2}{h_{ij}^2 s_{ij}^2} \quad (5)$$

Here \mathbf{Y} is the data matrix reconstructed from the PMF solution (i.e. $\mathbf{Y} = \mathbf{G}\mathbf{F}$), and h_{ij} are downweighting factors applied to the outliers according to the criteria:

$$h_{ij}^2 = 1 \quad \text{if } |e_{ij}/s_{ij}| \leq \alpha \quad (6a)$$

$$h_{ij}^2 = \left| \frac{e_{ij}}{s_{ij}} \right| / \alpha \quad \text{otherwise} \quad (6b)$$

In robust PMF, the h_{ij} are calculated for each iteration of the solution process. For pseudo-robust analysis, only a single calculation of the h_{ij} is performed. For each unique combination of p , instrument weight, fPeak and seed, PMF is applied twice. The first application is to the \mathbf{X} and \mathbf{S} matrices, and no downweighting of outliers is performed. From these results and Eq. (6), a new uncertainty matrix \mathbf{S}' (containing matrix elements s'_{ij}) is calculated as:

$$s'_{ij} = s_{ij} \quad \text{if } |e_{ij}/s_{ij}| \leq \alpha \quad (7a)$$

$$s'_{ij} = h_{ij} s_{ij} = \sqrt{\left| \frac{e_{ij} s_{ij}}{\alpha} \right|} \quad \text{otherwise} \quad (7b)$$

A second PMF calculation is then performed on \mathbf{X} and $\mathbf{S}'_{\text{inst}}$, yielding \mathbf{F} , \mathbf{G} , and \mathbf{E} for analysis.

2.2.2 Instrument weighting

Balanced weighting of the AMS and PTR-MS is implemented as follows. Unless otherwise noted, \mathbf{S} and \mathbf{S}' are

treated identically; for brevity we refer only to \mathbf{S} here. The constraint applied to the fit of each m/z in PMF is determined by \mathbf{S} . Instrument weight can, therefore, be controlled by the application of a scaling factor to selected components of \mathbf{S} . Here, the instrument relative weight is controlled by the application of the factor C_{PTR} to PTR-MS components of \mathbf{S} , yielding a new uncertainty matrix \mathbf{S}_{inst} (containing matrix elements $s_{\text{inst},ij}$) as follows:

$$s_{\text{inst},ij} = s_{\text{inst},ij} \quad \text{for } j = \text{AMSM}/z \quad (8a)$$

$$s_{\text{inst},ij} = s_{ij} / C_{\text{PTR}} \quad \text{for } j = \text{PTR-MSM}/z \quad (8b)$$

As C_{PTR} increases, the PTR-MS contribution to Q increases relative to the AMS contribution. This causes the PMF2 algorithm to find a solution that better represents the PTR-MS component of the dataset.

Solutions in which the AMS and PTR-MS datasets are balanced are determined by analysis of the scaled residuals for each instrument. For a balanced solution, the magnitude of the scaled residuals are required to be independent of the measuring instrument. This requirement is evaluated through the quantity $\Delta \bar{e}_{sc}$, defined as:

$$\Delta \bar{e}_{sc} = \overline{\left(\frac{|e_{ij}|}{s'_{ij}} \right)_{\text{AMS}}} - \overline{\left(\frac{|e_{ij}|}{s'_{ij}} \right)_{\text{PTR}}} \quad (9)$$

If $\Delta \bar{e}_{sc} = 0$, the AMS and PTR-MS data are balanced in the PMF solution. Values of $\Delta \bar{e}_{sc} < 0$ indicate that the AMS is overweighted (because the scaled residuals for the PTR-MS are larger than for the AMS), while $\Delta \bar{e}_{sc} > 0$ indicates that the PTR-MS is overweighted. Note that s'_{ij} is used in Eq. (9) rather than s_{ij} to prevent outlier domination of $\Delta \bar{e}_{sc}$. Further, s'_{ij} is used instead of $s'_{\text{inst},ij}$, i.e. C_{PTR} is removed. This is because the inclusion of C_{PTR} potentially affects $\Delta \bar{e}_{sc}$ without producing changes in the \mathbf{F} and \mathbf{G} matrices. For example, for very low or very high values of C_{PTR} , only one instrument is significantly considered by the PMF algorithm. In this scenario, a small change in C_{PTR} does not affect the solution because Q remains dominated by a single instrument. However, the change in C_{PTR} would affect $\Delta \bar{e}_{sc}$. To prevent such an artifact, the (unweighted) s_{ij} are used in Eq. (3).

3 Results and discussion

We first present results obtained from PMF analysis of the individual AMS and PTR-MS datasets. We then discuss the PMF analysis of the unified AMS/PTR-MS dataset in terms of (1) selection and evaluation of solutions, (2) physical interpretation of the extracted factors, and (3) comparison of the information yielded by the individual and unified analyses.

Table 2. Parameters describing the selected solution for each dataset.

	Dataset		
	AMS	PTR-MS	Unified
C_{PTR}	N/A	N/A	10
Number of factors (p)	5	5	6
fPeak	0	0	0
$Q_{\text{pseudo}}/Q_{\text{expected}}$	0.96	2.55	3.15
$Q_{\text{robust}}/Q_{\text{expected}}$	1.04	2.42	N/A
$Q_{\text{true}}/Q_{\text{expected}}$ (pseudo-robust method)	2.82	3.41	15.89
$Q_{\text{true}}/Q_{\text{expected}}$ (PMF2 robust mode)	7.19	3.51	N/A
Percent of data points classified as outliers (pseudo-robust method, $\alpha=4$)	1.71	1.82	4.79
Δe_{sc}	N/A	N/A	0.052

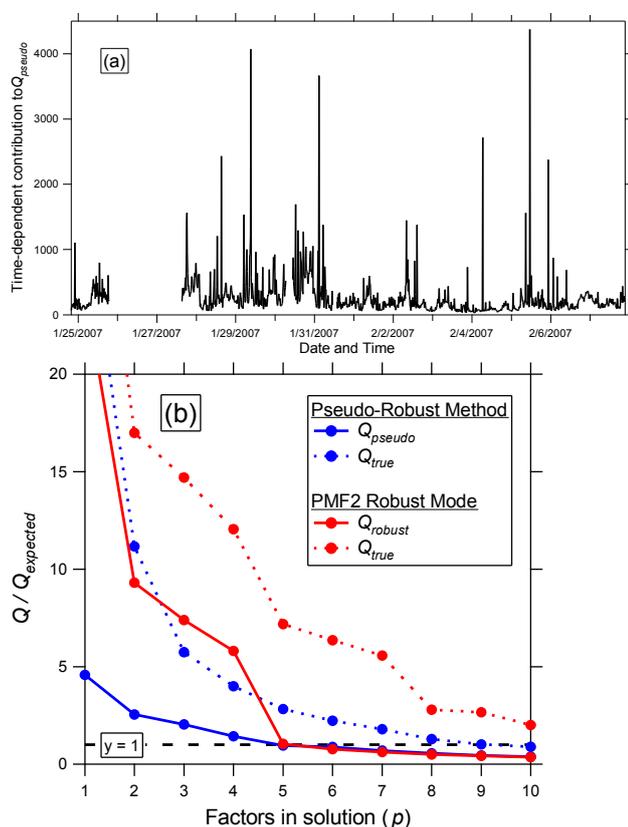


Fig. 2. (a) Time-dependent contribution to Q_{pseudo} for the AMS dataset at $p=5$. Ticks denote 00:00 of the specified day; this convention holds throughout the manuscript. (b) Ratio of Q/Q_{expected} as a function p for pseudo-robust method and robust mode.

3.1 AMS dataset

PMF analysis of an AMS organic mass spectral dataset has been previously described in detail (Lanz et al., 2007; Ulbrich et al., 2009a), and a similar approach was used in the

present study. A crucial consideration is the number of factors used in the PMF model (p). This number is somewhat subjective because the PMF model can be run with an arbitrary number of factors and no unambiguous method for determining the “correct” p exists. As discussed below, a solution at $p=5$ was selected based on the effects of the number of factors on the time-dependent contribution to Q (denoted Q_{cont}), correlations between the factor time series and external tracers and physical interpretation of the factor mass spectra. Summary statistics for this solution are presented in Table 2. The $p=5$ solution contains the following factors: oxygenated organic aerosol (OOA-1), hydrocarbon-like organic aerosol (HOA), charbroiling, biomass burning organic aerosol (BBOA) and an unidentified point source to the north of the measurement site.

As discussed in Sect. 2.2.1, the AMS dataset was analysed utilizing both the pseudo-robust method and the PMF2 robust mode. Figure 2a shows the pseudo-robust Q_{cont} time series where Q_{cont} is the summation over all m/z of the squares of the scaled residuals, that is:

$$Q_{\text{cont}} = \sum_j^m (e_{ij}/s'_{ij})^2 \quad (10)$$

Figure 2b shows Q/Q_{expected} as a function of p . In this figure, Q_{pseudo} and Q_{robust} are calculated from Eq. (2) using the outlier-downweighted uncertainties, while the Q_{true} values utilize the original s_{ij} . Despite the difference in Q -values between the pseudo-robust and robust modes, the \mathbf{G} and \mathbf{F} matrices are nearly identical. The discussion below focuses on the pseudo-robust analysis at fPeak=0 and seed=1. Results at other fPeak and seed are summarized in the Supplement (see Figs. S1 and S2 <http://www.atmos-chem-phys.net/10/1969/2010/acp-10-1969-2010-supplement.pdf>). The non-zero fPeak values yielded qualitatively similar solutions to those at fPeak=0, and did not lead to significant improvements in correlations of the factor time series with external tracers. In the absence of conflicting evidence, we focus our discussion on the solution with the lowest Q -values

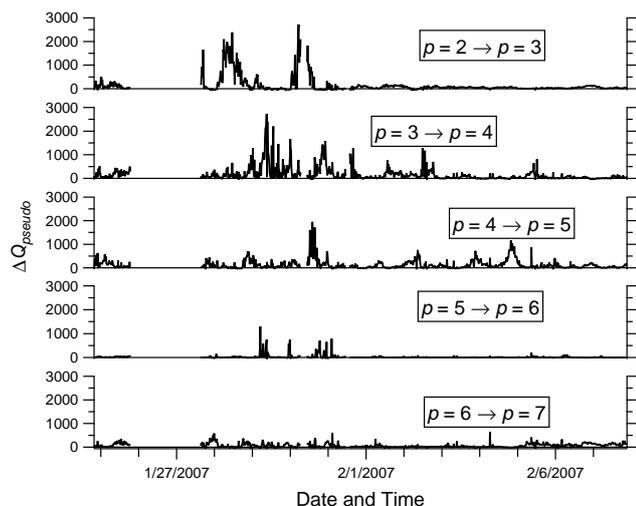


Fig. 3. Effect of the number of factors contained in a solution on the time-dependent contribution to Q_{pseudo} for the AMS dataset.

(fPeak=0). The set of solutions, encompassed by the fPeak range, provide one measure of the solution uncertainty.

Although the AMS data was averaged to 15 min time intervals for the PMF analysis, inspection of the original 1 min time series indicates that the spikes evident in the Q_{cont} time series correlate mostly with intense concentration spikes of <1 min duration. These spikes are due to emissions from nearby point sources, particularly a roadside hot dog stand and passing vehicles. Their presence is likely due to fluctuations in the emission profiles of these sources that cannot be fully represented by a single factor. During these periods, the scaled residuals are dominated by m/z corresponding to hydrocarbons.

Figure 3 shows the effect of the p on the Q_{cont} time series. Here, we plot ΔQ_{cont} between the p - and $(p+1)$ -factor solution, that is:

$$\Delta Q_{\text{cont}} = \left(\sum_j^m (e_{ij}/s_{ij})^2 \right)_p - \left(\sum_j^m (e_{ij}/s_{ij})^2 \right)_{p+1} \quad (11)$$

The structure in Fig. 3 indicates an improvement in the model fit (by transferring the signal from the residuals to the resolved factors) caused by increased p . The figure indicates that the solution is significantly improved as p increases to 5. Further increases yield significantly less improvement. Additionally, solutions at $p > 5$ include factors that cannot be validated through correlations with external tracers or reference spectra, unreliable low-mass factors and/or elements of factor mixing/splitting behaviour all of which suggest an excessive number of factors (Ulbrich et al., 2009a), particularly in the OOA-1 and charbroiling factors.

The factor mass spectra and time series of the AMS solution at $p=5$ are shown in Fig. 4a and b, respectively. Figure 4b also contains the time series for correlated tracer

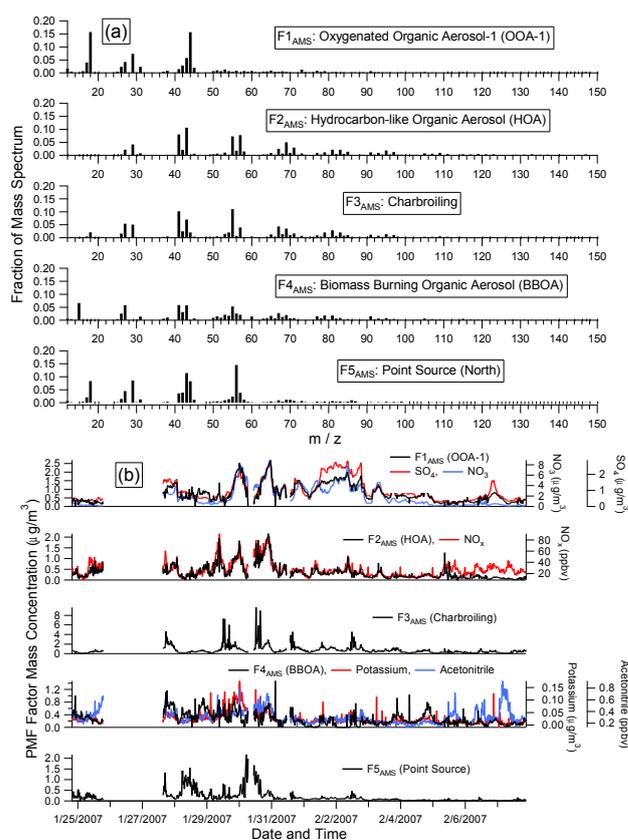


Fig. 4. Mass spectra (a) and time series (b) for the PMF solution to the AMS dataset. Figure 4b includes the time series both for the PMF factors (black traces, left axis) and selected tracer species (coloured traces, right axes).

species. The mass spectra are normalized so that the sum of each spectrum across all m/z 's is equal to 1. The time series are reported in terms of mass concentration ($\mu\text{g}/\text{m}^3$). All AMS reference spectra described below and in Sect. 3.3.2 were obtained from the AMS Spectral Database (Ulbrich et al., 2009b). Table 3 shows the fraction of the total mass apportioned to each factor, ratio of m/z 44/total organics (m/z 44 is the CO_2^+ ion, a marker for oxygenation) and the estimated O/C ratio (Aiken et al., 2008). Average mass fractions are calculated as the mean of the mass fraction time series for the designated factor ($g_{ih} f_{hj}/x_{ij}$), calculated as:

$$\overline{(g_{ih} f_{hj}/x_{ij})} = \left(\sum_i^n g_{ih} f_{ij}/x_{ij} \right) / n \quad \text{for a single } m/z \quad (12a)$$

$$\overline{(g_{ih} f_{hj}/x_{ij})} = \left(\sum_i^n \left(\sum_j^{j=\text{AMS}} g_{ih} f_{ij} / \sum_j^{j=\text{AMS}} g_{ih} f_{ij} x_{ij} \right) \right) / n \quad (12b)$$

for all AMS m/z

Here the i subscript is the matrix index and the calculation is performed for the h th factor and j th m/z (specified in the

Table 3. Factor properties for the AMS dataset: percent of total mass apportioned to each factor, ratio of m/z 44 (CO_2^+) to total organics, and estimated O/C ratio (Aiken et al., 2008). The reported mass fraction is the mean of the mass fraction time series, $(g_{ih}f_{hj}/x_{ij})$ (see Eq. 12), converted to a percentage. Values greater than 25% are bolded.

	OOA-1	HOA	Charbroiling	Biomass Burning	Point Source (North)
% Mass	0.36	0.15	0.32	0.14	0.07
m/z 44/org	0.16	< 0.01	0.02	< 0.01	0.08
Estimated O/C	0.69	< 0.08	0.16	< 0.08	0.39

table). For AMS data, the calculation is performed over all AMS m/z .

Factor $F1_{\text{AMS}}$ (oxygenated organic aerosol, OOA-1) is similar to OOA-1 factors obtained from AMS data in previous field studies (Zhang et al., 2007) using either the two-component deconvolution technique (Zhang et al., 2005) or PMF analysis ($R^2=0.96$ vs. Zurich winter OOA (Lanz et al., 2008a) and $R^2=0.90$ vs. both Zurich summer OOA-1; Lanz et al., 2007; and Pittsburgh OOA; Zhang et al., 2005). As shown in Table 3, it is the most oxygenated factor and a major component of the total mass ($\sim 33\%$). The OOA-1 factor correlates with particulate sulfate (Fig. 4b, $R^2=0.71$) and with back trajectories passing over industrial regions to the west/southwest of Toronto (trajectories were calculated using the NOAA HYSPLIT model; Draxler and Hess, 1998; Draxler and Rolph, 2003; Rolph, 2003).

Factor $F2_{\text{AMS}}$ (hydrocarbon-like organic aerosol, HOA) is similar to HOA factors previously observed in the field (Zhang et al., 2007) ($R^2=0.97$ vs. Zurich winter HOA (Lanz et al., 2008a), $R^2=0.95$ vs. Pittsburgh HOA; Zhang et al., 2005, and $R^2=0.91$ vs. Zurich summer HOA; Lanz et al., 2007). The organic functionality can be investigated by analyzing the delta (Δ) patterns (McLafferty, 1980), where $\Delta = m/z - 14n + 1$ (n is an integer). HOA is dominated by the $\Delta=0$ (i.e., m/z 27, 41, 55, 69, ...) and $\Delta=2$ (m/z 29, 43, 57, 71, ...) series, characteristic of alkenes and alkanes, respectively, although some contribution from oxygenated species such as $\text{C}_2\text{H}_3\text{O}^+$ is likely. The HOA time series correlates with NO_x (Fig. 4b) and is elevated overnight and during the morning rush hour (see Fig. S3 <http://www.atmos-chem-phys.net/10/1969/2010/acp-10-1969-2010-supplement.pdf>).

Factor $F3_{\text{AMS}}$ (charbroiling) is a major contributor to the total organic mass ($\sim 33\%$) with strong $\Delta=0$ and $\Delta=2$ series. As stated above, these series are characteristic of alkanes and alkenes. However, for charbroiling emissions, contributions from fatty acids and carbonyls are also likely (e.g. Schauer et al., 1999). The charbroiling factor mass spectrum is correlated with reference spectra for charbroiling emissions ($R^2=0.90$) and HOA ($R^2=0.90$ vs. Zurich winter HOA (Lanz et al., 2008a) and Pittsburgh HOA; Zhang et al., 2005). The difference between the HOA and char-

broiling factors is the relative strength of $\Delta=0$ and $\Delta=2$ series (26% each for HOA, 37% and 16%, respectively, for charbroiling). The diurnal profiles of the two factors are distinct, with charbroiling exhibiting strong signals around noon (see Fig. S3 <http://www.atmos-chem-phys.net/10/1969/2010/acp-10-1969-2010-supplement.pdf>). During 15 min intervals where the total organic mass is dominated by charbroiling, analysis of the original 1 min data shows the organic signal concentrated in intense spikes of <1 min duration, which occur exclusively during the operation of a roadside hot dog stand ~ 25 m from the sampling inlet. Day-to-day variation in the charbroiling signal is determined by the number of detected particles as measured by a fast mobility particle sizer (FMPS) (FMPS 3091, TSI, inc., Shoreview, MN, USA) and a condensation particle counter (CPC 3010, TSI, inc., Shoreview, MN, USA), rather than particle size, suggesting that the variation is driven by street-level mixing dynamics.

Factors $F4_{\text{AMS}}$ (biomass burning) and $F5_{\text{AMS}}$ (point source-north) are more difficult to validate due to their lower concentrations (see Table 3) and the absence of satisfactory tracer species. Identification of the biomass burning factor is tentative and this factor disappears in the unified dataset. Some features of the biomass burning time series are correlated with the AMS estimate of potassium (see Fig. 4b). However, AMS potassium measurements are not quantitative because of multiple ionization processes, high instrument background signal and interference from the C_3H_3^+ ion. The potassium event on 30 January correlates with high chloride concentrations due to road salt and may be partially influenced by this source. The biomass burning mass spectrum correlates only moderately well with previously extracted wood-burning factors (Lanz et al., 2007, 2008a) ($R^2 \sim 0.5$). However, burning signatures vary significantly with fuel type and burn conditions (Weimer et al., 2008). This is the only factor with a significant contribution from m/z 60, which is frequently used as a tracer for levoglucosan and an indicator of biomass burning (1.8% of the factor spectrum vs. 0.7% for OOA-1, for which m/z 60 has the next largest contribution). For the Lanz et al. wood-burning factors, m/z 60 comprises between 1.4% (winter) and 3.2% (summer) of the spectrum (Lanz et al., 2007, 2008a).

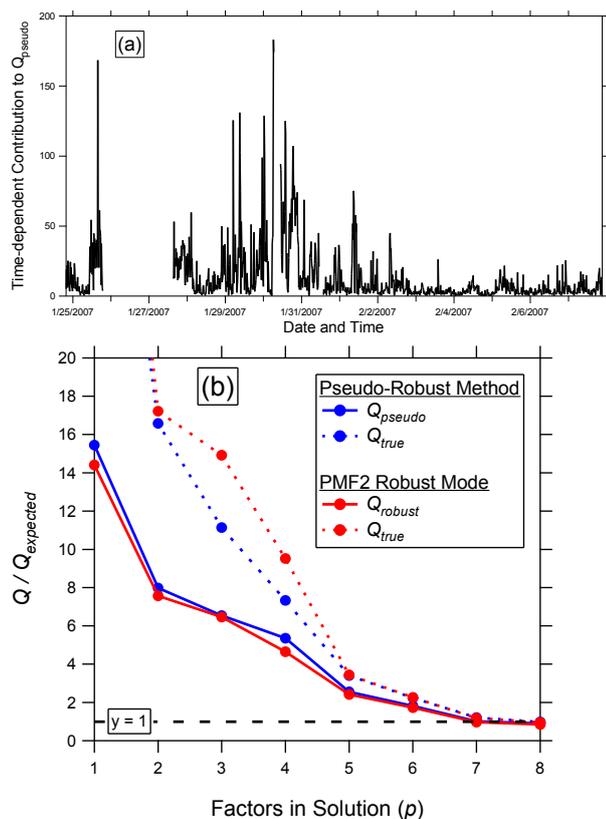


Fig. 5. (a) Time-dependent contribution to Q_{pseudo} for the PTR-MS dataset at $p=5$. (b) Ratio Q/Q_{expected} as a function of p .

A unique feature of the F5_{AMS} (point source-north) mass spectrum is the prominent signal at m/z 56 (16% of total). The presence of m/z 44 indicates oxygenation, suggesting that m/z 56 may be influenced by $\text{C}_3\text{H}_4\text{O}^+$ fragment, obtained from alkylcycloalkanones. However, contributions from C_4H_8^+ (cycloalkanes and branched alkenes), or $\text{C}_3\text{H}_6\text{N}^+$ (cyclic amines) cannot be ruled out. The point source-north time series does not correlate with any available tracer species or with the total organic mass (dominated by the charbroiling and OOA-1 factors), but is observed only during north/northeast winds suggesting a specific point source of primary emissions.

3.2 PTR-MS dataset

For the PTR-MS dataset, a solution at $p=5$ was selected using similar criteria to the AMS dataset. Summary statistics for this solution are presented in Table 2. The $p=5$ solution contains the following factors: (1) traffic, (2) long range transport (LRT)+local source, (3) LRT+painting, (4) local oxidation, and (5) oxygenates. Solutions were analysed using both the pseudo-robust method and robust mode, yielding similar Q -values and near-identical **F** and **G** matrices. The discussion below pertains to solutions obtained

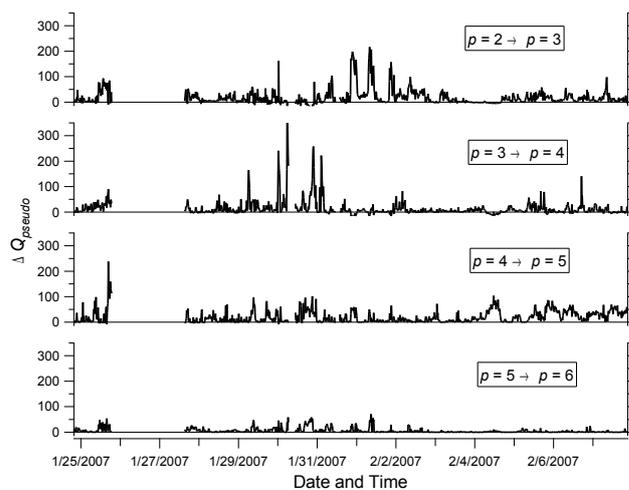


Fig. 6. Effect of the number of factors contained in a solution (p) on the time-dependent contribution to Q_{pseudo} for the PTR-MS dataset.

at $f_{\text{Peak}}=0$ and $\text{seed}=1$; solutions at other values are shown in Figs. S4 and S5 (<http://www.atmos-chem-phys.net/10/1969/2010/acp-10-1969-2010-supplement.pdf>). Similar to the AMS dataset, solutions at non-zero f_{Peak} did not significantly improve correlations with external tracer species and the solution with the lowest Q -value ($f_{\text{Peak}}=0$) is discussed below.

Figure 5a and b show the pseudo-robust Q_{cont} time series and Q/Q_{expected} as a function of p . As was the case for the AMS dataset, the time series contains significant temporal structure, denoting periods where the model description is imperfect. In Fig. 6, the ΔQ_{cont} time series (see Eq. 11) as a function of p is plotted. The most improvement is obtained as p increases to 5. Note also that the (smaller) improvements obtained at $p=6$ occur in periods that are described by preexisting factors (e.g. the structure on 1 February at $p=2 \rightarrow 3$ vs. $p=5 \rightarrow 6$), suggesting minor source variations and/or factor splitting. Similar results are obtained at $p > 6$.

The PTR-MS factor mass spectra and time series at $p=5$ are presented in Fig. 7a and 7b, respectively. Mass spectra are normalized so that the sum of each spectrum is equal to 1 and time series are reported in ppbv. Table 4 shows the fraction of signal apportioned to each factor on an m/z -by- m/z basis and the toluene/benzene ratio (see Eq. 12 for the mean mass fraction calculation). The toluene/benzene ratio can be used as a photochemical clock, because these two aromatics are typically emitted by similar sources, but toluene has a shorter lifetime (Roberts et al., 1984). In the present study, source emissions were estimated to have a toluene/benzene ratio of ~ 4.0 and the ratio decreases below 1 with increasing photochemical age. The source emission ratio is consistent with previous measurements of fresh traffic emissions (Kristensson et al., 2004; de Gouw et al., 2005).

Table 4. Factor properties for the PTR-MS dataset: toluene/benzene ratio and percent of signal at each m/z apportioned to the designated factor. The reported mass fractions are the mean of the mass fraction time series, $(g_{ih}f_{hj}/x_{ij})$ (see Eq. 12), converted to percentage. Values greater than 25% are bolded.

	Traffic	LRT+Local Source	LRT+Painting	Local Oxidation	Oxygenates
Tol/Benz	2.99	3.23	2.27	0.47	0.74
% m/z 31	8.5	7.6	9.4	22.3	52.2
% m/z 43	25.0	33.4	8.0	10.8	22.5
% m/z 45	7.8	6.0	9.0	60.8	16.4
% m/z 59	15.0	8.5	50.4	18.5	7.8
% m/z 61	5.6	52.4	6.0	20.2	16.3
% m/z 73	4.6	35.8	48.3	3.5	10.1
% m/z 79	37.0	9.3	11.4	25.2	16.5
% m/z 93	57.1	16.1	13.8	7.2	6.9
% m/z 107	63.6	13.8	9.6	4.7	8.8
% m/z 121	60.1	8.6	9.7	11.8	10.6

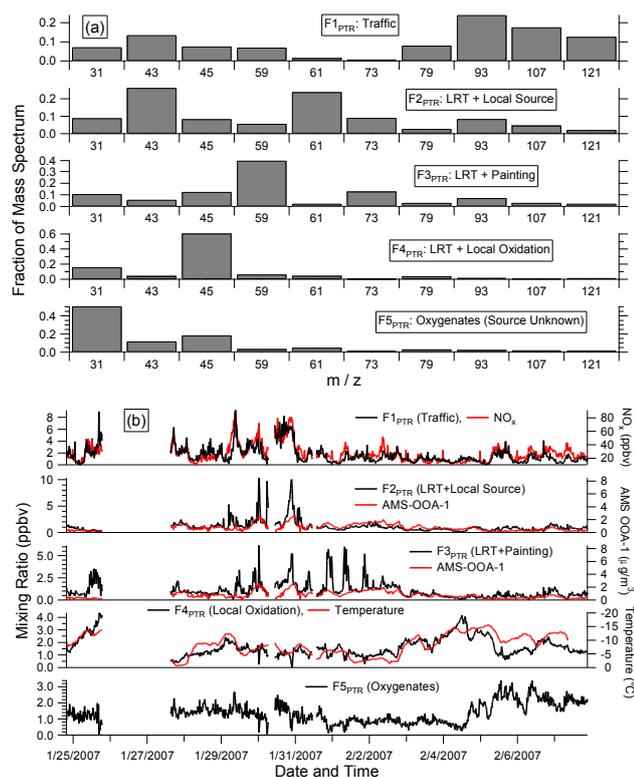


Fig. 7. Mass spectra (a) and time series (b) for the PMF solution to the PTR-MS dataset. Figure 7b includes the time series both for the PMF factors (black traces, left axis) and selected tracer species (coloured traces, right axes). Note that the temperature axis (vs. $F_{4\text{PTR}}$) is reversed.

Factor $F_{1\text{PTR}}$ (traffic) dominates the aromatic signal (~ 40 – 70% depending on m/z , see Table 4). The toluene/benzene ratio (2.99) indicates fresh emissions. The $F_{1\text{PTR}}$ factor peaks during the morning and evening rush hours (4–5 times nighttime values, see Fig. S6 <http://www.atmos-chem-phys.net/10/1969/2010/acp-10-1969-2010-supplement.pdf>) and is slightly elevated during the rest of the day. As shown in Fig. 7b, the factor correlates strongly with NO_x ($R^2=0.64$).

The $F_{2\text{PTR}}$ (LRT+local source) is dominated by signal at m/z 61 (acetic acid). The factor time series (Fig. 6b) correlates well with AMS-OOA-1 ($F_{1\text{AMS}}$) for most of the study, suggesting a contribution from transported, well-processed air. However, this correlation breaks down during the period of 29–31 January, where strong spikes in the factor time series are not reflected in the OOA-1 data. Such short-lived and intense features in the $F_{2\text{PTR}}$ time series likely indicate a local source. Further, the factor toluene/benzene ratio (3.23) is consistent with fresh emissions, although the aromatics are a minor component. Acetic acid is a product of ambient photochemical reactions, but has also been observed in emissions from spark-ignition engines (Zervas et al., 2001). The above data suggests that $F_{2\text{PTR}}$ is influenced by both LRT and local emissions and that these contributions cannot be decoupled through PMF using only the PTR-MS dataset. As discussed later, the effects of these sources can be largely decoupled in the unified AMS/PTR-MS dataset.

Factor $F_{3\text{PTR}}$ (LRT+painting) likewise results from inseparable sources. Similar to the LRT+local source factor, LRT+painting correlates with AMS-OOA-1, excepting spikes in $F_{3\text{PTR}}$ that correlate with local painting activity. Additionally, the toluene/benzene ratio (2.27) suggests some contributions from local emissions sources. The factor mass spectrum is dominated by acetone and constitutes more than half of the total acetone signal. Other major components

include m/z 45 (acetaldehyde) and m/z 73 (methyl ethyl ketone, methylglyoxal, butanal), though this is a small fraction of the total acetaldehyde (see Table 4). However, most of the non-residual m/z 73 is assigned to F3_{PTR}. Both acetone and methyl ethyl ketone have primary emission sources (including paint solvents), but are also generated as photochemical reaction products. Similar to F2_{PTR}, F3_{PTR} is affected by both local emissions and transported air and decoupling of the two effects is improved in the unified AMS/PTR-MS dataset.

F4_{PTR} (local oxidation) is dominated by acetaldehyde, a VOC oxidation product with a lifetime of less than a day. The signals at m/z 31, 59 and 61 are attributed to formaldehyde, acetone (with potential minor contributions from propanal and glyoxal) and acetic acid, respectively, which are all produced from VOC oxidation. In contrast to the factors described above, this factor has a low toluene/benzene ratio (0.47, see Table 4), indicating a greater extent of photochemical processing. However, the relatively short lifetimes of formaldehyde and acetaldehyde suggest local oxidation (as opposed to LRT). Although there are no correlated tracer species available, this is not surprising because all available tracers are expected to correlate with either direct emissions or transported, aged air. F4_{PTR} is anticorrelated with temperature ($R^2=0.50$), as shown in Fig. 7b.

Factor F5_{PTR} (oxygenates) consists of long- and short-lived oxygenated compounds, notably formaldehyde and acetaldehyde (though the factor contains only $\sim 15\%$ of the total acetaldehyde). Given the oxygenated nature of this factor, the signal at m/z 43 is probably from the CH_3CO^+ ion, which results from a variety of oxygenated compounds, including peroxyacetyl nitrate (PAN), acetone and acetic acid (de Gouw and Warneke, 2007). This factor does not correlate with any available tracers and exhibits no temperature dependence. The absence of distinct events in the time series and the low toluene/benzene ratio (see Table 4) suggest that the factor is not derived from a local point source.

3.3 Unified AMS/PTR-MS dataset

3.3.1 Selection and evaluation of solution

Due to the necessity of weighting the AMS and PTR-MS components of the unified dataset, the solution space must be explored in two dimensions: (1) p and (2) C_{PTR} . We recall that p denotes the number of factors in a solution, while C_{PTR} controls the relative weight of the AMS and the PTR-MS, with the PTR-MS weight increasing with C_{PTR} (see Eq. 8 and Sect. 2.2.2). Acceptable values of C_{PTR} are obtained when $\Delta\bar{e}_{sc} \sim 0$ (see Eq. 9 and Sect. 2.2.2). Values for p are evaluated using similar criteria to the individual datasets. The summary statistics for the selected solution ($p=6$, $\Delta\bar{e}_{sc}=0.052$, $C_{\text{PTR}}=10$) are shown in Table 2. The 6 factors obtained are: (1) charbroiling, (2) traffic, (3) aged secondary organic aerosol (SOA), (4) local SOA, (5) oxy-

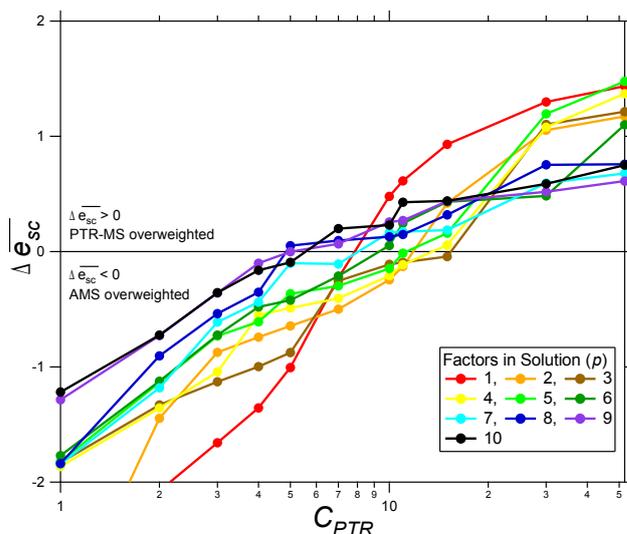


Fig. 8. Change in the mean-scaled residual between the AMS and PTR-MS ($\Delta\bar{e}_{sc}$) as a function C_{PTR} (x-axis) and p . Instruments carry their natural weight at $C_{\text{PTR}}=1$; the instruments are balanced at $\Delta\bar{e}_{sc}=0$.

genated primary organic aerosol (POA), and (6) local point source. The solutions discussed below were obtained at $f_{\text{Peak}}=0$ and $\text{seed}=1$; solutions at other parameter values are shown in Figs. S7 and S8 (<http://www.atmos-chem-phys.net/10/1969/2010/acp-10-1969-2010-supplement.pdf>). Similar to the AMS and PTR-MS datasets, solutions at non-zero f_{Peak} did not significantly improve correlations with external tracer species and the solution with the lowest Q -value ($f_{\text{Peak}}=0$) is discussed below.

An additional issue with respect to the unified dataset is that previous attempts to merge data from different instruments into a single matrix for PMF have sometimes failed due to multiplicative errors affecting entire rows of the data matrix of a single instrument (Paatero, 2009). This does not harm PMF analysis of a single instrument dataset, but may cause factor distortion in a unified dataset. The problem is indicated by data rows of a particular instrument, containing mostly negative or mostly positive residuals. As shown in the Supplement (Fig. S9 <http://www.atmos-chem-phys.net/10/1969/2010/acp-10-1969-2010-supplement.pdf>), the problem does not occur in the present study.

Figure 8 shows the value of $\Delta\bar{e}_{sc}$ as a function of C_{PTR} (x-axis) and p (coloured traces). From the figure, it is evident that $\Delta\bar{e}_{sc}$ depends on the combination of p and C_{PTR} , and the two cannot be considered separately. That is, for some values of C_{PTR} the solution is balanced at $p=1$ while at others it is balanced at $p=10$. However, for most values of C_{PTR} , only a few p yield balanced solutions. Such combinations of C_{PTR} and p can then be analysed similar to the individual datasets to determine whether the specified p appears correct. This analysis is discussed below for the set

Table 5. Factor properties for the unified dataset: factor-by-factor ratio of m/z 44 (CO_2^+) to total organics, estimated O/C ratio (Aiken et al., 2008), toluene/benzene ratio, and apportionment of total AMS mass and PTR-MS signal by m/z . The reported mass fractions are the mean of the mass fraction time series, $(g_{ih}f_{hj}/x_{ij})$ (see Eq. 12), converted to percentage. Values greater than 25% are bolded.

	Charbroil.	Traffic	Aged SOA	Local SOA	Ox. POA	Local Point Source
AMS % Mass	44.5	7.6	31.8	11.3	4.1	4.0
AMS m/z 44/org	0.02	< 0.01	0.14	0.10	0.095	< 0.01
AMS est. O/C	0.16	< 0.08	0.61	0.46	0.44	< 0.08
PTR-MS Tol/Benz	2.27	3.42	0.54	0.16	1.46	4.62
PTR-MS % m/z 31	6.4	2.1	18.7	25.8	44.7	2.4
PTR-MS % m/z 43	12.3	6.3	17.6	12.8	26.4	24.3
PTR-MS % m/z 45	< 0.1	13.5	< 0.1	64.0	18.0	4.4
PTR-MS % m/z 59	3.8	40.8	41.8	3.7	10.0	< 0.1
PTR-MS % m/z 61	0.1	< 0.1	23.9	22.5	21.0	33.0
PTR-MS % m/z 73	< 0.1	27.5	42.6	< 0.1	13.4	18.5
PTR-MS % m/z 79	19.3	15.8	9.4	23.1	23.1	10.9
PTR-MS % m/z 93	23.7	28.3	2.8	2.2	18.6	25.7
PTR-MS % m/z 107	27.7	27.6	< 0.1	< 0.1	20.1	26.5
PTR-MS % m/z 121	29.2	25.0	< 0.1	5.2	22.7	19.9

of solutions at $C_{\text{PTR}}=10$. The issue of what values of $\Delta\overline{e}_{sc}$ indicate a balanced solution is explored in Sect. 3.3.3.

Summary statistics for the solution at $C_{\text{PTR}}=10$, $p=6$ are shown in Table 2. Figure 9a shows the Q_{cont} time series in terms of the total Q and the Q derived from the individual instruments. Figure 9b shows Q/Q_{expected} as a function of p . For the Q calculations in Table 2 and Fig. 9b, C_{PTR} has been removed from the uncertainty matrix to facilitate comparison with the individual datasets. In Fig. 9a, C_{PTR} is left in place so that the $Q_{\text{cont,AMS}}$ and $Q_{\text{cont,PTR}}$ time series can be compared. The unified Q_{cont} time series (with C_{PTR} removed) are overlaid with those of the individual datasets in Fig. S10 (<http://www.atmos-chem-phys.net/10/1969/2010/acp-10-1969-2010-supplement.pdf>). The Q -value at $p=6$ is somewhat higher than those of the individual datasets. The time series indicate that this is mostly attributable to the AMS components. Similar to the individual dataset, the $Q_{\text{cont,AMS}}$ contains local source-derived concentration spikes. However, $Q_{\text{cont,AMS}}$ also resembles the AMS-biomass burning and AMS-point source-north factors (Fig. 4b), which, as shown below, are not resolved in the unified dataset. The $Q_{\text{cont,PTR}}$ time series is qualitatively similar to that obtained from the individual PTR-MS dataset (Fig. 5).

Figure 10 shows the $\Delta Q_{\text{cont,AMS}}$ and $\Delta Q_{\text{cont,PTR}}$ time series (see Eq. 11) as a function of p . The AMS solution significantly improves up to $p=4$, but not beyond. While the PTR-MS does not show a clear point beyond which $\Delta Q_{\text{cont,PTR}}$ decreases, beyond $p=6$ the structure in $\Delta Q_{\text{cont,PTR}}$ occurs mostly in periods that are described by pre-existing factors and splitting of the aged SOA factor occurs. However, the possibility of meaningful factors at higher p cannot be completely ruled out due to the higher-than-expected Q -values and absence of factors resolved in the individual datasets.

3.3.2 Physical interpretation of factors

Factor mass spectra and time series are shown in Fig. 11a and b, respectively. For presentation and intercomparison purposes, we do not directly report the f_{hj} and g_{ih} (here i and j are the matrix indices for time points and m/z , respectively; the calculation is performed for the h th factor). Instead, the mass spectra are re-normalized so that the sum of each spectrum for each instrument equals one, that is: $f_{hj} / \sum_j^{j=\text{AMS}} f_{hj}$ and $f_{hj} / \sum_j^{j=\text{PTR}} f_{hj}$ for the AMS and PTR-MS, respectively. The time series are scaled such that the mean concentration of each factor is one, that is: $g_{ih} / \overline{g_{ih}}$. We report $\overline{g_{ih,AMS}}$ and $\overline{g_{ih,PTR}}$ for each factor, calculated as:

$$\overline{g_{ih,AMS}} = g_{ih} \left(\frac{\sum_j^{j=\text{AMS}} f_{hj}}{\sum_j f_{hj}} \right) \quad (13a)$$

$$\overline{g_{ih,PTR}} = g_{ih} \left(\frac{\sum_j^{j=\text{PTR}} f_{hj}}{\sum_j f_{hj}} \right) \quad (13b)$$

In these figures, the AMS and PTR-MS time series (in $\mu\text{g}/\text{m}^3$ and ppbv) for the unified dataset are obtained as the product of either $\overline{g_{ih,AMS}}$ or $\overline{g_{ih,PTR}}$ with the displayed time series. Parameters tabulated for the individual datasets are reported for the unified datasets in Table 5 (see Eq. 12 for the mean-mass fraction calculation). Related factors in the individual and unified datasets are overlaid in the Supplement (Figs. S11 and S12 <http://www.atmos-chem-phys.net/10/1969/2010/acp-10-1969-2010-supplement.pdf>). Diurnal profiles are shown in Fig. S13. As previously stated, all

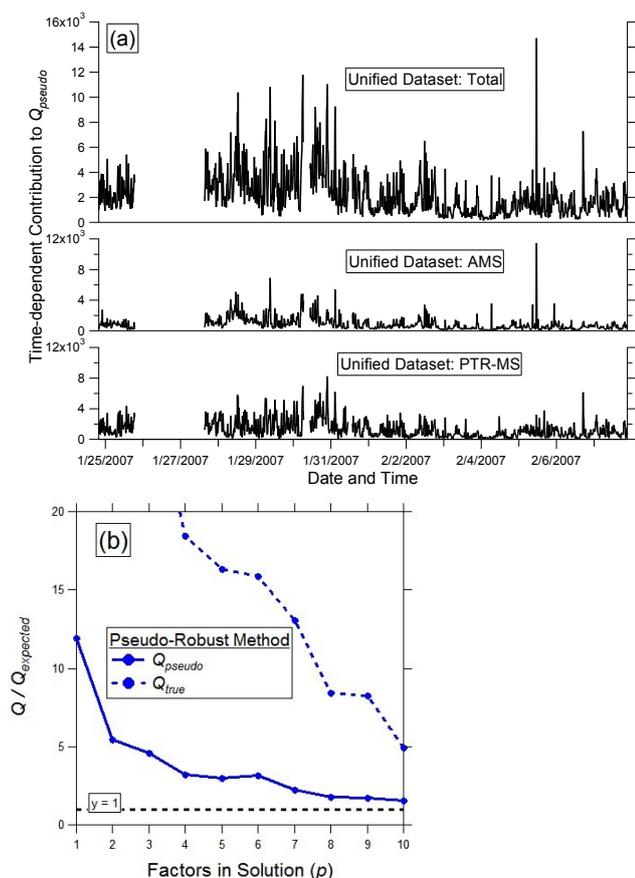


Fig. 9. (a) Time-dependent contribution to Q_{pseudo} for the unified dataset at $p=6$. Separate traces are shown for the total Q_{pseudo} and the contributions from the AMS and PTR-MS components. (b) Q/Q_{expected} as a function of p .

AMS reference spectra described below were obtained from the AMS Spectral Database (Ulbrich et al., 2009b).

Factor F1_{UN} (charbroiling) is very similar to that of the AMS-charbroiling factor. The time trend of UN-charbroiling also correlates with PTR-MS m/z 69 ($R^2=0.53$), excluded from the PMF analysis due to low signal-to-noise. Several compounds contribute to m/z 69, including furan, which is produced during meat cooking (Lee, 1999) and other combustion-related processes (Beychok, 1987; Andreae and Merlet, 2001). Aromatic VOCs are enhanced (see Table 5), consistent with combustion processes. Previous discussions of the AMS-charbroiling factor hold for UN-charbroiling, notably that the particle mass spectrum is characteristic of aliphatic hydrocarbons and that the time series is dominated by short-duration concentration spikes clustered in the early afternoon (see Fig. S13). The UN-charbroiling factor accounts for a significantly larger fraction of the particulate mass than AMS-charbroiling ($\sim 50\%$ vs. $\sim 33\%$).

Factor F2_{UN} (traffic) is correlated with NO_x (Fig. 11b), similar to AMS-HOA and PTR-traffic. The notable differ-

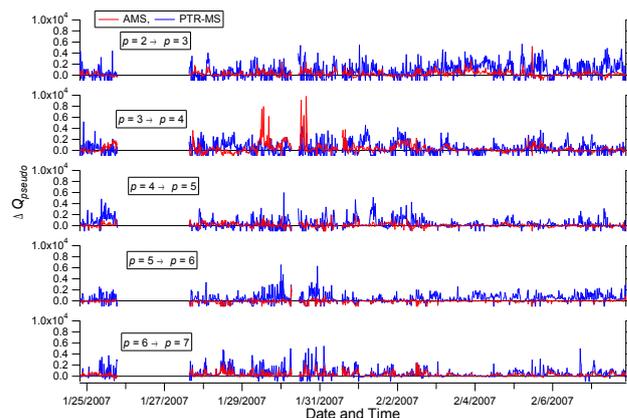


Fig. 10. Effect of the number of factors contained in a solution on Q_{pseudo} for the unified dataset. AMS (red) and PTR-MS (blue) contributions are plotted separately.

ences in the time series are lower particle mass (vs. AMS-HOA) and inclusion of some painting emissions (vs. PTR-traffic). The painting emissions cause an increase in the acetone and MEK contribution to the UN-traffic VOC mass spectrum. The UN-traffic and AMS-HOA particle mass spectra are similar (hydrocarbon-dominated). The toluene/benzene ratio (3.42) and contributions from aromatic VOCs indicate fresh emissions (see Table 5).

Factor F3_{UN} (aged SOA) is similar to AMS-OOA-1 ($R^2=0.995$, see Fig. S11 <http://www.atmos-chem-phys.net/10/1969/2010/acp-10-1969-2010-supplement.pdf>) and previously reported OOA-1 spectra (Lanz et al., 2007; Ulbrich et al., 2009a). Table 5 shows this to be the most oxygenated factor. Figure 11b shows correlation with AMS nitrate and sulfate. Back trajectory analysis indicates that the factor correlates with airflow over the industrialized regions west/southwest of Toronto. The PTR-MS mass spectrum is unique to the unified dataset and dominated by signals attributable to oxygenated species, particularly acetone (m/z 59). These species are consistent with secondary oxidation, though they also have direct emission sources. Strong correlation between acetone and aged particulate SOA is consistent with previous observations (Vlasenko et al., 2009). The factor mass spectra and correlations indicate regional transport of secondary organic aerosol. The apportionment of oxygenated VOCs such as acetone to secondary vs. primary factors is an important feature of the unified dataset solution and is discussed further in Sect. 3.3.5.

Factor F4_{UN} (local SOA) is dominated by acetaldehyde. Both the factor time series and VOC mass spectrum are similar to PTR-local oxidation (see Fig. S12). The dominant species in the UN-local SOA factor (acetaldehyde and to a lesser extent formaldehyde) have lifetimes of less than a day, while those in the UN-aged SOA factor (acetone, to a lesser extent acetic acid and MEK) have lifetimes in the order of weeks (Atkinson et al., 2006). Both the toluene/benzene

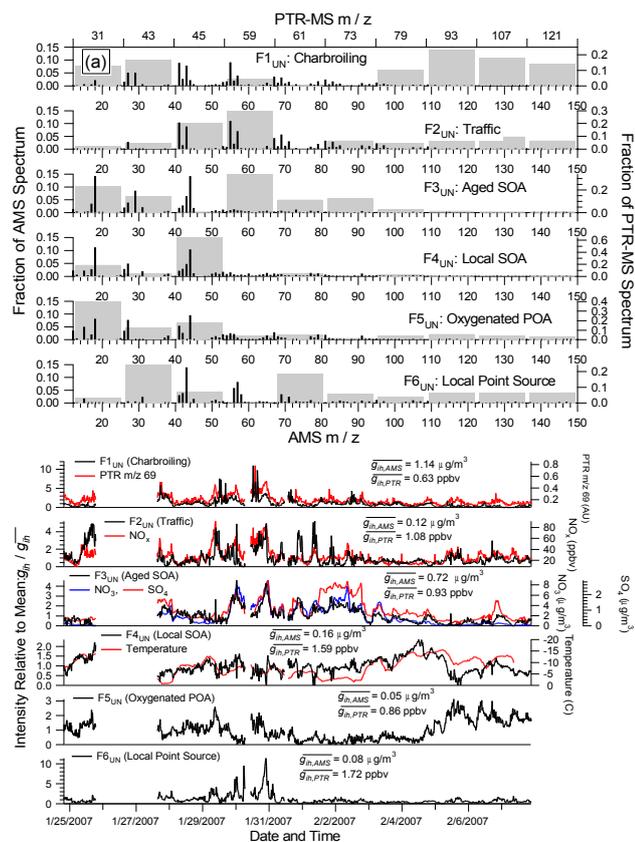


Fig. 11. Mass spectra (a) and time series (b) for the PMF solution to the unified dataset. Figure 11b includes the time series both for the PMF factors (black traces, left axis) and selected tracer species (coloured traces, right axis). Note the temperature axis (vs. F4_{UN}) is reversed. PTR-MS m/z 69 is plotted in arbitrary units.

ratio and the total aromatic VOC concentration are very low, suggesting secondary production. The AMS mass spectrum resembles AMS-OOA-1 and UN-aged SOA, but has proportionally less signal at m/z 44 (m/z 44/total organics=0.10, vs. 0.16 for AMS-OOA-1 and 0.14 for UN-aged SOA), indicating less oxygenation (Aiken et al., 2009). Similarly, the total spectral intensity at $m/z > 44$ relative to m/z 44 is higher for UN-local SOA (4.5 for UN-local SOA vs. 2.4 for UN-aged SOA). These trends are suggestive of the OOA-2 factors observed in Zurich (Lanz et al., 2007) and Pittsburgh (Ulbrich et al., 2009a), which were attributed to more volatile and/or fresher oxygenated organics, though the trends in the current study are less pronounced. However, the ratio of m/z 43 to m/z 44 is similar between the UN-aged SOA and UN-local SOA (~ 0.45), contrasting with previous measurements that show a higher 43/44 ratio for OOA-2 than OOA-1. The correlations of UN-local SOA with shorter-lived oxygenated VOCs and anticorrelation with temperature suggest that both the oxidation timescale and volatility contribute here to the factor time series. In the present study, the AMS component of this OOA-2-like factor can only be resolved through the unified dataset.

Factor F5_{UN} (oxygenated POA) includes a VOC spectrum with large contributions from both oxygenated and aromatic VOCs. The particle mass spectrum is nearly as oxygenated as that of the UN-local SOA factor, however the fraction of apportioned particle mass is below the $\sim 5\%$ threshold required for AMS factor resolution (Ulbrich et al., 2009a). The high aromatic VOC content and toluene/benzene ratio suggest primary emissions.

Factor F6_{UN} (local point source) occurs almost exclusively in a few discrete events and has a very high toluene/benzene ratio (4.62), suggesting a local primary emissions source. These events match those in the PTR-LRT+local source factor (see Fig. S12 <http://www.atmos-chem-phys.net/10/1969/2010/acp-10-1969-2010-supplement.pdf>) and are not connected with those of the AMS-point source-north factor. Compared to the PTR-LRT+local source factor, UN-local point source has a larger contribution from aromatic VOCs, a higher toluene/benzene ratio and a smaller contribution from oxygenates, suggesting that the unified dataset has an improved resolution between primary and secondary species. The particle spectrum is hydrocarbon-like, but falls below the $\sim 5\%$ reliability threshold.

3.3.3 Unified dataset solutions as a function of $\Delta \bar{e}_{sc}$

An important issue in the evaluation of the unified dataset is the extent to which the solution changes with $\Delta \bar{e}_{sc}$, i.e. how close to $\Delta \bar{e}_{sc}=0$ a solution must be to be balanced. This is analysed in the present dataset through the comparison of the solution discussed above ($\Delta \bar{e}_{sc}=0.052$, $p=6$) with $p=6$ solutions obtained at $\Delta \bar{e}_{sc} \sim \pm 0.25$, $\sim \pm 0.5$, and $\sim \pm 1$. These comparisons are shown in Figs. 12, S14, and S15, respectively.

The $\Delta \bar{e}_{sc} \sim \pm 0.25$ (Fig. 12) solutions are mostly similar to $\Delta \bar{e}_{sc}=0.052$, however, already some differences are apparent. At $\Delta \bar{e}_{sc}=0.246$, the painting emissions are transferred from UN-traffic to UN-aged SOA (see Fig. 12b and c). However, this does not occur at $\Delta \bar{e}_{sc}=-0.212$. (Note that the magnitude of the UN-local point source events (product of $\bar{g}_{ih,PTR}$ and the time series) is similar between these two solutions.) Additional differences are apparent as $\Delta \bar{e}_{sc}$ diverges farther from 0. For example, at $\Delta \bar{e}_{sc}=-0.481$, the AMS is sufficiently overweighted that the UN-oxygenated POA factor is replaced by a factor resembling AMS-point source-north (see Fig. S14). The solutions begin to approach those of the individual datasets beyond $\Delta \bar{e}_{sc} = \pm 1$ (see Fig. S15).

From this analysis, $\Delta \bar{e}_{sc} = \pm 0.25$ approximately corresponds to the point where significant deviations from the $\Delta \bar{e}_{sc}=0$ solution are observed. However, as this analysis has been conducted on only one dataset, it is not certain whether this is a general property or if it varies with the dataset analysed.

It should be further noted that the uncertainty weighting method discussed herein is not the only possible method for combining the datasets of different instruments. A possible alternative is the weighting of a dataset by including duplicate columns of data for that instrument. That is, in Fig. 1, the unified dataset constructed in this manner would contain many replicates of the PTR-MS columns. This approach has the advantage that the PMF2 robust mode may be used, but the disadvantage that tuning $\Delta\bar{e}_{sc}$ is more difficult without greatly increasing the size of the dataset. While a comprehensive investigation of this method is beyond the scope of the present study, a preliminary comparison of the $p=6$ solution is presented in Figs. S16 to S18 (<http://www.atmos-chem-phys.net/10/1969/2010/acp-10-1969-2010-supplement.pdf>). For this approach, which we term the PTR-redundancy method, $\Delta\bar{e}_{sc} \sim 0$ is obtained when the PTR-MS dataset is duplicated 50 times. The factor mass spectra and time series are qualitatively similar to the uncertainty weighting method, excepting that: (1) painting emissions move from UN-traffic to UN-local point source and (2) signal is transferred from UN-local point source to UN-aged SOA. A potential issue with the PTR-redundancy solution is shown in Fig. S17 (<http://www.atmos-chem-phys.net/10/1969/2010/acp-10-1969-2010-supplement.pdf>). The increase in PTR-MS rows containing mostly positive or mostly negative residuals may indicate factor distortion (Paatero, 2009). Such a distortion is hypothesized to result from multiplicative effects acting on entire rows of the individual instrument matrices, where these effects are not synchronized between instruments.

3.3.4 Q-values and the C3 parameter

As shown in Table 2, the some of the Q -values obtained for the AMS and unified datasets are higher than the theoretical values. In the case of the unified dataset, particularly for Q_{true} , the discrepancy is significant. One possible reason for the discrepancy is the underestimation of instrument and/or modelling errors. In this case, a parameter C3 may be incorporated in the error calculation such that $S_{C3} = S + C3 \times X$. The C3 parameter would then be empirically adjusted to yield $Q/Q_{expected} \sim 1$. However, the obtained Q -values may be higher than the expected Q for several reasons, including the presence of meaningful factors at higher p and the outlier treatment method (which strongly influences the relationship between Q_{pseudo} or Q_{robust} and Q_{true}). These issues are discussed in more detail below. The obtained $Q/Q_{expected}$ are also influenced by the choice of C_{PTR} . Because of these uncertainties, we do not have sufficient confidence in the expectation that $Q/Q_{expected} \sim 1$ to justify using C3 to tune the solution to a specific Q . Instead, s_{ij} are estimated from instrument parameters as discussed in Sect. 2.1, and the reasons for the significant deviations from $Q/Q_{expected} \sim 1$ are investigated. A brief exploration of the effects of $C3 > 0$ on Q , $\Delta\bar{e}_{sc}$ and the resolved factors follows.

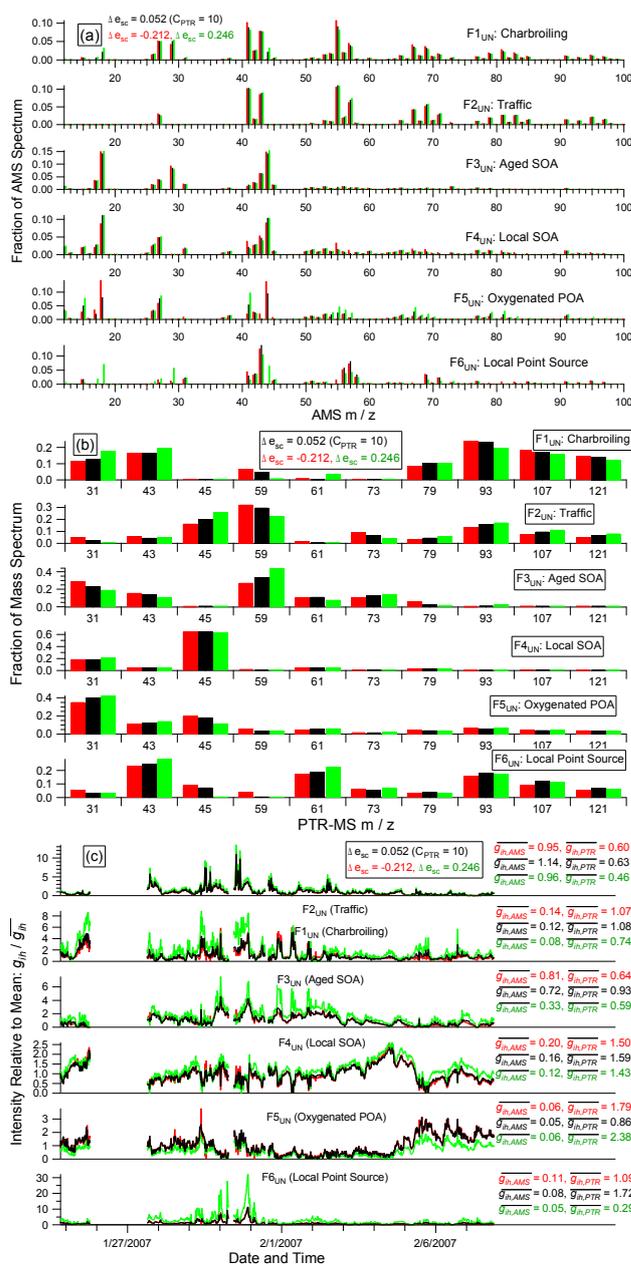


Fig. 12. Factor mass spectra (a, b) and time series (c) for the 6-factor solution to the unified dataset at $\Delta\bar{e}_{sc} = -0.212$ ($C_{PTR} = 7$, red), $\Delta\bar{e}_{sc} = 0.052$ ($C_{PTR} = 10$, black), and $\Delta\bar{e}_{sc} = 0.246$ ($C_{PTR} = 11$, green). Units for $g_{i,AMS}$ and $g_{i,PTR}$ are $\mu\text{g}/\text{m}^3$ and ppbv, respectively.

One possible reason for $Q/Q_{expected} > 1$ is that solutions with higher p contain meaningful information. Examples include factors representing additional distinct sources or multiple factors representing real variability in a source profile. The validation of such factors of either type is limited by (1) availability of appropriate external tracers, (2) a priori understanding of contributing sources, and (3) understanding of

the factor mass spectra. It is possible that with more detailed supporting information, additional factors could be identified, thereby decreasing Q/Q_{expected} .

A second issue is the difference between Q_{pseudo} and Q_{expected} evident in Table 2. The large differences between Q_{pseudo} and Q_{true} values for the AMS and unified datasets are related to outlier downweighting treatment, discussed in Sect. 2.2.1. Downweighting permits a looser fit to the selected data points. Because of the decreased outlier fit quality, Q_{true} is significantly higher than Q_{pseudo} (and higher than if no outlier downweighting is performed). This effect is pronounced in the AMS data, which is significantly influenced by outliers related to the charbroiling point source. In contrast, the PTR-MS dataset is not significantly influenced by outliers, therefore, Q_{pseudo} and Q_{true} are comparable (see Table 2). These observations suggest that outlier treatment and ambient characteristics significantly influence Q/Q_{expected} .

Because of the ambiguity in the correct p and meaningful differences between Q_{pseudo} and Q_{true} , it is not possible to obtain a single “correct” C3. A brief comparison is presented here between the selected C3=0 and C3=0.04. C3=0.04 yields $Q_{\text{pseudo}}/Q_{\text{expected}}=1.03$, $Q_{\text{true}}/Q_{\text{expected}}=1.54$ and $\Delta\bar{e}_{sc}=0.126$. Note, however, that these decreased Q are caused by increased s_{ij} rather than decreased e_{ij} . The C3=0 and C3=0.04 factors are generally similar. Differences in the C3=0.04 solution are the treatment of the UN-charbroiling spikes and the inclusion of local painting events in the UN-aged SOA factor (similar to the solution at $\Delta\bar{e}_{sc}=0.246$ in Fig. 12). This may suggest that for this C3 value the PTR-MS data is slightly overweighted and that a smaller C_{PTR} is needed (which will in turn influence Q and an empirically-determined C3). As stated above, Q_{pseudo} and Q_{true} values also decrease at larger p . For example, $p=10$ yields $Q_{\text{pseudo}}/Q_{\text{expected}}=1.54$, $Q_{\text{true}}/Q_{\text{expected}}=4.94$ and $\Delta\bar{e}_{sc}=0.230$. These Q ratios continue to decrease with increasing p and, as discussed above, the possibility of additional meaningful factors cannot be ruled out.

3.3.5 Assessment of unified vs. individual PMF solutions

For the first time, PMF has been applied to a unified AMS/PTR-MS dataset. The analysis provides the information on particle and VOC sources and atmospheric processing that cannot be obtained from the datasets of the individual instruments. This is discussed below in terms of understanding gas/particle factor coupling/decoupling scenarios, improved apportionment of VOCs to primary emissions vs. secondary reaction products and enhanced interpretation of particulate SOA.

It is informative to consider the circumstances under which joint particle/gas factors would be expected vs. the conditions that would cause them to separate. Note that the inclusion of both gas and particle elements in the unified

dataset does not require the assumption of mixed gas/particle factors: in the present analysis, the UN-oxygenated POA and UN-local point source factors contained negligible particulate mass. The above analysis can, therefore, be used as a tool for interrogating assumptions relating to gas and/or particle systems. The mixed gas/particle factors observed can be grouped into three scenarios:

1. Co-emission of VOCs and particles, coupled with a short source-to-receptor time relative to the atmospheric lifetimes of the emitted species. Examples: UN-charbroiling and UN-traffic.
2. Secondary generation of VOCs and particles through related atmospheric reactions, comparable lifetimes of the produced species. Example: UN-aged SOA.
3. Secondary generation of VOCs and particles through related chemical reactions, different atmospheric lifetimes, reaction-to-receptor time less than shortest atmospheric lifetime. Example: UN-local SOA.

The two VOC-dominated factors both appear to result from local sources, although the lack of sharp features in the oxygenated POA factor may indicate a diffuse source. This suggests that the lack of particle signal is driven by the emission source profile. Other possible scenarios in which VOC or particle-only factors might be expected include factors driven by gas/particle partitioning or by VOCs with very short lifetimes (e.g. biogenic molecules such as monoterpenes). However, no factor of either type was observed in the present study.

The unified dataset enables apportionment of the oxygenated VOCs as primary emissions or secondary reaction products, which was not possible with the individual PTR-MS dataset. Metrics that may be used to identify factors resulting from primary emissions are: (1) aromatic VOCs, which are exclusively primary emissions, (2) hydrocarbon-like particulate organics, (3) correlation between the factor time series and a primary tracer, and (4) diurnal patterns, where the primary source can be identified and has a distinct diurnal emission pattern. In the unified solution, the factors UN-charbroiling and UN-traffic satisfy all four metrics. UN-oxygenated POA and UN-local point source satisfy (1) and (2); (3) cannot be evaluated due to low signal in the AMS and lack of any correlated tracer; and (4) cannot be assessed due to the uncertain source identity. Further, the oxygenated character of the AMS and PTR-MS factor profiles for UN-aged SOA and UN-local SOA suggest these factors are secondary reaction products, as does the time series correlation of UN-aged SOA with nitrate and sulfate.

Such classification allows apportionment of each VOC to primary emissions (UN-charbroiling, UN-traffic, UN-oxygenated POA, UN-local point source) vs. secondary reactions (UN-aged SOA, UN-local SOA). For example, Table 5 shows that 60% of formaldehyde (m/z 31) is directly

emitted, compared to only 27% of acetaldehyde (m/z 45). The primary formaldehyde contribution is consistent with values reported for other urban locations. The best agreement is found with a winter study in Rome (Possanzini et al., 2002) (65–70%) and an April 2003 study in Mexico City (Garcia et al., 2006) (42 to 63%). Photochemical production makes a larger contribution during summer, as evidenced by lower primary contributions during campaigns in Rome (10–20%) (Possanzini et al., 2002), Houston (22%) (Friedfeld et al., 2002) and rural sites at rural sites outside of Toronto (17%) (Vlasenko et al., 2009) and Vancouver (33%) (Li et al., 1997).

Approximately 55% of both acetone and acetic acid are attributed to primary emissions, which is comparable to the 52% reported for acetone at a site outside Vancouver during summer (Li et al., 1997). All of the m/z 's corresponding to non-benzene aromatics have >95% attributed to primary sources. Benzene is more complex, with only ~70% attributed to direct emission factors. This may be because of interferences at m/z 79, or because the benzene lifetime is sufficiently long to be coupled into the time scale of the secondary factors. This consideration also holds for the longer-lived oxygenated species (i.e. acetone and acetic acid), meaning that the values presented above should be considered as lower limits for the direct emission contribution.

The primary/secondary VOC analysis presented above is only possible through the unified dataset. In the individual PTR-MS dataset, several factors were identified as having contributions from both primary and secondary sources, i.e. PTR-LRT+painting, PTR-LRT+local source. Inclusion of the AMS data, where the classification of factors is more closely related to primary vs. secondary sources, directs the PMF deconvolution of the unified dataset along these lines. Similarly, the AMS primary/secondary classification is enhanced by the inclusion of PTR-MS data. This is shown in Fig. 13, where the aerosol mass fraction attributed to OOA factors is plotted as a function of the toluene-to-benzene ratio for the AMS and unified datasets. Correlation between these two quantities is expected because (1) the toluene/benzene ratio is inversely related to photochemical age (Roberts et al., 1984), (2) the oxygen content of organic aerosol is known to increase with photochemical age, (3) at an urban site in winter, SOA precursors are dominated by anthropogenic emissions, notably VOCs. The figure shows a significantly tighter correlation between the SOA mass fraction and the toluene/benzene ratio for the unified dataset. Note that the enhanced correlation is not a function of a change in the UN-aged SOA factor ($R^2=0.05$, Fig. 13b) relative to AMS-OOA-1 ($R^2=0.03$, Fig. 13a) (which could conceivably result from the pulling of the AMS factor time series towards those of the PTR-MS), but rather the resolution of an additional SOA factor, UN-local SOA ($R^2=0.15$ Fig. 15c).

Through inclusion of the PTR-MS data in the unified dataset, the AMS SOA (i.e. AMS-OOA-1) is resolved into “OOA-1-like” aged SOA and “OOA-2-like” local SOA fac-

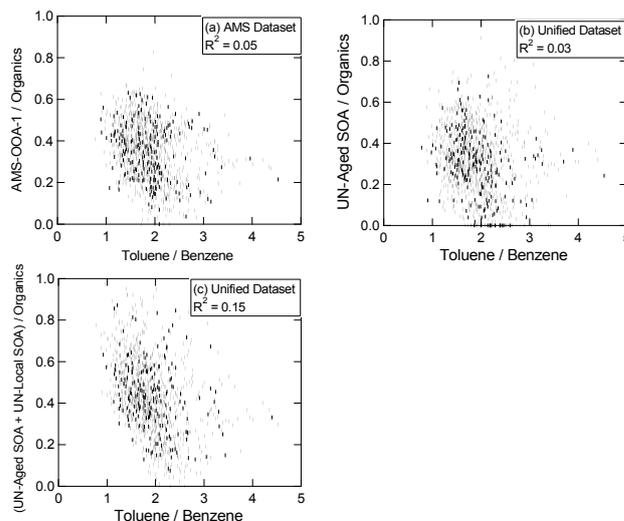


Fig. 13. Mass fraction of oxygenated aerosol as a function of the toluene/benzene ratio. The toluene/benzene ratio is inversely related to photochemical age.

tors. As noted in the previous section, both OOA-1 and OOA-2 factors have been extracted in several AMS datasets (Lanz et al., 2007; Ulbrich et al., 2009a). However, no OOA-2-like factor could be identified in the present dataset using only AMS data. The key factor, in resolving the OOA-2-like aged SOA factor in the unified dataset, is the distinction between short-lived photochemical reaction products (formaldehyde, acetaldehyde) vs. long-lived products (acetone, acetic acid). Similarly, an important feature of the OOA-1/OOA-2 distinction in other datasets has been to provide information about the relative age of the particulate organics: time periods with a higher ratio of OOA-1 to OOA-2 are generally more processed. Another example of a non-resolvable factor in PMF of AMS-only data was observed in Zurich (Lanz et al., 2008a). In the Zurich example, a wood-burning factor was extracted by forcing the mass spectrum of one factor towards the desired profile. The present study indicates that the inclusion of tracer species in the PMF analysis is useful in extracting hard-to-resolve factors in such cases.

Another important feature of the unified solution is the re-mixing or disappearance of several factors resolved in the individual datasets. For the PTR-MS factors, the prominent features in the time series are generally redistributed among factors in ways that enhance interpretation of the dataset, as described above in terms of primary vs. secondary species. However, different behaviour is observed with respect to the distribution of mass between $F1_{AMS}$ (charbroiling) vs. $F2_{AMS}$ (HOA) into $F1_{UN}$ and $F2_{UN}$, and the disappearance from the unified dataset solutions of $F4_{AMS}$ (biomass burning) and $F5_{AMS}$ (northeast point source). As these were minor factors in the AMS dataset, their disappearance is not surprising.

4 Conclusions

We present the first application of positive matrix factorization (PMF) to a unified AMS/PTR-MS dataset. The relative weights of the AMS and PTR-MS components in the PMF solution are balanced through the use of a residual based metric, $\Delta \bar{e}_{sc}$. This method can be directly applied to any dataset containing two mass spectrometers, and is readily generalized to account for three or more instruments. Analysis of the unified dataset complements that of the individual instrument datasets. In this study, the previously identified oxygenated aerosol factors OOA-1 and OOA-2 could be only distinguished within the unified dataset. Further, the unified dataset greatly enhanced interpretation of oxygenated VOC sources, apportioning them into primary sources vs. secondary reaction products. Minor factors in the individual dataset of one instrument lacking corresponding tracers in the other may not be resolvable in the unified dataset.

Acknowledgements. The authors gratefully acknowledge the NOAA Air Resources Laboratory (ARL) for the provision of the HYSPLIT transport and dispersion model. Partial operational funding was provided by the Canadian Foundation for Climate and Atmospheric Sciences (CFCAS) and the Natural Sciences and Engineering Council of Canada (NSERC). This project has also received funding support from the Ontario Ministry of the Environment. Such support does not indicate endorsement by the Ministry of the contents of this material. Funding for SOCAAR was provided by the Canadian Foundation for Innovation, the Ontario Innovation Trust, and the Ontario Research Fund.

Edited by: G. McFiggans

References

- Aiken, A. C., DeCarlo, P. F., Kroll, J. H., Worsnop, D. R., Huffman, J. A., Docherty, K. S., Ulbrich, I. M., Mohr, C., Kimmel, J. R., Sueper, D., Sun, Y., Zhang, Q., Trimborn, A., Northway, M., Ziemann, P. J., Canagaratna, M. R., Onasch, T. B., Alfarra, M. R., Prevot, A. S. H., Dommen, J., Duplissy, J., Metzger, A., Baltensperger, U., and Jimenez, J. L.: O/C and OM/OC Ratios of Primary, Secondary, and Ambient Organic Aerosols with High-Resolution Time-of-Flight Aerosol Mass Spectrometry, *Environ. Sci. Technol.*, 42, 4478–4485, doi:10.1021/es703009q, 2008.
- Aiken, A. C., Salcedo, D., Cubison, M. J., Huffman, J. A., DeCarlo, P. F., Ulbrich, I. M., Docherty, K. S., Sueper, D., Kimmel, J. R., Worsnop, D. R., Trimborn, A., Northway, M., Stone, E. A., Schauer, J. J., Volkamer, R. M., Fortner, E., de Foy, B., Wang, J., Laskin, A., Shutthanandan, V., Zheng, J., Zhang, R., Gaffney, J., Marley, N. A., Paredes-Miranda, G., Arnott, W. P., Molina, L. T., Sosa, G., and Jimenez, J. L.: Mexico City aerosol analysis during MILAGRO using high resolution aerosol mass spectrometry at the urban supersite (T0) –Part 1: Fine particle composition and organic source apportionment, *Atmos. Chem. Phys.*, 9, 6633–6653, 2009, <http://www.atmos-chem-phys.net/9/6633/2009/>.
- 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.: Technical note: A generalized 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., Jimenez, J. L., Williams, P. I., Alfarra, M. R., Bower, K. N., Jayne, J. T., Coe, H., and Worsnop, D. R.: Quantitative sampling using an Aerodyne aerosol mass spectrometer 1: Techniques of data interpretation and error analysis, *J. Geophys. Res.*, 108, 4090: doi:10.1029/2002JD002358, 2003.
- Andreae, M. O. and Merlet, P.: Emission of trace gases and aerosols from biomass burning, *Global Biogeochem. Cy.*, 15, 955–966, 2001.
- Atkinson, R., Baulch, D. L., Cox, R. A., Crowley, J. N., Hampson, R. F., Hynes, R. G., Jenkin, M. E., Rossi, M. J., Troe, J., and IUPAC Subcommittee: Evaluated kinetic and photochemical data for atmospheric chemistry: Volume II – gas phase reactions of organic species, *Atmos. Chem. Phys.*, 6, 3625–4055, 2006, <http://www.atmos-chem-phys.net/6/3625/2006/>.
- Beychok, M. R.: A Data Base for Dioxin and Furan Emissions from Municipal Refuse Incinerators, *Atmos. Environ.*, 21, 29–36, 1987.
- Buset, K. C., Evans, G. J., Leitch, W. R., Brook, J. R., and Toom-Saunty, D.: Use of advanced receptor modelling for analysis of an intensive 5-week aerosol sampling campaign, *Atmos. Environ.*, 40, 482–499, 2006.
- Buzcu, B. and Frazier, M. P.: Source Identification and Apportionment of Volatile Organic Compounds in Houston, TX, *Atmos. Environ.*, 40, 2385–2400, 2006.
- Canagaratna, M. R., Jayne, J. T., Jimenez, J. L., Allan, J. D., Alfarra, M. R., Zhang, Q., Onasch, T. B., Drewnick, F., Coe, H., Middlebrook, A. M., Delia, A., Williams, L. R., Trimborn, A. M., Northway, M. J., DeCarlo, P. F., Kolb, C. E., Davidovits, P., and Worsnop, D. R.: Chemical and microphysical characterization of ambient aerosols with the Aerodyne aerosol mass spectrometer, *Mass Spec. Rev.*, 26, 185–222, 2007.
- de Gouw, J. A., Goldan, P. D., Warneke, C., Kuster, W. C., Roberts, J. M., Marchewka, M., Bertman, S. B., Pszenny, A. A. P., and Keene, W. C.: Validation of proton transfer reaction-mass spectrometry (PTR-MS) measurements of gas-phase organic compounds in the atmosphere during the New England Air Quality Study (NEAQS) in 2002, *J. Geophys. Res.*, 108, 4682, doi:10.1029/2003JD003863, 2003.
- de Gouw, J. A., Middlebrook, A. M., Warneke, C., Goldan, P. D., Kuster, W. C., Roberts, J. M., Fehsenfeld, F. C., Worsnop, D. R., Canagaratna, M. R., Pszenny, A. A. P., Keene, W. C., Marchewka, M., Bertman, S. B., and Bates, T. S.: Budget of organic carbon in a polluted atmosphere: Results from the New England Air Quality Study in 2002, *J. Geophys. Res.*, 110, D16305, doi:10.1029/2004JD005623, 2005.
- de Gouw, J. A. and Warneke, C.: Measurements of volatile organic compounds in the Earth's atmosphere using proton-transfer-reaction mass spectrometry, *Mass Spec. Rev.*, 26, 223–257, 2007.
- Dockery, D. W. and Pope, C. A.: Acute respiratory effects of particulate air pollution, *Annual Review of Public Health*, 15, 107–132, 1994.
- Donahue, N. M., Robinson, A. L., Stanier, C. O., and Pandis, S. N.: Coupled Partitioning, Dilution, and Chemical Aging of Semivolatile Organics, *Environ. Sci. Technol.*, 40, 2635–2643,

- 2006.
- Draxler, R. R. and Hess, G. D.: An overview of the Hysplit_4 modelling system for trajectories, dispersion, and deposition, *Aust. Met. Mag.*, 47, 295–308, 1998.
- Draxler, R. R. and Rolph, G. D.: HYSPLIT (HYbrid Single-Particle Lagrangian Integrated Trajectory) Model access via NOAA ARL READY Website (<http://www.arl.noaa.gov/ready/hysplit4.html>), NOAA Air Resources Laboratory, Silver Spring, MD, 2003.
- Drewnick, F., Hings, S. S., DeCarlo, P. F., 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.
- Driscoll, C. T., Driscoll, K. M., Roy, K. M., and Mitchell, M. J.: Chemical responses of lakes in the Adirondack region of New York to declines in acidic deposition, *Environ. Sci. Technol.*, 37, 2036–2042, 2003.
- Friedfeld, S., Fraser, M., Ensor, K., Tribble, S., Rehle, D., Leleux, D., and Tittle, F.: Statistical analysis of primary and secondary atmospheric formaldehyde, *Atmos. Environ.*, 36, 4767–4775, 2002.
- Garcia, A. R., Volkamer, R., Molina, L. T., Molina, M. J., Samuelson, J., Mellqvist, J., Galle, B., Herndon, S. C., and Kolb, C. E.: Separation of emitted and photochemical formaldehyde in Mexico City using a statistical analysis and a new pair of gas-phase tracers, *Atmos. Chem. Phys.*, 6, 4545–4557, 2006, <http://www.atmos-chem-phys.net/6/4545/2006/>.
- Godri, K. J., Evans, G. J., Slowik, J., Knox, A., Abbatt, J., Brook, J., Dann, T., and Dabek-Zlotorzynska, E.: Evaluation and application of a semi-continuous chemical characterization system for water soluble inorganic PM_{2.5} and associated precursor gases, *Atmos. Meas. Tech.*, 2, 65–80, 2009, <http://www.atmos-meas-tech.net/2/65/2009/>.
- Hansel, A., Jordan, A., Holzinger, R., Prazeller, P., Vogel, W., and Lindinger, W.: Proton transfer reaction mass spectrometry: on-line trace gas analysis at ppb level, *Int. J. Mass Spectrom. Ion Processes*, 149/150, 609–619, 1995.
- Holzinger, R., Millet, D. B., Williams, B., Lee, A., Kreisberg, N., Hering, S. V., Jimenez, J. L., Allan, J. D., Worsnop, D. R., and Goldstein, A. H.: Emission, oxidation, and secondary organic aerosol formation of volatile organic compounds as observed at Chebogue Point, Nova Scotia, *J. Geophys. Res.*, 112, D10S24, doi:10.1029/2006JD007599, 2007.
- Jacobson, M. Z.: Strong radiative heating due to the mixing state of black carbon in atmospheric aerosols, *Nature*, 409, 695–697, 2001.
- Kanakidou, M., Seinfeld, J. H., Pandis, S. N., Barnes, I., Dentener, F. J., Facchini, M. C., Van Dingenen, R., Ervens, B., Nenes, A., Nielsen, C. J., Swietlicki, E., Putaud, J. P., Balkanski, Y., Fuzzi, S., Horth, J., Moortgat, G. K., Winterhalter, R., Myhre, C. E. L., Tsigaridis, K., Vignati, E., Stephanou, E. G., and Wilson, J.: Organic aerosol and global climate modelling: a review, *Atmos. Chem. Phys.*, 5, 1053–1123, 2005, <http://www.atmos-chem-phys.net/5/1053/2005/>.
- Kristensson, K., Johansson, C., Westerholm, R., Swietlicki, E., Gidhagen, L., Wideqvist, U., and Vesely, V.: Real-world traffic emission factors of gases and particles measured in a road tunnel in Stockholm, Sweden, *Atmos. Environ.*, 38, 657–673, 2004.
- Lanz, V. A., Alfarrá, M. R., Baltensperger, U., Buchmann, B., Hueglin, C., and Prvt, A. S. H.: Source apportionment of submicron organic aerosols at an urban site by factor analytical modelling of aerosol mass spectra, *Atmos. Chem. Phys.*, 7, 1503–1522, 2007, <http://www.atmos-chem-phys.net/7/1503/2007/>.
- Lanz, V. A., Alfarrá, M. R., Baltensperger, U., Buchmann, B., Hueglin, C., Szidat, S., Wehrl, 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, 2008a.
- Lanz, V. A., Hueglin, C., Buchmann, B., Hill, M., Locher, R., Staehelin, J., and Reimann, S.: Receptor modelling of C₂–C₇ hydrocarbon sources at an urban background site in Zurich, Switzerland: changes between 1993–1994 and 2005–2006, *Atmos. Chem. Phys.*, 8, 2313–2332, 2008b, <http://www.atmos-chem-phys.net/8/2313/2008/>.
- Lee, P. K., Brook, J. R., Dabek-Zlotorzynska, E., and Mabury, S. A.: Identification of the major sources contribution to PM_{2.5} observed in Toronto, *Environ. Sci. Technol.*, 37, 4831–4840, 2003.
- Lee, S. H.: Emissions from Street Vendor Cooking Devices (Charcoal Grilling), US EPA, Washington, DC, 1999.
- Lewis, C. W., Norris, G. A., Conner, T. L., and Henry, R. C.: Source apportionment of Phoenix PM_{2.5} aerosol with the UNMIX receptor model, *J. Air Waste Man. Assoc.*, 53, 325–338, 2003.
- Li, S.-M., Anlauf, K. G., Wiebe, H. A., Bottenheim, J. W., Shepson, P. B., and Biesenthal, T. A.: Emission ratios and photochemical production efficiencies of nitrogen oxides, ketones, and aldehydes in the Lower Fraser Valley during the Summer Pacific 1993 oxidant study, *Atmos. Environ.*, 31, 2037–2048, 1997.
- Lindinger, W., Hansel, A., and Jordan, A.: Proton-transfer reaction mass spectrometry (PTR-MS): on-line monitoring of volatile organic compounds at pptv levels, *Chem. Soc. Rev.*, 27, 347–354, 1998.
- McLafferty, F. W.: Interpretation of Mass Spectra, University Science Books, Mill Valley, CA, 1980.
- Owega, S., Khan, B. U. Z., D’Souza, R., Evans, G. J., Fila, M., and Jervis, R. E.: Receptor Modelling of Toronto PM_{2.5} Characterised by Aerosol Laser Ablation Mass Spectrometry, *Environ. Sci. Technol.*, 38, 5712–5720, 2004.
- Paatero, P.: Least squares formulation of robust non-negative factor analysis, *Chemom. Intell. Lab. Syst.*, 37, 23–35, 1997.
- Paatero, P.: Interactive Comment on “Simultaneous factor analysis of organic particle and gas mass spectra: AMS and PTR-MS measurements at an urban site” by J. G. Slowik et al., *Atmos. Chem. Phys. Discuss.*, 9, C114–C123, 2009.
- Paatero, P. and Tapper, U.: Positive Matrix Factorization: A non-negative factor model with optimal utilization of error estimates of data values, *Environmetrics*, 5, 111–126, 1994.
- Polissar, A. V., Hopke, P. K., and Poirot, R. L.: Atmospheric aerosol over Vermont: chemical composition and sources, *Environ. Sci. Technol.*, 35, 4604–4621, 2001.
- Pope, C. A. and Dockery, D. W.: Health effects of fine particulate air pollution: lines that connect, *J. Air Waste Man. Assoc.*, 56, 709–742, 2006.
- Possanzini, M., Palo, V. D., and Cecinato, A.: Sources and photodecomposition of formaldehyde and acetaldehyde in Rome ambient air, *Atmos. Environ.*, 36, 3195–3201, 2002.
- Quinn, P. K., Bates, T. S., Coffman, D., Onasch, T. B., Worsnop,

- D. R., Baynard, T., de Gouw, J. A., Goldan, P. D., Kuster, W. C., Williams, E., Roberts, J. M., Lerner, B., Stohl, A., Petersson, A., and Lovejoy, E. R.: Impacts of sources and aging on submicrometer aerosol properties in the marine boundary layer across the Gulf of Maine, *J. Geophys. Res.*, 111, D23S36, doi:10.1029/2006JD007582, 2006.
- Ramadan, Z., Song, X.-H., and Hopke, P. K.: Identification of sources of Phoenix aerosol by positive matrix factorization, *J. Air Waste Manag. Assoc.*, 50, 1308–1320, 2000.
- Ramanathan, V., Crutzen, P. J., Kiehl, J. T., and Rosenfeld, D.: Aerosols, climate, and the hydrologic cycle, *Science*, 294, 2119–2124, 2001.
- Roberts, J. M., Fehsenfeld, F. C., Liu, S. C., Bollinger, M. J., Hahn, C., Albritton, D. L., and Sievers, R. E.: Measurements of aromatic hydrocarbon ratios and NO_x concentrations in the rural troposphere: Estimates of air mass photochemical age and NO_x removal rate, *Atmos. Environ.*, 18, 2421–2432, 1984.
- Robinson, A. L., Donahue, N. M., Shrivastava, M. K., Weitkamp, E. A., Sage, A. M., Grieshop, A. P., Lane, T. E., Pierce, J. R., and Pandis, S. N.: Rethinking Organic Aerosol: Semivolatile Emissions and Photochemical Aging, *Science*, 315, 1259–1262, 2007.
- Rolph, G. D., Real-time Environmental Applications and Display sYstem (READY) Website (<http://www.arl.noaa.gov/ready/hysplit4.html>), NOAA Air Resources Laboratory, Silver Spring, MD, 2003.
- Schauer, J. J., Kleeman, M. J., Cass, G. R., and Simoneit, B. R. T.: Measurement of Emissions from Air Pollution Sources, 1. C₁ through C₂₉ Organic Compounds from Meat Charbroiling, *Environ. Sci. Technol.*, 33, 1566–1577, 1999.
- Schindler, D. W.: Effects of acid-rain on fresh-water ecosystems, *Science*, 239, 149–157, 1988.
- 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, <http://www.atmos-chem-phys.net/9/2891/2009/>.
- Ulbrich, I. M., Lechner, M., and Jimenez, J. L.: AMS Spectral Database, online available: <http://cires.colorado.edu/jimenez-group/AMSsd/>, 2009b.
- Vlasenko, A., Slowik, J. G., Bottenheim, J. W., Brickell, P. C., Chang, R. Y.-W., Macdonald, A. M., Shantz, N. C., Sjostedt, S. J., Wiebe, H. A., Leaitch, W. R., and Abbatt, J. P. D.: Measurements of VOCs by proton transfer reaction mass spectrometry at a rural Ontario site: Sources and correlation to aerosol composition, *J. Geophys. Res.*, 114, D21305, doi:10.1029/2009JD12025, 2009.
- Watson, J. G.: Visibility: Science and Regulation, *J. Air Waste Manag.*, 52, 628–713, 2002.
- Weimer, S., Alfarra, M. R., Schreiber, D., Mohr, M., Prevot, A. S. H., and Baltensperger, U.: Organic aerosol mass spectral signatures from wood-burning emissions: Influence of burning conditions and wood type, *J. Geophys. Res.*, 113, D10304, doi:10.1029/2007JD009309, 2008.
- Zervas, E., Montagne, X., and Lahaye, J.: C-C organic acid emissions from an SI engine: influence of fuel and air/fuel equivalence ratio, *Environ. Sci. Technol.*, 35, 2746–2751, 2001.
- Zhang, Q., Alfarra, M. R., Worsnop, D. R., Allan, J. D., Coe, H., Canagaratna, M. R., and Jimenez, J. L.: Deconvolution and Quantification of Hydrocarbon-like and Oxygenated Organic Aerosols Based on Aerosol Mass Spectrometry, *Environ. Sci. Technol.*, 39, 4938–4952, 2005.
- Zhang, Q., Jimenez, J. L., Canagaratna, M. R., Allan, J. D., Coe, H., Ulbrich, I., Alfarra, M. R., Takami, A., Middlebrook, A. M., Sun, Y. L., Dzepina, K., Dunlea, E., Docherty, K., DeCarlo, P. F., Salcedo, D., Onasch, T. B., Jayne, J. T., Miyoshi, T., Shimonono, A., Hatakeyama, S., Takegawa, N., Kondo, Y., Schneider, J., Drewnick, F., Borrmann, S., Weimer, S., Demerjian, K. L., Williams, P., Bower, K., Bahreini, R., Cottrell, L., Griffin, R. J., Rautiainen, J., Sun, J. Y., Zhang, Y. M., and Worsnop, D. R.: Ubiquity and dominance of oxygenated species in organic aerosols in anthropogenically-influenced Northern Hemisphere midlatitudes, *Geophys. Res. Lett.*, 34, L13801, doi:10.1029/2007GL029979, 2007.
- Zhao, W. X. and Hopke, P. K.: Source identification for fine aerosols in Mammoth Cave National Park, *Atmos. Res.*, 80, 309–322, 2006.