
2.4. Analysis software development

While the instrument was demonstrated to be fully capable of producing meaningful, quantitative data and the underlying principles needed for quantification presented (see section 2.3.) [Jayne *et al.*, 2000; Jimenez *et al.*, 2003a], the instrument was lacking a formalised structure for applying all of these principles and generating useful data. Also, while the instrument can be accurately set up and calibrated for any given instance in time, the performance is known to vary over time and a reliable method of accounting for these variations was needed. One of the main goals of the work presented in this thesis was to develop the generalised analysis tools required to systematically process the raw data produced by the AMS.

All the analysis code developed has used Igor Pro version 4 (Wavemetrics Inc., Lake Oswego, OR, USA) as a platform, as the logging software saves the data in the Igor text format, which can be read directly by this software. The initial development work started during SASUA 3, which concentrated on re-coding and expanding some pre-existing code originally written by P.I. Williams (UMIST) and further modified by P. Silva (then at ARI), designed to process MS data. In parallel, work as part of this thesis was also performed to further develop some code written by J.L. Jimenez (then at Caltech), designed to process TOF data.

After several months' development, it became apparent that the existing software tools were becoming overly cumbersome and too limited in scope to produce data of the desired quality assurance level. Therefore, the task of starting afresh and streamlining the code was undertaken during the summer of 2001. This entailed unifying the MS and TOF data analyses into one common toolkit, as many of the techniques employed require the interchanging of data from the two processes. Over time, as the sensitivity of the instruments in use improved, the typical averaging times decreased from around 30 to 5 minutes or less. This, coupled with improvements in instrument reliability, meant that the typical volumes of data from the individual campaigns vastly increased from tens to several hundreds of megabytes or more. This necessitated another major overhaul during the summer of 2002, which focused on the method of data handling, so the computers were still able to process the data efficiently.

The analysis toolkit produced is now in general use by most AMS-using groups worldwide. While it has been designed around the processing of ambient field data, it can also be used for laboratory studies. It automatically applies any necessary

calibration data and corrections, permitting near real time analysis and producing accurate, quantitative data. It also maximises the use of the entire mass spectra when making calculations, rather than simply using individual peaks. However, it has also been designed around user customisation, so the corrections and applications of techniques are user-definable, keeping it as flexible as possible. While the coding and version control of the analysis software has been performed as part of this work, it has been and continues to be an open-source project and has received much testing and feedback from the overall AMS community.

The current version of the code, along with other programmes and resources associated with it, is posted on the Internet and can be found at http://cloudbase.phy.umist.ac.uk/people/allan/ja_igor.htm. The functions, principles and mathematical details of the software are presented here. Parts of this were originally presented in *Allan et al.* [2003a; 2003b; 2004b].

2.4.1. Corrections

Before any results can be extracted from the raw data, several corrections must be applied. The problems that cause these artefacts are often unavoidable aspects of data collection and the methods to overcome the effects were developed as part of this work as they were encountered.

2.4.1.1. Sensitivity corrections

Over time, the performance of the electron multiplier slowly degrades, reducing the magnitude of the signal generated per ion detected. This deterioration is taken account of by the multiplier calibrations described in section 2.3.1., but as the multiplier degrades continuously, a correction must be applied as a function of time. The relative degradation can be assumed to be uniform for all parent species, so a correction factor can be calculated by inspecting the signals due to the air beam that enters the instruments with the aerosol sample (either m/z 28, which corresponds to N_2^+ or m/z 32, which corresponds to O_2^+), which would be constant in ambient air if the amplification of the ion signal did not decrease with time. While this correction is designed to take account of changes to the multiplier's amplification, it will also account for any changes in its ion collection efficiency and the transmission of the quadrupole.

The flow rate should also be included in the calculation, as this may vary during a sampling period (due to the inlet critical orifice becoming partially blocked with

debris) and alter the amount of air reaching the detection region. However, the flux of air passing the skimmer and ultimately reaching the ionization region is dependent on the morphology of the gas expansion at the nozzle and is therefore not necessarily directly proportional to the sample flow rate. Instead, the following approximation is used:

$$AB \propto Q + Q^* , \quad (2.4.1.1.i)$$

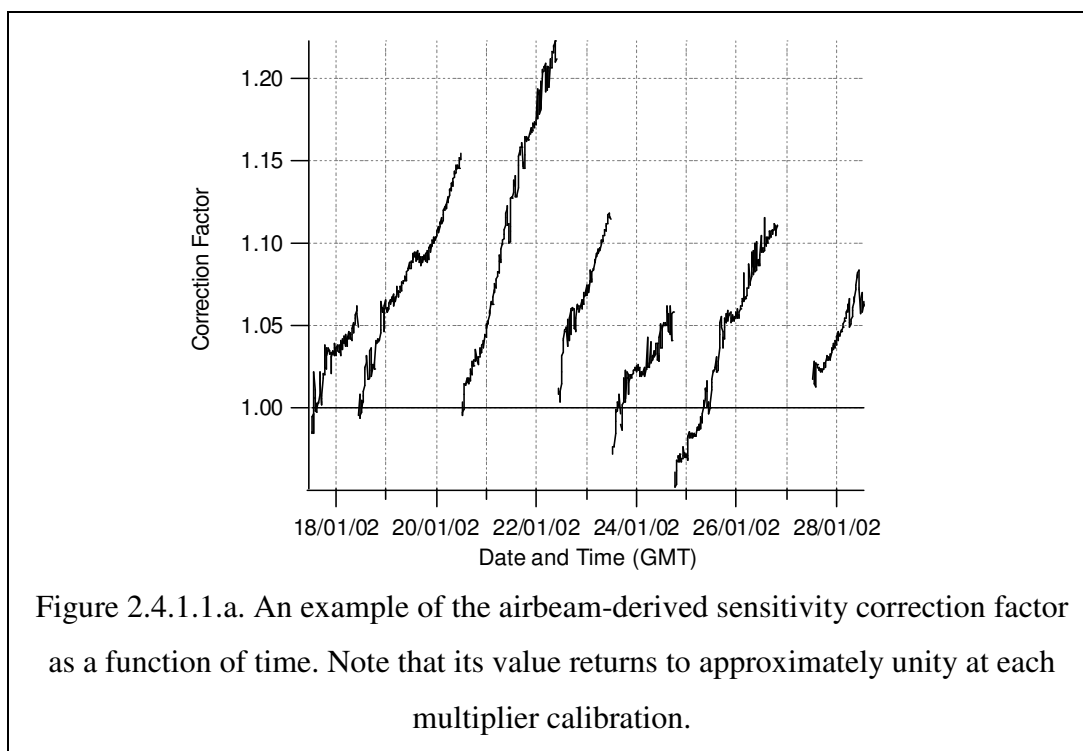
where AB is the signal due to the air beam (in detected ions per second), Q is the volumetric inlet flow rate and Q^* is a constant offset. This was evaluated in the laboratory by replacing the pinhole with a needle valve and measuring the airbeam strength as a function of measured flow rate. It was found that the approximation applied for flow rates above $0.5 \text{ cm}^3 \text{ s}^{-1}$ with Q^* equal to $0.68 \text{ cm}^3 \text{ s}^{-1}$. Below $0.5 \text{ cm}^3 \text{ s}^{-1}$, AB is not linear with flow rate. Using these assumptions, a dimensionless correction factor is obtained as a function of time (t), which is applied to all signals recorded by the AMS and is calculated using the following equation:

$$I_t^{\text{corrected}} = I_t^{\text{measured}} \cdot \frac{AB_0(Q_t + Q^*)}{AB_t(Q_0 + Q^*)}, \quad (2.4.1.1.i)$$

where AB_t and Q_t are the air beam strength and flow rate measured at time t and $t=0$ is considered to be the time of (or shortly after) a calibration. It must be stressed that this phenomenon is not the same as that described by *Bahreini et al.* [2003], as the changes experienced during their testing were caused by varying the upstream pressure to simulate changes in sampling altitude. Also note that this additional flow rate correction was developed for instruments that use a laminar pressure drop element upstream of the pinhole to measure the inlet flow; newer versions of the AMS estimate this quantity based on the pressure inside the lens, meaning that the correction may need to be applied differently. However, if the flow rate does not vary significantly during a campaign ($Q_t \approx Q_0$), this specific part of the correction is not needed.

The value of the correction factor can be seen as a ratio of the multiplier gain during calibration and the multiplier gain as a function of time. An example of this is shown in figure 2.4.1.1.a. The points where the calibration factor returns to unity are the occasions when electron multiplier calibrations were performed. The points where the correction factor does not return to exactly 1 may be due to either the tolerance of the calibration or variations in the ion collection efficiency (i.e. the fraction of available ions successfully detected rather than the strengths of the signals generated) through the

electron multiplier's working lifetime. Neither of these factors will affect the end results of the calculations, as the correction process will account for both phenomena.



As there is intrinsic noise on both the airbeam and flowrate time series, it may be necessary to numerically smooth them as a function of time to avoid artefacts when trying to correct for instrument performance. The amount of smoothing needed is dependent on both the sensitivity of the instrument and the averaging period chosen.

2.4.1.2. m/z corrections

As part of the routine calibration operations, the logging software must perform a calibration of the voltage sent to the quadrupole's RF generator as a function of the m/z desired by the software. This is approximated to a linear relationship with an intercept and a slope. It performs the calibration by inspecting the location of two known background peaks, m/z 28 (N_2^+) and 149 on the continuous mass spectrum and adjusting the calibration accordingly. m/z 149 is associated with softening agents used in plastics and is present in the background of all AMS mass spectra. It probably originates from either the O-rings used on the turbopump ports or the insulation on the wires used in the chopper assembly. This calibration is known to drift over time, especially if the cabin temperature changes. Also, the logging software applies the same calibration to both the open and closed spectra, whose calibrations can sometimes be

slightly different. The precise reasons for this are not known, but it is probably due to a minor electronic effect.

As part of the routine corrections, the analysis software re-analyses the continuous mass spectra to adjust the intercept and slope of the m/z calibration. It duplicates then smoothes the mass spectrum using a boxcar algorithm of an m/z width of 0.45 (9 data points), then locates the largest and most distinct peaks expected across the mass spectrum. The peak finding routine inspects the numerical first derivatives of the data and reports where these cross zero, using linear interpolation if necessary. The m/z values of the peaks that are searched for are normally 14 (N^+ and N_2^{++}), 18 (H_2O^+), 28 (N_2^+), 149 (the softening agent), 182, 184 and 186 (W^+ , from the filament). The m/z 's at which these peaks occur are compared with where they are expected and slope and intercept corrections to the m/z calibration are calculated using linear regression.

The discrete mass spectrum is then recalculated using this new calibration. Also, instead of simply taking the average of the nine data points inside the 0.45 m/z windows as the logging software does, the continuous spectrum is linearly interpolated, integrated within the window and then the area divided by 0.45. This means that the optimum location of the peak centre need not be located precisely on a data point. This correction is performed for each save on both the 'open' and 'blocked' mass spectra. The new 'difference' spectra are recalculated based on the new 'open' and 'blocked' spectra.

While it is safe to assume that all the ions encountered are of integer m/z when considering simple ions, this is not the case for those of m/z values of several hundred. For instance, the ^1H atom has an atomic mass of 1.0078 amu [*De Laeter et al.*, 2003]. While this variation from unity is not important when considering small numbers of hydrogen atoms (the maximum resolution of the data analysis software under normal operating conditions is 0.05), when a large molecule has several tens present, the positive discrepancy is of the order of the peak width at the default quadrupole resolution (0.45), so the expected peak location is misplaced by the analysis routine and the ion rate misrepresented. Likewise, heavy transition metals also exhibit non-integer atomic weights (e.g. the atomic weight of ^{208}Pb is 207.977 amu), although in these cases the discrepancy is negative.

The solution to the non-integer mass problem has been to specify a non-integer value for each m/z where the centre of the peak is expected. For all the low m/z , this is a value close to the integer but calculated based on the expected composition of the majority of the ions expected at that m/z (e.g. m/z 64 is expected to be SO_2^+ , which is

63.9619). For the higher m/z , all the peaks are assumed to be organic. As the precise molecular weights cannot be evaluated, they are estimated based on the assumption that their mass is made of carbon and hydrogen atoms in a 1:2 ratio. This procedure also means that the m/z values specified in the fitting procedure must also be specified as non-integer, so the values above become 14.003, 18.011, 28.006, 149.174, 181.948, 183.951 and 185.954.

The user must alter the list of non-integer m/z values if species not covered by the standard list (e.g. heavy metals) are expected, which may occur during laboratory studies or when sampling unusual environments such as flue emissions. The method relies on the assumption that each m/z channel either has a single species of ions present or that the multiple ion species at that m/z do not have significantly different molecular weights. Fortunately, for most data encountered during ambient sampling, at least one of these conditions holds true. This is an unfortunate limitation of the quadrupole mass spectrometer; any further inspection would require an m/z resolution of much less than unity.

2.4.1.3. Other MS corrections

When processing mass spectra, a correction described by the equation below should in theory also be applied to take account of varying instrument response for different m/z values. This is independent of chemical species and is taken from *Jimenez et al.* [2003a].

$$I_{m/z}^{Corrected} = \frac{I_{m/z}^{Measured}}{T_{m/z} G_{m/z}}, \quad (2.4.1.3.i)$$

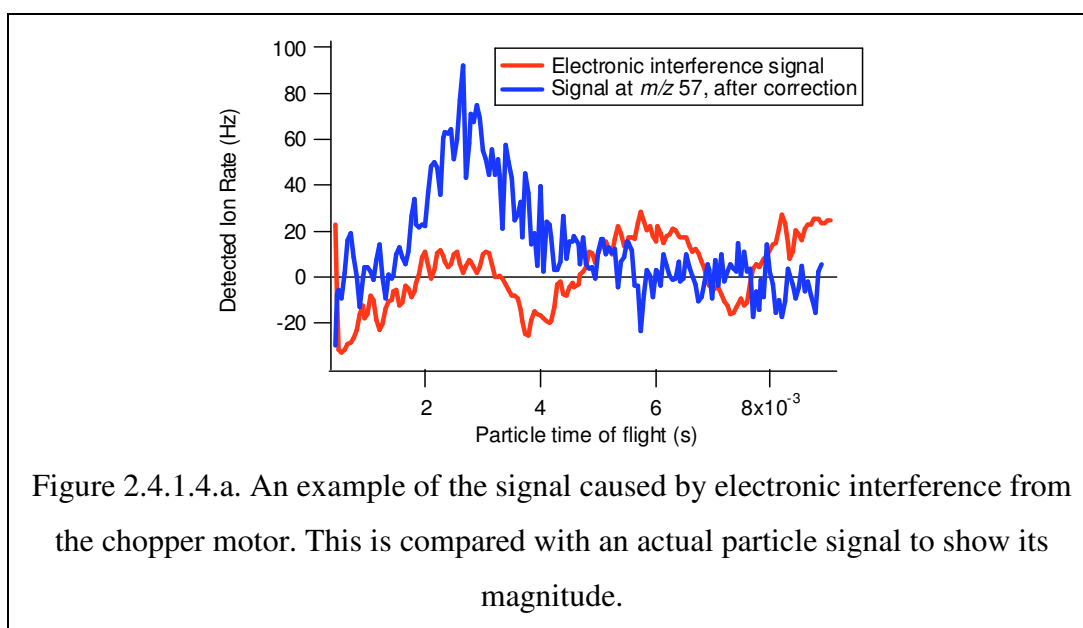
where $I_{m/z}$ is the signal intensity as a function of m/z in Hz, $T_{m/z}$ is the relative transmission of the quadrupole and $G_{m/z}$ is the relative gain and detection efficiency of the electron multiplier. However, this correction is not applied routinely, as the relative transmission of the quadrupole as a function of m/z has not yet been measured but should only be a minor correction for m/z 's less than 300, according to the manufacturer. Similarly, changes in the relative gain and detection efficiency of the multiplier are not deemed significant.

A problem was also encountered during NAMBLEX where the baselines of the mass spectra were being reported incorrectly. Normally, the logging software calculates the baseline (zero) of each mass spectrum by taking a mean average of some low m/z data points where signal is not expected (normally around 3-9), the mean of some data

points at the high extreme of the mass spectrum (around m/z 290-300) and linearly interpolating between the two points. Unfortunately, due to technical problems associated with the quadrupole during that experiment, the mass spectral data sometimes featured single point spikes in the data, which caused errors in the mean averages that in turn manifested as erroneous MS data across the board during the affected saves. This problem was remedied by recalculating the baselines using numerical medians rather than mean averages and adjusting the MS data accordingly.

2.4.1.4. TOF filtering

Due to temperature variations during operation, the chopper frequency can vary slightly during the course of an averaging period, meaning that some of the raw TOF samples may contain different numbers of data points compared to others. The result of this is the last few data points of each averaged trace can contain mathematical artefacts that normally manifest as having considerably higher or lower values than the rest. These can create problems later on during the data processing, so are filtered. This is performed by removing points one by one from the end of the trace until the last point is no further from zero than the penultimate point.



A common problem experienced with the UMIST instrument has been associated with the chopper motor. An electronic three-coil brushless motor has been used in all the data presented in this thesis (except for SASUA 3), which sometimes causes electromagnetic interference on the signal from the mass spectrometer. This is not a problem in MS mode, as the successive averaging removes this, but in TOF mode it is an issue, as the data collection is intrinsically synchronised to the motor's rotation.

This manifests as a periodic signal of three times the chopper frequency being superimposed on the data. Figure 2.4.1.4.a shows an example of the interference signal, which can be significant when compared to an actual particle signal. While it has been found that this can be mitigated by rearranging the cabling of the instrument, the problem is often only detected at the analysis stage. This is remedied by routinely monitoring m/z 11 during sampling. This m/z corresponds only to boron, which is not expected in any measurable quantities in ambient particles and should therefore give a flat trace. If it does contain a pattern, it is considered an artefact and is subtracted from all the other traces, removing any common interference.

2.4.2. Mass concentration calculation

2.4.2.1. Basic theory

The basic method for the conversion of an ion rate into a mass concentration is presented in section 2.2.2. However, as molecules of most species undergo fragmentation during the electron impact ionisation process, the detected ion rates at multiple m/z channels must be considered in practice. For example, there are two major nitrate peaks in the mass spectrum of ammonium nitrate aerosol at m/z 30 (NO^+) and m/z 46 (NO_2^+). To measure the total mass concentration of nitrate (NO_3) using the method described by *Allan et al.* [2003a], both fragments must be summed, as follows:

$$C_{\text{NO}_3} = 10^{12} \cdot \frac{MW_{\text{NO}_3}}{IE_{\text{NO}_3} Q N_A} (I_{30} + I_{46}) \quad (2.4.2.1.i)$$

This approach can be easily applied to ambient data for species such as nitrate and sulphate, where their major peaks are normally distinct in the ensemble mass spectrum and contributions from interfering species (e.g. organics) are low. However, the technique of simply summing the ion rates for individual m/z channels in the ensemble mass spectrum is not sufficient for most species. For example, the calculation of ammonium concentrations is not possible using this technique alone and the ability to provide this measurement is greatly desired from a scientific point of view.

Through laboratory studies using particles generated from an aqueous solution, it is known that particulate ammonium vaporises as ammonia (NH_3) and fragments under electron impact ionization to produce three major ions: m/z 15 (NH^+), 16 (NH_2^+) and 17 (NH_3^+). The relative sizes of these peaks in the mass spectrum have been found to be independent of the anion used in the solution, which is due to the vaporisation and

ionisation processes being decoupled. However, it is not possible to use these signals from the ensemble mass spectrum directly, as they all receive significant interferences from other species in the ensemble mass spectrum, most notably the $^{15}\text{N}^+$, CH_3^+ , O^+ , O_2^{++} and OH^+ ions.

In practice, m/z 15 is often ruled out as a useful channel for the calculation of ammonium concentrations because it is a minor peak in the ammonia mass spectrum (around 10 % of the size of the m/z 16 peak), with a consequently low signal to noise ratio. Also, it is difficult to determine quantitatively how much of the signal present is made up of CH_3^+ ions from certain organic species. Instead, it was found that it is possible to use the ammonia signals at m/z 16 and 17 for the calculation of ammonium concentrations, provided that the fractions of the signals due to interfering chemicals are predicted correctly. The first application of this specific technique will be presented in *Delia et al.* [In preparation].

The O^+ and O_2^{++} contributions originating from gas-phase oxygen can be predicted based on the signals from other major gas phase signals, as the ratio between these should be invariant as long as the ionisation, ion extraction and mass analysis conditions do not change. It has been found that using the signal at m/z 14 (N^+ and N_2^{++}) multiplied by a factor is the least susceptible to artefacts created by minor interferences by other species and changes in the quadrupole or ioniser performance. The EI fragmentation of water also produces large numbers of O^+ ions, but these can be predicted based on the signal at m/z 18, which is almost entirely due to H_2O^+ ions. The fragmentation pattern of water molecules has been observed to be independent of their source, whether they are from water vapour in the air beam, particle phase water in aqueous particles or from thermal decomposition of other chemical species on the heated surface. Similarly, the contribution of the OH^+ ion from water to the signal at m/z 17 can also be predicted based on the signal at m/z 18.

To summarise, the work by *Delia et al.* [In preparation] concluded that the signals due to vaporised ammonia at m/z 16 and 17 ($I_{\text{NH}_4,16}$ and $I_{\text{NH}_4,17}$) can be calculated from the raw mass spectrum as follows:

$$\begin{aligned} I_{\text{NH}_4,16} &= I_{16} - a_1 I_{14} - a_2 I_{18} \\ I_{\text{NH}_4,17} &= I_{17} - a_3 I_{14} - a_4 I_{18} \end{aligned} \quad (2.4.2.1.ii)$$

where a_1 , a_2 , a_3 and a_4 are multipliers whose values are based on the measured fragmentation patterns of air and water. The subtraction applied to $I_{\text{NH}_4,17}$ based on I_{14} is

due to the $^{17}\text{O}^+$ and $^{16}\text{O}^{18}\text{O}^{++}$ ions. As these isotopes make up less than 0.3 % of the total oxygen, a_3 will be small. The ammonium signal at m/z 15 ($I_{\text{NH}_4,15}$) is based on $I_{\text{NH}_4,16}$, as follows:

$$I_{\text{NH}_4,15} = a_5 I_{\text{NH}_4,16} = a_5 I_{16} - a_1 a_5 I_{14} - a_2 a_5 I_{18} \quad (2.4.2.1.iii)$$

where a_5 is based on laboratory measured fragmentation patterns of vaporised ammonia. These three calculated ammonia signals can then be summed and converted to a mass concentration of ammonium in a similar manner to equation 2.4.2.1.i. When using this technique, the quantitative capability of the AMS at a given m/z channel is limited by two main factors; the signal to noise ratio at that channel due to instrument background levels, electronic noise and the signal intensity being measured (see below) and interference between multiple species when one species' abundance is significantly greater than that of another.

2.4.2.2. Matrix calculation method

While the above technique has been used to generate ammonium mass concentrations effectively, it highlighted a more general problem of interferences within the mass spectra, which is applicable to all species. Previous work that used the summation technique of equation 2.4.2.1.i, such as that featured in *Allan et al.* [2003a], could only reliably calculate mass concentrations for sulphate and nitrate when their corresponding peaks were sufficiently distinct. Additionally, the calculated organic data were frequently subject to artefacts caused by minor inorganic peaks that were not accounted for (e.g. $^{34}\text{SO}^+$). From the ammonium exercise, it also became apparent that the study of species with even smaller contributions to the overall mass spectrum (e.g. chloride) would be more difficult still.

For these reasons, a new generalised method was developed with a view to being universally applied to all species. The underlying approach is to extract a 'partial' mass spectrum corresponding to a particular chemical species from the ensemble mass spectrum using known relationships between the peaks of all of the species involved. For example, in the case of ammonium described above, the magnitude of the signals at m/z 15, 16 and 17 due to vaporised ammonia is calculated based on the signals at m/z 14, 16, 17 and 18 in the ensemble spectrum. For the purposes of the general analysis, the following approach was used to process the ensemble mass spectrum:

$$\vec{I}_s = \mathbf{M}_s \vec{I}_{en} \quad (2.4.2.2.i)$$

where \vec{I}_s is the partial mass spectrum due to species s , \vec{I}_{en} the ensemble mass spectrum retrieved from the instrument and \mathbf{M}_s , a square conversion matrix specific to the species. The recorded mass spectrum and the deconvolved partial mass spectra are treated as vectors of ion rates with dimensions equalling the number of discrete m/z channels scanned, e.g. a 300 dimension vector if the m/z range 1 to 300 was scanned during operation. Using this approach, any single peak intensity in any species' partial mass spectrum can be estimated as a linear combination of the peaks in the ensemble mass spectrum. The conversion matrices are generated using laboratory-derived fragmentation ratios of all the species and knowledge of isotopic ratios and instrument performance. In practice, the contributions from the majority of channels in the ensemble spectrum to a given channel in a species' partial spectrum will be insignificant and so the conversion matrix, \mathbf{M}_s , will be sparse and a non-circular solution can be found.

When calculating the total mass concentrations of particular chemicals, the total ion rates are calculated by summing all of the peaks in their partial mass spectrum, giving the following generalised formula:

$$C_s = 10^{12} \cdot \frac{MW_s}{IE_s Q N_A} \sum_{\text{all } i} I_{s,i} \quad (2.4.2.2.\text{ii})$$

2.4.2.3. Fragmentation tables

In order to generate the conversion matrices (\mathbf{M}_s), a system of fragmentation tables has been developed. Operationally, each table represents a different chemical species, with each row being an m/z in that species' partial mass spectrum. Within the table, the user defines which peaks exist in each species' partial mass spectrum and their dependencies on other peaks in their own mass spectrum, the mass spectra of other species or the ensemble mass spectrum.

Examples of such tables, formatted as used in the software, are shown in tables 5.4.I-III, in the appendix section. The species-specific column names are preceded with 'frag_' for identification purposes. Each entry consists of a comma-separated list of the components that must be summed to obtain the value of that channel in the species' partial mass spectrum. These components can be peaks in the ensemble mass spectrum, which in this note are denoted by an integer number in square brackets (in the actual software, the square brackets are omitted, but are included here for clarity). Alternatively, a cross-reference to another entry in the fragmentation tables can be

given, which is denoted by the name of the table in question followed by a number in square brackets, identifying the referenced m/z in the table. The references can also be specified as having a negative contribution (to remove interferences) by the inclusion of a minus sign, or can include multipliers if a fractional contribution is required.

The development of extensive and reliable fragment tables, applicable to all instruments, is the result of ongoing collaborative work by several groups within the AMS community. The examples shown in this thesis are part of the tables used for analysing ambient AMS data, current at the time of submission, although these are continually being developed and are frequently changed to suit specific laboratory or field applications and many different versions are already in existence.

There are certain peaks that arise due to minor isotopes, for example, the peak at m/z 29 due to $^{15}\text{N}^{14}\text{N}^+$. These peaks are denoted by an ^a flag in the tables and the ratios were predicted based on the IUPAC recommended isotopic abundances [Coplén *et al.*, 2002; De Laeter *et al.*, 2003]. Patterns due to molecular fragmentation for the main inorganic species were obtained by various groups by analysing laboratory-generated single component particles and are denoted by a ^b flag. Detailed information on this work will be presented in future publications, but typical experiments used carrier gases of controlled compositions, for example pure argon, to minimise the interferences to the particle-phase signals. This type of analysis now permits the use of more peaks within a species' mass spectrum. For instance, rather than the two m/z signal contributions used in equation 2.4.2.1.i, the table for nitrate actually contains eight contributions from vaporised nitric acid, which include the parent ion at m/z 63, minor isotope ions at m/z 31, 32, 47, and 48 and the N^+ fragment at m/z 14. Also, ammonium sulphate vaporises as NH_3 , H_2O , SO_3 and H_2SO_4 . The amount of ammonium and water present in the particles can shift the vaporisation equilibrium between H_2SO_4 and SO_3 plus H_2O , therefore it is necessary to handle these separately, as they have different individual mass spectra. As stated before, these phenomena are currently under investigation and details will be presented in future publications [e.g. Hogrefe *et al.*, 2004; Middlebrook *et al.*, In preparation].

There are certain aspects of the fragmentation tables that are known to change between instruments, instrument configurations and even deployments. For instance, the value of the O^+ and O_2^{++} to N^+ and N_2^{++} ratio mentioned above is known to vary slightly depending on the precise setup of the ionizer and quadrupole. Therefore, this value must be carefully set when analysing a dataset if reliable ammonium concentrations are to be

calculated, as the O^+ and O_2^{++} signal is typically large compared to that of NH_2^+ . Also, the values of certain quantities are dependent on the sampling environment, for example the contribution of gas-phase CO_2 to the m/z 44 peak. This contribution must be set by the user and in certain circumstances, for example when sampling low organic concentrations or when sampling combustion sources directly, may have to be time-dependent. These values can be obtained by analysing the mass spectra of filtered air, which should be taken routinely during sampling. If the calculated particle signals at m/z 16 and 44 are non-zero, the values in the fragmentation tables are altered accordingly, as the only significant signals observed through the filter are from the gas phase.

The organic fragmentation table is based on the assumption that any detected matter that cannot be accounted for by known inorganic components must be organic. There are certain peaks in the ensemble mass spectrum that contain both inorganic and organic mass signals. In most cases, a minor inorganic contribution is calculated based on larger peaks in that species' spectrum and laboratory measured fragmentation patterns. Where this is not possible, for example at m/z 64, the organic contribution is instead based on the other signals in the organic mass spectrum. In this case, the numerical average of the signals at m/z 50 and 78 are used, as these correspond to the addition and subtraction of a CH_2 group from hydrocarbon fragments; long chained saturated hydrocarbons will fragment under standard 70 eV electron impact ionization by loss of C_nH_{2n+1} alkyl groups [McLafferty and Turecek, 1993, ch. 5]. While this is only an approximation, it has been found to be fairly robust during episodes where the high m/z organic component of ambient aerosol is composed of long chained, saturated hydrocarbons such as lubricating oils, which is typical of urban environments (see section 3.1.5.2.).

The analysis software performs the conversion from a fragmentation table column to a matrix in two stages. The first stage is to iteratively evaluate all the cross-references until only references to m/z channels in the ensemble mass spectrum remain. This generates an intermediate table. The second stage is to generate the matrix from this new table, and an example is shown in table 2.4.2.3.I, which is the conversion matrix for ammonium (M_{NH_4}). The i^{th} row of a matrix corresponds to the i^{th} m/z channel of the species' partial mass spectrum, or the i^{th} row of the fragmentation table. The j^{th} column of the matrix corresponds to the j^{th} m/z channel of the ensemble mass spectrum. The individual contributions (i.e. the fully evaluated multipliers) specified in the

intermediate table are added to the matrix according to their row in the table and column corresponding to the m/z channel in the ensemble mass spectrum that each entry refers to. As mentioned earlier, the matrices generated using this technique tend to very sparse, so only part of it is shown for clarity. The values not shown in this example are zero or otherwise less than 10^{-5} in magnitude.

	14	15	16	17	18	...	29	30	...	46
15	-3.53e-02	0	0.1	0	-4.00e-03		-3.11e-05	1.41e-03		1.41e-03
16	-3.53e-01	0	1	0	-4.00e-02		-3.10e-04	1.41e-02		1.41e-02
17	3.85e-05	0	0	1	-2.50e-01		3.39e-08	-1.54e-06		-1.54e-06

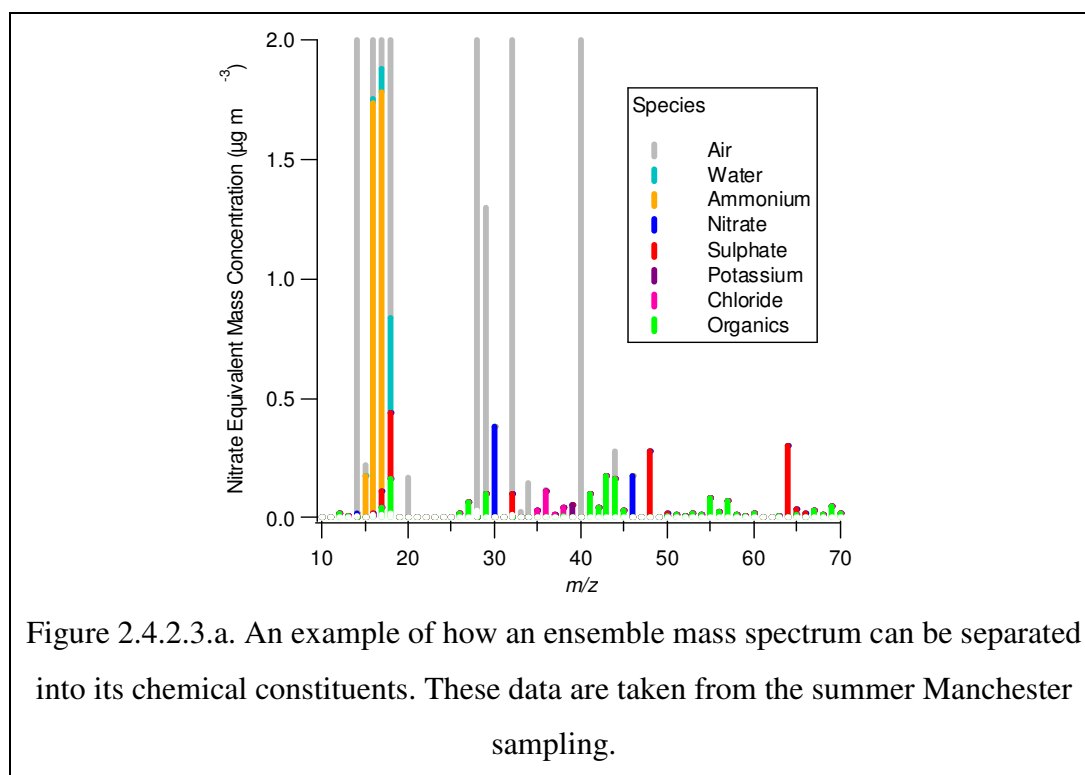
Table 2.4.2.3.I. Part of the conversion matrix used for the extraction of the ammonium mass spectrum (M_{NH_4}). Values not shown are zero or less than 10^{-5} in magnitude, i.e. contribute less than 10 parts per million to the total. Data are shown to 3 significant figures for clarity.

The matrix contains unity values at [16][16] and [17][17], which shows that the calculated values of the ammonium partial mass spectrum at m/z 16 and 17 include those of the ensemble mass spectrum. The various negative values in columns 14 and 18 show that subtractions are made based on the signals at m/z 14 and 18. These are, respectively, manifestations of the oxygen (O^+ and O_2^{++}) and water (O^+ and OH^+) fragment subtractions detailed in equation 2.4.2.1.ii. Note the exception is the value at row 17, column 14, which is slightly positive. Equation 2.4.2.1.ii deals with a minor subtraction at m/z 17 due to the $^{17}\text{O}^+$ and $^{16}\text{O}^{18}\text{O}^{++}$ ions but an additional correction is also made to the water signal at m/z 18 due to the $^{18}\text{O}^+$ ion from gas phase oxygen, which manifests as a net positive influence on 17 that largely cancels out the former correction.

The zero value at [15][15] shows that there is no contribution to the calculated value at m/z 15 from the corresponding signal in the ensemble mass spectrum; instead, it is based on the signals at m/z 14, 16 and 18, in the same way as equation 2.4.2.1.iii. The entries at columns 30 and 46 come about because the aforementioned subtraction of the O^+ and O_2^{++} signal is based on the N^+ and N_2^{++} signal from air, which in turn must include a correction due to the N^+ fragment from nitrate (NO_3) in the particle phase. This correction is based on the signals at m/z 30 (NO^+) and 46 (NO_2^+). m/z 30 also contains a minor contribution from organic signal fragments containing a single ^{13}C isotope, which is predicted based on the signal at m/z 29, hence the small dependency there. There are other dependencies elsewhere in the matrix, but these are smaller still. While the extra accuracy achieved through the smaller of these corrections is probably

small compared to other errors or uncertainties associated with the instrument, they do demonstrate the thoroughness of the technique.

As well as being able to calculate mass concentrations, another advantage of this approach is that parts of peaks within graphed mass spectra can be coloured according to their chemical sources, facilitating data interpretation. An example of this is shown in figure 2.4.2.3.a, which uses data from sampling ambient air in Manchester during January 2002 (see section 3.1.2.) and the example tables presented in the appendix. As a diagnostic tool, the peak ratios in the partial mass spectra can be compared with expected values and the partial mass spectra summed and compared with the ensemble mass spectra they are derived from. Any inconsistencies are normally indicative of either one or more incorrectly set values within the fragmentation tables or a species that is not correctly accounted for within the tables used.



Since its original development, this technique has so far proven to be a very powerful and versatile calculation method. Not only has it facilitated the calculation of ammonium, organic and other chemical concentrations, it also means that the evaluation of species can now be performed more thoroughly and systematically, and in a manner that is self-consistent across different AMS groups. This should mean that more accurate results can be obtained that are less prone to artefacts and chemical interferences. It has also allowed for the study of a multitude of other chemical types and sub-groups within types, such as studying different categories of organic groups

such as polycyclic aromatic hydrocarbons, oxygenated and non-oxygenated hydrocarbons and markers for identifying biomass burning products. The details of these and the scientific conclusions arising will again be the subject of future publications.

While the example tables presented in this thesis were current at the time of submission, work is continuously being performed by various groups within the worldwide AMS users community to update, expand and validate these tables, a process that is likely to be ongoing for some time. For the most recent fragmentation tables and information on the current work of the various groups, the reader is directed to http://cloudbase.phy.umist.ac.uk/people/allan/ja_igor.htm and <http://cires.colorado.edu/jimenez/ams.html>.

2.4.2.4. Relative ionisation efficiency

In the equations given in sections 2.4.2.1. and 2.4.2.2., an ionisation efficiency IE_s is required, which is specific to the chemical species being measured. However, it is impractical to experimentally determine specific values of IE for each species and each experiment. Normally, only the value for nitrate, IE_{NO_3} is routinely measured. The calibrated value of IE_{NO_3} can also be applied to other chemical species, negating the need to perform IE calibrations for multiple chemicals routinely. The assumption is that the ionization cross-section of the parent molecules is proportional to the number of electrons present [Jimenez *et al.*, 2003a]. If it can be assumed that this in turn is to a close approximation proportional to the molecular weight, the following generalization can be made about a species s when compared to nitrate:

$$\frac{IE_s}{MW_s} = RIE_s \frac{IE_{NO_3}}{MW_{NO_3}}, \quad (2.4.2.4.i)$$

where RIE_s is the relative ionisation efficiency as described by Alfarra *et al.* [2004], a dimensionless constant specific to the chemical species, which can be determined in the laboratory. This means that the mass calculation formula for nitrate can be applied to other chemical species by dividing by RIE_s and using the values MW_{NO_3} and IE_{NO_3} , providing that the chemical's partial mass spectrum has been correctly evaluated. This gives the following generalised formula:

$$C_s = 10^{12} \cdot \frac{MW_{NO_3}}{RIE_s IE_{NO_3} Q N_A} \sum_{\text{all } i} I_{s,i} \quad (2.4.2.4.ii)$$

The values of *RIE* for the various chemical species are determined in the laboratory or during routine calibrations. The values used for the data presented in this thesis are listed in appendix 5.3. The value for sulphate is derived from laboratory measurements by many groups [Hogrefe *et al.*, 2004; Middlebrook *et al.*, In preparation], but the analysis of this experimental work is currently ongoing, and the most appropriate value is still being evaluated. The value of 1.15 has been used in all the data presented in this thesis (D.R. Worsnop, Aerodyne Research, Personal communication, 2003).

The value for ammonium has been found to vary somewhat between instrument configurations and an evaluation of this number is included as part of the standard ammonium nitrate calibrations as described in section 2.2.2. Some of the data presented was taken before this part of the calibration procedure was instituted, so instead, the value is inferred through an ion balance using the sulphate and nitrate concentrations, providing that nitrate loadings are significant. If ionic nitrate is present in the particles, they must be pH neutral because otherwise, it would evaporate as nitric acid. Likewise, they cannot be basic, as that would cause the ammonium to evaporate as ammonia. Therefore under these circumstances, the charge-weighted molar concentration of ammonium must balance those of the nitrate and sulphate. This works on the assumption that only nitrate, sulphate and ammonium contribute to the pH of the bulk of the particles. While other groups have reported large ambient chloride concentrations in certain environments (N. Takegawa, University of Tokyo, Unpublished field data 2002) this is not present in significant quantities in any of the data presented here. Also, while there is evidence to suggest the presence of internally mixed n-carboxylic acids and possibly amines, these are assumed to be too weak to affect the balance significantly. Finally, basic dust particles composed of calcium compounds may also possess a surface layer of nitrate with no ammonium (see section 3.3.1.2.), but these tend to exist in the coarse mode, so are not detected efficiently and do not contribute significantly to the total mass observed by the AMS.

Somewhat counter-intuitively, the value for nitrate is normally 1.1, which is the value used in all the data presented in this thesis. This is because the logging software that performs the nitrate IE calibration only takes account of the *m/z* 30 and 46 signals, whereas the analysis software is able to include several peaks arising from ¹⁵N and ¹⁸O isotopes, as well as the HNO₃⁺ ion. This means that the logging software-derived *IE*_{NO₃} is actually an underestimation of the 'true' *IE*_{NO₃}. The value for organic chemicals has

been shown to vary according to the type of chemical encountered (M. R. Alfarra, UMIST, unpublished laboratory data, 2002; P. J. Silva, Aerodyne Research Inc., now at Utah State University at Logan, unpublished laboratory data, 2001) but the general value of 1.4 is used for ambient organics. However, it must be stressed that this is only an approximation, as there are potentially many different types of chemicals present in the particles in varying quantities. Also, the molecules may thermally decompose on the vapouriser to produce some inorganic vapour molecules (see section 2.4.1.) in addition to the organic molecules, which will have different *RIEs*. Efforts to evaluate this further are ongoing.

2.4.2.5. Collection efficiency

As discussed before in section 2.2.2., some particle types exhibit non-unity collection efficiencies. It has been found that during most ambient sampling, accumulation mode particles are dried by the heat of the instrument, so that on average only half of the particles reach and stick to the heater. For this reason, it has been necessary to include an additional correction factor to take account of this, which is taken to be the reciprocal of a species' collection efficiency. This quantity is the fraction of the total number of particles sampled that are successfully vaporised on the heater, as discussed in section 2.2.2.

For species that normally exist in the accumulation mode, such as nitrate, sulphate and ammonium, this value is taken to be 0.5. The value for organics is also taken to be 0.5 if it is exclusively contained within the accumulation mode, but a suitable value has yet to be evaluated for occasions when significant amounts are contained within primary exhaust particles. These particles will be geometrically more irregular than accumulation mode particles, so produce a more divergent beam [Huffman *et al.*, 2004], but also possess a less volatile organic liquid coating, which may promote sticking. There is also as yet no suitable correction for collection efficiency applied to coarse mode particles, as these are also subject to losses due to the lens transmission efficiency. It must be noted however, that simply assuming a collection efficiency is by no means ideal. To evaluate this value, it may be necessary to use temperature cycling on the inlet (see section 1.4.2.1.), employ a beam shape probe (see section 1.4.2.2.) or compare with data from another source (see section 1.7.2.). The effect can also be mitigated by humidifying the particles (see section 1.3.2.2.), but this should be used in conjunction with a method of measuring the effect, to verify that it is

working for the particles being measured. Support laboratory work using particles of known compositions is also ongoing [*Matthew and Middlebrook, 2003*].

2.4.2.6. Error Analysis

Because a finite number of randomly generated ions are being detected, there will be an intrinsic variability in the average number detected. As the number of available molecules is high but the probability of successfully ionizing and detecting a specific molecule is low, it can be assumed that the probable distribution of numbers detected for a given population can be modelled as a Poisson distribution (the probability of producing and detecting multiple ions from a single molecule is considered lower still). The standard deviation of this distribution is equal to the square root of the product of the number of available molecules and the probability of detection [*Stroud, 1987, pp. 893*]. This product should be approximately equal to the average number of ions detected during the sampling period. Using this approach, the error associated with a signal I , in counted ions per second, can be estimated as follows:

$$\Delta(I t_s) = \alpha \sqrt{I t_s} \therefore \Delta I = \alpha \frac{\sqrt{I}}{\sqrt{t_s}}, \quad (2.4.2.6.i)$$

where t_s is the amount of time, in seconds, spent sampling a particular m/z channel. As the mass spectrum is scanned continuously (rather than dwelling on specific m/z) the calculations used to evaluate this time will have to include the averaging period, the scanning rate (1000 s^{-1}) and the m/z window size (0.45). α is a factor applied to account for the fact that the signals from single ions are not identical but arise from a Gaussian distribution of pulse areas, which adds an extra uncertainty to the ion measurements. The relative standard deviation of this distribution has been determined in the laboratory to be about 0.68 for one type of multiplier detector and can be readily determined for a different detector by the data acquisition software. If this Gaussian is convolved with the Poisson distribution, the latter's width is effectively increased by 20%, so therefore α is given the value 1.2.

Note that this calculated error applies to a single 'open' or 'blocked' signal. In order to calculate the error on the 'difference' signal, which is the arithmetic subtraction of the 'blocked' from the 'open', the errors must be summed in quadrature, as follows:

$$\Delta I_d = \sqrt{\Delta I_o^2 + \Delta I_b^2} = \alpha \frac{\sqrt{I_o + I_b}}{\sqrt{t_s}} \quad (2.4.2.6.ii)$$

A final allowance is also made to account for electronic effects associated with sampling, mainly due to unavoidable electromagnetic interference. This can be related to random (white) noise, periodic noise and artefacts associated with the resolution of the analogue to digital converter. Providing it is not synchronised with the mass spectrometer, periodic noise should be removed over successive scans through the averaging process, its magnitude being inversely proportional to the amount of averaging time. White noise, on the other hand, will decrease in the same manner as the error associated with the ion counting statistics, so will reduce inversely proportional to the square root of the sampling time. For this reason, only the white noise is considered here, but it must be noted that periodic noise may be important if the instrument is ever used for very high time resolution sampling. Errors associated with the voltage resolution of the data acquisition system are considered small compared to the white noise.

This necessitates the inclusion of an extra parameter, q_I , whose units are $\text{Hz s}^{0.5}$. This is the standard error of a population of measured signal data points, in Hz, multiplied by the square root of the sampling time per data point in seconds. This can be measured by monitoring the signal from the mass spectrometer when the filament is turned off and no ion signals should be present and calculating the standard deviation of the data points and dividing by the root of the sampling frequency. Note that any random noise directly affects the measured voltage and not the ion rate as perceived by the analysis software, so as this error parameter is measured in $\text{Hz s}^{0.5}$, it varies linearly with the measured single ion strength. The final version of the error calculation is therefore as follows:

$$\Delta I_d = \frac{\sqrt{\alpha^2 I_o + \alpha^2 I_b + 2q_I^2}}{\sqrt{t_s}} \quad (2.4.2.6.iii)$$

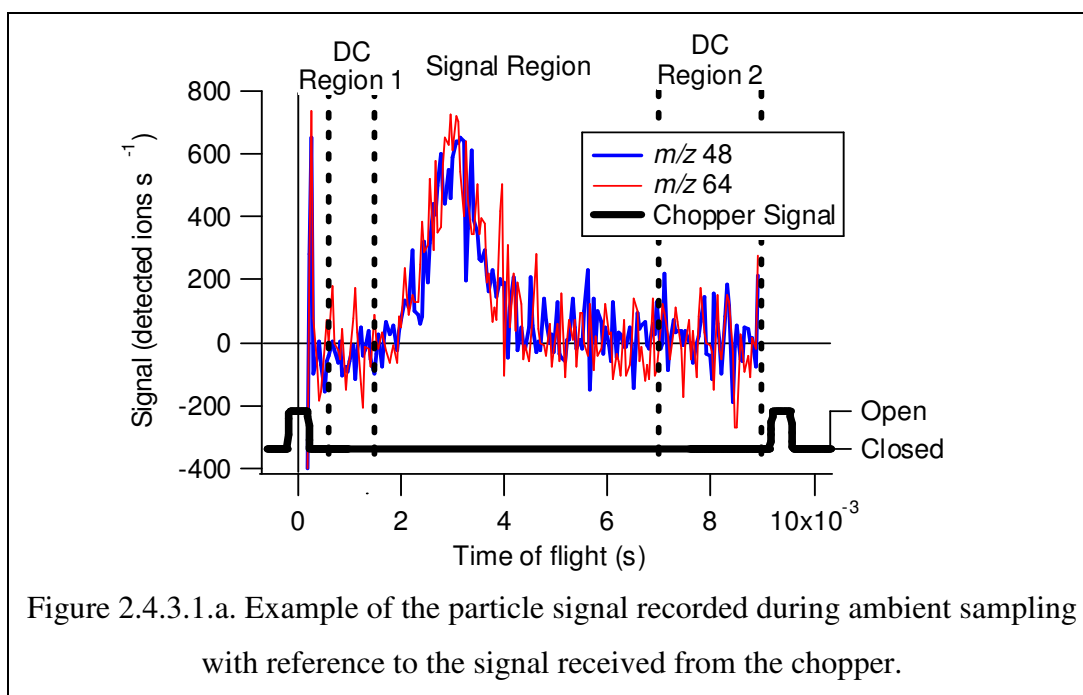
These calculated errors can be propagated to the chemical species mass concentrations, providing that the appropriate arithmetic is used with respect to the matrix operations.

2.4.3. Size-resolved mass distributions

2.4.3.1. Inversion from time of flight to diameter space

When the instrument is run in time of flight (TOF) mode and the detector response is displayed as a function of particle time of flight at a fixed m/z , a size

distribution for a particular chemical species can be obtained such as that presented in figure 2.4.3.1.a. The area under the graph represents the average number of ions detected per chopper cycle and can be converted to total mass rate, using the same mathematical method as before but including the slit width information to compensate for the smaller fraction of the total particle beam sampled in TOF mode compared to MS mode.



The zero in a TOF trace is evaluated in two regions, shown in 2.4.3.1.a as the ‘DC regions’, where no particles are expected to be detected. These correspond to velocities that are too low or too high to represent the particles that are transmitted into the instrument. This evaluation of the baseline removes the contribution to the signal from background gas phase material in the detection region. The need for a particle-free region at the end of the TOF period determines the maximum chopper frequency at which the instrument can be operated.

The ion rate as a function of time of flight, I_p , can be transformed to \log_{10} diameter space, $I_{\log(Dva)}$ using equation 2.4.3.1.i. This approach treats the detected ion rate as dN_I/dt_p , where N_I is the average total number of ions detected per TOF period (i.e. the area under the curve in figure 2.4.3.1.a) and is maintained during the transformation. This uses the fitted parameters from the velocity calibration described in section 2.3.3.2.

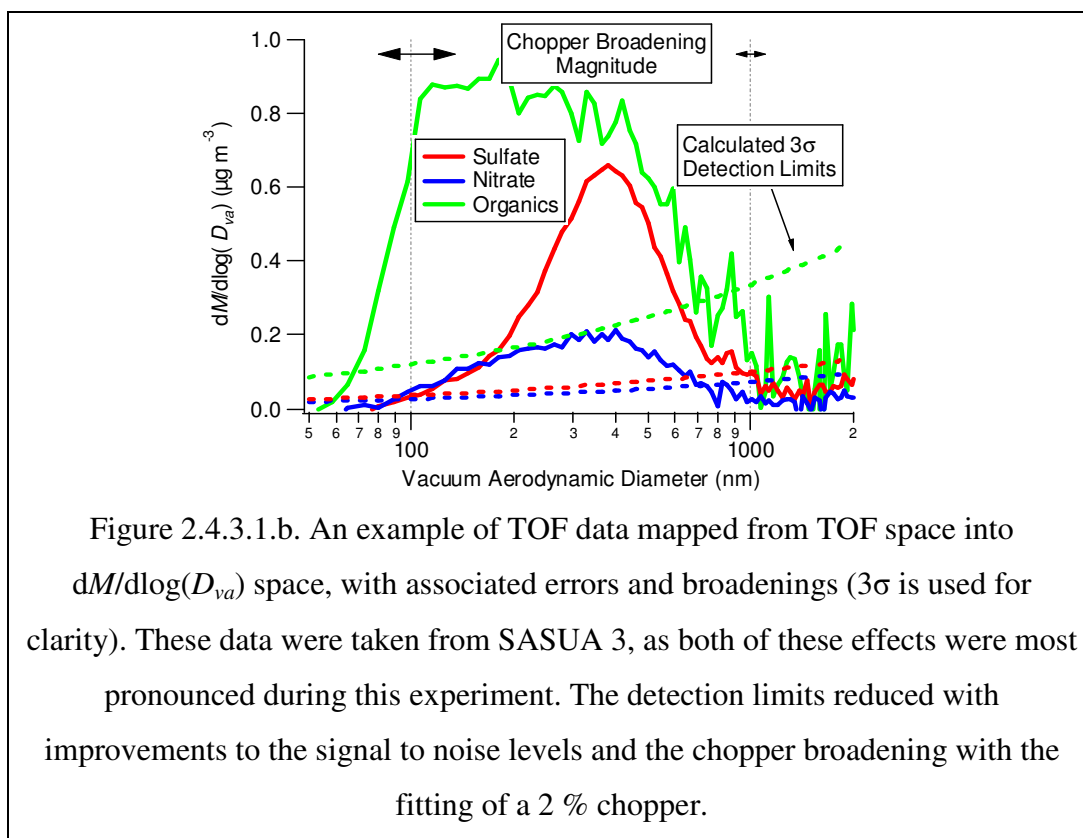
$$\log(D_{va}) = \log(D^*) + \frac{1}{b} \log\left(\frac{v_g - v_l}{(L_c/t_p) - v_l} - 1\right)$$

$$I_{\log(D_{va})} = \frac{dN_I}{d\log(D_{va})} = \frac{dN_I}{dt_p} \cdot \frac{dt_p}{d\log(D_{va})} = 2.3026 I_{t_p} b (A_0 + A_1 t_p + A_2 t_p^2)$$

$$\text{where } A_0 = -\frac{L_c}{v_g - v_l} \quad A_1 = 1 + \frac{2v_l}{v_g - v_l} \quad A_2 = -\left(1 + \frac{v_l}{v_g - v_l}\right) \frac{v_l}{L_c}$$

(2.4.3.1.i)

The detected ion distribution in \log_{10} diameter space can then be easily converted to a mass distribution ($dM/d\log(D_{va})$) by normalizing to the total mass concentrations calculated from the MS mode data; figure 2.4.3.1.b shows an example of this. An indication of the 3σ detection limit associated with the signals has also been included. This method used to calculate this is described in section 2.4.3.3.



The normalization serves to correct the integral of the trace if not all of the m/z channels corresponding to a chemical species are being scanned in TOF mode. It also removes any errors associated with the fact that at any one time, the precise m/z selected during TOF mode may not be positioned at the optimal location within the peak in the mass spectrum. Normally, this is set to be between -0.15 to -0.2 relative to the centre of the peak in mass spectrum mode. This is needed because due to a lag in the electronics,

the actual m/z selected during scanning is a fraction of a millisecond behind the scanned voltage sent to the RF generator. Sometimes the peaks drift with variations in the m/z calibration (see above), which the logging software does not always take account of properly. Also, peaks in the mass spectra sometimes have been observed to possess some fine structure to their shape, related to the time the energy of the ions during filtering (if the ion energy is too high, the ions do not spend enough time in the quadrupole to be resolved correctly, but if it is too low, the transmission is affected, so a trade-off has to be made). In MS mode, the process of averaging over the 0.45 m/z window largely removes these structures, so the mass concentrations derived from this mode are deemed more reliable.

This distribution should, if a particle density can be assumed or derived from the composition, be comparable with a volume ($dV/d\log(D_p)$ or $n_v(D_p)$) distribution that can be derived from other sizing instruments [*Seinfeld and Pandis*, 1998, pp. 414-416]. However, it is expected that the mode aerodynamic diameter measured by the AMS will not necessarily agree with the diameters obtained using different sizing techniques such as OPCs or DMPSs. Perfect size agreement with other instruments that employ aerodynamic methods such as aerodynamic particle sizers (APSs) or cascade impactors is also not expected, due to the differences in the sizing methods discussed in section 1.2.1.

2.4.3.2. Weighting m/z 's

As discussed above, when calculating the total mass concentration for a species, many m/z channels may contribute to the total loading. In the case of organics, hundreds of channels are normally used. It is impractical to use all m/z 's in TOF mode, so the channels used for a particular species must represent the overall chemical distribution when summed. This is not a problem for species such as sulphate, as this fragments consistently within the ionisation region and traces from all m/z channels are normally identical in shape, if not in magnitude. However, because there are many different types of chemicals present within the organic fraction, the relative intensities of the peaks within the mass spectrum change according to which are present in the sampled particles. If there is more than one species of organic particle present, as is routinely seen, particles of one size may be significantly different chemically to those of another. This means that the signals at different organic m/z 's are not necessarily the same shape.

The m/z channels chosen to study organic chemicals in TOF mode typically include 41, 43, 44, 45, 55, 57, 69, 71 and 91 based on the fact that these are prominent organic peaks in ambient mass spectra and are expected to arise from chemical species such as alkanes, carboxylic acids, alcohols and aromatics [McLafferty and Turecek, 1993], which have been observed in ambient particles using other methods [e.g. Fuzzi *et al.*, 2001; Saxena and Hildemann, 1996]. In urban environments in particular (as is discussed in section 3.1.), the organic fractions form bimodal distributions, where the mode vacuum aerodynamic diameters are the same between m/z channels, but the relative sizes of the modes aren't. When trying to generate an overall organic distribution, representative of all organic chemicals and not just those strongly accounted by the channels chosen, it is not sufficient to simply sum the traces.

The frequent observations of bimodal size distributions (as discussed in more detail in section 3.1.5.1.) suggests that at any one time, the total observable organic mass distribution in log diameter space, $I_{org, \log(D_{va})}$, can be approximated as the sum of two parameterised modes. This can be summarized using the following equation:

$$I_{org, \log(D_{va})} \approx B_1 f_{\log(D_{va})}^1 + B_2 f_{\log(D_{va})}^2 \quad (2.4.3.2.i)$$

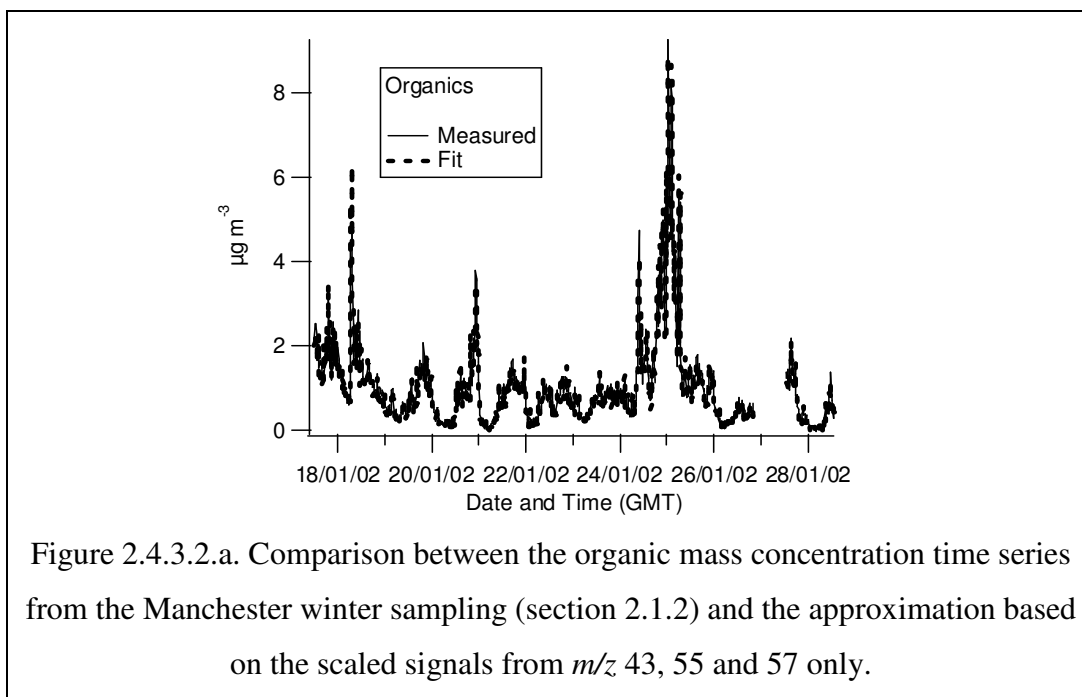
where f^1 and f^2 are two modes in log diameter space of unit area and B_1 and B_2 are the sizes of the modes. It is also assumed that the signal (ion rate) distributions from individual m/z channels also obey the same approximation using the same mode functions f^1 and f^2 .

The m/z channels chosen for the determination of the total organic distribution must have high signal to noise and when combined with the appropriate weighting factors, must correctly represent the behaviour of the total organic loading. In the example shown in figure 2.4.3.2.a (taken from the Manchester January 2002 sampling presented in section 3.1.2), the first requirement of high signal to noise is best fulfilled by m/z 43, 55 and 57. The m/z 44 and 45 channels have too low a signal to noise to be suitable for this analysis. The ability of the three channels to represent the total organic loading is checked by doing a linear least squares fit to the following equation:

$$I_{org} \approx p_{43} I_{43} + p_{55} I_{55} + p_{57} I_{57} \quad (2.4.3.2.ii)$$

where I_{org} is the total organic mass loading, $I_{m/z}$ are the total signals from the respective m/z channels obtained as a time series from MS mode, and $p_{m/z}$ are the weighting factors for each m/z channel determined from the fit. The fit produced using this method is shown in figure 2.4.3.2.a. The fit is very good and the weighting factors

p_{43} and p_{57} relative to p_{55} are found to be 1.46 and 0.57 respectively. On average, these fragments represented 17.6 % of the total organic mass measured, although this can be expected to vary between field projects.



The weighting factors are indicative of how much of the behaviour of the total organics is represented by the activity in a particular m/z channel. Note that it is the relative sizes and not the total magnitude of these factors that is important, as the final TOF distribution will be normalized to the total organic loading after summation. The area under the total organic distribution will be a sum of the areas of the two modes, as follows:

$$I_{org} \approx B_1 + B_2 \quad (2.4.3.2.iii)$$

If it can be assumed that the weighting factors, $p_{m/z}$, are constant within a measurement period, then the total organic mass distribution can be derived as a function of time using the following equation:

$$I_{org, \log(D_{va})} \approx p_{43} I_{43, \log(D_{va})} + p_{55} I_{55, \log(D_{va})} + p_{57} I_{57, \log(D_{va})} \quad (2.4.3.2.iv)$$

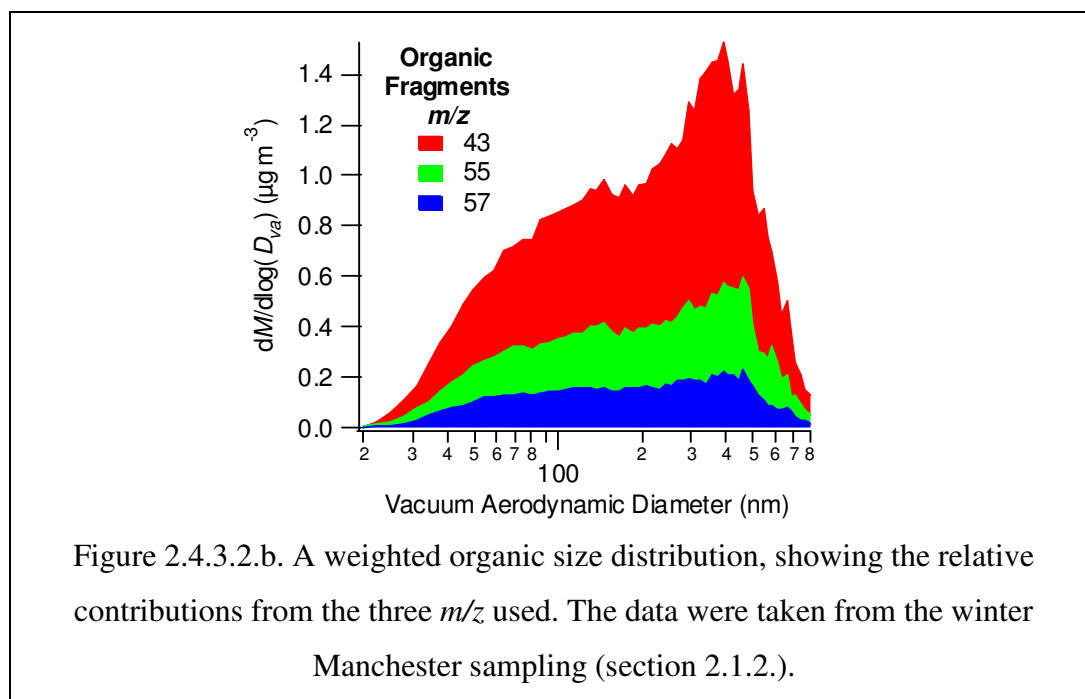
where $I_{m/z, \log(D_{va})}$ is the signal distribution of channel m/z . If the peak sizes within the individual distributions ($a_{m/z}$ and $b_{m/z}$) are separated, the following generalizations for B_1 and B_2 can be made, based on the previous equations:

$$\begin{aligned} B_1 &= p_{43} a_{43} + p_{55} a_{55} + p_{57} a_{57} \\ B_2 &= p_{43} b_{43} + p_{55} b_{55} + p_{57} b_{57} \end{aligned} \quad (2.4.3.2.v)$$

Note that changes to the modal diameter or indeed the shape of the modelled distributions (f^1 and f^2) over time will not affect this approach, providing that any change is reflected equally in all m/z channels.

The assumption that the weighting factors do not change within a sampling period seems to hold reasonably well, based on the quality of the fit in figure 2.4.3.2.a. However, this is based on the assumption that any changes in the chemical makeup of the organic fraction can be accounted for in three degrees of freedom as three m/z channels are used to model the total. While this assumption will not strictly be true, this approach should be more than adequate as a first order approximation providing that the model described by equation 2.4.3.2.i holds and the chemistry within each mode does not vary significantly with time. These leave the sizes of the two modes as the only independent variables and therefore at least two degrees of freedom are necessary. Also, there is no reason why more organic channels cannot be used for this method, providing that their signal to noise ratios are high enough. This will not be a problem for heavily polluted environments or well-optimised instruments.

An example of the final weighted summation of the signal distributions, after normalizing to the total organic loading, is given in figure 2.4.3.2.b. Also shown are the fractional contributions from the different m/z channels. All the size-resolved data presented in this thesis have been calculated using this method, with new weighting factors being calculated for each campaign. The values used are presented in appendix 5.3.



2.4.3.3. TOF error analysis

There will be noise on the signal due to the background ion signal in any particular m/z channel, which is subject to the random nature of the ion detection as discussed in section 2.4.2.6. The noise can be evaluated by calculating the standard deviation of the signal data points within the DC regions in time of flight space. Note that in some m/z channels such as 16, which have a contribution from both a gas-phase signal ($^{16}\text{O}^+$ in this case) and a particle signal (e.g. NH_2^+ from ammonium), the front DC region is contaminated by the gas-signal and is excluded from the analysis.

It must be pointed out that this method of error analysis does not robustly calculate uncertainties of the data points where there is mass present in the distribution, as these data points will have additional noise due to the ion counting statistics of the particulate signals, as discussed in the MS mode error analysis section. However, it is very useful for calculating detection limits for TOF data. The detection limits can also be transformed into particle diameter space along with the rest of the data, to help validate the processed data, as was shown in figure 2.4.3.1.b.

When presenting size resolved data for a given chemical species as a function of time, it can be displayed as an image, where the colour represents the $dM/d\log D_{va}$ values, with violet equalling zero, going through the rainbow to red equalling the highest values. This scheme is used for graphs featured in part 3 of this thesis. Those points where the value is below the 1σ detection limit are coloured white. Significant negative values should not be ignored, as these can be indicative of systematic errors in the data, so negative data points less than -3σ are coloured black, with those between -3σ and -1σ shades of grey according to their value. If the noise on the data is purely random, which should be the case, the blacks and greys should be randomly distributed. However, if there is a systematic error such as a baseline problem or electronic interference, this will manifest as large areas of black.

2.4.3.4. Broadening effects

2.4.3.4.1. Chopper broadening

There are intrinsic limitations to the methods used for collecting the time of flight distributions that cause the trace acquired to be a more broadened version of the 'true' distribution [Jimenez, In preparation]. The first is an unavoidable uncertainty in the 'start' time of a particle's time of flight, caused by the fact that the chopper wheel's slits are of a finite width. This effect can be mitigated at the measurement stage by

reducing the temporal width of the chopper slit. This can be achieved by reducing the chopper's duty cycle (as discussed in section 2.2.1.3.), but this reduces the amount of material delivered to the detection region. The chopper frequency can also be increased, but this reduces the time spent on a time of flight scan, so it may cause problems with larger particles interfering with the second DC region.

Mathematically, the measured distribution in time of flight space can be thought of as the actual distribution convolved with a top hat distribution of the same temporal width as the chopper and of unit area. In effect, it is akin to boxcar smoothing in time of flight space. In theory, it may be possible to recover the original distribution.

A straight deconvolution method has been attempted, but the intrinsic random noise causes instabilities in the operation and the inverted distribution tends to contain artefacts and increased random noise (J.L. Jimenez, U. Colorado at Boulder, personal communication, 2001). Also, an attempt was made at UMIST to adapt and use a smooth-Twomey algorithm [Markowski, 1987] developed by C. Brock at the NOAA Aeronomy Laboratory for the inversion of size-resolved CPC data (personal communication, 2003). Variations of this method have been used by many people to invert cascade impactor data with some success [Winklmayr *et al.*, 1990], but again, the intrinsic noise of the AMS caused problems and artefacts. The smooth-Twomey algorithm can eliminate these by smoothing the data during iterations, but unfortunately, the smoothing itself causes broadening of the order of that which was trying to be removed, so the idea was abandoned.

Ultimately, the effect of the chopper broadening isn't as critical as originally thought, as the improvements to the instrument performance have allowed a reduction in the chopper duty cycle, meaning that the effect on the results is much less pronounced than when the instrument was first used. In some of the more polluted environments, 1 % chopper wheels can now be used with some of the newer instruments.

2.4.3.4.2. Single particle broadening

Single particle broadening is mathematically more complex and a much less predictable form of broadening. This is due to the fact that the particles take a finite time to completely vaporise when they hit the heated surface. For particles of less than a few hundred nm in diameter, this is not a problem as they often completely vaporise faster than the data acquisition software can discriminate, so appear as a single point

spike to the logging software. However, larger particles take longer, due to their higher heat capacity and the larger volume of material that has to be ejected from the surface. This causes the spike seen by the logging software to be wider, with a decaying ‘tail’, which means some mass from a particle is detected after its initial impact, so the averaged trace acquired is drawn out towards the high end of the distribution. During operation, this effect can be mitigated by choosing a heater temperature high enough to quickly vaporise the particles that are being sampled.

This is a difficult effect to model or predict, as it is dependent on the particle size, volatility and the heater temperature. However, a method has been developed as part of this work to eliminate this effect from the distributions. During operation, the instrument is set to detect single particles as pulses in the multiplier output during TOF mode as described in section 2.1.3. There are also inevitably some false counts due to random noise.

The method uses the single particle data, integrating the pulses and recording their areas, the m/z channels they appeared in and the precise time at which they were recorded with reference to the chopper. The single particle events are then converted to Dirac delta functions and then added together to form a new time of flight distribution. False counts are not a problem, as if they were created by random noise, they have minimal areas so do not affect the total distribution. To account for the small particles that are not counted, the new distribution is added to the distribution recorded by the logging software averaged from TOF scans where no particles were detected.

An example of the effect this inversion has on the raw data is shown in figure 2.4.3.4.2.a., which uses data taken from ACE-Asia. The results are encouraging, showing that the low diameter end of the distribution is largely unaffected and the high diameter end is less broadened. However, as can be seen, the reconstructed data only differs slightly from the original, so this method is not routinely applied to ambient datasets and is not used for any of the results presented in this thesis. However, it may yet still prove useful for future applications where particularly large or relatively involatile particles are being studied or when accuracy is critical.

