



This discussion paper is/has been under review for the journal Atmospheric Measurement Techniques (AMT). Please refer to the corresponding final paper in AMT if available.

# Peak fitting and integration uncertainties for the Aerodyne Aerosol Mass Spectrometer

J. C. Corbin<sup>1,\*</sup>, A. Othman<sup>2,\*\*</sup>, J. D. Haskins<sup>3</sup>, J. D. Allan<sup>4,5</sup>, B. Sierau<sup>1</sup>,  
D. R. Worsnop<sup>6</sup>, U. Lohmann<sup>1</sup>, and A. A. Mensah<sup>1</sup>

<sup>1</sup>ETH Zurich, Institute for Atmospheric and Climate Science, Zurich, Switzerland

<sup>2</sup>University of Zurich, Institute for Clinical Chemistry, Zurich, Switzerland

<sup>3</sup>Massachusetts Institute of Technology, Boston, Massachusetts, USA

<sup>4</sup>National Centre for Atmospheric Science, University of Manchester, Manchester, UK

<sup>5</sup>School of Earth, Atmospheric and Environmental Sciences, University of Manchester, Manchester, UK

<sup>6</sup>Aerodyne Research Inc., Billerica, Massachusetts, USA

\* now at: Laboratory for Atmospheric Chemistry, Paul Scherrer Institute, Villigen, Switzerland

\*\* now at: Institute of Experimental and Clinical Pharmacology and Toxicology, University of Lübeck, Lübeck, Germany

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Received: 10 February 2015 – Accepted: 19 February 2015 – Published: 2 April 2015

Correspondence to: J. C. Corbin (joel.corbin@psi.ch)

Published by Copernicus Publications on behalf of the European Geosciences Union.

**AMTD**

8, 3471–3523, 2015

---

**AMS peak-integration errors**

J. C. Corbin et al.

---

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## Abstract

The errors inherent in the fitting and integration of the pseudo-Gaussian ion peaks in Aerodyne High-Resolution Aerosol Mass Spectrometers (HR-AMS's) have not been previously addressed as a source of imprecision for these instruments. This manuscript evaluates the significance of these uncertainties and proposes a method for their estimation in routine data analysis.

Peak-fitting uncertainties, the most complex source of integration uncertainties, are found to be dominated by errors in  $m/z$  calibration. These calibration errors comprise significant amounts of both imprecision and bias, and vary in magnitude from ion to ion. The magnitude of these  $m/z$  calibration errors is estimated for an exemplary data set, and used to construct a Monte Carlo model which reproduced well the observed trends in fits to the real data.

The empirically-constrained model is used to show that the imprecision in the fitted height of isolated peaks scales linearly with the peak height (i.e., as  $n^1$ ), thus contributing a constant-relative-imprecision term to the overall uncertainty. This constant relative imprecision term dominates the Poisson counting imprecision term (which scales as  $n^{0.5}$ ) at high signals. The previous HR-AMS uncertainty model therefore underestimates the overall fitting imprecision.

The constant relative imprecision in fitted peak height for isolated peaks in the exemplary data set was estimated as  $\sim 4\%$  and the overall peak-integration imprecision was approximately 5%. We illustrate the importance of this constant relative imprecision term by performing Positive Matrix Factorization (PMF) on a synthetic HR-AMS data set with and without its inclusion. Finally, the ability of an empirically-constrained Monte Carlo approach to estimate the fitting imprecision for an arbitrary number of known overlapping peaks is demonstrated. Software is available upon request to estimate these error terms in new data sets.

AMTD

8, 3471–3523, 2015

## AMS peak-integration errors

J. C. Corbin et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## 1 Introduction

The Aerodyne High-Resolution Aerosol Mass Spectrometer (AMS; Jayne et al., 2000; DeCarlo et al., 2006) can provide continuous, time- and size-resolved measurements of particulate-matter (PM) composition (Canagaratna et al., 2007; Kimmel et al., 2011).

5 In its different configurations, the AMS has been used to investigate the composition and evolution of organic PM (Aiken et al., 2007; Jimenez et al., 2009; Zhang et al., 2011), the internal mixing state and composition of black carbon (Cappa et al., 2012; Corbin et al., 2014), and the external mixing state of atmospheric aerosols (Freutel et al., 2013; Lee et al., 2014) among many other applications (Canagaratna et al., 2007).

10 High-resolution AMS signals are routinely quantified with the free, open-source “PIKA” software (Sueper et al., 2011), written in Igor Pro (WaveMetrics, OR USA). PIKA allows the user to perform both basic data processing and complex analyses, including high-resolution peak analysis (DeCarlo et al., 2006), mass quantification (Allan et al., 2004), and elemental analysis (Aiken et al., 2007).

15 PIKA is widely used to prepare AMS data for least-squares statistical models such as Positive Matrix Factorization (PMF) (Paatero and Tapper, 1994; Ulbrich et al., 2009; Zhang et al., 2011), or simple linear regression (e.g. Lee et al., 2010). These statistical models are solved by minimizing the residuals between uncertainty-weighted data and the model, so both the data and their uncertainties must be quantified in order to obtain meaningful results.

20 The current version of PIKA (1.10C) estimates AMS uncertainties from the square root of the number of estimated ion counts (DeCarlo et al., 2006), following techniques developed for unit-mass-resolution (UMR) versions of the AMS (Allan et al., 2003). These UMR AMS’s estimate the signals at a given integer  $m/z$  by summing all ion counts near that signal. In contrast, PIKA estimates the signals of a given ion by fitting a Gaussian-like (pseudo-Gaussian) function to background-subtracted data and integrating the fitted peak. This additional complexity introduces additional uncertainties,

AMTD

8, 3471–3523, 2015

### AMS peak-integration errors

J. C. Corbin et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



but is necessary and useful because HR-AMS mass spectra are frequently comprised of incompletely-resolved (overlapping) isobaric peaks, each represented by relatively few data points. The significance and magnitude of peak-integration uncertainties in PIKA-analyzed AMS data has not been previously addressed.

This manuscript addresses peak-integration uncertainties in PIKA by using a test data set to explore and understand the origins of peak-integration uncertainties in PIKA, using methods that are intended to be applicable to any AMS or other mass spectrometer. The results of this empirical analysis are then used to construct a Monte Carlo model of the PIKA peak-fitting procedure, which is then used to estimate the magnitude of peak-integration uncertainties for isolated peaks and for overlapping peaks. This empirically-based approach allows several assumptions behind uncertainty estimation to be directly evaluated, and may be easily repeated for any new data set.

The manuscript is structured as follows. First, Sect. 2 outlines the sources of uncertainty in HR-AMS analysis and describes the details of a PIKA analysis. Section 3 then addresses peak-integration errors for isolated peaks, beginning with an empirical analysis and ending with the Monte Carlo model mentioned above. Section 4 extends the Monte Carlo approach to the case of overlapping peaks. The resulting overall AMS uncertainty is discussed in Sect. 5. Finally, Sect. 6 places our results in the context of previous work, discusses limitations, and demonstrates the importance of the proposed error model during PMF analysis.

Throughout this manuscript, the terms “imprecision” and “bias” are used when referring respectively to random and independent errors (averaging to zero), and to errors of constant value. The distinction between these two concepts varies naturally at different stages of the analysis: if each PIKA peak integration is biased by an independent and varying amount, then an imprecision (relevant to PMF) may be introduced into the entire data set.

A number of mathematical symbols and abbreviations are used throughout; a list is provided in the Appendix. The Appendix also provides technical details of the test data set discussed below.

## AMS peak-integration errors

J. C. Corbin et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## 2 Background

### 2.1 Conceptual basis

Four conceptual categories contributing to HR-AMS uncertainties can be defined: interpretation, counting, instrumental, and analysis uncertainties. These uncertainties may be defined and addressed as follows.

1. Interpretation uncertainties arise when a given signal may arise from sources other than the analyte. For example, in the HR-AMS,  $\text{H}_2\text{O}^+$  ions may form when either gas-phase water, particle-bound water, or the thermal decomposition of oxygenated functional groups in organic PM (OM) (Aiken et al., 2008), inorganic PM (Chen et al., 2011), or BC (Corbin et al., 2015b). The interpretation of an  $\text{H}_2\text{O}^+$  ion as originating from any one of these sources therefore requires additional information or assumptions to be made. For example, a fraction of the overall  $\text{H}_2\text{O}^+$  signal may be attributed to OM according to laboratory-measured fragmentation patterns (Allan et al., 2004) as detailed by Chen et al. (2011). This so-called fragmentation-table approach may introduce biases but not imprecision to the data set.

Interpretation uncertainties may also occur if an ion is misidentified or omitted in the PIKA software. Such errors are not considered in this work.

2. Counting uncertainties estimate the degree to which a count of  $n$  ions would vary if that count were repeated for the same system and the same time period (Taylor, 1997). They are therefore a measure of imprecision. Poisson uncertainties are included in the standard UMR and HR-AMS error models and are determined by translating mass-spectral peak areas to ion signal rates and applying the Poisson distribution to determine  $\sigma_n$  as  $\sqrt{n}$  for a count of  $n$  ions over a specific time period (Allan et al., 2003; DeCarlo et al., 2006).

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



**AMS peak-integration errors**

J. C. Corbin et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



3. Instrumental uncertainties may arise due to electronic noise or to changes in the performance of various instrumental components. In the latter case, which for example may reflect changes in detector sensitivity or long-term performance fluctuations, the significance of such variations can be evaluated by Allan-variance analysis (Allan, 1966; Werle et al., 1993; Ng et al., 2011; Onasch et al., 2012). Whereas the Allan variance decreases with increasing averaging time for an ideal mass spectrometer (where only ion-counting uncertainties exist), it eventually increases with longer averaging times in real instruments. This increase reflects the introduction of additional error terms from slowly-varying contributors such as electronics temperature or detector stability (Allan, 1966; Werle et al., 1993). Conversely, the minimum in a plot of Allan variance against averaging time indicates the maximum timescale over which instrumental uncertainties can be considered negligible.

For an SP-AMS, Onasch et al. (2012) have shown that  $\sigma_{\text{counting}}$  dominates instrumental uncertainties for averaging times below 100 s in a filtered-air sample. Since this is longer than typical AMS or SP-AMS averaging times, instrumental uncertainties should therefore be small relative to counting uncertainties. However, this may not be true for instrumental uncertainties that are dependent on signal loadings, such those due to background measurements or detector saturation.

4. Finally, analysis uncertainties reflect the confidence with which mass-spectral peak areas can be determined. These uncertainties may comprise both biases and imprecisions. The remainder of this manuscript focusses on these uncertainties in PIKA, with an emphasis on the peak-integration imprecision which is most relevant to PMF (and other least-squares-minimization techniques). The next section describes the PIKA fitting procedure to provide a basis for this discussion.

## 2.2 Constrained peak fitting in PIKA

In PIKA (up to the current version, 1.11C), the signal intensity  $y$  due to a specific ion is considered to arise from a peak  $\phi$ , together with some noise  $\varepsilon$ , for each measured ion-time-of-flight  $x$  (corresponding to different  $m/z$ ):

$$y(x) = \phi(x) + \varepsilon. \quad (1)$$

To a first approximation, the peak  $\phi$  has a Gaussian shape, and may be modelled by

$$G = h \cdot \exp \left[ \frac{-(x - \mu)^2}{w^2} \right], \quad (2)$$

with  $h$  the peak amplitude (peak height),  $\mu$  the mode and mean of the Gaussian (peak location), and  $w$  the SD (peak width). These parameters may be estimated by least-squares fitting to the data.

In practice, the Gaussian model is modified to account for peak broadening, skewness, tailing, or other instrumental non-idealities (DeCarlo et al., 2006) by defining a peak-shape function  $v$ . The  $v$  is determined empirically by averaging a large number of normalized, isolated peaks from a given data set, and may be visualized as a one-dimensional vector of correction factors in normalized  $x$  space ( $x_{\text{norm.}} = [x - \mu]/w$ ). Including  $v$  in Eq. (2) defines the pseudo-Gaussian function (DeCarlo et al., 2006),

$$f = v \cdot h \cdot \exp \left[ \frac{-(x - \mu)^2}{w^2} \right]$$
$$f = v \cdot G \quad (3)$$

To improve the robustness of the PIKA fitting routine against poorly-resolved peaks and noisy data, some of the parameters in Eq. (3) are constrained during peak fitting. The  $v$  is determined as described above,  $\mu$  is obtained from an  $m/z$  calibration and  $w$  from a peak-width parameterization,  $w(m/z)$ . These are obtained as follows.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## AMS peak-integration errors

J. C. Corbin et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



The  $m/z$  calibration is achieved by fitting Eq. (2) to user-selected background peaks. The fitted peak locations  $\mu$  in time of flight space are calibrated to their known  $m/z$  following the procedure described by DeCarlo et al. (2006). Separate  $m/z$  calibrations are obtained for every mass spectrum. In practice, these calibrations are applied to predict peak locations in ion time-of-flight space before fitting the raw data. We therefore refer below to “ $\mu$  prediction errors” rather than “ $m/z$  calibration errors”, except when specifically discussing  $m/z$  calibrations. Although all peaks were fitted in ion-time-of-flight space below, the discussion refers to peak  $m/z$  for clarity.

The  $w(m/z)$  calibration is also obtained by fitting Eq. (2), but to a separate set of peaks. This set of peaks is carefully selected to ensure that only single, isolated peaks of unambiguous composition are considered. The fitted peak widths  $w$  are averaged before being parameterized by a (typically linear) calibration curve, as described by DeCarlo et al. (2006). In the present analysis, the  $w(m/z)$  calibration procedure was slightly modified to improve its robustness against poor-quality peak fits, as described in detail in Corbin et al. (2014) and briefly in the Appendix.

Thus, the three inputs to Eq. (3) are defined from three different calibrations on the data. These calibrations are fundamentally different: the  $w(m/z)$  and  $v$  calibrations are determined once for the entire data set, while the  $m/z$  calibration determined for each individual mass spectrum. Variability in the accuracy of the  $m/z$  calibration may therefore lead to variability in the analysis. Also, different peaks are used for all three calibrations, so that their prediction errors may also differ.

With the inputs described above, Eq. (3) is fitted to estimate  $h$ . An example is shown in Fig. 1, showing the  $\text{CO}_2^+$  signal at  $m/z$  44 in the lower panel and the fit residuals in the upper panel. The peak spans only 2–3 detector bins (corresponding to 4–6 ns), which was typical for  $m/z$  in the range 15–75 for this specific instrument. Other studies have reported slightly-lower  $m/z$  resolutions, and correspondingly a slightly-higher number of detector bins representing each peak (e.g., DeCarlo et al., 2006; Sun et al., 2010; Ortega et al., 2013).

The fitted  $h$  is used by PIKA to estimate the area of each peak via the Gaussian integral,  $A(G) = hw\sqrt{\pi}$ , which leads to (Sueper et al., 2011)

$$A = hw\sqrt{\pi} \cdot \frac{A_{f_0}}{A_{G_0}} \cdot k_{DC} \quad (4)$$

where  $A_{G_0}$  and  $A_{f_0}$  are the respective integrals of the standard Gaussian and pseudo-Gaussian distributions. These standard integrals are defined by Eqs. (2) and (3) with unit  $h$  and  $w$ , and  $\mu = 0$ . The factor  $k_{DC}$  is a correction for the mass-spectrometer duty cycle, to account for the fact that lighter ions will be transported faster than slower ions into the mass spectrometer prior to time-of-flight measurement and consequently, reside in the extraction region for shorter times (Hings, 2006). The ratio  $A_{f_0}/A_{G_0}$  is typically close to unity.

After estimating  $A$ , PIKA currently estimates the uncertainty in  $A$  from the square root of the number of ions associated with the signal, according to an assumed Poisson distribution of errors (Allan et al., 2003).

In the AMS, more than one peak is typically observed at each integer  $m/z$ . These peaks are modelled as a linear superposition  $f_T$  of pseudo-Gaussians with similar  $w$  (Sueper et al., 2011),

$$f_T = \sum_{i=1}^j h_i f_{0,i}(\mu_i, w) \quad (5)$$

The uncertainty of fits to Eq. (5) with  $j > 1$  (multiple overlapping peaks) is much more complex than the case where  $j = 1$  (single, isolated peaks).

The next section therefore focusses on the case of isolated peaks, before Sect. 4 discusses the overlapping-peak case.

AMS peak-integration errors

J. C. Corbin et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



### 3 Single-peak fitting errors

#### 3.1 Integration uncertainty

The integration of fits to isolated pseudo-Gaussian peaks via Eq. (4) will lead to an uncertainty

$$\left(\frac{\sigma_A}{A}\right)^2 = \left(\frac{\sigma_h}{h}\right)^2 + \left(\frac{\sigma_w}{w}\right)^2, \quad (6)$$

which, if the fractions  $\sigma_h/h$  and  $\sigma_w/w$  are sufficiently large, may contribute significantly to the overall AMS uncertainty.

In Eq. (6), the uncertainties in  $A_{f_0}$  and  $k_{DC}$  are considered negligible, since the integral  $A_{G_0}$  is known exactly; the integral  $A_{f_0}$  is also known exactly for a given  $\nu$ ; and the uncertainty in  $k_{DC}$  depends only on the stability of the mass-spectrometer voltages, which fluctuate negligibly for the time periods over which data are collected (Sect. 2.1). In addition,  $h$  and  $w$  have negligible covariance as they are determined via separate and independent calibrations (Sect. 2.2); the  $w(m/z)$  calibration requires, and can be used to validate, the assumption that  $w$  is independent of  $h$ .

The value of  $\sigma_w$  may be directly estimated from the peak-width calibration procedure via the uncertainties in the calibration-fit coefficients. For the test data set discussed below, this approach indicated  $\sigma_w/w \approx 2.5\%$ . Such a percentage uncertainty is important as it scales with signal differently to the typically-used Poisson uncertainty, increasing linearly with the number of ion counts rather than as the square root.

The estimation of  $\sigma_h$  is much more complex than  $\sigma_w$ . Whereas  $\sigma_w$  may be estimated directly from the calibration fit,  $\sigma_h$  reflects the imprecision of the constrained PIKA fitting procedure (Sect. 2.2), the constraints of which are predicted by two separate calibrations (peak width and location) as well as an empirically-defined peak shape  $\nu$ .

## AMS peak-integration errors

J. C. Corbin et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## 3.2 Fitting errors

Since the width, location, and shape of the PIKA pseudo-Gaussian fit function (Eq. 3) are predefined by different calibrations, the exponential term of Eq. (3) reduces to a transformation of the  $x$  variable:

$$f = hf_0(x, w, v, \mu) \quad (7)$$

where  $f_0$  is defined by Eq. (3) with  $h = 1$  and with  $\mu$ ,  $w$ , and  $v$  given by the calibrations described in Sect. 2.2.

A fit to Eq. (7) is equivalent to a simple linear regression. An uncertainty estimate for  $h$  could thus be obtained directly from the regression, but this estimate would not account for bias in the fit. Such bias must be accounted for, because a variable bias in Eq. (7) produces a variable error in  $h$  – that is, an imprecision in  $h$ . Moreover, a variable bias in Eq. (7) is to be expected because a new  $f_0$  is defined for each fitted ion. Variations in the  $m/z$  calibrations which are obtained for each mass spectrum, for example, may cause such variations in  $f_0$ .

The influence of such variations on the imprecision of the fitted  $h$  values cannot be estimated directly from a single fit, but must be considered in the context of the whole data set. It can be expected, however, that errors in  $h$  due to  $f_0$  will be proportional (scale linearly with  $h$ ) since  $f_0$  is scaled by  $h$  during fitting.

To quantify the error in a given fit of  $f$ , we use the SD of the fit residuals, the fit root-mean-square error (RMSE),

$$\text{RMSE} = \sqrt{\frac{1}{j-1} \sum_{i=1}^j [y(x_i) - f(x_i)]^2} = \sqrt{\frac{\text{SSR}}{j-1}}. \quad (8)$$

For a sufficiently large sample, the expected value of the squared RMSE is the sum of the model variance and the squared model bias (Wilks, 2011). As the RMSE of a single peak is the main quantity of interest below, the range  $1 \leq i \leq j$  was limited to

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



two SDs on either side of the peak centre ( $\mu \pm 2w$ ), representing the trimmed variance of the residuals (e.g. Wilks, 2011).

The RMSE is not used to directly infer the error in the fitted  $h$ , but as a diagnostic with which the causes of errors in  $f_0$  may be elucidated and understood. Based on this understanding, the imprecision in  $h$  is then estimated.

### 3.3 Case study

The RMSE of a number of isolated peaks in the test data set are plotted in Fig. 2. Each point in the figure represents the RMSE of a fit of Eq. (7) to the data. The figure includes seven isolated peaks based on their being both well-resolved from any neighbours and present at a wide range of signal intensities. The mass spectra have been background subtracted before fitting as described in the Appendix. Background subtraction was performed to remove inconsequential differences between the ions, for example due to signals from background CO<sub>2</sub> gas. Other ions had very small background signals regardless.

The peaks in Fig. 2 span a range of different  $m/z$ , as indicated by the integer  $m/z$  values shown in the legend, and represent a range of different species. For example, C<sub>3</sub><sup>+</sup> ions formed when BC vaporizes at close to 4000 K, whereas C<sub>2</sub>H<sub>3</sub>O<sup>+</sup> and C<sub>3</sub>H<sub>7</sub><sup>+</sup> ions formed when organic PM vaporizes below 873 K (600 °C) (Corbin et al., 2014). In contrast to all other ions, these latter two organic ions are found at the same integer  $m/z$  and were therefore fitted simultaneously as a linear sum in PIKA (Eq. 5 with  $j = 2$ ). This point is returned to in Sect. 4.

Two distinct trends are evident in Fig. 2. First, the RMSE approaches an asymptotic value for low peak heights ( $\sim 1$  Hz). This constant RMSE ( $\sigma_y \approx 0.2$  Hz bin<sup>-1</sup>) is representative of noise in the data  $y$ . Second, the RMSE shows a constant slope for high-signal peaks, indicating a constant relative RMSE (RMSE/ $h \approx 3\%$ ). This suggests an importance of fit-function errors, for the reasons given in Sect. 3.2. The constant RMSE and constant relative RMSE are respectively depicted by the dotted and dashed lines in Fig. 2, as well as their quadratic sum.

## AMS peak-integration errors

J. C. Corbin et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



The constant  $\text{RMSE}/h$  for high signals suggests that  $\sigma_h/h$  may also be constant. Thus  $\sigma_h$  may also scale linearly with signal, as was the case for  $\sigma_w/w$  (Sect. 3.1), and may significantly influence the overall peak-integration uncertainty for high signals. The next sections therefore aim to obtain an understanding and quantification of  $\sigma_h/h$  for incorporation into the overall AMS uncertainty.

### 3.4 Impacts of fit-parameter errors

The source of the constant relative RMSE in Fig. 2 was argued via Eq. (7) to be most likely due to errors in the pseudo-Gaussian  $f_0$ . To elucidate the potential causes of these errors, noise or biases were manually added to each of several input parameters in the PIKA analysis procedure as follows.

The original magnitude of each error was estimated directly from the data. Based on these estimates, significantly-larger uncertainties were added to the data, as specified in Table 1. In most cases, the errors specified in the table were used to define the SD of a Gaussian probability distribution from which a new error was sampled for each analyzed peak.

Figure 3 plots the effects of these errors on the resulting RMSE for one exemplary ion,  $\text{C}_3\text{H}_7^+$ . This ion was chosen simply because it was observed over a large range of signals; the behaviour of all ions was similar. The magnitudes of the errors used to generate the figure are highlighted in boldface in Table 1, and were chosen so as to give a visible change in the RMSE graphs (where possible), and not to represent realistic errors.

With this approach, multiple potential sources of the constant relative RMSE term can be eliminated: noise in the predicted peak width  $w$ , errors in the slope of the subtracted linear baseline, and the use of a different AMS vaporizer did not alter the magnitude of the relative RMSE.

Conversely, two potential sources of the constant relative RMSE can be identified: noise in the predicted peak location  $\mu$ , and errors in the peak-shape  $v$ . Figure 3f–h shows a relative RMSE of  $\sim 15\%$ . However, the relative  $\mu$  prediction errors necessary

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## AMS peak-integration errors

J. C. Corbin et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



to achieve this  $\sim 15\%$  relative RMSE were four orders of magnitude smaller than the errors in  $v$  (ppm vs. % scale), suggesting that  $m/z$  calibration uncertainties may play a larger role than peak-shape uncertainties in a real data set. This difference in sensitivity is not surprising, given that  $\mu$  is part of the exponential term in Eq. (3) while  $v$  is not. In the next section, the fitting procedure was altered to prove that the majority of the fitting error was due to  $\mu$  prediction errors.

### 3.5 Impacts of fit constraints

To explore the impact of errors in  $\mu$  prediction on the fit RMSE, the fitting procedure was altered to allow  $\mu$  and/or  $w$  to be varied during fitting. The peak shape  $v$  could not be meaningfully varied.

Allowing  $\mu$  to vary by  $\pm 20$  ppm during fitting reduced the fit RMSE by almost an order of magnitude at high signals (Fig. 4, red triangles), indicating that the majority of the fitting error was due to errors in the predicted  $\mu$  (which result from both imprecisions and biases in  $m/z$  calibration). The RMSE at low signals was virtually unchanged, since it was dominated by noise in the data (Sect. 3.3). While Fig. 4 illustrates this result using  $C_3H_7^+$ , these conclusions were verified for all of the isolated ions discussed above.

Further relaxing the  $\mu$  constraint to  $\pm 50$  ppm did not further reduce the RMSE. Conversely, reducing the constraint to  $\pm 10$  ppm resulted in a change intermediate between the 0 ppm (fully-constrained) and  $\pm 20$  ppm constraints, suggesting that the true accuracy of the  $m/z$  calibration was on the order of  $\pm 15$  ppm. This value does not correspond to the imprecision of the  $m/z$  calibration, as discussed in the next section.

Allowing  $w$  to vary during the fit by  $\pm 5\%$  ( $2\sigma_w$ ) had no observable effect on the RMSE (Fig. 4). Although this result may appear to contradict the fact that a significant uncertainty in the  $w$  calibration was estimated during calibration (Sect. 5), it is fully consistent with the only minor changes in RMSE observed when large errors were added to  $w$  in Sect. 3.4 (Fig. 3b). That is, although uncertainties in  $w$  do not strongly

influence the fitted  $h$ , they lead to uncertainties during the peak integration by Eq. (4) (cf. Eq. 6).

When both  $w$  and  $\mu$  were allowed to vary during fitting, the RMSE behaved similarly to the case where only  $\mu$  was allowed to vary (i.e., the red triangles in Fig. 4). In this case, some linear dependence of the RMSE on the fitted peak height remained, suggesting that  $f_0$  still had some influence on the RMSE. This remaining dependence is attributed to errors in the peak-shape factor  $\nu$ . Since the empirically-defined  $\nu$  could not be allowed to vary during fitting as  $\mu$  and  $w$  were in Fig. 4, its influence on the RMSE trends could not be explored in this context. If the remaining RMSE trend was due to errors in  $\nu$ , these errors had a much-smaller impact on the RMSE (and therefore the goodness of fit of Eq. 3) than the errors in  $\mu$  prediction, so they are not discussed further.

### 3.6 Estimation of $\mu$ prediction ( $m/z$ ) errors

When the predicted peak location  $\mu_{\text{predicted}}$  was allowed to vary during fitting (Sect. 3.5) the relative RMSE was significantly reduced in the high-signal regime. The final fit therefore represented a significantly better model of the data, which is interpreted as an improvement due to a reduced error in the  $\mu$ . The fitted peak location,  $\mu_{\text{fitted}}$ , is therefore interpreted as a good approximation to the true peak location,  $\mu_{\text{true}}$ , and used to estimate errors in  $\mu$  prediction as  $e_{\mu} \approx (\mu_{\text{fitted}} - \mu_{\text{predicted}})$  in the current section. Limitations of this estimation approach are discussed in Sect. 6.

The dark-shaded data in Fig. 5 show the distribution of the estimated  $\mu$  prediction errors for peaks outside of the noise regime, defined as  $h_{\text{fitted}} > 20 \text{ Hz bin}^{-1}$  from Fig. 2. For this analysis,  $\mu$  was allowed to vary by  $\pm 40$  ppm. If the aggregated  $e_{\mu}$  are fitted to a Gaussian function, the mean appears to be virtually zero (fitted value:  $-0.5 \pm 0.8$  ppm), suggesting zero bias in  $\mu_{\text{predicted}}$ .

However, the bias for each individual ion were not zero, as shown by the light-shaded data in the same figure. The  $\mu_{\text{fitted}}$  for each ion showed significant biases (mean different from zero) as well as imprecisions (significant spread relative to the mean). The

## AMS peak-integration errors

J. C. Corbin et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



magnitude of these biases and imprecisions were estimated by a Gaussian fit in each case, with the fitted values shown in the figure. The imprecisions in  $\mu_{\text{predicted}}$  are generally of similar magnitude to the biases, and the biases show both positive and negative values.

The observed  $m/z$  calibration biases are thought to arise from non-idealities in the ion trajectories within the mass spectrometer, and are not unique to this instrument. Similar positive- and negative-valued biases have been reported by DeCarlo et al. (2006), who compared the known  $m/z$  of perfluorinated ions with their observed peak location, and by Müller et al. (2011) for a lower-resolution AMS.

Although it is readily apparent that a bias in  $\mu$  prediction may lead to a bias in the fitted  $h$  for a given ion, it is less apparent that such a bias also affects the imprecision in fitted  $h$ . This is because the slope of a peak increases with increasing distance from its mode (for small distances), such that the impact of a  $\mu$  prediction imprecision depends also on the  $\mu$  prediction bias. With increasing  $\mu$  prediction bias, the imprecision in  $h$  typically increases, as demonstrated below. Biases in  $\mu$  must therefore be considered when estimating the magnitude of  $\sigma_h$ .

### 3.7 Monte-Carlo model

The  $\mu$  prediction errors estimated in Sect. 3.6 were shown, in Sect. 3.5, to be the major cause of fitting errors. On this basis, an empirically-constrained Monte Carlo model of the PIKA fitting routine was constructed, in which peaks of the expected shape and width were simulated and fitted with some small error in  $\mu$  prediction.

The model was constructed by simulating peaks via Eq. (3) using the  $m/z$  axis of a mass spectrum from the test data set. For each simulated peak, a different  $m/z$  axis was randomly selected. The simulated peaks were generated using the peak shape  $v$  and width  $w$  expected for this data set, and the Gaussian noise described in Sect. 3.3 was added to the data. The simulated peaks were then fitted to Eq. (3) with  $\mu$  intentionally constrained to erroneous values. These erroneous  $\mu$  constraints were obtained by sampling from a Gaussian distribution with mean and SD given by the fits in Fig. 5.

## AMS peak-integration errors

J. C. Corbin et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## AMS peak-integration errors

J. C. Corbin et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



The RMSE of these fits to the simulated data are plotted in Fig. 6 for  $C_3H_7^+$ . Each simulated peak height was fitted 100 times, each time with a different sample of  $\mu$  prediction and data-noise errors. Increasing this number of samples did not affect the results. The figure shows the resulting mean (white line) and SD (light-blue shading) of the RMSE, which are in excellent agreement with the data.

It was noted in Sect. 3.5 that some small errors in  $\nu$  were likely present. Although no physical basis for the functional form of such errors was available, the impact of errors in  $\nu$  was investigated by adding arbitrary functions to the true peak shape. This allowed the trends in Fig. 4 for the relaxed- $\mu$  constraint case to be reproduced, but made no visible difference to Fig. 6. Errors in  $\nu$  were therefore omitted from the model.

### 3.8 Imprecision and bias in $h$

The model described above was used to simulate the bias and imprecision when fitting peaks with the erroneous  $\mu$  constraints estimated in Sect. 3.6 for all seven of the isolated peaks discussed above. The simulations were performed as described above, except that only signals much larger than the noise level were represented and a larger number of iterations (5000) were performed for each peak, so that histograms could be plotted in detail.

Figure 7 shows the resulting distribution of fitting errors. The mean fitting error (bias) has been subtracted from each distribution to emphasize the different distributions of errors (imprecisions). The errors are normalized to the absolute peak height, as they were a constant fraction of the height (as expected from Sect. 3).

It might be hypothesized that the different physical mechanisms behind the detection of different ions might influence the imprecisions shown in Fig. 7. However, there is no evidence for such a difference: the imprecisions of  $C_3^+$  (vaporization temperature  $\sim 4000$  K) and  $C_2H_3O^+$  (vaporization temperature  $< 873$  K) are similar.

Each probability distribution in Fig. 7 shows a very different skewness. These differences are largely controlled by the  $\mu$  prediction bias and imprecision, which may cause the predicted peak location to be consistently away from the peak mode. Away from the

mode, the peak itself slopes more steeply and a small  $\mu$  prediction imprecision leads to a large variability in the fitted height. This fact means that a greater imprecision results from a greater bias in the  $m/z$  calibration. The  $\mu$  prediction bias is therefore an important component of the imprecision of the fitted  $h$ .

Repeating the analysis shown in Fig. 7 without including  $\mu$  prediction biases led to significantly different estimates of  $\sigma_h$  for these peaks, as shown in Table 2. Accurately simulation of the peak-fitting imprecision therefore requires an estimate of  $m/z$  calibration biases. Such biases are not unique to the present instrument or analysis (DeCarlo et al., 2006; Müller et al., 2011) but are rather a limitation of the  $m/z$  calibration.

The  $\mu$  prediction biases discussed above also mean that the magnitude of the imprecision depends on the absolute peak width, since broader peaks are less steeply sloped. Broader peaks are also represented by more points in the mass spectrum, which has a major impact on the imprecision: for the case of  $C_3H_7^+$ , increasing the peak width by a factor of 1.5 (which doubled the number of data points representing > 1% of the peak height from 4 to 8) decreased the imprecision from 1.6 to 0.2% (Table 2). This broadening would, however, increase the influence of the peak-overlap errors discussed below. It would therefore be ideal to increase the number of points representing the peak without reducing the mass-spectral resolution, as proposed by Hilmer and Bothner (2011).

This increase in width was performed to represent mass spectrometers with a lower  $m/z$  resolution or higher ion-time-of-flight resolution, and implies that it may be advantageous to operate the present instrument with a lower resolution in future studies, to as to reduce fitting imprecision.

Figure 7 represents only the seven isolated peaks for which the  $\mu$  prediction bias and imprecision could be estimated. These seven peaks present a very limited sample, and are unlikely to adequately represent the following important uncertainty. In particular, during exploratory data simulation, it was observed that the modelled fitting errors were extremely sensitive to two important real-world parameters: (1) the number of data points comprising each peak, which corresponds to the mass-spectral resolution and

## AMS peak-integration errors

J. C. Corbin et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion





## AMS peak-integration errors

J. C. Corbin et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



This section therefore uses the empirically-based model described in Sect. 3.7 to evaluate how fitting errors may be increased in the case of overlap. The generic term “error” is used, rather than the specific terms “imprecision” and “bias”, since the distinction between the two specific terms is much less clear in the multiple-peak case. For example, if only one of two overlapping peaks is of interest and if variability in the intensity of the second peak causes variable fitting errors, then a non-random imprecision is imparted to the first peak.

#### 4.1 Evaluation of a modified fitting procedure

Given that  $\mu$  prediction errors ( $e_{\mu}$ ) were shown above to be the major source of peak-fitting errors in PIKA, and to be biased by different amounts for different ions, the possibility of modifying the PIKA fitting procedure to directly correct for  $e_{\mu}$  at each integer  $m/z$  was evaluated by updating Eq. (5) to

$$f_T = \sum_{i=1}^j h_i f_{0,i}(w, \mu_i \pm \delta\mu), \quad (10)$$

where  $\delta\mu$  defines the  $e_{\mu,i}$  for each peak in a set of overlapping peaks to be equal; that is, nearby peaks are assumed to have the same  $e_{\mu}$ . The alternative, allowing each  $e_{\mu,i}$  to vary separately, would double the number of fit parameters (and thus degrees of freedom) in a model where the number of data constraining each peak is already very low (cf. Fig. 1).

The implicit assumption in Eq. (10) is that all peaks at a given integer  $m/z$  share approximately the same  $m/z$  calibration error. We evaluated this assumption directly using the two ions at  $m/z$  43,  $\text{C}_2\text{H}_3\text{O}^+$  ( $m/z$  43.0184) and  $\text{C}_3\text{H}_7^+$  ( $m/z$  43.0548). These two ions were well-resolved, with a separation of  $\sim 7$  SDs, such that their estimated  $e_{\mu}$  (Sect. 3.6) are considered reliable.

The  $e_\mu$  of these two ions are compared in Fig. 8, where each data point represents different  $e_\mu$  of the same mass spectrum. Only data where both of these ions were fitted with  $h > 20$  are compared in the figure.

Rather than the expected 1 : 1 relationship, a slight negative trend is evident in Fig. 8 ( $r^2 = 0.07$ ). This unexpected result may be related to the sensitivity of the fits to the location of the mode of a peak relative to the nearest detector bin, noted in Sect. 3.8. Fitting the two ions with the same  $\delta\mu$  would have biased the fit towards one ion or the other, in a manner no better (or worse) than the simple  $\mu$  prediction-constrained case. Since allowing different  $e_\mu$  for each peak is infeasible, as noted above, we conclude that Eq. (10) provides an unsuitable solution to the  $m/z$  calibration errors highlighted in Sect. 3.

## 4.2 Multiple-peak modelling

Exploratory simulations were first performed to explore the fitting errors for two simulated overlapping peaks as a function of peak width  $w$ , peak shape  $v$ , fit-constraint errors, noise in the data, and relative peak intensity. All of these parameters were found to influence the shape and magnitude of fitting errors. In particular, the fitting error for overlapping peaks increased nonlinearly as two peaks are brought closer together in  $m/z$  space, as shown in Fig. 9.

Figure 9 shows the errors in  $h$  for two peaks, a larger reference peak (blue lines) and a smaller peak (red). Three different relative heights of the smaller peak are illustrated. The larger peak is kept at a fixed position, and the smaller peak is moved along the  $x$  axis to explore the influence of peak separation on the fitting errors. Ion-counting imprecisions, as addressed by Cubison et al. (2014), are not included for simplicity.

The error in Fig. 9 is the difference between the simulated peak height and the height retrieved by a fit of Eq. (5) with  $i = 2$  to the data, with  $\mu$  fixed at a value +10 ppm from its true value. (Changing this value to +5 ppm, had a negligible influence on the plot. Changing it to -10 ppm had the effect of mirroring the plot about the line  $w = 0$ .) Note that, in contrast to all other simulations, the number of data points representing each

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



peak in these simulations was chosen to be arbitrarily high, to avoid the real-world effects described in Sect. 3.8. Using a real mass spectrum resulted in minor periodic fluctuations in the plot, as the distance from the peak mode to the nearest bin changed.

As shown in the top-right corner of Fig. 9, the error for peaks with relative heights of 1 : 1 (solid lines) begins to increase at about 1 SD of separation. For relative heights of 1 : 10, this increase begins at roughly 2 SDs, and similarly for the 1 : 100 case. When the two peaks are well-separated, the interaction of the two peaks becomes negligible. This is consistent with the observation in the test data set that  $C_2H_3O^+$  and  $C_3H_7^+$  at  $m/z$  43, separated by  $\sim 7$  SDs, displayed the same trends as truly-isolated peaks in Fig. 2.

As the two peaks become very close, the error increases rapidly, because the signal of one peak is assigned to the other during fitting. For example, the very-high error of the larger peak for slightly-negative separations is a result of the fitting algorithm assigning all of the larger-peak signal to the smaller peak. In a real data set, such an error may be identifiable to the analyst by considering the mass spectrum as a whole, or by observing that the fitted height varies randomly between positive and zero values if the  $\mu$  prediction error varies about zero (due to imprecision). However, if the  $\mu$  prediction bias is greater than its imprecision, as may be the case (Fig. 5), such an identification would not be possible. This special case is beyond the scope of this work.

### 4.3 Case study

An exhaustive description of the dependence of fitting errors on the five parameters listed above would be impractically complex. However, such a description is not necessary, because many of these parameters are sufficiently-constrained for the corresponding fitting errors to be estimated based on a modelling approach similar to that described in Sect. 3.

With the assumption that the peaks  $i$  have been correctly identified based on their exact  $m/z$  and/or on mass-spectral interpretation, the separation between two peaks can normally be well-constrained. For example, the width of  $C_3H_7^+$  in this data set was

## AMS peak-integration errors

J. C. Corbin et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



137 ppm, an order of magnitude larger than its estimated  $\mu$  prediction errors ( $-10 \pm 7.5$ , Fig. 5). The peak width and shape are similarly well-constrained (Sect. 3.4). In general,  $\mu$  prediction errors are the least-constrained variable, but a reasonable estimate of their magnitude can be made. Finally, the relative peak heights can be estimated by fitting Eq. (5) to the data.

As a detailed example, Fig. 10 shows the fitting errors for overlapping  $C_5^+$  and  $C_2H_4O_2^+$  peaks across a range of relative peak heights. Data noise was omitted from the simulation for simplicity and generality, although it would be important in a real case.

Figure 10 shows that the height of the larger peak may be generally considered as accurately estimated, as also concluded from Fig. 9. The uncertainty in this peak may be estimated similarly to that of an isolated peak.

The figure also shows that  $C_2H_4O_2^+$  must be present with a signal of at least  $\sim 1/300$  relative to  $C_5^+$  (in the absence of noise), if it is to be fitted with an uncertainty better than 50%. If  $C_2H_4O_2^+$  is present with a relative signal between  $1/300$  and 1, its fitted value can be used to estimate its relative height and therefore its fitting error. In this case, the critical relative height of  $\sim 1/300$  was estimated for a specific  $\mu$  prediction error, which would not be known in a real case. However, as the true  $\mu$  prediction error of any one peak cannot be known a priori, it is difficult to improve this estimate.

Thus, in practice, a meaningful (though not ideal) estimate of the peak-fitting error for overlapping peaks may be obtained by simply fitting the initial signals in  $m/z$  space, using the fit results together with the independently-estimated uncertainties described above, and estimating the fitting error by fits to synthetic peaks simulated with realistic  $\mu$  prediction errors.

## 5 Overall AMS uncertainty

With  $\sigma_w/w$  determined as  $\sim 2.5\%$  (Sect. 3.1) and  $\sigma_h/h$  shown to be constant for a given peak and estimated at  $\sim 4\%$  (Sect. 3.8), the overall imprecision in the peak

## AMS peak-integration errors

J. C. Corbin et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



integration discussed in Sect. 3.1 can be estimated from an equation similar to Eq. (6) as

$$\left(\frac{\sigma_A}{A}\right)^2 = \left(b\frac{\sigma_h}{h}\right)^2 + \left(\frac{\sigma_w}{w}\right)^2 \quad (11)$$

where the factor  $b$  has been added to represent the increase in  $\sigma_h/h$  due to overlap, which may be estimated as described in Sect. 4. For isolated peaks,  $b = 1$ . As estimated above,  $\sigma_h/h \approx 4\%$ ,  $\sigma_w/w \approx 2.5\%$  for the present data set, so,

$$\sigma_A \approx (4.7\%)A \quad (12)$$

for isolated peaks in this data set.

$\sigma_A$  is independent of the Poisson counting uncertainties  $\sigma_p$  described in Sect. 2.1, so that the overall AMS uncertainty may be expressed in units of ion counts per detector bin as

$$\sigma_{\text{AMS}} = \sqrt{\sigma_A^2 + \sigma_p^2} \quad (13)$$

Since the first term in Eq. (13) scales linearly with signal, whereas the second term scales with the square-root of signal, the first term dominates at high signals. This is shown in Fig. 11a, which has been obtained by programming Eq. (13) into PIKA using the value for  $\sigma_A$  given above. As is standard in PIKA, the calculations have accounted for additional details related to sampling time, baseline noise, and mass-spectral duty cycle (Allan et al., 2003; Sueper et al., 2011).

The essential difference between fitting and counting errors is shown by Fig. 11b. Whereas the relative uncertainty of Poisson errors ( $\sim \sqrt{n}$ ) falls to zero as the signal increases, the relative uncertainty of fitting errors ( $\sim n$ ) tends to an asymptotic value as signal increases. This feature strongly affects the relative importance of high-signal data during uncertainty-weighted fitting. This importance is even greater for peaks affected by overlap, for which fitting errors may increase rapidly and nonlinearly.

AMS peak-integration errors

J. C. Corbin et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## 6 Discussion

### 6.1 Context and previous work

The results presented above represent a case study of a data set from a single mass spectrometer. The absolute values of the estimated imprecisions and biases are therefore likely to change for other instruments, and should be determined on a case-by-case basis. The Igor Pro computer code which has been used in the current work is available from the authors. In general, the trends and behaviours displayed above are expected to be robust.

The largest uncertainty in this analysis is due to the estimation of the biases and imprecisions in  $\mu$  prediction (which reflect errors in  $m/z$  calibration) directly from the data (Sect. 3.6). We therefore emphasize that the magnitude of our estimated  $\mu$  prediction imprecisions are consistent with those of DeCarlo et al. (2006) and, similarly, that significant  $\mu$  prediction biases have also been reported by different workers and for different instruments (DeCarlo et al., 2006; Müller et al., 2011). As these  $\mu$  prediction biases may significantly increase the peak-fitting imprecision, they must be included in estimates of the peak-fitting imprecision, as discussed above. These biases, as well as the importance of the detector-bin temporal spacing noted above (and further discussed by Hilmer and Bothner, 2011) are also likely to influence the conclusions of numerical studies, such as the recent work by Cubison et al. (2014).

Cubison et al. (2014) also proposed a parameterization to estimate the imprecision for systems of two overlapping peaks. We suggest that our proposed may provide more accurate estimates of imprecision, as it is not limited to two overlapping peaks and directly accounts for the influence of the detector-bin temporal spacing in the data under consideration.

## AMS peak-integration errors

J. C. Corbin et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## 6.2 Fitting procedure

The above conclusions that the major causes of fitting errors in PIKA are errors in peak-location prediction raises the question of whether the fit procedure itself might be improved. One attempt was made in Sect. 4.1, however, a more sophisticated approach would be desirable.

A more sophisticated approach might allow a priori knowledge of calibration uncertainties to be incorporated into the fitting procedure, following Bayesian theory (Gelman et al., 2013). In such an approach, realistic uncertainty distributions could be directly applied to the constraint of  $\mu$ , and translated directly into uncertainty distributions for the resulting  $h$ . The development of such an approach is beyond the scope of the present work, in part because the Igor Pro software in which PIKA has been developed over the past ten years does not provide the necessary framework. We note that although the relaxation of the  $\mu$  constraint in Sects. 3.5 and 3.6 allowed for uncertainties in  $\mu$  to be accounted for, the implicit probability distribution assigned to these uncertainties was an unphysical, uniform distribution, with unrealistic, discontinuous edges. For the typical case where multiple overlapping peaks are fitted, this approach is not considered robust.

In addition to an improved fitting procedure, an improved calibration procedure would be an obvious recommendation for reducing  $\mu$  prediction errors. An accurate calibration requires both a consistent calibrant signal and a consistent peak-shape. These requirements were not met by any of the mass-spectral peaks: the only consistent signals were those of background ions from the gas phase, which were relatively few and displayed significantly-different peak shapes than particulate signals owing to their originating from generally different regions of the AMS ionization chamber. An improved calibration might make use of an internal standard, for example a polyfluorinated organic (DeCarlo et al., 2006), to be fitted by the same pseudo-Gaussian function used for data analysis (Eq. 3). However, at some point the limited temporal resolution of the

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



ion detector (Hilmer and Bothner, 2011) and the variability of ion flight paths also limit the accuracy of the  $m/z$  calibration.

### 6.3 Other sources of uncertainty

While only peak-integration and ion-counting uncertainties were addressed in the discussion above, a number of other AMS-specific uncertainties can be identified.

Craven et al. (2012) discussed and addressed a number of AMS uncertainties with regard to background signals and overlapping ions. Craven et al. (2012) also developed a data-smoothing approach to evaluate data quality for PMF, rejecting noisy variables. Such an approach may remain useful for the very-slowly-changing systems studied by Craven et al. (2012), as it may account for imprecisions other than those caused by peak integration, however, it did not result in improved PMF uncertainty estimates but rather the selective rejection of low signal-to-noise variables. More generally, smoothing approaches cannot be applied to the rapidly-changing signals observed during roadside studies, aircraft studies, or laboratory studies on rapidly-changing sources, all of which may entail the measurement of the high PM loadings for which peak-integration uncertainties are most important.

Other AMS-specific uncertainties include the fundamental uncertainty involved in converting electronic signals at the detector to ion counts. This conversion is performed after estimating the signal intensity of a single ion, a process complicated by the signal-thresholding applied by the data acquisition software (version 4.0.9). We performed this single-ion measurement procedure on each of the measurement days described herein and obtained results varying by  $\sim 20\%$ . As it is not clear whether this range represents instrumental or procedural variability (the standard procedure resulted in different ions being used for single-ion measurement on each day), it has not been included in Eq. (13). In addition, it is worth noting that the newest version of the AMS includes significant hardware improvements which address this issue.

A second AMS-specific scenario arises when the intensity of isotopic-daughter-ion peaks are predicted based on their respective isotopic mothers. This procedure propa-

## AMS peak-integration errors

J. C. Corbin et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



gates fitting errors across integer  $m/z$ , and the corresponding uncertainty should also be propagated to these and neighbouring peaks, as discussed in Corbin et al. (2014).

Finally, not all ions follow the peak shape established by the PIKA calibration procedure. In particular, thermally-generated ions such as  $K^+$  (Drewnick et al., 2006; Corbin et al., 2014) or other slowly-evaporating species (Salcedo et al., 2010; Craven et al., 2012) may lead to additional uncertainties for this reason.

## 6.4 Influence on Positive Matrix Factorization results

The arguments presented above clearly show that  $m/z$  calibration limitations lead to a linear (fractional) imprecision term during the constrained peak integration procedures of PIKA. This imprecision significantly increases the overall imprecision in a resulting data matrix; in particular, linear imprecisions become more significant than ion-counting uncertainties for high signals, and overlap errors may be larger still. Thus, omitting this imprecision term from the uncertainty provided to PMF leads to an over-weighting of higher-signal data, which may bias the PMF solution (Paatero and Tapper, 1994; Paatero and Hopke, 2003).

The importance of high signals means that fitting errors may be especially important for the high aerosol concentrations that may be measured at the roadside (e.g. Dallmann et al., 2014), within combustion plumes (e.g. Cubison et al., 2011), in highly-polluted cities like Beijing or Mexico City (e.g. Zhang et al., 2007), or during source studies (e.g. Elsasser et al., 2013; Timko et al., 2014; Corbin et al., 2014).

Given the variable importance of high signals within different data sets, and the importance of instrument-specific parameters to the peak-integration imprecision as discussed above, an absolute statement of the relevance of these uncertainties in PMF is impossible. Nevertheless, as an example we performed PMF on a synthetic data matrix to demonstrate the significance of addressing peak-fitting errors.

The synthetic data matrix was constructed using the PMF solution of the present data set discussed in Corbin et al. (2014). The factors (mass spectra) and loadings (time series) of the PMF solution reported in that manuscript were recombined into an

## AMS peak-integration errors

J. C. Corbin et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



**AMS peak-integration errors**

J. C. Corbin et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



error-free synthetic data matrix. Simulated noise was added to this matrix to represent Poisson and peak-fitting imprecisions (given by Eq. 13) by sampling from Poisson- and Gaussian-noise generators (Igor Pro, version 6.3) with a SD of 5 % of the peak height. Peak-overlap errors were not included. PMF was performed as described in Corbin et al. (2014). In particular, low signal-to-noise variables were downweighted (Paatero and Hopke, 2003) and the robust mode of PMF was used (Paatero, 1997), consistent with standard practice in the AMS community (Zhang et al., 2011).

When this synthetic matrix was factorized using only a Poisson imprecision term to weight the data (i.e. with the largest signals overweighted), the residual matrix showed significant outliers for the highest signals. For example, the largest residual outliers for  $\text{CO}^+$ , which was the highest-signal ion in the synthetic matrix, corresponded to spikes in the signal of that ion. These residual outliers were removed when the synthetic matrix was factorized using the correct imprecision model, Eq. (13).

The most significant observation in this experiment is that the highest  $\text{CO}^+$  signals appeared as spikes in the residual matrix when using the incorrect imprecision model. Under normal circumstances, the analyst may have regarded such spikes as “outliers” reflecting transient signals or data-analysis problems, even though (in this case) they were purely the result of underestimated uncertainties. Properly weighting these spikes caused the  $r^2$  between the output and input time series of the lowest-signal factor to increase from 0.54 to 0.74, with similar but smaller increases for the higher-signal factors, which were retrieved with  $r^2 > 0.91$  in the wrongly-weighted case and with  $r^2 > 0.999$  in the correctly-weighted case.

## 7 Conclusions

Peak-integration uncertainties in the AMS PIKA software originate from uncertainties in peak-width prediction and in peak-height fitting. The former uncertainty may be estimated from the peak-width calibration procedure; the latter by fits to simulated peaks in an empirically-constrained Monte Carlo approach.

## AMS peak-integration errors

J. C. Corbin et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Peak-fitting uncertainties depend most strongly on errors in the  $m/z$  calibration used to predict the peak location  $\mu$  in ion-time-of-flight space. For well-resolved peaks, fitting uncertainties are sensitive to imprecisions in  $\mu$ , to biases in  $\mu$ , to the number of data points representing a given peak, and to the position of the peak centre relative to the nearest data point in  $m/z$  space.

Since peaks are fitted in PIKA by linearly scaling a predefined function, peak-fitting errors also scale linearly with peak height, for well-resolved peaks. This leads to a constant-relative-imprecision term in the overall peak-integration uncertainty. Since a constant relative imprecision scales linearly with the ion count  $n$  as  $n^1$ , but counting uncertainties scale as  $n^{0.5}$ , this constant-relative imprecision term dominates counting uncertainties at high signal intensities. For example, in an exemplary data set with an estimated  $\sim 4\%$  imprecision in fitted peak height and  $5\%$  imprecision in integrated peak area, the relative imprecision term dominated counting uncertainties for well-resolved ions with areas of  $\sim 1$  kHz. Peak-integration uncertainties will therefore be especially important for the high aerosol concentrations that may be measured near pollution sources or within highly-polluted cities.

In a synthetic data set, including the constant-relative-imprecision term during PMF led to a significantly improvement in the accuracy of the solution. The dependence of the relative imprecision on the  $m/z$  calibration and mass-spectral resolution indicate that it should be estimated for each new data set, or at least each new voltage configuration of a given mass spectrometer. The software used to perform this estimation in the present study, written in Igor Pro, is available upon request.

Finally, peak-fitting errors may also increase rapidly when peaks overlap significantly, potentially becoming much larger than the uncertainties of well-resolved peaks. The magnitude of this increase may be estimated by a Monte Carlo approach similar to that applied to isolated peaks, and a corresponding imprecision obtained by simulating the expected distribution of  $m/z$  calibration errors.

*Acknowledgements.* The open-source nature of the PIKA software was essential to this work, making it possible to read and understand the details of the existing AMS analysis procedures.

Discussions with P. Lowdon, M. Tanadini, and M. R. Canagaratna led to significant improvements in this work.

## Appendix A: Description of test data set

The data set used for evaluating and testing fitting uncertainties represents the mass spectra of fresh, aged, and filtered-and-aged aerosols emitted from a beech-wood combustion stove. Up to six batches of wood were burnt consecutively on three consecutive days in these experiments. A complete description of the experimental setup and instrument configuration is given in Corbin et al. (2014).

The wood aerosols were vaporized and ionized in an Aerodyne HR-AMS equipped with a Soot-Particle (SP) vaporization module (Onasch et al., 2012). The majority of signals presented in the paper reflect SP-AMS-mode measurements, which were used because signals were consistently higher with the SP laser on and the analysis in this work applies mainly to higher signals. SP signals were higher because (i) the majority of PM emitted by the stove was refractory BC (Corbin et al., 2015a) and (ii) the AMS is more sensitive to organic coatings when they are vaporized by the SP laser, due to its physical position (Willis et al., 2014). As shown in Fig. 3e, the laser state had no effect on the peak-fitting results discussed herein.

## Appendix B: Analysis of test data set

All data were analyzed in Igor Pro (Version 6.2, Wavemetrics, OR, USA) using a modified version of PIKA, derived from PIKA 1.10H, and custom code. The modifications to PIKA consisted of: improvements to the peak-width calibration procedure, the selective introduction of errors to the analysis, attempted improvements to the peak-fitting procedure, and the implementation of peak-integration uncertainties as part of the overall PIKA error calculation. The first modification improved the robustness of the peak-width

## AMS peak-integration errors

J. C. Corbin et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## AMS peak-integration errors

J. C. Corbin et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



calibration procedure by replacing the mean peak width with a trimmed-mean, followed by a weighted fit to the data, as detailed in Corbin et al. (2014).

In this manuscript, “Diff” HR-AMS data have been presented. Diff data represent the difference of “Open” measurements (comprising signals from particulate, gaseous, and background species) and “Closed” measurements (comprising background species from gases and slowly-evaporating material). The same trends seen in the Diff data were seen in analogous plots for the Open and Closed data, however the noise regime of the RMSE was much less noticeable in these data. Diff data were used to allow the two regimes of the RMSE to be clearly highlighted, and to remove inconsequential differences due to different background levels, for example of  $\text{CO}_2^+$  due to gaseous carbon dioxide.

The following peaks were used for  $m/z$  calibration:  $\text{CH}^+$ ,  $\text{Ar}^+$ ,  $\text{CO}_2^+$ ,  $^{182}\text{W}^+$ ,  $^{184}\text{W}^+$ ,  $^{186}\text{W}^+$ . While this list is not ideal, no other peaks were consistently present with sufficient signal for use as calibrant ions. Some peaks in the test data were isolated but were excluded for having anomalous peak shapes. These peaks included gas-phase ions present at high signals (e.g.  $\text{N}_2^+$  or  $\text{O}_2^+$ ), ions which were present at very low  $m/z$  and were therefore represented by only two detector bins ( $\text{C}^+$  and  $\text{CH}^+$ ) such that reliable fits could not be performed, and ions which were known to follow anomalous vaporization-ionization physics (e.g.  $\text{K}^+$ , discussed further in Corbin et al., 2014).

## References

- Aiken, A. C., DeCarlo, P. F., and Jimenez, J. L.: Elemental analysis of organic species with electron ionization high-resolution mass spectrometry, *Anal. Chem.*, 79, 8350–8358, doi:10.1021/ac071150w, 2007. 3474
- 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., and Sueper, D.: 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. 3476

## AMS peak-integration errors

J. C. Corbin et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



- Allan, D. W.: Statistics of atomic frequency standards, *Proc. IEEE*, 54, 221–230, doi:10.1109/PROC.1966.4634, 1966. 3477
- 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. 3474, 3476, 3480, 3495
- Allan, J. D., Delia, A. E., Coe, H., Bower, K. N., Alfarra, M. R., Jimenez, J. L., Middlebrook, A. M., Drewnick, F., Onasch, T. B., and Canagaratna, M. R.: A generalised method for the extraction of chemically resolved mass spectra from aerodyne aerosol mass spectrometer data, *J. Aerosol Sci.*, 35, 909–922, doi:10.1016/j.jaerosci.2004.02.007, 2004. 3474, 3476
- Canagaratna, M., Jayne, J., Jimenez, J., Allan, J., Alfarra, M., Zhang, Q., Onasch, T. B., Drewnick, F., Coe, H., Middlebrook, A., Delia, A., Williams, L., Trimborn, A., Northway, M., DeCarlo, P., Kolb, C., Davidovits, P., and Worsnop, D.: Chemical and microphysical characterization of ambient aerosols with the aerodyne aerosol mass spectrometer, *Mass Spectrom. Rev.*, 26, 185–222, doi:10.1002/mas.20115, 2007. 3474
- Cappa, C. D., Onasch, T. B., Massoli, P., Worsnop, D. R., Bates, T. S., Cross, E. S., Davidovits, P., Hakala, J., Hayden, K. L., and Jobson, B. T.: Radiative absorption enhancements due to the mixing state of atmospheric black carbon, *Science*, 337, 1078–1081, doi:10.1126/science.1223447, 2012. 3474
- Chen, Q., Liu, Y., Donahue, N. M., Shilling, J. E., and Martin, S. T.: Particle-phase chemistry of secondary organic material: modeled compared to measured O:C and H:C elemental ratios provide constraints, *Environ. Sci. Technol.*, 45, 4763–4770, doi:10.1021/es104398s, 2011. 3476
- Corbin, J. C., Keller, A., Sierau, B., Lohmann, U., and Mensah, A. A.: Black-carbon-surface oxidation and organic composition of beech-wood soot aerosols, accepted for publication in *Atmos. Chem. Phys. Discuss*, 2015. 3474, 3479, 3483, 3499, 3500, 3502, 3503, 3515
- Corbin, J. C., Keller, A., Sierau, B., Lohmann, U., and Mensah, A. A.: Wood-stove and pellet-burner organic aerosol emissions aged in a continuous-flow photoreactor, submitted to *Aerosol Sci. Tech.*, 2015a. 3502
- Corbin, J. C., Sierau, B., Gysel, M., Laborde, M., Keller, A., Kim, J., Petzold, A., Onasch, T. B., Lohmann, U., and Mensah, A. A.: Mass spectrometry of refractory black carbon particles from six sources: carbon-cluster and oxygenated ions, *Atmos. Chem. Phys.*, 14, 2591–2603, doi:10.5194/acp-14-2591-2014, 2014. 3476

## AMS peak-integration errors

J. C. Corbin et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



- Craven, J. S., Yee, L. D., Ng, N. L., Canagaratna, M. R., Loza, C. L., Schilling, K. A., Yatavelli, R. L. N., Thornton, J. A., Ziemann, P. J., Flagan, R. C., and Seinfeld, J. H.: Analysis of secondary organic aerosol formation and aging using positive matrix factorization of high-resolution aerosol mass spectra: application to the dodecane low-NO<sub>x</sub> system, *Atmos. Chem. Phys.*, 12, 11795–11817, doi:10.5194/acp-12-11795-2012, 2012. 3498, 3499
- 5 Cubison, M. J., Ortega, A. M., Hayes, P. L., Farmer, D. K., Day, D., Lechner, M. J., Brune, W. H., Apel, E., Diskin, G. S., Fisher, J. A., Fuelberg, H. E., Hecobian, A., Knapp, D. J., Mikoviny, T., Riemer, D., Sachse, G. W., Sessions, W., Weber, R. J., Weinheimer, A. J., Wisthaler, A., and Jimenez, J. L.: Effects of aging on organic aerosol from open biomass burning smoke in aircraft and laboratory studies, *Atmos. Chem. Phys.*, 11, 12049–12064, doi:10.5194/acp-11-12049-2011, 2011. 3499
- 10 Cubison, M. J., Sueper, D., and Jimenez, J. L.: Statistical precision of the intensities retrieved from constrained fitting of overlapping peaks in high-resolution mass spectra, *Atmos. Meas. Tech. Discuss.*, 7, 12617–12647, doi:10.5194/amtd-7-12617-2014, 2014. 3492, 3496
- 15 Dallmann, T. R., Onasch, T. B., Kirchstetter, T. W., Worton, D. R., Fortner, E. C., Herndon, S. C., Wood, E. C., Franklin, J. P., Worsnop, D. R., Goldstein, A. H., and Harley, R. A.: Characterization of particulate matter emissions from on-road gasoline and diesel vehicles using a soot particle aerosol mass spectrometer, *Atmos. Chem. Phys.*, 14, 7585–7599, doi:10.5194/acp-14-7585-2014, 2014. 3499
- 20 DeCarlo, P. F., Kimmel, J. R., Trimborn, A., Northway, M. J., Jayne, J. T., Aiken, A. C., Gonin, M., Fuhrer, K., Horvath, T., and Docherty, K. S.: Field-deployable, high-resolution, time-of-flight aerosol mass spectrometer, *Anal. Chem.*, 78, 8281–8289, doi:10.1021/ac061249n, 2006. 3474, 3476, 3478, 3479, 3487, 3489, 3496, 3497
- 25 Drewnick, F., Hings, S. S., Curtius, J., Eerdekens, G., and Williams, J.: Measurement of fine particulate and gas-phase species during the New Year's fireworks 2005 in Mainz, Germany, *Atmos. Environ.*, 40, 4316–4327, doi:10.1016/j.atmosenv.2006.03.040, 2006. 3499
- Elsasser, M., Busch, C., Orasche, J., Schön, C., Hartmann, H., Schnelle-Kreis, J., and Zimmermann, R.: Dynamic changes of the aerosol composition and concentration during different burning phases of wood combustion, *Energy Fuels*, 27, 4959–4968, doi:10.1021/ef400684f, 2013. 3499
- 30 Freutel, F., Drewnick, F., Schneider, J., Klimach, T., and Borrmann, S.: Quantitative single-particle analysis with the Aerodyne aerosol mass spectrometer: development of a new clas-

## AMS peak-integration errors

J. C. Corbin et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



sification algorithm and its application to field data, *Atmos. Meas. Tech.*, 6, 3131–3145, doi:10.5194/amt-6-3131-2013, 2013. 3474

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B.: *Bayesian Data Analysis*, CRC Press, Florida, USA, 2013. 3497

5 Hilmer, J. K. and Bothner, B.: Physical signal modulation of time-of-flight mass analyzers increases precision and decreases noise, *Rapid Commun. Mass Sp.*, 25, 795–805, doi:10.1002/rcm.4923, 2011. 3489, 3496, 3498

Hings, S. S.: *Characterization and Field Deployment of a Novel Quantitative Time-of-Flight Aerosol Mass Spectrometer (ToF-AMS)*, Ph.D. thesis, Johannes Gutenberg-Universität Mainz, Mainz, 2006. 3480

10 Jayne, J. T., Leard, D. C., Zhang, X., Davidovits, P., Smith, K. A., Kolb, C. E., and Worsnop, D. R.: Development of an aerosol mass spectrometer for size and composition analysis of submicron particles, *Aerosol Sci. Tech.*, 33, 49–70, doi:10.1080/027868200410840, 2000. 3474

15 Jimenez, J. L., Canagaratna, M. R., Donahue, N. M., Prévôt, A. S. H., Zhang, Q., Kroll, J. H., DeCarlo, P. F., Allan, J. D., Coe, H., Ng, N. L., Aiken, A. C., Docherty, K. S., Ulbrich, I. M., Grieshop, A. P., Robinson, A. L., Duplissy, J., Smith, J. D., Wilson, K. R., Lanz, V. A., Hueglin, C., Sun, Y. L., Tian, J., Laaksonen, A., Raatikainen, T., Rautiainen, J., Vaattovaara, P., Ehn, M., Kulmala, M., Tomlinson, J. M., Collins, D. R., Cubison, M. J., E, Dunlea, J., Huffman, J. A., Onasch, T. B., Alfarra, M. R., Williams, P. I., Bower, K., Kondo, Y., Schneider, J., Drewnick, F., Borrmann, S., Weimer, S., Demerjian, K., Salcedo, D., Cottrell, L., Griffin, R., Takami, A., Miyoshi, T., Hatakeyama, S., Shimono, A., Sun, J. Y., Zhang, Y. M., Dzepina, K., Kimmel, J. R., Sueper, D., Jayne, J. T., Herndon, S. C., Trimborn, A. M., Williams, L. R., Wood, E. C., Middlebrook, A. M., Kolb, C. E., Baltensperger, U., and Worsnop, D. R.: Evolution of organic aerosols in the atmosphere, *Science*, 326, 1525–1529, doi:10.1126/science.1180353, 2009. 3474

20 Kimmel, J. R., Farmer, D. K., Cubison, M. J., Sueper, D., Tanner, C., Nemitz, E., Worsnop, D. R., Gonin, M., and Jimenez, J. L.: Real-time aerosol mass spectrometry with millisecond resolution, *Int. J. Mass Spectrom.*, 303, 15–26, doi:10.1016/j.ijms.2010.12.004, 2011. 3474

25 Lee, A. K. Y., Willis, M. D., Healy, R. M., Onasch, T. B., and Abbatt, J. P. D.: Mixing state of carbonaceous aerosol in an urban environment: single particle characterization using the soot particle aerosol mass spectrometer (SP-AMS), *Atmos. Chem. Phys.*, 15, 1823–1841, doi:10.5194/acp-15-1823-2015, 2015. 3474

**AMS peak-integration errors**

J. C. Corbin et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Lee, T., Sullivan, A. P., Mack, L., Jimenez, J. L., Kreidenweis, S. M., Onasch, T. B., Worsnop, D. R., Malm, W., Wold, C. E., and Hao, W. M.: Chemical smoke marker emissions during flaming and smoldering phases of laboratory open burning of wildland fuels, *Aerosol Sci. Tech.*, 44, i–v, 2010. 3474

5 Müller, M., George, C., and D'Anna, B.: Enhanced spectral analysis of C-TOF aerosol mass spectrometer data: iterative residual analysis and cumulative peak fitting, *Int. J. Mass Spectrom.*, 306, 1–8, doi:10.1016/j.ijms.2011.04.007, 2011. 3487, 3489, 3496

Ng, N. L., Herndon, S. C., Trimborn, A., Canagaratna, M. R., Croteau, P., Onasch, T. B., Sueper, D., Worsnop, D. R., Zhang, Q., and Sun, Y.: An Aerosol Chemical Speciation Monitor (ACSM) for routine monitoring of the composition and mass concentrations of ambient aerosol, *Aerosol Sci. Tech.*, 45, 780–794, doi:10.1080/02786826.2011.560211, 2011. 3477

10 Onasch, T. B., Trimborn, A., Fortner, E. C., Jayne, J. T., Kok, G. L., Williams, L. R., Davidovits, P., and Worsnop, D. R.: Soot particle aerosol mass spectrometer: development, validation, and initial application, *Aerosol Sci. Tech.*, 46, 804–817, doi:10.1080/02786826.2012.663948, 2012. 3477, 3502

15 Ortega, A. M., Day, D. A., Cubison, M. J., Brune, W. H., Bon, D., de Gouw, J. A., and Jimenez, J. L.: Secondary organic aerosol formation and primary organic aerosol oxidation from biomass-burning smoke in a flow reactor during FLAME-3, *Atmos. Chem. Phys.*, 13, 11551–11571, doi:10.5194/acp-13-11551-2013, 2013. 3479

20 Paatero, P.: Least squares formulation of robust non-negative factor analysis, *Chemometr. Intell. Lab.*, 37, 23–35, doi:10.1016/S0169-7439(96)00044-5, 1997. 3500

Paatero, P. and Hopke, P. K.: Discarding or downweighting high-noise variables in factor analytic models, *Anal. Chim. Acta*, 490, 277–289, doi:10.1016/S0003-2670(02)01643-4, 2003. 3499, 3500

25 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, doi:10.1002/env.3170050203, 1994. 3474, 3499

Salcedo, D., Onasch, T. B., Aiken, A. C., Williams, L. R., de Foy, B., Cubison, M. J., Worsnop, D. R., Molina, L. T., and Jimenez, J. L.: Determination of particulate lead using aerosol mass spectrometry: MILAGRO/MCMA-2006 observations, *Atmos. Chem. Phys.*, 10, 5371–5389, doi:10.5194/acp-10-5371-2010, 2010. 3499

**AMS peak-integration errors**

J. C. Corbin et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



- Sueper, D., Jimenez, J. L., Aiken, A., and DeCarlo, P.: PIKA ToF-AMS High Resolution Analysis Software, available at: [cires.colorado.edu/jimenez-group/ToFAMSResources/ToFSoftware/PikalInfo](http://cires.colorado.edu/jimenez-group/ToFAMSResources/ToFSoftware/PikalInfo), 2011. 3474, 3480, 3495
- Sun, Y. L., Zhang, Q., Anastasio, C., and Sun, J.: Insights into secondary organic aerosol formed via aqueous-phase reactions of phenolic compounds based on high resolution mass spectrometry, *Atmos. Chem. Phys.*, 10, 4809–4822, doi:10.5194/acp-10-4809-2010, 2010. 3479
- Taylor, J. R.: *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*, University Science Books, Sausalito, California, USA, 1997. 3476
- Timko, M. T., Albo, S. E., Onasch, T. B., Fortner, E. C., Yu, Z., Miake-Lye, R. C., Canagaratna, M. R., Ng, N. L., and Worsnop, D. R.: Composition and sources of the organic particle emissions from aircraft engines, *Aerosol Sci. Tech.*, 48, 61–73, doi:10.1080/02786826.2013.857758, 2014. 3499
- 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, doi:10.5194/acp-9-2891-2009, 2009. 3474
- Werle, P., Mücke, R., and Stelm, F.: The limits of signal averaging in atmospheric trace-gas monitoring by tunable diode-laser absorption spectroscopy (TDLAS), *Appl. Phys. B*, 57, 131–139, 1993. 3477
- Wilks, D.: *Statistical Methods in the Atmospheric Sciences*, International Geophysics, Elsevier Science, Oxford, UK, 2011. 3482, 3483
- Willis, M. D., Lee, A. K. Y., Onasch, T. B., Fortner, E. C., Williams, L. R., Lambe, A. T., Worsnop, D. R., and Abbatt, J. P. D.: Collection efficiency of the soot-particle aerosol mass spectrometer (SP-AMS) for internally mixed particulate black carbon, *Atmos. Meas. Tech.*, 7, 4507–4516, doi:10.5194/amt-7-4507-2014, 2014. 3502
- 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., 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. 3499

Zhang, Q., Jimenez, J. L., Canagaratna, M. R., Ulbrich, I. M., Ng, N. L., Worsnop, D. R., and Sun, Y.: Understanding atmospheric organic aerosols via factor analysis of aerosol mass spectrometry: a review, *Anal. Bioanal. Chem.*, 401, 3045–3067, doi:10.1007/s00216-011-5355-y, 2011. 3474, 3500

**AMTD**

8, 3471–3523, 2015

---

**AMS peak-integration errors**

J. C. Corbin et al.

---

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## AMS peak-integration errors

J. C. Corbin et al.

**Table 1.** Effects of manually-introducing errors to the PIKA analysis procedure. **Emphasized** cases are plotted in Fig. 3. Bracketed values are the estimated errors for the reference case.

Quantity	Error	Effect on constant RMSE	Effect on constant RMSE/ <i>h</i>
Data <sup>a</sup> ( <i>x</i> )	[~ 0.2], 1, <b>5</b> , 50 Hz	Increased	Negligible
Width <sup>a</sup> ( <i>w</i> )	[< 2.5], 5, <b>25</b> , 50, 250 %	Negligible	Increased spread
Baseline height <sup>b</sup>	[< 0.1], 1, <b>5</b> , 50 Hz	Increased	Negligible
Baseline slope <sup>a</sup>	[10 <sup>-5</sup> ], <b>10<sup>-3</sup></b> , 10 <sup>-1</sup> Hz ( <i>m/z</i> ) <sup>-1</sup>	Negligible	Negligible
Location <sup>a</sup> ( <i>μ</i> )	[~ 20], <b>50</b> , 100, 200 ppm	Negligible	Increased and increased spread
Peak shape <sup>c</sup> ( <i>v</i> )	[Empirical.] <b>positive</b> or <b>negative</b> skewness	Negligible	Increased

<sup>a</sup> Added as Gaussian noise.<sup>b</sup> Added as the absolute value of Gaussian noise.<sup>c</sup> Skewed by multiplying the empirical peak-shape function with a normal cumulative distribution function, resulting in a 25 % smaller peak area (see Fig. 3g and h).

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion





## AMS peak-integration errors

J. C. Corbin et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

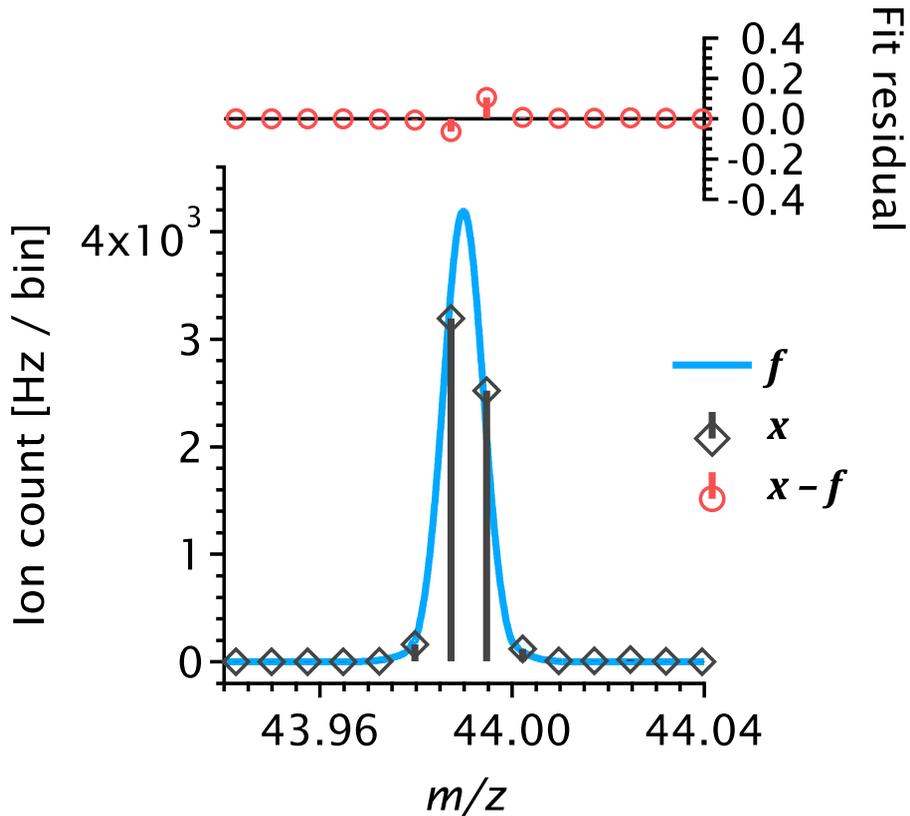
Printer-friendly Version

Interactive Discussion



**Table A1.** List of symbols and abbreviations. Symbols used only once in the text are omitted from this list.

AMS	Aerodyne aerosol mass spectrometer
HR-AMS	High-resolution AMS
PIKA	HR-AMS analysis software
PMF	Positive Matrix Factorization
RMSE	root mean square error (Eq. 8)
UMR	unit (integer) mass resolution
$A$	the area of a peak
$e$	an error (imprecision, bias, or both)
$f$	a pseudo-Gaussian function (Eq. 3)
$f_0$	a unit-height peak function (Eq. 3 with $h = 1$ )
$f_T$	a linear superposition of peaks $f$ (Eq. 5 with $h = 1$ )
$G$	a Gaussian function (Eq. 2)
$h$	the height of a peak
$k_{DC}$	mass-spectrometer duty cycle
$m/z$	mass-to-charge ratio
$v$	the peak-shape function (Eq. 3)
$w$	the width of a peak
$x_i$	the $i$ th $m/z$ bin of a mass spectrum
$\mu$	the mode of a peak
$\sigma$	an imprecision



**Figure 1.** Example of a PIKA fit. The raw data ( $x$ ), fit ( $f$ ), and residuals ( $x - f$ ) are shown.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

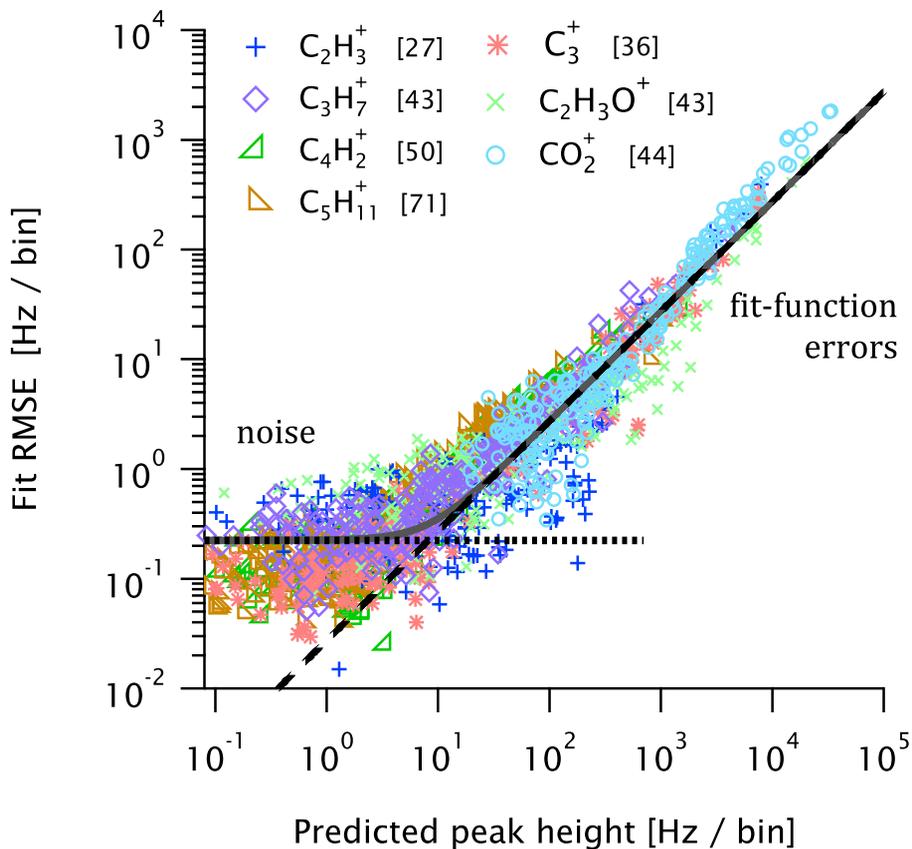
Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion





**Figure 2.** The RMSE of standard PIKA fits to isolated peaks in the test HR-AMS dataset. Numbers in the legend show the integer *m/z* of these peaks. The dotted line indicates the lower limit of constant RMSE, the dashed line the upper limit of constant relative RMSE, and the solid line the quadratic sum of the two.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

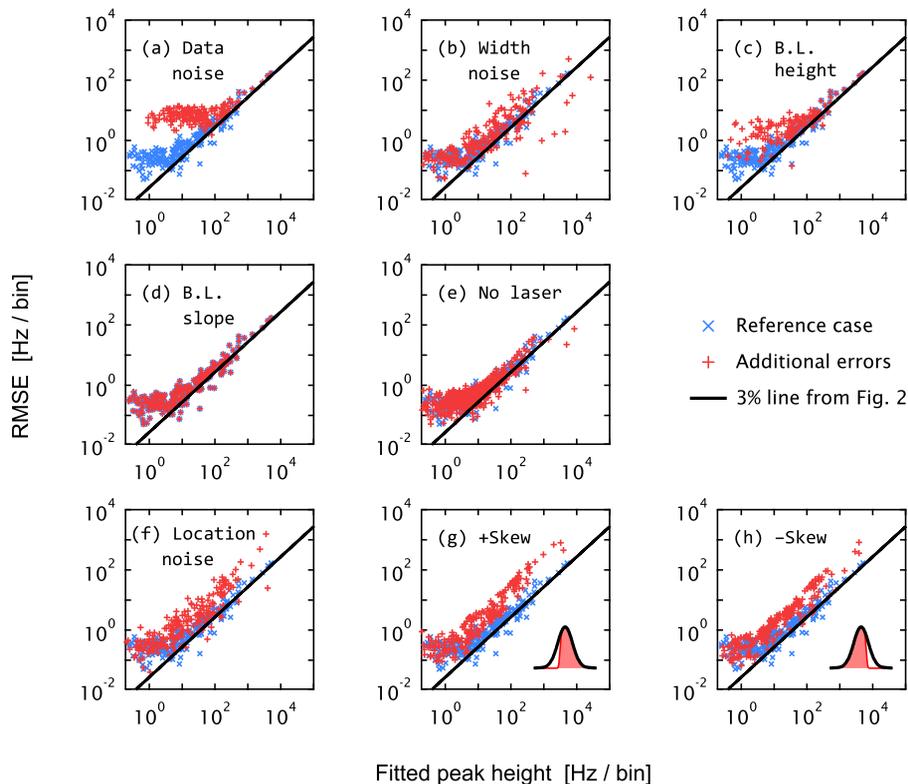
Printer-friendly Version

Interactive Discussion



## AMS peak-integration errors

J. C. Corbin et al.



**Figure 3.** Response of the RMSE of a representative ion ( $C_3H_7^+$ ) to the addition of very large errors at various stages of analysis (**a–d**), (**f–h**). The added errors are defined by the highlighted values in Table 1. The laser vaporizer of a dual-vaporizer SP-AMS (see Corbin et al., 2014) was on for all data except those of panel (**e**). B.L. = baseline.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

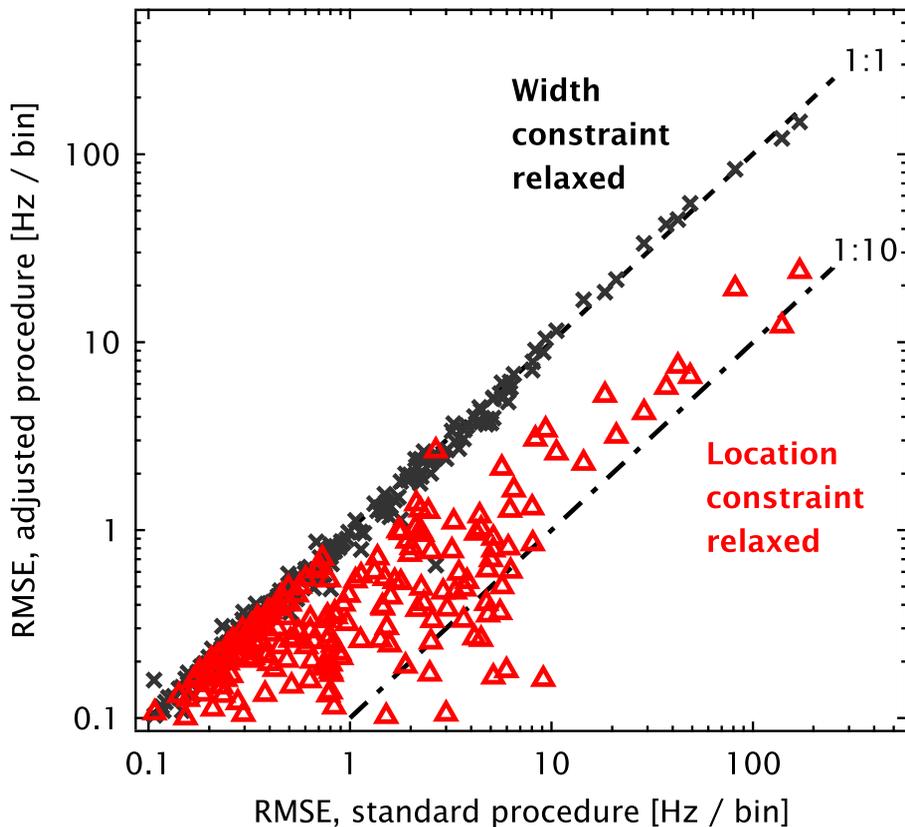
Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

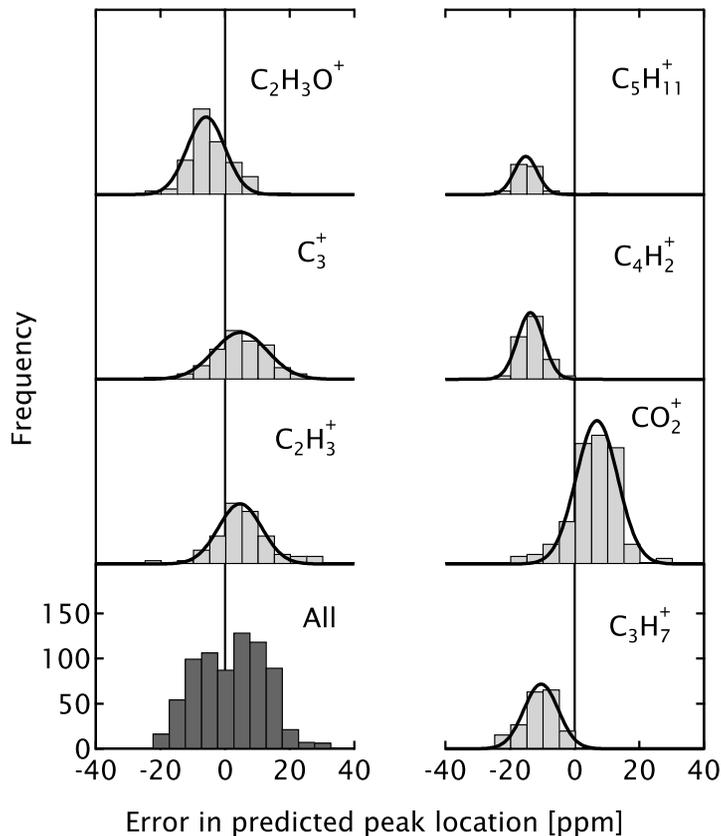




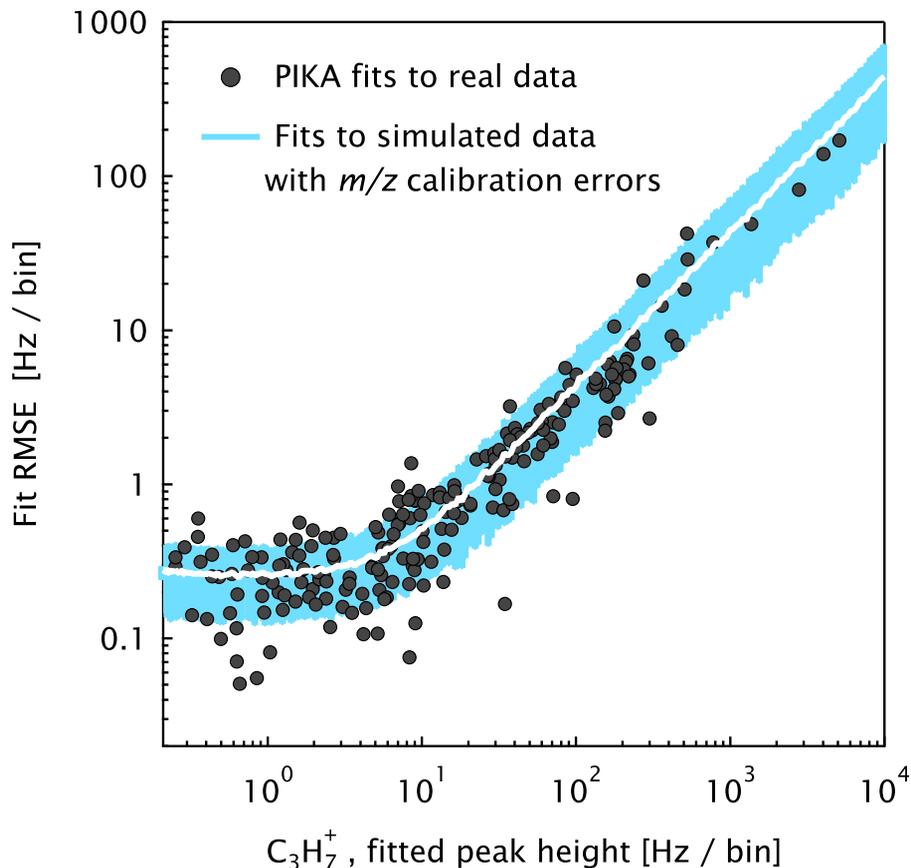
**Figure 4.** Changes in the RMSE of  $C_3H_7^+$  in response to changes in the fitting procedure. The normal PIKA fitting procedure (abscissa) is compared to procedures where  $\mu$  was allowed to vary by  $\pm 20$  ppm (red  $\Delta$ ) and where  $w$  was allowed to vary by  $\pm 5\%$  (grey +). Lines highlight the trends expected for no decrease in the RMSE (1 : 1) and a factor-of-ten decrease (1 : 10).

Title Page	
Abstract	Introduction
Conclusions	References
Tables	Figures
◀	▶
◀	▶
Back	Close
Full Screen / Esc	
Printer-friendly Version	
Interactive Discussion	





**Figure 5.** Distribution of prediction errors in  $\mu$  for each of the ions shown in Fig. 2, including only peaks outside of the noise regime ( $h_{\text{fitted}} > 20 \text{ Hz bin}^{-1}$ ). Light-shaded data represent specific ions, dark-shaded data the aggregate of the other data. The light-shaded data are scaled similarly, so that their relative areas represent the relative number of data used to construct them. The mode and widths of the Gaussian fits displayed are given in Table 2.



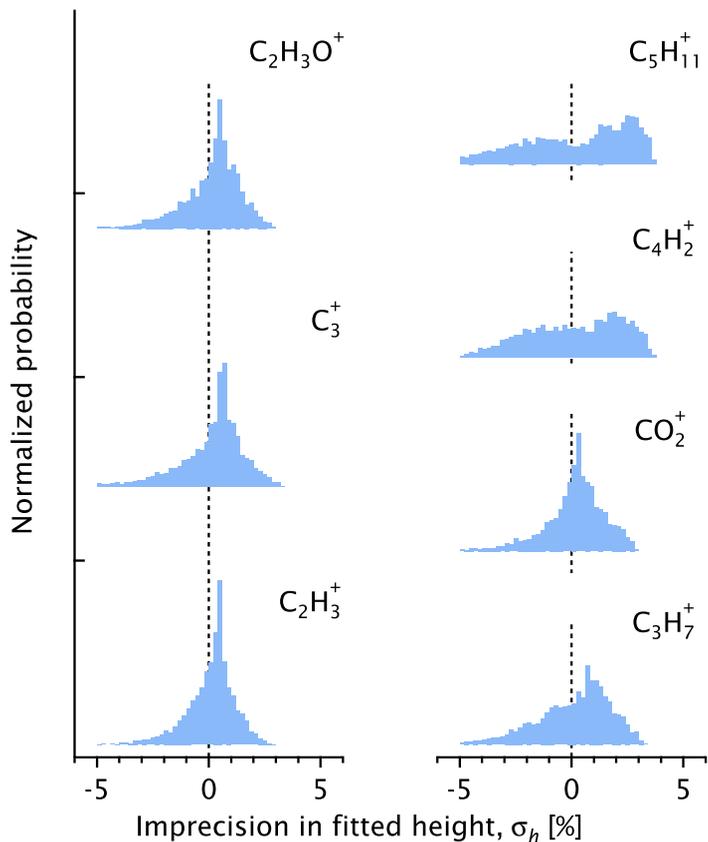
**Figure 6.** The RMSE of  $C_3H_7^+$  fits in the test data set (symbols) and to the simulated data (line, mean; shading, SD). Simulated data was generated and fitted using Eq. (3) with input parameters, noise, and simulated errors in peak-location prediction determined from the data (see text).

AMS peak-integration errors

J. C. Corbin et al.

Title Page	
Abstract	Introduction
Conclusions	References
Tables	Figures
◀	▶
◀	▶
Back	Close
Full Screen / Esc	
Printer-friendly Version	
Interactive Discussion	





**Figure 7.** Normalized probability distributions of the simulated imprecision in fits to isolated peaks from the test data set, using the  $\mu$  prediction errors given in Table 2 and randomly-selected  $m/z$  axes from the data set.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

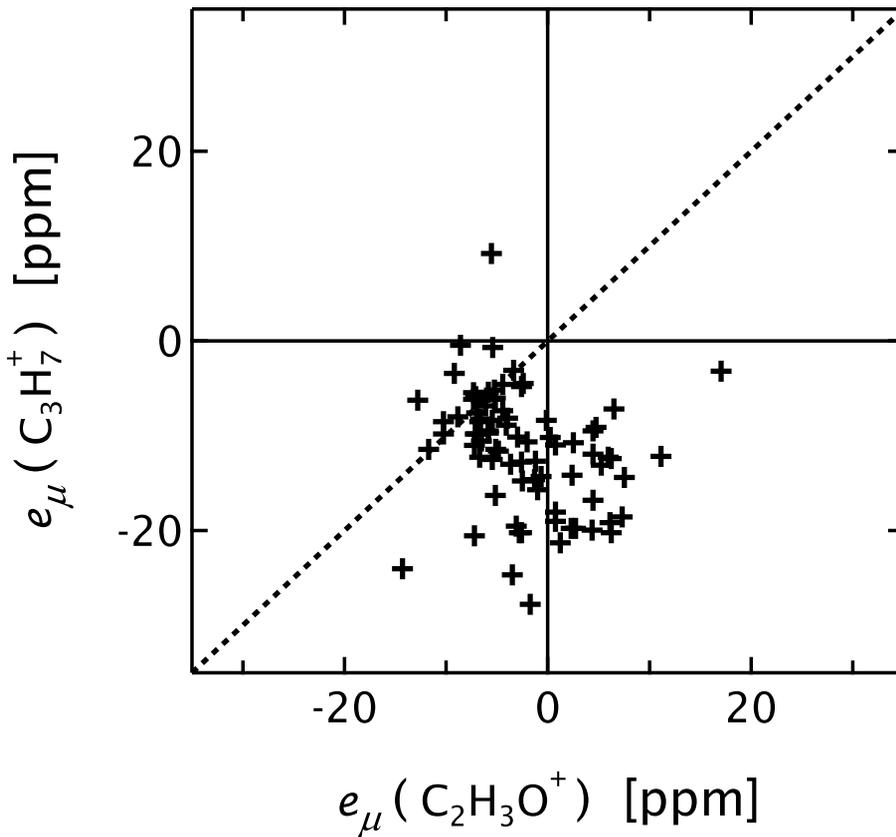
Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion





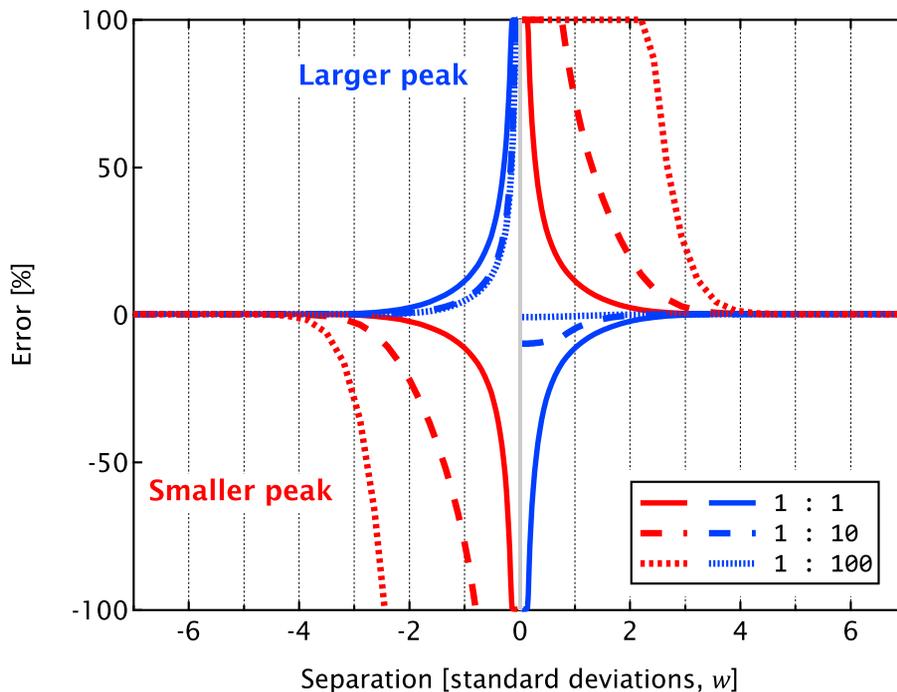
**Figure 8.** Scatterplot of the estimated errors in peak-location prediction,  $e_{\mu}$ , for the ions  $C_2H_3O^+$  and  $C_3H_7^+$ , both found at  $m/z$  43.

Title Page	
Abstract	Introduction
Conclusions	References
Tables	Figures
◀	▶
◀	▶
Back	Close
Full Screen / Esc	
Printer-friendly Version	
Interactive Discussion	



## AMS peak-integration errors

J. C. Corbin et al.

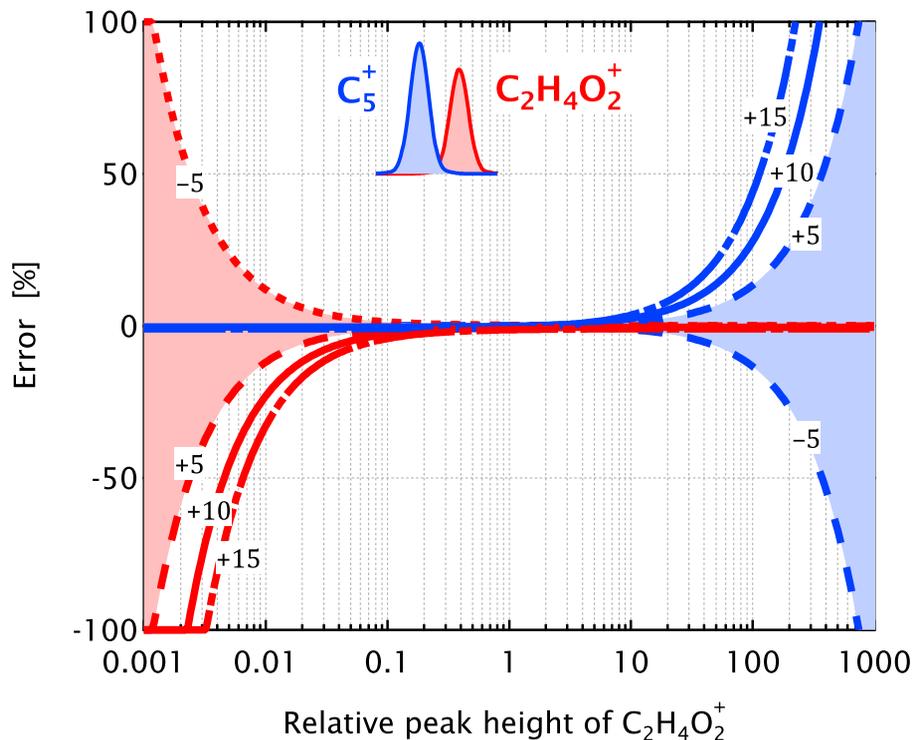


**Figure 9.** The increase in fitting error when  $\mu$  is wrongly constrained by +10 ppm, as a function of the separation of two peaks. The peaks were modelled with a width corresponding to that of at  $m/z \sim 44$  on an arbitrary 100-point  $m/z$  basis. Noise was not included.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

## AMS peak-integration errors

J. C. Corbin et al.



**Figure 10.** Influence of relative peak height on the modelled fitting errors of  $C_5^+$  and  $C_2H_4O_2^+$  at  $m/z$  60. Errors in the  $\mu$  constraint were represented by the values shown in the figure. The shaded regions show the range of fitting errors encompassed by  $\mu$  prediction errors ranging from  $-5$  to  $5$  ppm.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

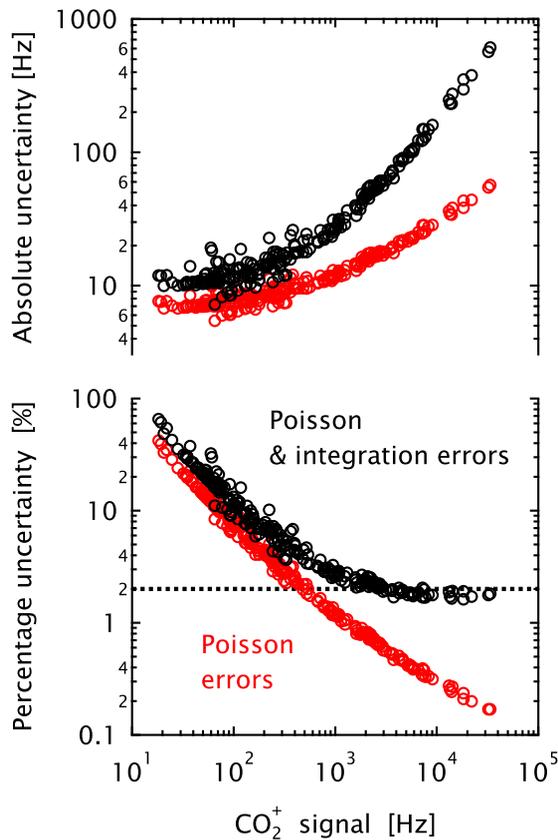
Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion





**Figure 11.** Example of the AMS uncertainties obtained when considering only Poisson (red) or both Poisson and fitting imprecision (black), Eq. (13). The upper panel shows the absolute imprecision; the lower panel the relative imprecision. The horizontal dotted line illustrates the different asymptotic behaviours of the two approaches.

**AMS peak-integration errors**

J. C. Corbin et al.

Title Page	
Abstract	Introduction
Conclusions	References
Tables	Figures
◀	▶
◀	▶
Back	Close
Full Screen / Esc	
Printer-friendly Version	
Interactive Discussion	

