Fully automatic and precise data analysis developed for time-of-flight mass spectrometry

Stefan Meyer, Andreas Riedo, Maike B. Neuland, Marek Tulej and Peter Wurza

Scientific objectives of current and future space missions are focused on the investigation of the origin and evolution of the solar system with the particular emphasis on habitability and signatures of past and present life. For in situ measurements of the chemical composition of solid samples on planetary surfaces, the neutral atmospheric gas and the thermal plasma of planetary atmospheres, the application of mass spectrometers making use of time-of-flight mass analysers is a technique widely used. However, such investigations imply measurements with good statistics and, thus, a large amount of data to be analysed. Therefore, faster and especially robust automated data analysis with enhanced accuracy is required. In this contribution, an automatic data analysis software, which allows fast and precise quantitative data analysis of time-of-flight mass spectrometric data, is presented and discussed in detail. A crucial part of this software is a robust and fast peak finding algorithm with a consecutive numerical integration method allowing precise data analysis. We tested our analysis software with data from different time-of-flight mass spectrometers and different measurement campaigns thereof. The quantitative analysis of isotopes, using automatic data analysis, yields results with an accuracy of isotope ratios up to 100 ppm for a signal-to-noise ratio (SNR) of 10. Furthermore, we observe that the accuracy of isotope ratios is inversely proportional to the mass resolution. Additionally, we show that the accuracy of isotope ratios is depending on the sample width $T_s$ by $T_s^{-0.5}$. Copyright © 2017 John Wiley & Sons, Ltd.

Keywords: automatic spectrum analysis; analysis software; TOF mass spectrometry; isotope analysis; space research

Introduction

ESA’s ExoMars1 and JUICE2 mission, and NASA’s curiosity mission3 are only three of several present and future planetary missions that are designed for exploration of the surfaces and atmospheres of different planetary objects in our solar system. One of the key scientific objectives of these missions is the investigation of the formation of the solar system and its evolution, including the investigation of the processes, which are driving the exchange of material between planetary surface and exosphere, the alteration of both, thus all kind of space weathering processes. Measurements of the chemical composition of both, the atmospheres and the surface material, are crucial information to expand our knowledge about the origin of the solar system and for a better understanding of processes in planetary evolution.4,5 These data allow additionally the reconstruction of planetary climate, which may be of primary importance in the search for past or present life activities.

Laser ablation ionisation time-of-flight (TOF) mass spectrometry (LIMS) is regarded as one of the most promising analytical techniques for application in space research. Several miniature laser ablation ionisation mass spectrometers (with instrument name LAMS, LMS, LASMA) have been developed with the aim to investigate the chemical composition of solid materials on planetary surfaces in situ.5–8 LIMS allows high measurement rates with high sensitivity and sufficiently high mass resolution. It also allows probing with high spatial resolution, in lateral as well as in vertical direction.9,10 The simplicity of operation, low power consumption and the robust design is of additional advantage for a space instrument. LIMS is a well-established analytical method for investigations of the chemical composition of solid material and applied in various research fields, including, e.g. geology, metallurgy, environmental and biological science.10 Furthermore, TOF mass spectrometry is already used on several space missions. An example is ESA’s Rosetta mission, where the reflectron time-of-flight mass spectrometer (RTOF) instrument of the ROSINA experiment is built to analyse the chemical composition of cometary gas.11,12 The neutral gas mass spectrometer (NGMS) was selected to detect and analyse the volatile species in the lunar soil on the Russian Luna-Resurs mission.13 The Neutral gas and Ion Mass spectrometer (NIM), part of the Particle Environment Package on-board ESA’s upcoming L-class mission JUICE, is intended for composition measurements of neutral gas and thermal plasma of Jupiter’s moons.14 With these instruments, i.e. the LMS, NGMS and NIM, new capabilities are opened to study planetary exospheres and solid samples on the surface. By precise measurements of, e.g. Pb, isotopes on a planetary surface dating analyses by radioisotope geochronology are feasible, or the investigation of mineralogy of rocky planetary materials due to very sensitive elemental analysis. However, such...
investigations imply high amounts of measurements, often with different instrument settings. For example, establishing mineralogical maps of the surface of a solid sample requires measurements in the defined region of interest on the sample with a sufficiently high spatial resolution. Depending on sampling rate, measurement time available and the spatial resolution required, these can be about \( \sim 10^2 \) measurements for the investigation of a micrometre-sized to millimetre-sized region.\(^{[15,16]} \)

For such large data sets (~Gbytes of data), manual data reduction and processing is not feasible. While being very time consuming, it is furthermore important that reproducibility and robustness of the results is provided, also for different analysts working on the same project. Therefore, faster and especially automated data analysis with high accuracy is required. Based on these needs, we developed a robust automatic data analysis software, which allows the guided data analysis of TOF data. The analysis follows the sequence of data read-in and accumulation procedure up to the quantitative results of the abundances of elements and isotopes within minutes. Thereby, we present a robust and fast numerical integration method, which allows all in all a precise and fast data analysis. We tested our data analysis software in several different measurement campaigns with the LMS instrument.\(^{[15–22]} \) In addition, it was also used for data analysis with the NGMS instrument prototype.\(^{[13]} \) Currently, this automatic data analysis software is used in development and testing of the NIM instrument.\(^{[14]} \)

**Experimental methods**

**NIM prototype instrument**

The results presented in Section on Data Analysis Results and Discussion are obtained with the NIM prototype instrument. NIM is a TOF instrument with heritage from the RTOF instrument of the ROSINA experiment on the Rosetta mission\(^{[11,12]} \) and the Polar Balloon Atmospheric Composition Experiment instrument.\(^{[23]} \) NIM is designed to operate in three different modes: thermal mode, neutral mode and ion mode. In neutral mode, the ambient gas enters the ionisation region of the ion source directly (open source); in thermal mode, the ambient gas enters through an antechamber (equilibrium sphere) where the gas is thermally accommodated with the chamber wall before it proceeds to the ionisation region of the ion source (closed source). In ion mode, the ambient thermal ions are measured directly. A detailed description of the NIM prototype instrument and the results from an extensive test and measurement campaign are given in Meyer et al.\(^{[14]} \)

For the herein presented results, measurements conducted in thermal mode are used. In thermal mode, the neutral gas is decelerated from beam velocity down to thermal energies in an equilibrium sphere and passed on into the ion source in thermal state, where it is ionised by an electron beam and stored,\(^{[24]} \) until the subsequent guidance through the ion optics to the multi-channel plate (MCP) detector. This mode (also referred to as closed source in the literature) is used for neutral gas measurements at any JUICE mission phase but mainly during Europa torus crossing and all other flybys. The so measured neutral gas beam consisted of a gas mixture of 89.5% H\(_2\), 7.5% Ne, 1.5% Ar and 1.5% Kr with a beam velocity of 2.4 km/s.

**Data acquisition**

The analogue signal from the MCP detector of the NIM prototype instrument is acquired and digitised by standard high-speed analogue-to-digital converter (ADC) data acquisition card U1084A from Keysight. This card can digitise with up to 4 GS/s with an analogue bandwidth of up to 1.5 GHz and a vertical resolution of 8 bit for a single waveform. The averager firmware of the data acquisition card allows the on-board accumulation of single waveforms in real time. In standard acquisition mode, the waveforms from about 1000 sequences are accumulated on board and stored on the host computer as one measurement file. Typically, 100 measurement files, summed up to one spectrum, were recorded, each for 20 \( \mu \)s TOF, which corresponds to the mass range from 0 to approximately 800 u/e.

**Data reduction and mass spectrometric analysis**

**Data acquisition process**

The mass spectrometric analysis of a TOF spectrum involves waveform accumulation, spectrum calibration and appropriate representation of the spectral intensities for determination of the quantitative information. Accumulation of waveforms is performed to statistically enhance the dynamic range: the spectrum calibration to obtain the mass scale from the TOF spectrum. For acceleration of the mass calibration and for making this analysis step independent from the individual experimenters, we developed an algorithm for automatic mass calibration. This algorithm is based on the identification of mass peaks occurring at approximately multiples of the integer atomic mass unit and the identification of isotope patterns of elements of interest. Finally, all these operations allow the quantitative analysis of isotope ratios by comparing the number of measured ions of different isotopes within one particular element or the quantitative elemental analysis by comparing the number of measured ions for isotopes of different elements.

In spite of the very explicit data analysis steps to be applied, the following points are important features of any measured TOF spectrum.

**Background**

In general, a background slowly varying over the TOF spectrum, mainly caused by DC voltage drifts on the anodes, can be observed in every measurement. This leads to a shifting and distortion of the baseline, where mass peaks can occur in an ascending region, a straight region or a descending region, respectively. Therefore, the background needs to be subtracted prior to peak area determination. An example of such a background is illustrated in Fig. 1 by its grey shaded area in a measurement of different noble gases with the NIM prototype. The shape of the background can be different for different instrument settings, including, e.g. voltages of the ion-optical system and ionisation parameters.

**Peak asymmetry**

Theoretically, the mass peaks of ions produced by electron impact ionisation from a thermal gas with a normal kinetic energy distribution are expected to be of Gaussian shape. However, in practice, also asymmetric peaks can occur. This asymmetry can be caused during the ionisation process itself or by the ion-optical system that is not fully optimised. In addition, at high ion densities, the mass peaks can be broadened because of mutual Coulombic repulsion within an ion packet.\(^{[17]} \) Also, bandwidth limitations in the detector circuit and acquisition electronics can lead to an exponentially decreasing peak shape.
Noise

Basically, in every electronic measurement system, a certain noise amplitude has to be considered. Figure 1 shows an example of the noise level in a measurement of different noble gases, with an inset on the noise band.

Generally, one measurement contains several waveforms to enhance dynamic range by statistical means, which are automatically generated and saved by the data acquisition system. These consecutively acquired files, with the effective measurement data generated and saved by the data acquisition system. These enhance dynamic range by statistical means, which are automatically acquired files, with the effective measurement data as a digital value for each sampling point, are read-in by the data analysis software.

Thereof, the intensity for one waveform, which is the back conversion of the digital values to the corresponding measured voltage $U(\text{TOF})$, can be calculated as the detected (counted) electron rate $N_e$, which is the number of electrons per time increment with the unit (electrons/s), using

$$N_e(\text{TOF}) = \frac{U(\text{TOF})}{Re}$$

with $R$ the input impedance (here $R = 50 \, \Omega$) of the acquisition circuit and the elementary charge $e$.

For linking the peaks in the TOF spectrum to the corresponding masses, conversion from TOF to mass scale has to be performed. In TOF mass spectrometry, separation between ions of different masses is achieved along the field free drift path, after accelerating the ions to equal kinetic energy. Then, the mass-per-charge ratio $m/z$ can be written as

$$\frac{m}{z} = k_o(t - t_o)^2$$

with the spectrometer constants $k_o$ and $t_o$ which are instrument dependent. $k_o$ and $t_o$ can be determined by a linear fit of the TOF for identified isotopes against the square root of their respective masses. With a suitable set of isotopes, mass scale accuracies better than 500 ppm can be achieved even with very compact instruments.

Peak area determination

Quantitative element and isotope analysis generally deal with abundances of particular species. Therefore, the number of ions of such species, i.e. elements and their isotopes, is of interest. However, using Eqn (1), the number of electrons of a particular isotope $A_e^{(a,b)}$, with its mass peak distributed in the TOF interval $[a,b]$ can be calculated by its area as

$$A_e^{(a,b)} = \int_a^b N_e(\text{TOF}) \, d\text{TOF}$$

$A_e^{(a,b)}$ is proportional to the number of ions of this particular isotope $A_{ion}$, assuming constant gain for the MCP detector. This is well the case, because the MCP voltage is kept at a constant potential during measurements.

Peak area determination according Eqn (3) can be accomplished by different approaches such as numerical (direct) integration and peak fitting either peak by peak or multiple peaks at the same time. With respect to a robust automatic algorithm, the peak area determination is to be kept as simple as possible. Furthermore, single peak analysis is preferred, thus calculating the peak areas one after another. It has been observed that multiple peak fitting lacks of robustness for these types of data, because of the growing number of parameters to be estimated. Furthermore, intermediate and interfering isotopes, which have to be assigned correctly, make fitting of multiple peaks even more problematic for an automated analysis of many data sets.

Peak window finding

Nevertheless, also for single peak analysis, the appropriate peak window needs to be determined, i.e. the TOF interval $[a,b]$ in which the corresponding mass peak is located. Therefore, possible mass peaks have to be identified within the mass spectrum, and their starting and ending points $a$ and $b$ have to be determined accurately. Currently, most algorithms used in mass spectrometry are based on local maxima searching methods for peak detection. On the other hand, several algorithms using wavelet transform methods were proposed. The continuous wavelet transform-based pattern matching algorithm presented in Du et al. seems to be very powerful, in this case, for mass spectra obtained by surface enhanced laser desorption ionisation TOF spectroscopy. Yet, it has to be mentioned that these mass spectra consisted of a dynamic range of about one to two decades and peaks were identified for signal-to-noise ratio (SNR) $> 3$. The requirement for data analysis to detect and accurately quantify major elements as well as trace elements (down to lowest SNR) within a dynamic range of about $10^5$ makes the situation more complex. Furthermore, wavelet transform methods are only used for peak identification; the interval boundaries $a$ and $b$ are not defined by this process.

For determining the correct peak windows, the raw measurement data are preprocessed with an adaptive moving average filter. In our analysis software, we apply a first-derivative zero-crossing algorithm to the filtered data for finding the peaks and the valleys between the peaks. The peak areas are calculated by using the raw, thus non-filtered data to avoid possible influences of the preprocessing of the data.

The implemented process of finding the peak windows is illustrated in Fig. 2. The upper panel of Fig. 2 shows the mass range of Ne with the detected isotope peaks and the detected boundaries of the peaks. The definition of peak boundaries is most complicated for peaks that are close to the noise limit. As an example for an extreme case, the detail of Fig. 2 shows the mass region of the $^{21}$Ne isotope with the details of the different steps of finding the peak boundaries of this low-intensity mass peak near the noise limit. First, the local maximum located nearest to the theoretical mass.
of the analysed mass peak is selected as the centre of the peak. Then, a linear fit of the detected valleys within an expected mass range around this peak centre is executed (Fig. 2; linear fit for valleys). The two valleys next to the peak centre, which are below the linear fit added with $0.35 \sigma$ error of fit and $0.3 \sigma$ of noise, are selected as potential peak window boundaries (linear fit for invalid valleys threshold in Fig. 2). This linear fit and the threshold defined by the fit avoid the false allocation of valleys in case of possible notches in the peak. In the case of $^{21}$Ne, shown in the lower panel of Fig. 2, it can be seen that a detected valley was identified to be a false one, because of the procedure described earlier. The potential peak window boundaries are linearly fitted again (linear fit for chosen valleys in Fig. 2); i.e. the linear interpolation with additional $0.1 \sigma$ of noise is used to reduce the peak window boundaries to the nearest sample point next to the interceptions with the filtered data (Fig. 2, linear fit for effective peak span), thus giving exact starting point $a$ and ending point $b$ of the peak. In addition, this linear interpolation with additional $1 \sigma$ of noise serves as a threshold for peak detection (Fig. 2; linear fit for peak max threshold).

**Background correction**

Because of the varying background, the slopes of the linear fits for effective peak span are not the same (Figs 1 and 2) for all isotopes, respectively, all elements in the mass spectrum. This needs to be considered by using a suitable background model for precise peak area calculation.

We tested different background correction models in the past (constant, linear, quadratic and exponential) and found that best quantitative results are obtained with a linear correction model for the local background of each mass peak. Furthermore, the background area error is about two orders of magnitude lower than the peak area calculation error. Therefore, a linear background correction is implemented peak-wise from starting point $a$ to ending point $b$ of the determined peak window.

**SNR and mass resolution**

The SNR of a mass peak is calculated as

$$SNR = \frac{I_{\text{max}}}{\text{std}(I_b)}$$

where $I_{\text{max}}$ denotes the background-corrected maximum intensity of a mass peak and std($I_b$) describes the standard deviation of the baseline intensity over an equivalent mass range as the peak itself. As this value is determined from a region, where no mass peaks are present, usually at the very end of the mass spectrum, the value is representative for the noise level within the complete spectrum. Equation (4) is in accordance with a technical note from Agilent Technologies about signal, noise and detection limits in mass spectrometry.\(^{[27]}\)

Mass resolution $m/\Delta m$ is the ability to distinguish between two mass peaks (i.e. between ions with adjacent mass-to-charge ratio) and can be written as

$$m/\Delta m = \frac{1}{2} \frac{t}{\Delta t} = \frac{1}{2} \frac{\mu}{\text{FWHM}}$$

where $t$ denotes the TOF with $\mu$, the centre of the intensity peak. $\Delta t$ is determined experimentally from the TOF spectrum as the width of the intensity peak at full width at half maximum.\(^{[10]}\) This peak width includes the uncertainty of the time of ion formation in the source (duration of ion formation in the source, ion location in the source, various initial conditions of the ions), broadening effects caused by ion optics during the ion flight (TOF dispersion, Coulomb repulsion), time-focusing limitations in the reflectron (e.g. field imperfections) and duration of the signal generation by the MCP.
The resolving power of the mass spectrometer is a function of all these effects, and choice of the experimental instrument settings can be critical for achieving an optimum mass resolution.

With the choice of the peak-wise area determination and especially its peak window finding, a minimum mass resolution is necessary such that the mass peaks are completely separated. Otherwise, the peak window and thus the area of interfering mass peaks would be underestimated by falsely cutting off too much of background. Hence, the minimum mass resolution ($m/Δm)_{\text{min}}$ is defined with a maximum span of 1 u to sufficiently separate two peaks separated by 1 u without interference, corresponding to approximately ±3 $σ_{\text{max}}$, of an adequate Gaussian peak. This means that 1 u corresponds to 6 $σ_{\text{max}}$ and leads to

$$\left(\frac{m}{Δm}\right)_{\text{min}} = \frac{μ_m}{FWHM_{\text{max}}} = \frac{μ_m}{2\sqrt{2\ln(2)σ_{\text{max}}}} = \frac{6μ_m}{2\sqrt{2\ln(2)}}$$

with the centre of the mass peak $μ_m$. This implies that the minimum mass resolution linearly increases with the peak’s mass. This has to be kept in mind especially for the interpretation of the calculated accuracies of the isotope ratios, which will be discussed in Section on Dependence of Accuracy of Isotope Ratios on Mass Resolution.

### Peak area calculation

For mass spectrometric analyses, the area of each individual mass peak needs to be derived. The question remains, if peak area calculation by applying a fit or by numerical integration is to be prefered. Already in the past, we tested our analysis software with different numerical integration methods and fitting functions, such as Gaussian and EMG fit, numerical integration using spline interpolation and numerical Monte Carlo (MC) acceptance/rejection integration to compare the performances.[19] Therein, we presented our results on peak areas derived by different integration methods for measurements on standard materials. We showed that results with increased accuracy are obtained using numerical integration methods compared with least-squares fitting methods. These observations are in agreement with theoretical considerations discussed in Bevington and Robinson,[28] where the authors state that the least-squares method consistently underestimates the area of a fitted curve in case of Poisson statistics. Moreover, by assuming the peak shape matching a predefined mathematical function, a preselection is made that is not necessarily in accordance with the measured peak shape. The disadvantage of numerical integration using spline interpolation is that it does not give any error estimation. Testing peak integration with MC integration was the measured peak shape. The disadvantage of numerical integration, denoted as $OT(Δf)^4$ in Eqn (7), depends on the sample width $T_s$ and the fourth-derivative of the intensity function $f(x)$ and can be written as

$$e = \left| \frac{3}{80} T_s^5 f^{(4)}(ζ) \right|, \quad ζ \in [a, b]$$

(8)

With Eqn (8), a maximal error $e$ for the numerical integration can be estimated by taking the maximum of the fourth-derivative of the measured intensity in interval $[a,b]$ as an upper limit,

$$e \leq \frac{3}{80} T_s^5 \max_{a \leq x \leq b} |f^{(4)}(x)|$$

(9)

Computing the fourth-derivative numerically, using finite differences,[30] the maximum numerical error estimate $ε_{\text{Simpson}}$ is

$$ε_{\text{Simpson}} \leq \frac{3}{80} T_s^5 \max_{a \leq x \leq b} \left| f(x_0) - 4f(x_{i+1}) + 6f(x_{i+2}) - 4f(x_{i+3}) + f(x_{i+4}) \right|$$

(10)

As described earlier, with the time integral over the electron rate in Eqn (3), the peak area equals the counted and accumulated number of electrons, corresponding to the number of detected ions, over a certain TOF interval $[a,b]$. This process is subject to the Poisson distribution. Therefore, the following area and error considerations are valid, in accordance with Bevington and Robinson,[28]: assuming three independent functions, $f_p(t)$ describing the effective peak signal (from ions only), $f_{bg}(t)$ a background signal and a function $f_n(t)$ for the remaining noise, composing the total signal to be measured $f_{tot}(t)$,

$$f_{tot}(t) = f_p(t) + f_{bg}(t) + f_n(t)$$

(11)

The area $A_p$ within the peak interval $[a,b]$ is given by

$$A_p = \int_a^b f_{tot}(t) \, dt - f_{bg}(t) - f_n(t) \, dt$$

$$= \int_a^b f_{tot}(t) \, dt - f_{bg}(t) \, dt - f_n(t) \, dt = A_{tot} - A_{bg}$$

(12)

where the integral over the noise function $f_n(t)$ should evaluate to zero, assuming random distributed noise. Using Poisson statistics, where the variance of a stochastic variable is equal to its expectation value[31] and Gaussian error propagation, the variance of the peak area $σ_{A_p}^2$ can be derived with the statistical variance of the total area $σ_{tot}^2$ and $σ_{bg}^2$ the variance of the background area. This results in an estimate $σ_p$ for the error of peak area calculation including $ε_{\text{Simpson}}$ from Eqn (10)

$$σ_p = \sqrt{σ_{tot}^2 + σ_{bg}^2 + ε_{\text{Simpson}}^2} = \sqrt{A_{tot} + A_{bg} + ε_{\text{Simpson}}^2}$$

(13)

with $A_{bg}$ the portion of the total peak area $A_{tot}$ attributed to the background. With these considerations, it is possible to automatically calculate the number of measured electrons, i.e. the area $A_p ± σ_p$ for all peak intervals detected by the algorithm described in the beginning of this section.
Quantitative isotope analysis

Precise and accurate determination of the abundances of isotopes is of high interest in various fields of research, e.g. environmental science, analysis of plants or agricultural products, medical cell analyses, radioactive waste control, forensic analyses or geochemical analyses, including radioisotope geochronology.[10] The investigations of isotope abundances in materials of solar system bodies are of particular importance to cosmochemistry and planetary science.[32] The quantitative measurements of the element and isotope abundances in matter from planetary objects provide critical constraints on events and their chronology in the different stages of the solar system, from its formation to its evolution.

Every element with more than one stable isotope has its own isotope abundance distribution. The isotope distribution can provide insight into different physical and chemical processes, as has been mentioned earlier. When carrying out quantitative isotope analyses, special attention has to be paid to possible isobaric interferences of isotopes of different elements.

The isotope abundances are obtained by calculating the ratios of the isotopes' areas. To quantify the accuracies of the calculated isotope abundances, generally certified reference materials are used, or in case of a terrestrial sample, natural (terrestrial) isotope abundance can be assumed. The relative accuracy of the isotope ratio is defined as

$$\text{relative accuracy of isotope ratio} = \frac{\text{experimental isotope ratio}}{\text{reference isotope ratio}}$$

The reference isotope ratio is from a certified material, e.g. a sample from National Institute of Standards and Technology or US Geological Survey.

Automatic analysis software

For a fully automatic data analysis software, the individual analysis steps that have been presented earlier have to be linked in a structured way. A flow-chart of the individual processes and presented algorithms is illustrated in Fig. 3.

Figure 3. Flow-chart of the implemented functions for fully automated data analysis of a TOF measurement.
As described in Section on Data Acquisition Process, the read-in and accumulation process is executed to generate the TOF spectrum from the measurement data. Subsequently, the automatic mass calibration of the TOF spectrum is initiated, providing the mass spectrum for further analysis. If the automatic mass calibration is not successful, a manual mass calibration is initiated. In this procedure, the software user is asked to provide a list of isotopes for calibration. Afterwards, automatic peak area determination by Simpson integration (Section on Section on Peak Area Determination) of all mass peaks is executed, leading to the files containing all information that is needed for quantitative composition analysis. For keeping our software as universally applicable as possible, the possibility for applying other integration algorithms (MC integration, Gauss fit, EMG fit, Voigt fit and Lorentz fit) was kept. In the presented flow-chart, this is described as advanced peak area determination, where the user can select the desired algorithm. Based on the results from the automatic or the advanced peak area determination, either isotope analysis or element analysis can be carried out for a set of user defined elements and/or isotopes. This includes the computation of quantitative element abundances or isotope ratios and the according graphic accounts as an output.

With our software tool, quantitative data analysis of a measured data set can be carried out fully automated by the guidance of a

**Figure 4.** Accuracy of Ne and Kr isotope ratios relative to terrestrial abundances from Meija et al., with respect to SNR. Missing data points indicate that the corresponding fitting method failed for this specific isotope. [Colour figure can be viewed at wileyonlinelibrary.com]

**Figure 5.** Relative accuracies of isotope ratios, derived by comparison with terrestrial abundance from Meija et al., versus SNR for different isotopes of a noble gas measurement conducted with the NIM instrument. [Colour figure can be viewed at wileyonlinelibrary.com]
trained user within minutes. In addition, the presented algorithms and processes can easily be adapted to every other TOF mass spectrometer (which has been carried out successfully in Paardekooper et al.\cite{34}), because all functions have been implemented in a modular way with universal interfaces and all used constants editable on top level.

Data analysis results and discussion

Peak area calculation method

To compare the described numerical integration methods to other implemented fitting functions, the relative accuracy of the isotope ratios with respect to the terrestrial abundance [Eqn (14); Section on Quantitative Isotope Analysis] has been calculated for several isotopes of Ne and Kr, measured with the NIM instrument. To cover a wide variety of mass peaks, the $^{78}$Kr isotope with a low SNR and high mass, the $^{80}$Kr isotope with a medium SNR and high mass, as well as the $^{20}$Ne isotope with a high SNR and low mass, and the $^{22}$Ne isotope with a low SNR and low mass have been evaluated. Because $^{21}$Ne is below the noise limit (compare Fig. 2) and the mass resolution in this measurement is too low for accurate automatic analysis of $^{82}$Kr, $^{83}$Kr, $^{84}$Kr and $^{86}$Kr, terrestrial abundances have been assumed for these isotopes. The corresponding analysis of different fitting functions and numerical integration methods, using the previously described peak window finding algorithm, is illustrated in Fig. 4.

It can be seen from Fig. 4 that the results with the highest accuracy are obtained with the numerical methods: MC integration and Simpson integration. In each of the cases analysed, the deviation of the isotope abundance resulting from our analysis from the terrestrial abundance is smallest for these two methods. Both methods, MC integration and Simpson integration, approximately yield the same results. For the fitting models, we find that the Voigt function matches least to the measured mass peaks and the EMG function, as well as Voigt function failed to fit for some of the isotope peaks. Furthermore, the calculation time has to be considered, which is the shortest for the Simpson integration with < 0.5 s for all listed isotopes. About 3 s is needed for the calculation by the algorithms using a fitting method. In contrast, the MC integration takes considerably more time, namely, 8 h for all listed isotopes. From this investigation, it can be concluded that numerical Simpson integration is the most suitable method for our purpose of automatic peak area calculation. Not only is it very accurate, but it is also extremely robust and fast in comparison with all other methods discussed.

Dependence of accuracy of isotope ratios on SNR

The analysis of the relative accuracy of isotope ratios as a function of the SNR shows an inverse proportional relationship between the relative accuracy of isotope ratios and SNR. This correlation is expected. Under the assumption of constant peak widths, the SNR is proportional to the peak areas. In Fig. 5, the relative accuracy of isotope ratios, assuming terrestrial abundances, is shown as a function of the SNR. Blue data points in Fig. 5 represent the

Figure 6. Relative accuracy of isotope ratios of $^{17}$O, $^{22}$Ne, $^{84}$Kr and $^{124}$Xe calculated for different mass resolutions. The results were obtained from simulated spectra with Gaussian peaks. [Colour figure can be viewed at wileyonlinelibrary.com]
measured data. In addition, the results from simulated spectra with Gaussian peaks, using SNR, mass resolution and a uniform distributed noise level equivalent to the measurement, are illustrated by the red squares. The measurement results, as well as the simulation results, are in good agreement with the trend line, which indicates a proportionality SNR\(^{1.5}\). Figure 5 demonstrates that relative accuracies for isotope ratios from worst case 1, for mass peaks close to the noise level (SNR = 1), down to 10 ppm can be achieved for a dynamic range of 10\(^5\), i.e. SNR = 10\(^{-5}\).

**Dependence of accuracy of isotope ratios on mass resolution**

For a Gaussian peak, the SNR is proportional to the area of the peak and inversely proportional to the width of the peak, if we consider the maximum of the Gaussian peak at its mean value. Regarding the mass resolution \(m/\Delta m\), \(\Delta m\) can be represented by the width of the peak in the case of a Gaussian function. Therefore, for a constant peak area, i.e. a constant number of detected ions, the relative accuracy of isotope ratios also is inversely proportional to the mass resolution. This correlation is shown in Fig. 6. The graph shows the analysis of relative accuracies of the isotope ratios of \(^{17}\)O, \(^{22}\)Ne, \(^{84}\)Kr and \(^{124}\)Xe for different mass resolutions. For this analysis, simulated mass spectra with Gaussian mass peaks using SNR and a uniform distributed noise signal equivalent to the measurement were used. From Fig. 6, it can be seen that the proportionality only holds above the minimum mass resolution that is needed by the analysis software to resolve two adjacent mass peaks. As described in Eqn (6) in Section on Peak Area Determination, this minimum mass resolution is proportional to the mass itself; higher masses therefore need also higher mass resolution.

**Dependence of accuracy of isotope ratios on sample width**

The error \(\varepsilon_{\text{Simpson}}\) introduced by integration of the mass peaks, depends on the SNR but is, in addition, related to the sample width \(T_s\) [Eqn (10)]. This leads to the presumption that the accuracy of the peak calculation depends on \(T_s\) as well. For testing this correlation, the noble gas measurements have been resampled several times and analysed with respect to the accuracy of the Ne and Kr isotopes. The original measurement data were acquired with \(T_s = 0.25\) ns (equivalent to a sampling rate of 4 GS/s). In addition, also simulated spectra with Gaussian peaks, using SNR, mass resolution and a uniform distributed noise signal equivalent to the measurement, have been analysed. The results are presented in Fig. 7. From this graph, the correlation of the relative accuracy of isotope ratios and the sample width is obvious. In all cases, the correlation of \(T_s^{-0.5}\), marked by a corresponding trend line, fits both the measured and resampled data, as well as the simulated spectra.

A possible explanation for this trend can be given by considering oversampling and quantisation noise. It is known, see, e.g. SILICON.

![Figure 7](https://wileyonlinelibrary.com/journal/jms)

**Figure 7.** Relative accuracy of isotope ratios calculated for different sample widths, for \(^{20}\)Ne (a), \(^{22}\)Ne (b), \(^{78}\)Kr (c) and \(^{80}\)Kr (d). Panels on the left hand side show results of simulated spectra with Gaussian peaks using SNR, mass resolution and a uniform distributed noise level equivalent to the measurement. [Colour figure can be viewed at wileyonlinelibrary.com]
Automatic data analysis of mass spectra

LABS [36] that for each bit of additional resolution, the signal has to be oversampled by a factor of four. Thus, the sampling frequency \( f_s \) can be written as

\[
f_s = 4^{n_1} f_{Nyq}
\]

(15)

with \( f_{Nyq} \) representing the Nyquist frequency and \( n_1 \) the additional number of bits due to oversampling. Assuming a minimum quantisation noise of one digit, the maximum possible signal-to-quantisation noise ratio \( SQNR_{max} \) covers the full range of resolution with \( N \) the total number of bits of

\[
SQNR_{max} = 2^N = 2^{n_1 + n_0}
\]

(16)

where the total number of bits is a sum of the effective ADC bits \( n_0 \) and the additional number of bits due to oversampling \( n_1 \). Finally, by combining Eqns (15) and (16), the sampling frequencies can be expressed by the corresponding Nyquist sample width \( T_{Nyq} \) and the sample width \( T_s \), leading to

\[
SQNR_{max} = 2^{n_1} \left( \frac{f_s}{f_{Nyq}} \right)^{0.5} = 2^{n_0} \left( \frac{T_s}{T_{Nyq}} \right)^{0.5}
\]

(17)

which explains the correlation observed in Fig. 7.

Summary and conclusions

A fully automated mass spectrum analysis software has been developed for measurement data obtained with a TOF mass spectrometer. With the help of the automatic mass calibration algorithm, data input, reduction and analysis, to the point of peak area determination of all detected mass peaks, can be performed in a robust, reliable and reproducible way. Additionally, implemented routines for isotope and element analyses allow a trained user to derive quantitative results from the measurement data within minutes. This automatic spectrum analysis software enables the analysis of hundreds of measurements in short time but also yields very accurate results as well as extremely high reproducibility. In addition, the presented algorithms and processes can easily be adapted to every other TOF mass spectrometer system, because all functions have been implemented in a modular way with universal interfaces and all used constants can be edited on top level.

By applying a numerical Simpson integration and developing the corresponding error estimation, we were able to carry out accurate isotope abundance analyses with up to 1 ppm relative accuracy. We showed that Simpson integration yields more accurate results than common curve fitting models. Additionally, the Simpson integration method is more suitable concerning the robustness of the algorithm compared with the common curve fitting models. Furthermore, Simpson integration is of advantage because of short computation times. It allows the calculation of peak areas for a whole spectrum in a few seconds, whereas the equally accurate and robust MC integration needs several hours for the same analysis. Therefore, the Simpson integration discussed in this article is currently considered to be the most suitable method for automatic peak area determination.

A peak window finding algorithm, a prerequisite for accurate isotope abundance calculation, was developed. For this algorithm, we found a robust local minimum and maximum searching algorithm being most suitable. The algorithm is capable to successfully cover a range from SNR > 1 up to maximum SNR, which is about 10^5 for the NIM instrument used for the measurements presented here. Together with the according routines for background correction, SNR and mass resolution calculation, the automatic peak area calculation algorithm based on Simpson integration has been implemented in the analysis software. The automated and robust peak area calculation algorithm not only allows very accurate data analysis but also guarantees unbiased data processing and highest reproducibility of the results.

Based on the results obtained with simulated mass spectra, we conclude that the quantitative isotope analysis using automatic peak area detection yields results with a relative accuracy of up to 1 per mille for the most abundant isotopes and up to 100 ppm for a SNR of 10^5. We showed that the relative accuracy of isotope ratios is in fact proportional to SNR^{-0.5}, because the calculated peak area is linearly coupled with SNR. This turns out to be useful for the estimation of relative accuracy of isotope ratios at a certain abundance level, for example to predict an estimation error for radioisotope geochronology of material, as presented in Riedo et al. [19]. However, the minimum mass resolution is essential for obtaining highest accuracies. Moreover, we observed that relative accuracy of isotope ratios is inversely proportional to the mass resolution, as SNR is. Mass resolution is linked to the SNR, because a better mass resolution means that the mass peaks are narrower and higher, thus have larger SNR, under the assumption of constant peak area, i.e. constant number of detected ions.

Additionally, we found that the relative accuracy of isotope ratios is proportional to the sampling width \( T_{Nyq}^{-0.5} \). This has been shown for measured and resampled data, as well as for simulated data. Concluding from this result, better accuracies can be achieved by higher sampling frequency of the used ADC system. Nevertheless, dynamic range should be given priority by choosing an appropriate ADC, i.e. data acquisition system, because the dependence of the relative accuracy on SNR is stronger than the dependence on the sampling width.

Acknowledgements

The authors would like to acknowledge the financial support from Swiss National Science Foundation and the PRODEX programme of the Swiss Space Office.

References
