Atmos. Meas. Tech. Discuss., 8, 1629–1669, 2015
www.atmos-meas-tech-discuss.net/8/1629/2015/
doi:10.5194/amtd-8-1629-2015

Atmospheric
Measurement
Techniques
Discussions
Open Access

PTRwid: a new widget-tool for processing PTR-TOF-MS data

R. Holzinger

Institute for Marine and Atmospheric Research, Utrecht, IMAU, Utrecht University, Utrecht, the Netherlands

Received: 5 December 2014 – Accepted: 11 January 2015 – Published: 5 February 2015

Correspondence to: R. Holzinger (r.holzinger@uu.nl)

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

Abstract

PTRwid is a fast and user friendly tool that has been developed to process data from proton-transfer-reaction time-of-flight mass-spectrometers (PTR-TOF-MS) that use HTOF time-of-flight mass-spectrometers from Tofwerk AG (Switzerland). PTRwid is designed for a comprehensive evaluation of whole laboratory or field based studies. All processing runs autonomously and whole laboratory or field campaigns can, in principle, be processed with a few mouse clicks.

Unique features of PTRwid include (i) an autonomous and accurate mass scale calibration, (ii) the computation of an "Unified Mass list" that – in addition to an uniform data structure – provides a robust method to determine the precision of attributed peak masses, and (iii) fast data analysis due to well considered choices in data processing.

1 Introduction

The development of PTR-TOF-MS (Jordan et al., 2009; Graus et al., 2010) has leaped forward this technology mainly through two key features: (i) the high mass resolution allows distinguishing between ions with a fractional mass difference such as, for example, isoprene and furan that are both detected on the same integer mass (69 Da), and (ii) the recording of full mass spectra is an inherent feature of time of flight (TOF) mass spectrometers and the ions to be monitored do not need to be selected beforehand. While the PTR-TOF-MS community is readily exploiting the former feature (e.g. Liu et al., 2013; Veres et al., 2013), there are only a handful publications that evaluate full mass spectra (e.g. Brilli et al., 2014; Park et al., 2013a). One reason for this is that a full evaluation of TOF mass spectra is still challenging and currently only one community based evaluation tool is available (Mueller et al., 2013; https://sites.google.com/site/ptrtof/).

The data processing package that is introduced here pursues different (and unconventional) approaches in several aspects. For example, the mass scale calibration does not rely on peaks of known ion masses and subsequent statistical analysis of ion counts as in Mueller et al. (2013), but on maximising the matches with a library of ion masses. The underlying idea is that $C_xH_yO_{0\text{–}5}$ compounds are always abundantly detected either as constituents in the sample or due to contamination in the experimental setup. Another example is the method to determine the peak shape, which is not retrieved from selected peaks that can be entirely attributed to ions of the same composition (Mueller et al., 2013), but from all detected ion peaks – disregarding the fact that many peaks are distorted due to signal from ions with different composition. The underlying idea is that the superposition of all detected peaks will envelop the true peak shape. Third and last example worthwhile mentioning here is the determination of the precision of detected peak masses, which is not based on statistics and propagation of errors from calibration peaks (Mueller et al., 2013), but on a top-down evaluation of typically hundreds of peak lists obtained from entire field or lab campaigns. Besides all statistical uncertainties, these precisions include also uncertainties associated with the contribution of ions of different compositions to a detected peak (unresolved ions). These three examples constitute essential parts of the backbone of PTRwid that allows autonomous operation with a minimum of user input (e.g. no calibration peaks need to be defined, no peaks to retrieve the peak shape need to be identified).

In order to reduce the computational cost a large fraction of the required analysis – most importantly peak detection and mass scale calibration – is done on the *SumSpectrum*, which is the total of all mass spectra saved in a HDF5 data file. This approach works best when new data files are created every 30–120 min. The core of PTRwid performs basic data processing such as peak detection, mass scale calibration, baseline correction, determination of peak shape and mass resolution, signal corrections for ion peaks that are not fully separated due to the limited mass resolving power of the mass spectrometer, and the computation of raw signals (counts per second) and volume mixing ratios (nmol mol^{-1}) for all detected peaks. In the current version two “extended processing” tools for general use are available: (i) a tool for averaging, categorizing, and merging data, and (ii) a tool to attribute possible chemical formulas to detected ion masses. In future releases dedicated analysis tools can be integrated as additional "extended processing" tools.

PTRwid has been designed for an efficient and consistent analysis of full TOF mass spectra obtained over time periods that range from single lab measurements (> hour) to field campaigns (weeks) and long term monitoring (year). PTRwid is programmed in IDL (Interactive Data Language) and also runs on the free IDL Virtual Machine. The code is based on routines that have been used in our earlier publications (e.g. Holzinger et al., 2010a, 2013; Park et al., 2013b), but has now been cast into a user-friendly graphical interface and is ready to be shared with a broader community. The full source code is available at http://www.imau.nl/ptrwid/. The following sections describe the algorithms and evaluate their performance. A second aim of this paper is familiarizing new users with the PTRwid framework so that the brief information in the appendices is sufficient to successfully start using PTRwid. In that sense also the organization of this paper is unconventional but, in my opinion, efficiently serves the aim of this work.

2 The key features of PTRwid

During the development of PTRwid maximizing autonomous processing and minimizing required user input was emphasised. The organization of the data processing framework is modular: different tasks are typically performed by dedicated procedures. A list of the key procedures is provided in Appendix D. The modular design facilitates adjustments and extensions, e.g. the development and assimilation of new add-on tools that perform dedicated tasks.

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

2.1 Basic data processing

The basic data processing is controlled by the PTRwid main widget window (Fig. 1). Quick start instructions and a brief description of interactive elements of PTRwid are provided in Appendices A and B, respectively.

2.1.1 Organization and recommended data structure

Many parameters that are needed for basic data processing are not calculated from individual measurements, but from the *SumSpectrum* (i.e. the total of all measured mass spectra that are stored in the respective HDF5 file) which is always available as separate entry in HDF5 data files. This strategy largely increases computing speed, but it is not possible to capture drifts in these parameters that might have occurred during the period that is covered by the data in the file. However, under typical operating conditions such drifts should be minor on time scales of 1–2 h. Individual HDF5 files should contain data that cover data of 30–120 min periods, to ensure optimal performance of PTRwid. However the code also works on files that that contain data of longer or shorter periods.

PTRwid creates "*\ w_data\" and "*\ w_data\FL\" sub-directories in the directory containing the HDF5 raw data. All data containing parameters on individual files are stored in "*\ w_data\FL\". Parameter files or processed data can be linked to the raw data through their specific file names that contains the creation time in the format "JJJJ.MM.DD-HHhMMmSSs". The suffix after the creation time indicates the type of data: "PAR" files contain parameter for mass scale calibration, "FL" files contain the m/z values of the detected peaks, "PeakShape" files contain the exact shape of the peaks, and "Baseline" files specify the level of noise.

2.1.2 Peak detection

The *DetectPeaks* routine applies Savitzky-Golay smoothing filters (Vetterling et al., 1992) on the *SumSpectrum* and scans through the smoothed *SumSpectrum* and its derivative (referred to as *Spec* and *dSpec* hereafter), respectively. Depending on the settings of the TOF mass spectrometer, the *SumSpectrum* contains several 10^5 data points which correspond to the signal detected at the TOF-bins (typically 10^{-10} s per TOF-bin). Following strategy is applied to identify peaks:

1. A potential start of a peak is detected at a TOF-bin where *dSpec* exeedes 6 times the median of the absolute value of *dSpec* in the neighbouring range (±1 Da). This TOF-bin marks the point where the signal emerges from the noise and is labelled as *“PeakStart”*.

2. The corresponding TOF-bins of the peak maximum ($dSpec = 0$), and the position at which the slope becomes positive again after the peak maximum ($dSpec > 0$), are identified and labelled *“PeakMax”* and *“PeakEnd”*, respectively[1]

After scanning through *dSpec*, the list with Peak parameters is used to convert the TOF-bin scale to an ion-mass-scale (see next Sect. 2.1.3). Using the ion masses of *“PeakStart”, “PeakMax”* and *“PeakEnd”* the *“peak broadness”* is calculated, i.e. $m[PeakMax]/(m[PeakEnd] - m[PeakStart])$. Peaks of which *“peak broadness”* is outside the range 20–10 000 are dismissed. As a second criterion the signal at *“PeakMax”* must exceed the background by 8 times the variability of the background. Here, the background in the vicinity of the peak is calculated as the median signal around *“PeakMax”* in the range ±0.5 Da. The variability of the background is calculated as the SD of all values in the vicinity of *“PeakMax”* that are below the background. Figure 2 shows a small section of *Spec* and *dSpec* (TOF-bins 100 294–100 644) that corresponds to

---

[1] Note that algorithm determining *PeakEnd* relies on the fact that some level of noise is always present in a real spectrum. In an idealized spectrum the slope would approach zero but not become positive before a new peak emerges.

the mass range 142.6–143.6 Da. According to the procedure described above, five peaks were detected in this section. The range *"PeakStart"* to *"PeakEnd"* is indicated by horizontal red lines below and above the peak in the upper chart, and the red ticks on these lines mark the position of "*PeakMax*".

2.1.3 Mass scale calibration

The conversion of the TOF-bin scale to an ion-mass-scale is done by three different methods, the quality of which is assessed in Fig. 3. All methods convert the scales according to Eq. (1)

$$t_{\mathrm{TOF}} = a * m^{\mathrm{ex}} + t_0, \tag{1}$$

where m and t_{TOF} represent the ion mass and TOF-bin, respectively, and a, t_0, and ex are parameters that can be optimised. Since the ions enter the TOF region with the same kinetic energy, t_{TOF} should be proportional to the square root of m. Thus ex in Eq. (1) should be equal to 0.5. However, in practice better results are obtained if small deviations from 0.5 are allowed for this parameter.

The first method (*CalCrude*, Appendix D) builds on the assumption that the peaks corresponding to H_3O^+ (19.018 Da) and $H_2O{\bullet}H_3O^+$ (37.028 Da) are among the eight largest signals in the mass spectrum. An example spectrum is plotted in the top chart of Fig. 3. The eight largest peaks are plotted in red. For this method the parameter ex is kept constant at 0.5. Parameters a and t_0 are calculated by combining any two of the largest eight peaks and assuming that these peaks correspond to 19.018 and 37.028 Da. Using these parameters the TOF-bin scale is converted to a mass scale and it is checked if the scale conversion produces peaks that can be attributed $H_3^{18}O^+$ (21.022 Da) and $C_3H_6OH^+$ (59.049 Da). These two species have been chosen because they are normally present in any PTR-TOF-MS spectrum – as isotope peak of primary H_3O^+ ions and the signal of protonated acetone (an virtually omnipresent impurity). If positively identified, the parameters a and t_0 are re-calculated based on 21.022 and 59.049 Da. In the case that both, 21.022 and 59.049, are identified for more than one combination of the eight largest peaks (i.e. one must be a false peak attribution!), all detected peaks are matched with a compound library (~ 2400 compounds, see *masslib* function in Appendix D). The correct time to mass conversion yields more matches with the library. Note that this procedure does not require any user input and nevertheless provides a robust method that allows for a reasonable time to mass conversion. The second chart in Fig. 3 shows the result of comparing the obtained peak masses with library compounds. The x axis represents the mass attributed to all detected ion peaks, and the y axis shows the deviation of the attributed mass to the closest match with a library compound. The points plotted in red are considered to be matched as their smallest deviation from a library compound is less than 20 ppm. As can be seen clearly in Fig. 3, for ion masses above 90 Da the deviation starts drifting towards positive values, thereby indicating the limitations of this method that relies on constant ex and two ions with relatively low masses.

The second method (*CalFine*, Appendix D) uses the parameters a, t_0, and ex from the first method and performs a variation of constants. The parameters are optimized by maximising the number of matches with library compounds. The improved parameters resulting from this procedure (third chart of Fig. 3) no longer show the obvious drifts as in chart two of Fig. 3. The number of matches with library compounds increased from 113 to 200.

The third method (*Cal3pt*, Appendix D) allows the user to specify ion masses (see Appendix C for a list with adjustable parameters) in the low, middle, and high region of the mass scale, respectively. In a first step the parameters a, t_0, and ex are calculated by using the specified ion masses and the corresponding TOF-bins of the corresponding "*PeakMax*" values. In a second step the specified ion masses are allowed to vary within given boundaries and the number of matches with library compounds is maximized by performing a variation of constants. The improved parameters resulting from this procedure further increased the number of matches with library compounds from 200 to 232 (bottom chart of Fig. 3).

These three methods yield three sets of parameters a, t_0, and ex. The set that yields most matches with library compounds is chosen for all subsequent analysis.

2.1.4 Baseline signal

The baseline signal consists of electronic noise that is detected in absence of ions. Basically the baseline can be considered the signal between regions where ions are expected to be detected, i.e, the regions between integer masses. This signal has to be subtracted in order to determine the correct peak shape (Sect. 2.1.5) and to calculate the signal that is attributed to an ion (Sect. 2.1.7). Figure 4 shows the baseline of a *SumSpectrum*. The baseline varies over several orders of magnitude and remains enhanced for several 10^{-7} seconds after ion peaks of high intensity.

The baseline is computed according to the following procedure. First, the *SumSpectrum* is divided into segments of 90 nanoseconds. In each segment the following two steps are repeated 7 times: (1) the position with the highest signal is identified and (2) nine nanoseconds of data around this position are symmetrically removed (unless asymmetry is forced by the edges of the 90 ns segment). These steps remove 70 % of the original data in each segment. The baseline values of the segments are calculated as the mean of the remaining 30 % of the data. In Fig. 4 the baseline values of individual segments can be easily seen at TOF-bin numbers in the range $3.5–5.5 \times 10^4$ (blue data row). If the baseline value for an individual segment is calculated to be zero, it is replaced by the minimum of all non-zero segment values. The baseline is calculated by smoothing over the individual segment values (red data in Fig. 4).

Calculating the baseline signal on basis of the *SumSpectrum* can cause high and low biases when high count rates in the *SumSpectrum* are not caused by a constant high signal at a certain mass but by fluctuating signals ranging from low to very high. During the periods with low count rates the baseline is calculated too high, and during the episode with high count rates the baseline is calculated too low. These artefacts may even result in negative values for the computed volume mixing ratio, however, under most circumstances these artefacts are very minor and hardly noticeable. On the other hand, baseline calculations that rely on single measurements (or shorter periods) can become inaccurate due to poor counting statistics.

2.1.5 Peak shape and resolution

Peak shape analysis is needed for three fundamental reasons: (i) to determine the mass resolution, (ii) to set correct boundaries for peak integration (Sect. 2.1.7), (iii) and for corrections of overlapping peaks (Sect. 2.1.7).

In order to determine the peak shape we consider all peaks the maximum signal (in counts per TOF-bin) of which is within following range: (a) a predefined minimum (the default value is 800 counts, see Appendix C), and (b) a maximum which is the larger of either 10 times the minimum signal or 1 % of the maximum signal of the entire *SumSpectrum*. The lower boundary is set to avoid distortion of the peak shape due to poor counting statistics, and the upper boundary is set to avoid distortion by the largest signals that cause saturation in the counting electronics of the mass spectrometer.

The peak shape analysis starts by estimating the mass resolution (defined as full width at half maximum, FWHM) as mean of all calculated "*peak broadness*" values (Sect. 2.1.2). In the following the baseline signal is subtracted and all individual peaks are rescaled (relative to FWHM) and normalised. The left hand chart of Fig. 5 shows the relative (normalised) intensity of all considered individual peaks (137) in the range −4 to 4 times FWHM. Not surprisingly, for a number of peaks the signal does not decrease before and/or after the peak maximum because overlapping peaks interfere. However, all individual peaks together envelop the true peak shape. Technically, the true peak shape is retrieved by calculating the 10 % quantile of all considered peaks and using a Savitzky–Golay filter to obtain a smoothed true peak shape. The right hand chart of Fig. 5 shows that there is little difference between the individual peak at 59.049 Da and the true peak shape (smoothed in blue, unsmoothed in red). In this example the mass resolution has been calculated to be 3371 (FWHM). Assuming quasi-Gaussian peak shape the ratio of the mass resolution and the SD (sigma) of the distribution is constant (i.e., $2 \cdot (2 \cdot \ln(2))^{0.5} = 2.35$). Using this relationship we calculated the fraction of the total signal that is within and outside the boundaries of ±2 and ±4 sigma (right hand chart of Fig. 5). Typically more than 90 % of the total signal is expected within the ±2 sigma range and these are the standard boundaries for peak integration (2.1.7).

2.1.6 Unified mass list

For extended lab and field studies the problem arises that it is unlikely to retrieve exactly the same set of peaks for different measurement intervals. The main reason is trivial, that is, different samples yield different results. Other reasons include the possibility that the operating conditions of the PTR-MS instrument changed over time, and that – within the levels of precision – the same ion may be attributed to a slightly different mass. The "Unified mass list" routine produces an uniform peak list that is obtained from the file peak lists, which in turn are calculated from the *SumSpectra* of individual HDF5 data files (Sect. 2.1.2).

The "Unified mass list" routine executes all routines described above. However, if the maximum signal of a *SumSpectrum* does not exceed a certain threshold value (the default value is 10^6 counts, Appendix C), the *SumSpectra* of the subsequent files are added. The summing of *SumSpectra* is interrupted when the threshold value is exceeded, or the gap between the end of the last measurement in the current file and the start of the first measurement in the following file exceeds 10 min (default value, Appendix C).

A preparatory step is the computation of a bin-mass scale with bin widths of 1 mDa for ion masses below and 8 ppm bins above 125 Da. Thus, above 125 Da the bin with increases steadily and corresponds to 1.6, 3.2, and 4.8 mDa for ion masses of 200, 400, and 600 Da, respectively. All individual file peak lists (one for each HDF5 raw data file) are subsequently loaded and the peak count value of a mass-bin is increased by one for every peak that is detected within the mass range covered by the bin. Figure 6 shows a small section of the detections-per-mass-bin data (DpB) corresponding to the range 142.92–143.22 Da. The data are from a ∼ 2 week field campaign with our thermal-desorption proton-transfer-reaction mass-spectrometer (TD-PTR-MS, Holzinger et al., 2010a) at a rural site in the Netherlands. More than 300 peak lists from individual files have been included each covering a period of 75 min.

In the following the DpB data, the smothed DpB data (by a running mean of five points, DpB_sm), and its derivative (dDpB_sm) are evaluated in order to create the unified mass list. First, all masses where dDpB_sm crosses from positive to negative are added to the "Unified mass list". Second, all peaks are removed that do not fulfil following two criteria: (i) DpB_sm must be larger than both, 0.55 and 5 % of the maximum value of DpB_sm; (ii) the minimum distance between two peaks must be at least 6 bins on the bin mass scale. This corresponds to 6 mDa or 48 ppm for ion masses below or above 125 Da, respectively. The precision of the mass attributed to the peak is calculated as sigma (SD) of a Gaussian fit including 11 data points (i.e. a range of at least 11 mDa or 88 ppm) around a detected peak. An example fit is shown in Fig. 6, where the precision of the peak at 143.070 Da has been calculated to be ±11 ppm. The unified mass list and associated parameters (such as precision and integration boundaries) are saved in the data directory together with a list of possible molecular formulas within the ±2 sigma boundaries for every peak.

2.1.7 Peak integration, applied corrections, calculation of mixing ratios

The *Export* procedure is activated by clicking on one of the *"Export"* button, and implements (together with the associated sub-routines *CorrPoissDead*, *Integrate*, *Overlap*, and *Calcppb*) peak integration, several corrections, and the computation of mixing ratios. The *"file mass list"* and the *"unified mass list"* and different formats to store data can be optionally chosen for the data export (Appendices C and D).

First, the *CorrPoissDead* routine corrects for physical limitations of the ion detection system. The reason for these limitation are recovery times (dead times up to 2×10^{-8} s) that follow an ion detection event, and the inability to discriminate between single and multiple ion detections in a data acquisition interval of typically $1–2 \times 10^{-10}$ s. These effects have been thoroughly studied and appropriate corrections have been developed and tested (Titzmann et al., 2010; Cappellin et al., 2011). PTRwid corrects for these issues using the formula for a combined Poisson and Dead time correction developed by Titzmann et al. (2010). Since the effect on lower signals is very minor, the corrections are applied only for signals that exceed one detection per 1000 data acquisition periods (of typically 35–60 µs). The instrument specific extending and non-extending dead times can be adjusted (Appendix C), 1.3 and 15 ns are used as default values, respectively.

Second, the *integrate()* routine calculates the raw signal for all ions included in the chosen mass list (*unified* or *file mass list*). The signal is calculated as the total signal within two SDs around the peak maximum. In case of overlapping integration boundaries, the boundaries are set at equal distance between the peak maxima (see upper chart of Fig. 2). Note that a fraction of the signal within the integration boundaries of a particular peak may originate from neighbour peaks, if these neighbour peaks are close enough. Furthermore, the fraction outside the ±2 sigma range is not accounted for in the raw signal data.

Third, the *overlap()* routine corrects for the abovementioned deviancies of the raw signal data. Mathematically this represents the problem of solving a system of coupled linear equations that has to be solved for each measurement. Knowledge of peak shape and the distance of a close-by neighbour peak allows to calculate the fraction of the neighbour's total signal that is expected within the integration boundaries. Solving the system of linear equations results in a correct attribution of (i) signal (ion counts) from neighbours that are expected within the integration boundaries, and (ii) ion counts outside the integration boundaries that have been missed by the integrate routine. The result of the *integrate()* and the *overlap()* routines is a dataset that contains the measured signal for every detected peak in units of counts per second (cps). Mixing ratios are subsequently calculated from this dataset.

Fourth, the *calcppb()* routine calculates mixing ratios according to the method outlined in Holzinger et al. (2010b). Input parameter are temperature, pressure and the electrical potential across the PTR-MS reaction chamber (drift tube), and the electrical potential between the last ring of the drift tube and the entrance lens to the TOF region.

The mass-dependent transmission efficiency of ions trough the TOF mass spectrometer is the result of the increasing duty cycle for larger ion masses (Chernushevich et al., 2001). The transmission efficiency is corrected using the simple parameterization that has been suggested and tested by Cappellin et al. (2012). Ageing of the front Multi-Channel-Plate (MCP, the device detecting ions in the TOF) or an insufficient electrical potential between the MPCs produces non-linear mass dependent discrimination that cannot be corrected (Müller et al., 2014). Therefore, a proper tuning and monitoring of TOF parameters during the measurements (according to the recommendations given in Müller et al., 2014) is essential to obtain correct mixing ratios. As default reaction rate constant for the reactions with both H_3O^+ and $H_2O{\cdot}H_3O^+$ a value of $3 \times 10^{-9}\,cm^3\,s^{-1}$ is used. Note, that the reaction rate constant for organics with $H_2O{\cdot}H_3O^+$ can be set to zero (Appendix C), if the reaction with the water-hydronium cluster should not be considered. The primary ion signal is not measured directly but calculated from signals at isotope peaks using the natural abundance of stable oxygen isotopes: the H_3O^+ signal is calculated from the signal measured at 21.022 (mostly $H_3^{18}O^+$) by multiplying with 487, and the $H_2O{\cdot}H_3O^+$ signal can be either calculated from the signal measured at 38.033 (mostly $H_2^{17}O{\cdot}H_3O^+$ or $H_2O{\cdot}H_3^{17}O^+$) by multiplying with 1350 or from the signal measured at 39.033 (mostly $H_2^{18}O{\cdot}H_3O^+$ or $H_2O{\cdot}H_3^{18}O^+$) by multiplying with 242 (see Appendix C).

2.2 Extended processing

Extended processing is activated by clicking on the “Extended Processing” button in PTRwid (Fig. 1). A pop-up menu shows the available tools for extended analysis – currently the “Average and Merge Data” tool, the “Attributing Formulas” tool, and the “Filter sampling” tool. The first two tools are of general use and will be briefly described below. The “Filter sampling” tool is more specialized since it is dedicated to process aerosol filter data obtained with our offline-TD-PTR-MS method (Timkovsky et al., 2015) and will not be discussed here. In the future more add-on analysis tools can be integrated here. Note that, so far, all extended analysis tools use data exports that are based on the *unified mass list* (and cannot be used on data based on the file mass list).

2.2.1 The "Average and Merge Data" tool

The main purpose of this tool is categorizing data and reducing data size by averaging over pre-defined periods. This tool opens a new window (Fig. 7a) that (i) allows displaying data from multiple selected files, (ii) creating index files to categorize data, and (iii) export merged, averaged and categorized data. As an example, Fig. 8a shows a 7.5 h timeline of benzene mixing ratios in ambient air at the CESAR observatory (http://www.cesar-observatory.nl/) recorded with 5 s time resolution, which corresponds to 6 selected data files. Two periods of background measurement (ambient air cleaned through a Pt-catalyst operated at 350 °C) are visible around 1496.7 and 1496.95, and a plume with mixing ratios up to 1 $\mathrm{nmol\,mol^{-1}}$ was observed shortly after 1496.8. The time-period for averaging, and the index that separates background and ambient data can be set in a new window by clicking on the "Create New Index file" button (label 7 in Fig. 7). The index file name and averaging period (in seconds, i.e. 300 s in Fig. 7b) are set in the first row. In the following rows up to ten conditions can be defined that – if met – assign index values to individual measurements. In the example shown in Figs. 7b and 8b only one condition is defined: if the data $m/z = 79.053$ (benzene) is larger than 0.085 $\mathrm{nmol\,mol^{-1}}$ the index value is 100 (ambient air), otherwise the index value is 300 (zero air). Besides ion count data, all collected engineering data can be used to specify conditions. Note that false attributions due to statistical noise are supressed by considering neighbouring data points. Figure 8b shows that the above defined condition (in combination with supressing noise attributions) is sufficient to separate between ambient and zero air. After clicking the "Export" button (Fig. 7a) all data files (specified by file numbers below the "Export" button) are loaded and individual measurements are averaged until (i) the time-period for averaging is reached, or (ii) the index value changes.

Besides defining (up to eight) conditions up to four index values can be manipulated in the "Hardcopy section" (Fig. 7b). This feature has been used, for example, to separate the different desorption temperatures of the two aerosol collectors of our TD-PTR-MS system during the SOAS[2] campaign in Alabama, USA, in summer 2013. The temperature program of the aerosol collectors was externally controlled, started at 50 °C, and was increased every 3 min by 50 °C until the final temperature of 350 °C was reached. Since the temperature program was externally controlled, the PTR-MS instrument recorded only if the PTR-MS was connected to aerosol collector A or B, to which index values of 100 and 200 were assigned, respectively. In the hardcopy section we added 1 to the first 36 measurements (i.e. 3 min since a measurement was recorded every 5 s), 2 for the following 36 measurements and so on (see Fig. B1 and reddish and bluish index values in Fig. 9), whenever the original index value was 100 or 200, respectively. In this way the hardcopy section could be used to separate different desorption temperatures even though this parameter was not directly monitored by the PTR-MS.

The last part in the "New index" window (Fig. 7b) allows to create a "sampling" index for up to three different inlets, which is useful for all in-situ PTR-MS applications that involve a sampling/pre-concentration step. The sampling index can take 2 values (10 or 20) which are typically associated with sampling of ambient or zero air, respectively.

An example for advanced categorization is shown in Fig. 9, which depicts an almost 4 h period of raw data measured at m/z 83.085 Da during the SOAS campaign in summer 2013. During the campaign the PTR-MS connected to three inlets: aerosol sampler A (blue index values, 100–108), aerosol sampler B (red index values, 200–208), and a thermal desorption 3-stage denuder inlet (green index values, 300–308) to sample semivolatile compounds in the gas phase. The corresponding sampling times are indicated as horizontal lines (and matching colors) below the zero-line. The lower lines indicate sampling of background (i.e. particle free ambient air for the aerosol inlets and

---

[2] More information at http://climate.envsci.rutgers.edu/SOAS/SOAS_White_Paper_final.pdf and http://wiki.envsci.rutgers.edu/index.php/Main_Page

zero air for the das phase inlet), and the upper lines indicate sampling of ambient air. The corresponding start and end times of sampling are saved in the exported dataset.

2.2.2 The "Attribute formulas" tool

The "Attribute formulas" tool lists possible empirical formulas in the range ±150 ppm around a detected peak. Since the precision of the detected mass peaks is typically in the range 10–25 ppm, this list contains many more compounds than the list saved along with the unified mass list (Sect. 2.1.6). Constraints are provided through the selected dataset, which is used to check for consistency with the natural abundance of the two stable carbon and nitrogen isotopes, i.e. $^{13}C/^{12}C \approx 0.011$ and $^{15}N/^{14}N \approx 0.0037$. Optionally, the selected dataset can be reduced by providing index values, in which case the analysis is limited to the specified subset.

The tool loops through all possible formulas (referred to as candidates hereafter) and performs following tasks: (i) the stable isotope distribution of the candidate is calculated; (ii) The candidate's major isotopologues are matched with the unified mass list; (iii) in case a match could be found (in the range ±150 ppm) the ratio of the two data rows (corresponding to the major isotopologues) is calculated. In order to reduce the uncertainty due to low signal/high noise ratio values are dismissed when the corresponding values of the data assigned to the candidate are below the 0.6 percentile. Of the remaining 40 % of the ratios the median is calculated and used for further evaluations. A candidate is labelled as "dismissed" if (i) the median suggests that the candidate accounts for less than 10 % of the measured signal, or (ii) if a more abundant isotopologue (at least by a factor of 3) could not be matched with the unified mass list (labelled *"no parent peak"* in Fig. 10). The tool returns a report that lists possible formulas, the fractional mass difference with the detected mass, the maximum fraction of the total signal that can be attributed to the compound (only if constrained by isotopologues), and a statistical summary (mean and median) of the dataset (Fig. 10).

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

3 Performance

The performance is demonstrated by processing 54 h (17–20 June 2013) of measurements with our PTR-TOF 8000 instrument (Ionicon Analytik GmbH, Austria) during the SOAS campaign. Throughout the campaign, ions were injected into the time of flight region of the mass spectrometer every 60 µs and the detector was operated at 10 GHz (0.1 ns time bins). A 5 s time resolution was obtained by internally totalling the signal of 83 333 initial mass spectra. So, throughout the 54 h period more than 3.8×10^4 mass spectra (6×10^5 data points per spectra) were collected and processed. The data were stored in 43 HDF5 data files each containing 900 measurements or 75 min of data.

Processing was done on a laptop computer (purchased in 2013; Dell Latidude E6430) with an Intel (Core™ i7-3720QM CPU @2.60 GHz, 4 GB RAM) processor and 32 bit operating system (Windows 7). The analysis based on the *SumSpectrum* of individual files (Peak detection, mass scale calibration, baseline signal, peak shape, mass resolution, unified mass list) was started by clicking the “*Unified Mass List”* button (Fig. 1) and completed in less than 15 min. The *Unified Mass List* contained 843 ion peaks in the range 16–820 Da. The extraction of these 843 ion signals from the raw data (including Poisson and dead time corrections, correction of overlapping peaks, and the computation of mixing ratios) was started by clicking the “Export” button (Fig. 1) and completed in 46 min. Note that the processing time of this step is largely independent of the number of ion peaks, since a large fraction of the processing time is consumed by opening the compressed HDF5 raw data.

The extended analysis tools require less processing power: averaging and merging according to the categories shown in Fig. 9 is completed in less than 2 min, and creating a report with the “Attribute formulas” tool is a matter of ∼ 20 s.

4 Conclusions and outlook

PTRwid processes PTR-TOF-MS data and runs under IDL or the free IDL Virtual Machine. Much of the analysis is done on the *SumSpectrum*; which reduces the processing time significantly. The main innovative features are the autonomous mass scale calibration, and the computation of a uniform *"unified mass list"*, which also provides a robust method to determine the precision of attributed ion masses.

The modular design allows for flexible adjustments and easy integration of extended processing tools that can be add-on procedures dedicated to specialized tasks. An example of this is the "Filter" tool that allows efficient processing of data collected with our offline-TD-PTR-MS setup (Timkovsky et al., 2015). Several future extensions are already underway and anticipated for the near future. These include (i) a peak modelling tool that allows identifying low intensity peaks the signal of which is completely submerged by high intensity peaks, (ii) a tool to create plots such as van Krevelen diagrams, and presentations of desorption thermograms, or carbon oxidation state, which are useful for the interpretation of TD-PTR-MS data (e.g. Holzinger et al., 2013), and (iii) a tool for exploiting high frequency datasets, for example, to calculate BVOC fluxes according to Park et al. (2013b).

Appendix A: Quick start instructions

A1 IDL virtual machine

- The IDL virtual machine can be freely downloaded after creating an user account at http://www.exelisvis.com/. Note that the "Download" section is blocked until the account verification is completed, which requires ∼ 24 h and thoroughly completed forms.
- Start the IDL virtual machine and load the *PTRwid.sav* file.

- Starting up takes ∼ 10 s, time which is needed to compute the extended version of the compound library (∼ 41 000 entries). After this, the empty PTRwid control window (Fig. 1) appears on the screen.

A2 IDL

- Runs on IDL Version: 7.0.0 or higher.
- Open the *PTRwid.pro* file. Compile the code (two times!). Type PTRwid in the command line, press enter.
- Starting up takes ∼ 10 s, time which is needed to compute the extended version of the compound library (∼ 41 000 entries). After this, the empty PTRwid control window (Fig. 1) appears on the screen.

Appendix B: Brief description of the elements of the PTRWID

PTRwid control window (Fig. 1)

Select data directory field (label 1 in Fig. 1): A click with the right mouse button activates a context menu with two options: (i) “browse” to select a new directory with HDF5-data, and (ii) “last settings” to load the directory used in the previous session.

File list field (label 2 in Fig. 1): After the directory has been specified, all HDF5-data files (including files in sub-directories!) are listed. Note that rather than displaying file names the file-creation-times are displayed. The files are chronologically ordered (by creation-time) and an identifier number (file index) is attributed to each file. Individual files can be selected in the list. Available engineering data and detected mass peaks (File mass list) are displayed in the engineering data and the mass peaks fields, respectively (labels 3 and 4 in Fig. 1).

Engineering data field (label 3 in Fig. 1): Individual engineering data can be selected and are plotted.

Mass peaks field (label 4 in Fig. 1): Individual mass peaks can be selected and are plotted. A droplist (label 5 in Fig. 1) allows to select different presentations.

Unified mass list button (label 6 in Fig. 1): Calculates/loads the unified mass list. Calculating requires basic evaluation of all data files and takes some time. Loading is completed within 2–3 s. On the first (3rd, 5th, etc.) click on the unified mass list button, the mass peaks of the unified mass list are displayed in the *Mass peak field* and can be selected and reviewed. On the second (4th, 6th, etc.) click on the unified mass list button, an integer mass list is displayed in the *Mass peak field* and all mass peaks on the selected integer mass can be reviewed.

Pulldown menue (label 7 in Fig. 1): allows to select the file mass list (of the file selected in the *file list field*) or the unified mass list which are displayed in the *Mass peak field*, respectively.

Export button (label 8 in Fig. 1): Implements bulk data processing as described in Sect. 2.1.7. The files to be exported can be selected in the text fields below the button (label 9 in Fig. 1). The export is based on the file list selected by the pulldown menue (label 7 in Fig. 1).

Extended processing button (label 10 in Fig. 1): opens a new window from which available extended data processing tools can be started.

“Average and Merge Data” tool widged windows (Fig. 7)

A short description of all interactive elements of the main window (Fig. 7a) is given below. Additional information on the *New index window* is provided in Fig. B1.

Select data directory field (label 1 in Fig. 7a): A click with the right mouse button activates a context menu which allows to select a data directory with data exports (usually the “w_data” subdirectory of the HDF5 raw data directory).

File list field (label 2 in Fig. 7a): After the directory has been specified, all available data files are listed. Multiple files can be selected in the list. Available engineering data and the mass peaks of the *Unified mass list* are displayed in the engineering data and the mass peaks fields, respectively (labels 3 and 4 in Fig. 7a).

Engineering data field (label 3 in Fig. 7a): Individual engineering data can be selected and are plotted.

Mass peaks field (label 4 in Fig. 7a): Individual mass peaks can be selected and are plotted.

Droplists (label 5 in Fig. 7a): These four droplists allow to specify the data for the parameters (pdrift, udrift, udx, Tdrift) which are needed to calculate the mixing ratios from ion counts (cps). Note that volume mixing ratios are re-calculated here based on averaged counts of primary and product ion signals.

Index file field (label 6 in Fig. 7a): A click with the right mouse button activates a context menu which allows to select an index file (*.ind). When an index file is selected the displayed data are categorized according to the index parameter.

New index button (label 7 in Fig. 7a): This button opens a new window (Fig. 7b) in which a new index can be created and saved.

Export button (label 8 in Fig. 7a): Averages and merges all data selected in the text fields below the button (label 9 in Fig. 7a) into one data file.

Appendix C: List of adjustable parameters

When PTRwid is started for the first time a file (C:\PTRwid\ parfile.txt) is created the only content of which is the path and name of the parameter file (e.g. C:\PTRwid\ v001_Nov_13_2014_par.txt). A new parameter file is created whenever a new version of PTRwid is started for the first time. Every row in the parameter file contains only one variable, but there are also empty rows and header rows. Rows containing variables do not contain whitespaces and start with the "_" character, followed by the variable name, the "=" character, and the value of the variable. The default values are useful for most of our studies with the Utrecht PTR-TOF-MS, however different instruments and/or different applications may require adjustments to these parameters. This can be easily done in any text editor. E.g., the default for the variable "aMin" is 16 000 and the entry in the corresponding row is "_aMin=16000". If this value is to be changed to 17 000 the entry has to be changed to "_aMin=17000". Table C1 provides a list of variables that can be adjusted in the parameter file.

Appendix D: Overview of routines

PTRwid uses more than 80 custom made routines the code of which is compiled in the PTRwid.pro file (available at www.imau.nl/ptrwid/). Table D1 provides some information on the key routines. Routines not included in Table D1 perform secondary tasks (such as saving and opening files, plotting data into graph windows) or auxiliary tasks (such as accessing HDF5 data, providing time formats, and many more).

Acknowledgements. Valuable and appreciated feedback on draft manuscripts has been provided by Pawel K Misztal. PTRwid relies on important research on PTR-TOF-MS. I greatly appreciate the work done by various groups, notably the work presented by Cappellin et al. (2011, 2012), Müller et al. (2013, 2014), and Titzmann et al. (2010). I used the quantile() function written and provided by Jarle Brinchmann from Leiden University, NL, which is greatly appreciated. I also want to thank numerous colleagues and students for spotting inconsistencies and bugs which helped to improve the code over the years. Funding of our TD-PTR-MS by the Netherlands Organization for Scientific Research (NWO, Grant 834.08.002) is gratefully acknowledged. The measurements at the CESAR observatory were funded by ACTRIS (FP7-262 254), and measurements for the SOAS campaign were funded by the Electric Power Research Institute (EPRI), grant agreement 00-10000358.

References

Brilli, F., Gioli, B., Ciccioli, P., Zona, D., Loreto, F., Janssens, I. A., and Ceulemans, R.: Proton Transfer Reaction Time-of-Flight Mass Spectrometric (PTR-TOF-MS) determination of volatile organic compounds (VOCs) emitted from a biomass fire developed under stable nocturnal conditions, Atmos. Environ., 97, 54–67, doi:10.1016/j.atmosenv.2014.08.007, 2014.

Cappellin, L., Biasioli, F., Schuhfried, E., Soukoulis, C., Märk, T. D., and Gasperi, F.: Extending the dynamic range of proton transfer reaction time-of-flight mass spectrometers by a novel dead time correction, Rapid Commun. Mass Sp., 25, 179–183, 2011.

Cappellin, L., Karl, T., Probst, M., Ismailove, O., Winkler, P. M., Soukoulis, C., Aprea, E., Märk, T. D., Gasperi, F., and Biasioli, F.: On quantitative determination of volatile organic compound concentrations using proton transfer reaction time-of-flight mass spectrometry, Environ. Sci. Technol., 46, 2283–2290, doi:10.1021/es203985t, 2012.

Chernushevich, I. V., Loboda, A. V., and Thomson, B. A.: An introduction to quadrupole-time-of-flight mass spectrometry, J. Mass Spectrom., 36, 849–865, 2001.

Graus, M., Muller, M., and Hansel, A.: High resolution PTR-TOF: quantification and formula confirmation of VOC in real time, J. Am. Soc. Mass Spectr., 21, 1037–1044, 2010.

Holzinger, R., Kasper-Giebl, A., Staudinger, M., Schauer, G., and Röckmann, T.: Analysis of the chemical composition of organic aerosol at the Mt. Sonnblick observatory using a novel high mass resolution thermal-desorption proton-transfer-reaction mass-spectrometer (hr-TD-PTR-MS), Atmos. Chem. Phys., 10, 10111–10128, doi:10.5194/acp-10-10111-2010, 2010a.

Holzinger, R., Williams, J., Herrmann, F., Lelieveld, J., Donahue, N. M., and Röckmann, T.: Aerosol analysis using a Thermal-Desorption Proton-Transfer-Reaction Mass Spectrometer (TD-PTR-MS): a new approach to study processing of organic aerosols, Atmos. Chem. Phys., 10, 2257–2267, doi:10.5194/acp-10-2257-2010, 2010b.

Holzinger, R., Goldstein, A. H., Hayes, P. L., Jimenez, J. L., and Timkovsky, J.: Chemical evolution of organic aerosol in Los Angeles during the CalNex 2010 study, Atmos. Chem. Phys., 13, 10125–10141, doi:10.5194/acp-13-10125-2013, 2013.

Jordan, A., Haidacher, S., Hanel, G., Hartungen, E., Mark, L., Seehauser, H., Schottkowsky, R., Sulzer, P., and Mark, T. D.: A high resolution and high sensitivity proton-transfer-reaction time-of-flight mass spectrometer (PTR-TOF-MS), Int. J. Mass Spectrom., 286, 122–128, 2009.

Liu, Y. J., Herdlinger-Blatt, I., McKinney, K. A., and Martin, S. T.: Production of methyl vinyl ketone and methacrolein via the hydroperoxyl pathway of isoprene oxidation, Atmos. Chem. Phys., 13, 5715–5730, doi:10.5194/acp-13-5715-2013, 2013.

Müller, M., Mikoviny, T., Jud, W., D'Anna, B., and Wisthaler, A.: A new software tool for the analysis of high resolution PTR-TOF mass spectra, Chemometr. Intell. Lab., 127, 158–165, 2013.

Müller, M., Mikoviny, T., and Wisthaler, A: Detector aging induced mass discrimination and non-linearity effects in PTR-ToF-MS, Int. J. Mass Spectrom., 365, 93–97, 2014.

Park, J. H., Goldstein, A. H., Timkovsky, J., Fares, S., Weber, R., Karlik, J., and Holzinger, R.: Active atmosphere-ecosystem exchange of the vast majority of detected volatile organic compounds, Science, 341, 643–647, 2013a.

Park, J.-H., Goldstein, A. H., Timkovsky, J., Fares, S., Weber, R., Karlik, J., and Holzinger, R.: Eddy covariance emission and deposition flux measurements using proton transfer reaction – time of flight – mass spectrometry (PTR-TOF-MS): comparison with PTR-MS measured vertical gradients and fluxes, Atmos. Chem. Phys., 13, 1439–1456, doi:10.5194/acp-13-1439-2013, 2013b.

Timkovsky, J., Dusek, U., Henzing, J. S., Kuipers, T. L., Röckmann, T., and Holzinger, R.: Offline thermal-desorption proton-transfer-reaction mass spectrometry to study composition of organic aerosol, J. Aerosol Sci., 79, 1–14, 2015.

Titzmann, T., Graus, M., Müller, M., Hansel, A., and Ostermann, A.: Improved peak analysis of signals based on counting systems: illustrated for proton-transfer-reaction time-of-flight mass spectrometry, Int. J. Mass Spectrom., 295, 72–77, 2010.

Veres, P. R., Faber, P., Drewnick, F,. Lelieveld, J., and Williams, J.: Anthropogenic sources of VOC in a football stadium: assessing human emissions in the atmosphere, Atmos. Environ., 77, 1052–1059, doi:10.1016/j.atmosenv.2013.05.076, 2013.

Vetterling, W. T., Press, W. H., Teukolsky, S. A., and Flannery, B. P.: Numerical Recipes Example Book C (The Art of Scientific Computing), Chapter 14, Cambridge University Press, ISBN-10: 0521437202, 1992.

Table C1. List of adjustable parameters.

Variable name	Default value	Associated with	Comment
Desired_Min_Signal	1.0×10^6	Unified mass list	If the maximum value (unit is counts per TOF-bin) in the SumSpectrum is lower, the SumSpectrum of the following file is added.
Max_Time_Gap	10	Unified mass list	If the gap between the last measurement of the current file and the first measurement in the following file is larger (unit is minutes), no SumSpectrum is added even if the signal is lower than the value of "Desired_Min_Signal".
Min_Mass	10	Peak search	Peaks below this value (unit is Da) are deleted
aMin	16 000	Mass scale calibration	lowest allowed value for *a*
aMax	19 000	Mass scale calibration	highest allowed value for *a*
t0Min	0	Mass scale calibration	lowest allowed value for t_0
t0Max	100 000	Mass scale calibration	highest allowed value for t_0
M1a	21.0221	3-point mass scale calibration	Low mass, option 1
M1b	42.034	3-point mass scale calibration	Low mass, option 2
M2a	59.0491	3-point mass scale calibration	Medium mass, option 1
M2b	116.906	3-point mass scale calibration	Medium mass, option 2
M3a	205.195	3-point mass scale calibration	High mass, option 1
M3b	355.0731	3-point mass scale calibration	High mass, option 2
tol_ppm	300	3-point mass scale calibration	Boundaries for the variation of constants, i.e. The low, medium, and high mass are varied by the specified value (in ppm)
MinSig	800	Peak shape	Minimum signal (unit is counts per TOF-bin) of the peak maximum to be included for peak shape computation
k19	3	Compute volume mixing ratio	Reaction rate constant for reaction with H_3O^+ (unit $10^{-9}\,cm^3\,s^{-1}$).
k37	3	Compute volume mixing ratio	Reaction rate constant for reaction with $H_2O{\cdot}H_3O^+$ (unit $10^{-9}\,cm^3\,s^{-1}$).
m38	1	Compute volume mixing ratio	If 1 the signal at $m/z = 38.033$ is used to calculate the $H_2O{\cdot}H_3O^+$. Set this value to 0 to use $m/z = 39.033$ instead.
ExportOne	0	Export	If set to 1, only one data file is exported (existing files are overwritten) and jpg pictures depicting the entire dataset (similar to Figure D2 in Holzinger et al., 2010a) are saved for quality control purposes.

Table C1. Continued.

Variable name	Default value	Associated with	Comment
CorrectOverlap	1	Export	If set to 0, no corrections for the expected signal outside the integration boundaries and from nearby neighbour peaks are applied.
SaveCsv	1	Export	If set to 0, exported files are only saved in *.fdt files[a] (no comma separated values, *.csv files)
JunkSize	1.5×10^8	Export	Maximum number of values that are loaded as one junk from the HDF5 raw data file. Low numbers increase the processing time, but if the value is too high insufficient virtual memory will cause the program to crash.
CorrectionON	1	Export	If set to 0, no poisson and dead time corrections will be performed.
NonExtendingDeadTime	15	Export	Unit is nanoseconds. More information in Titzmann et al. (2010) and Müller et al. (2013).
ExtendingDeadTime	1.3	Export	Unit is nanoseconds. More information in Titzmann et al. (2010) and Müller et al. (2013).

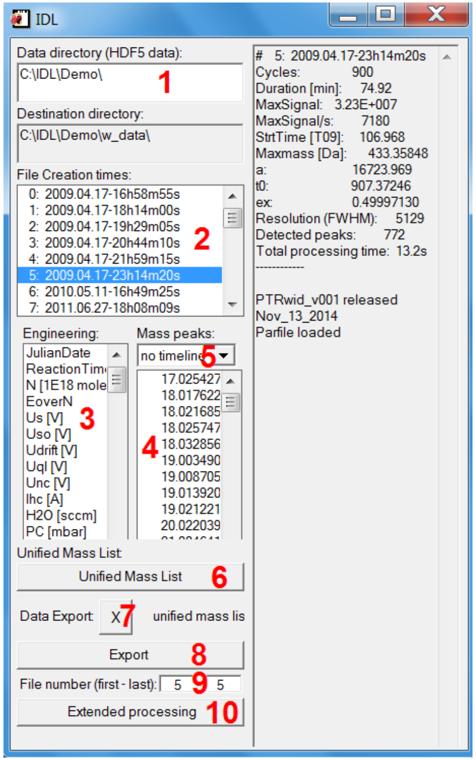
[a] The float data format (*.fdt) saves multi-dimensional float-arrays into a one-dimensional file (one value per row). The first value gives the dimension and the following values give the number of elements in each dimension. Below this “header” follow the values of each element of the array.

Table D1. Key routines used by PTRwid.

Procedure name	Activated by	Essential processing
EXPORT	PTRWID, button (label 8 Fig. 1)	Calls CORRPOISSDEAD, INTEGRATE, OVERLAP, CALCPPB
SINGLEFILE	PTRWID, list (label 2 Fig. 1)	Calls PS1
UNIMASSLIST	PTRWID, list (label 6 Fig. 1)	Calls PS1, PS2
EXPRT	Average and Merge Data tool, button (label 8 Fig. 7a)	Calls INDEX, AVERAGING, CALCPPB
IDENTIFICATION	Attribute formulas tool	Computes possible formulas and evaluated consistency with natural 13C/12C ratio
AVERAGING	EXPRT	averages data according to categorisation provided by INDEX
CAL3PT	PS1	Performs a 3 point mass scale calibration and optimisation.
CALCPPB	EXPORT, EXPRT	Computation of volume mixing ratios.
CALCRUDE	PS1	Performs a crude mass scale calibration based on m/z 21.022 and 59.049.
CALFINE	PS1	Optimizes the crude mass scale calibration.
CORRPOISSDEAD	EXPORT	Correction of Poisson statistics and detector dead-times

Table D1. Continued.

Procedure name	Activated by	Essential processing
CORRTR	CALCPPB	Mass dependent transmission efficiency of the TOF according to Cappellin et al. (2012). Other custom transmission efficiency can be hard-coded here.
DETECTPEAKS	PS1	Scans the *SumSpectrum* for peaks.
INDEX	EXPRT	Provides index vectors according to the parameters specified in the *New index window*
INTEGRATE	EXPORT	Integration of junks of raw HDF5 data.
MASSLIB	PTRWID, PS1	Computes library including $C_{1-40}H_{0-82}O_{0-5}$ (~ 2400 compounds) and $C_{1-40}H_{0-82}O_{0-16}N_{0-2}$ if the/EXTENDED keyword is set (~ 41 000 compounds).
OVERLAP	EXPORT	Correction of overlapping peaks.
PS1	SINGLEFILE, UNIMASSLIST	Calls DETECTPEAKS, CALCRUDE, CALFINE, CAL3PT. Retrieves baseline signal, peak shape and resolution (FWHM).
PS2	UNIMASSLIST	Computes parameters of the unified mass list, most importantly the precision of the computed ion mass.



Data directory (HDF5 data): C:\IDL\Demo\

Destination directory: C:\IDL\Demo\w_data\

File Creation times:

```
0:  2009.04.17-16h58m55s
1:  2009.04.17-18h14m00s
2:  2009.04.17-19h29m05s
3:  2009.04.17-20h44m10s
4:  2009.04.17-21h59m15s
5:  2009.04.17-23h14m20s
6:  2010.05.11-16h49m25s
7:  2011.06.27-18h08m09s
```

Engineering:

```
JulianDate
ReactionTim
N [1E18 mole
EoverN
Us [V]
Uso [V]
Udrift [V]
Uql [V]
Unc [V]
Ihc [A]
H2O [sccm]
PC [mbar]
```

Mass peaks: no timeline

```
17.025427
18.017622
18.021685
18.025747
18.032856
19.003490
19.008705
19.013920
19.021221
20.022039
```

Data Export: X unified mass lis

File number (first - last): 5 5

```
#   5:  2009.04.17-23h14m20s
Cycles:             900
Duration [min]:      74.92
MaxSignal:    3.23E+007
MaxSignal/s:        7180
StrtTime [T09]:     106.968
Maxmass [Da]:        433.35848
a:                  16723.969
t0:                 907.37246
ex:                 0.49997130
Resolution (FWHM):      5129
Detected peaks:        772
Total processing time:  13.2s
------------


PTRwid_v001 released
Nov_13_2014
Parfile loaded
```

Figure 1. Main PTRwid control window. In the top-left text field the user selects a data directory. The creation times (i.e. start of measurement) of all HDF5 files in the data folder (including sub directories) are listed in the field labelled "File creation times". Individual files can be selected and first processing results are displayed in the session log on the right hand side of the window. The names of available engineering data and the detected mass peaks of the selected file are displayed in separate lists, respectively. The interactive elements are labelled with bold red numbers and briefly described in Appendix B.

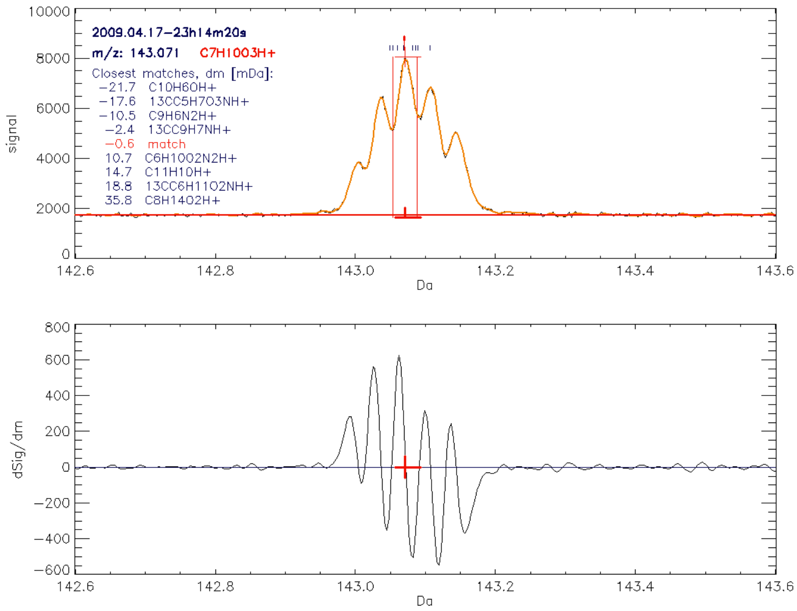


Figure 2. A small section of *Spec* (upper chart, smoothed in orange, unsmoothed in black) and *dSpec* (lower chart, smoothed only). The baseline signal (Sect. 2.1.4) is plotted as red line in the upper chart. The peak search routine detected 5 peaks at 143.002, 143.036, 143.071, 143.105, and 143.142, respectively. The range "*PeakStart*" to "*PeakEnd*" is indicated by a horizontal red lines above the x axis and in the upper chart area, respectively. The red ticks on these lines mark the position of "*PeakMax*". The vertical red lines in the upper chart indicate the integration boundaries (Sect. 2.1.7.) that correspond to a range of ±2 sigma of the quasi-Gaussian peak shape (Sect. 2.1.5.). The black and red ticks above the selected peak in the upper chart correspond to ion masses of the closest matches in the compound library (see masslib() function in Appendix D). The largest peak at 143.071 has been matched to $C_7H_{10}O_3H^+$; other compounds in the vicinity are plotted in the chart area together with their deviation from the detected peak.

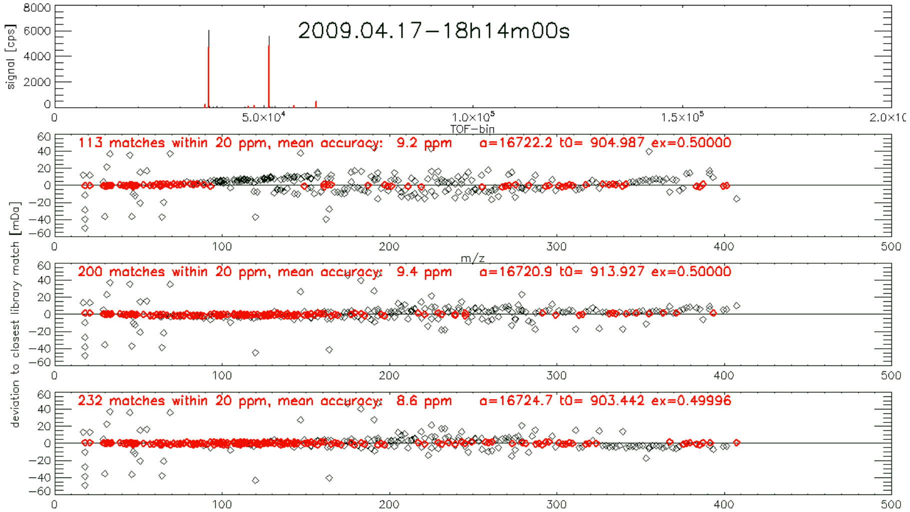


Figure 3. The top chart shows the SumSpectrum, the largest eight peaks are marked in red. Charts 2–4 demonstrate the quality of the tree methods of mass scale calibration (see text for details): each point represents a detected peak (ion mass, Da, on the *x* axis). The peaks were compared with the compound library. The *y* axis shows deviation to the closest library match (in mDa). The quality of the scale conversion is assessed by the number of matches (red data) with library compounds. The mass scale calibration is the best for method 3 (chart 4).

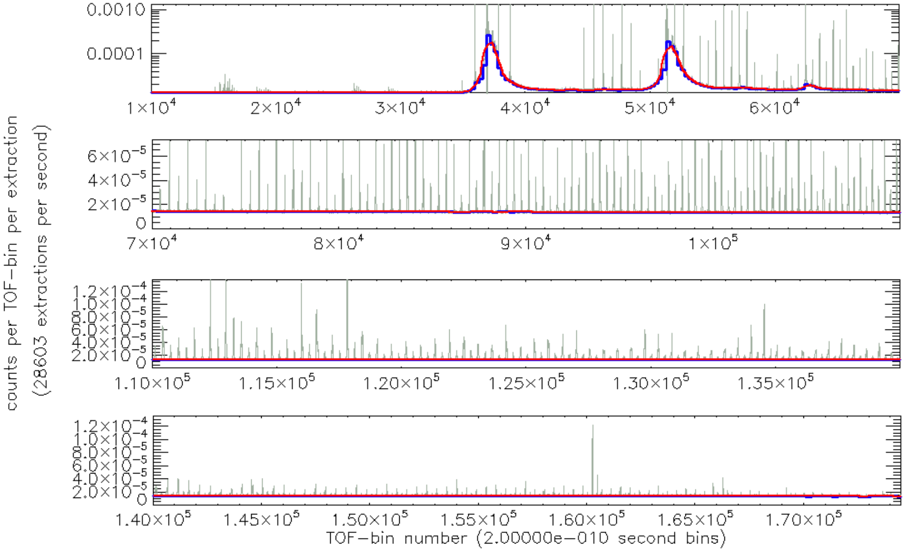


Figure 4. An example *SumSpectrum* zoomed in to the baseline (grey data). The blue stepwise data show the baseline values of individual 90 nanosecond segments (see text for details). The baseline is obtained by smoothing over these individual baseline values (red data).

AMTD
8, 1629–1669, 2015

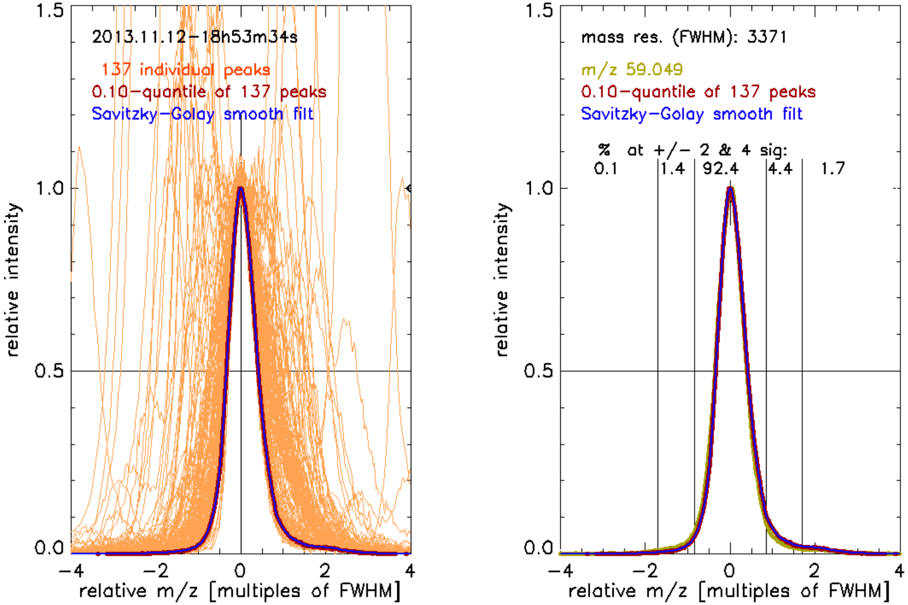


Figure 5. The left hand chart shows 137 normalized and re-scaled individual peaks retrieved from an example SumSpectrum. The individual peaks envelop the true peak shape (blue line) which is retrieved by smoothing the 0.1 percentile of the relative intensity of all 137 peaks (brown line, barely visible because superimposed by the blue line). The right hand chart shows the true peak shape (as in the left hand chart) together with an example peak detected at m/z 59.049. The calculated mass resolution (FWHM) and the signal fractions expected within and outside the ±2 and ±4 sigma boundaries are labelled in the plot area. 92.4 % of the total signal is expected within the ±2 sigma boundary. A small shoulder is apparent at the high mass end and therefore 1.7 % of the signal are detected outside the +4 sigma boundary.

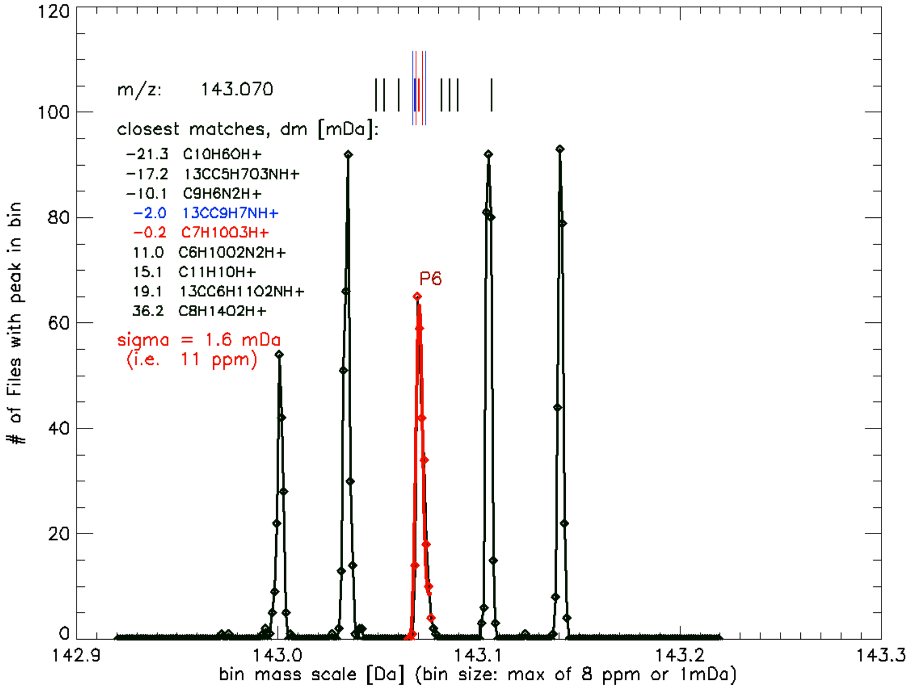
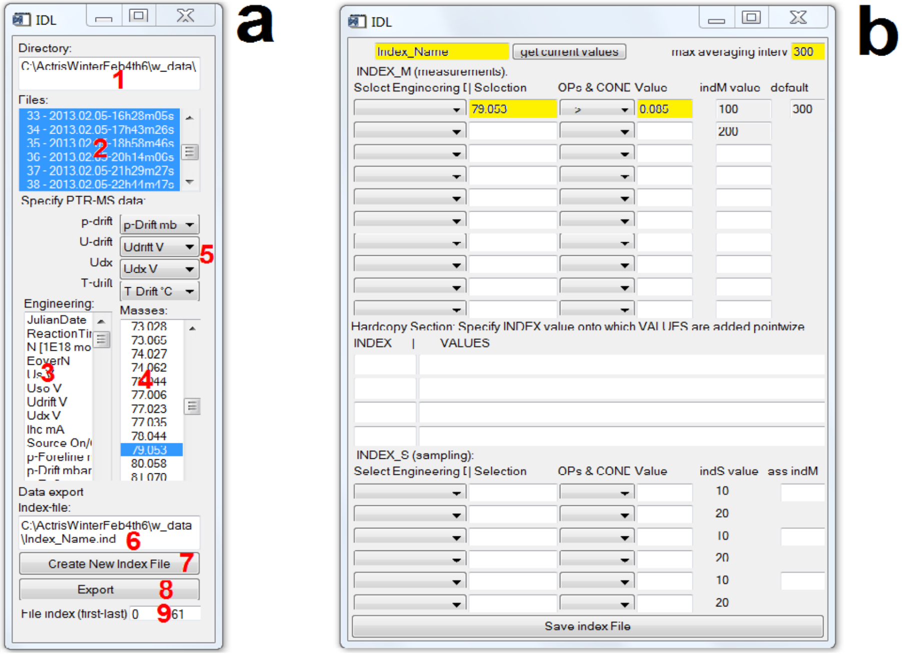


Figure 6. Histogram of peak detections in 8 ppm mass bins in the range 142.92–143.22 Da. More than 300 peak lists from individual files each covering a period of 75 min have been included (i.e. the data are from a ∼ 2 week field campaign). The precision of a peak at 143.070 has been determined as the sigma (SD) of a Gaussian fit. The red markers are the 11 points that have been fitted, and the red line shows the Gaussian fit. The long red and blue vertical lines above the peak mark the ±1 and ±2 sigma boundaries, respectively. The short vertical lines mark the position of close matches with compounds in the library (also plotted in the chart area). Compounds plotted in red and blue correspond to compounds that are within ±1 and ±2 sigma, respectively.



Directory: C:\ActrisWinterFeb4th6\w_data\

Files:
33 - 2013.02.05-16h28m05s
34 - 2013.02.05-17h43m26s
35 - 2013.02.05-18h58m46s
36 - 2013.02.05-20h14m06s
37 - 2013.02.05-21h29m27s
38 - 2013.02.05-22h44m17s

Specify PTR-MS data:
p-drift: p-Drift mb
U-drift: Udrift V
Udx: Udx V
T-drift: T Drift °C

Engineering: JulianDate, ReactionTi, N [1E18 mo, EoverN, Us, Uso V, Udrift V, Udx V, Ihc mA, Source On/, p-Foreline r, p-Drift mbar

Masses: 73.028, 73.065, 74.027, 74.062, 74.044, 77.006, 77.023, 77.035, 78.044, 79.053, 80.058, 81.070

Data export
Index-file: C:\ActrisWinterFeb4th6\w_data\Index_Name.ind
File index (first-last) 0 61

Index_Name
max averaging interv 300

INDEX_M (measurements).

Select Engineering [	Selection	OPs & COND	Value	indM value	default
	79.053	>	0.085	100	300
				200	

Hardcopy Section: Specify INDEX value onto which VALUES are added pointwize
INDEX | VALUES

INDEX_S (sampling):

Select Engineering [	Selection	OPs & COND	Value	indS value	ass indM
				10	
				20	
				10	
				20	
				10	
				20	

Figure 7. The "Average and Merge Data" tool widged windows. The main window **(a)** allows selecting and displaying data (example in Figs. 8 and 9) in order to produce and test useful data categorizations (Figs. 8b and 9) by creating an index file. The interactive elements are labelled with bold red numbers and briefly described in Appendix B. The "New Index File" widget window **(b)** allows creating complex data categories by using up to 10 conditions. Up to 4 index values can be manipulated in the hardcopy section, and up to 3 sampling indices can be created for in situ PTR-MS applications.

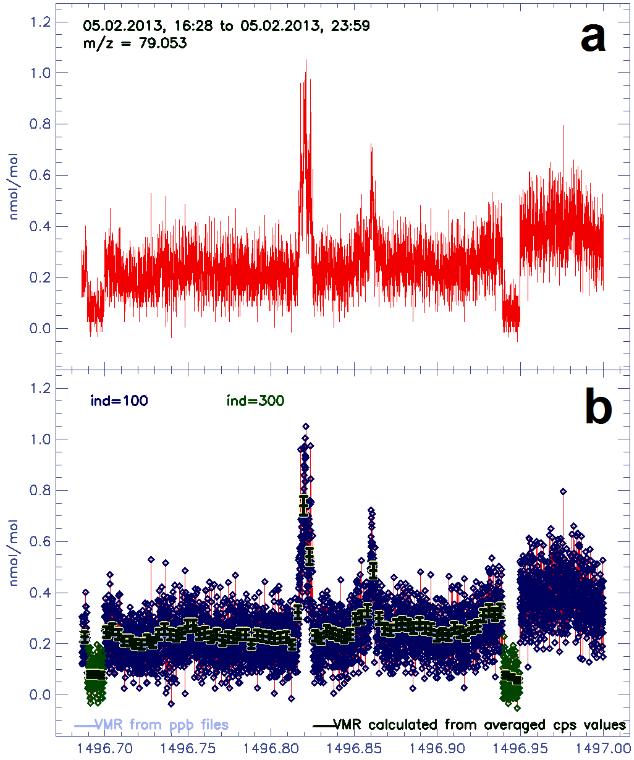


Figure 8. A seven hour timeline of $m/z = 79.053$ (benzene) collected at the CESAR observatory (http://www.cesar-observatory.nl/) during the ACTRIS winter campaign in January/February 2013. The time axis displays days since 1 January 2009; the displayed period (in ordinary date format) is labelled on the top. The upper chart **(a)** displays the raw data. In the lower chart, markers of different colors correspond to index values as defined in the "New Index File" widget window (Fig. 7b). The black lines correspond to the 300 s means and the error bars indicate the standard error due to counting statistics.

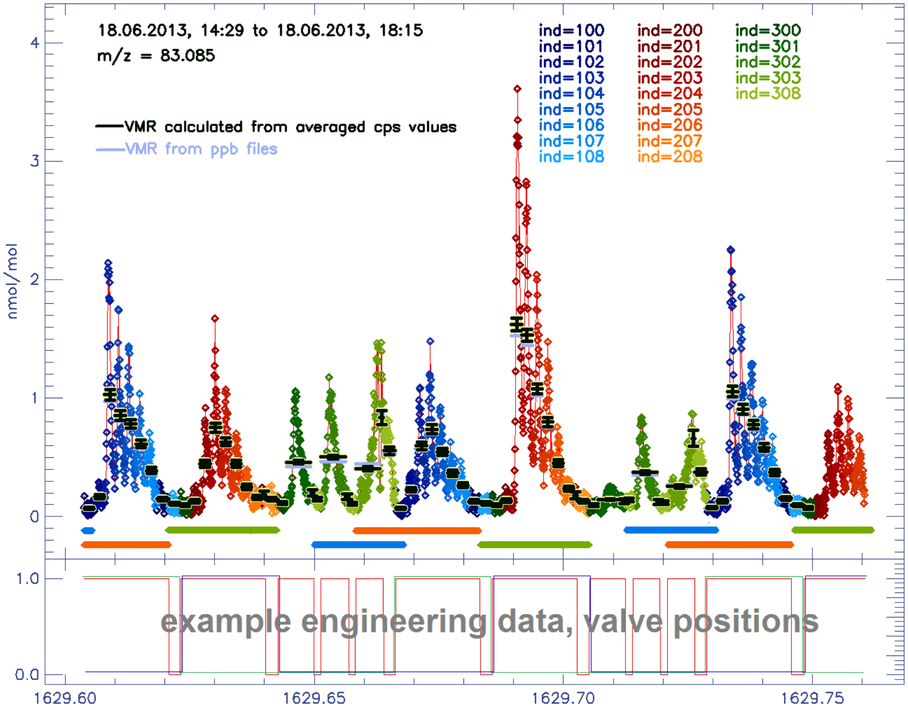
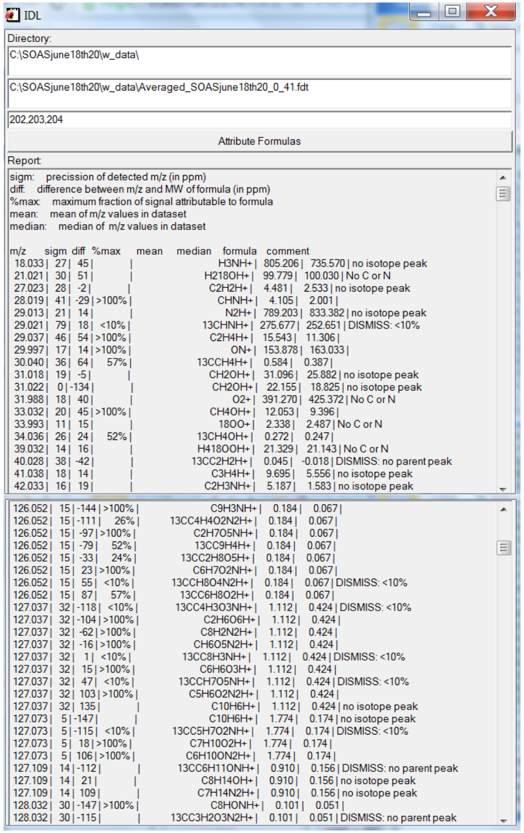


Figure 9. Mixing ratios detected at m/z 83.085 over 4 h period during the SOAS campaign in summer 2013. The data have been categorised using the index parameters depicted in Fig. B1. The PTR-TOF-MS was connected to three inlets: aerosol sampler A (blue index values, 100–108), aerosol sampler B (red index values, 200–208), and a thermal desorption 3-stage denuder inlet (green index values, 300–308) to sample semivolatile compounds in the gas phase. The corresponding sampling times are indicated as horizontal lines (and matching colors) below the zero-line. The lower lines indicate sampling of background (i.e. particle free ambient air for the aerosol inlets and zero air for the das phase inlet), and the upper lines indicate sampling of ambient air.



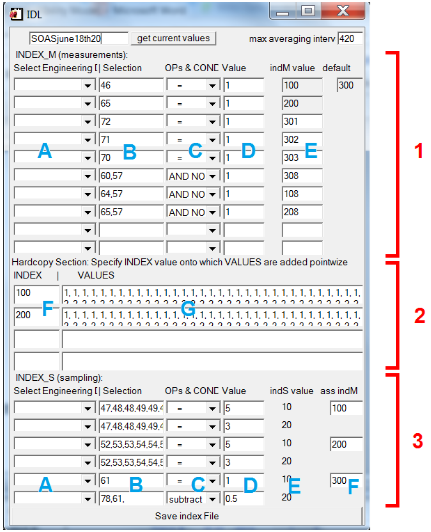
```
C:\SOASjune18th20\w_data\
C:\SOASjune18th20\w_data\Averaged_SOASjune18th20_0_41.fdt
202,203,204
```

```
sigm:    precission of detected m/z (in ppm)
diff:    difference between m/z and MW of formula (in ppm)
%max:    maximum fraction of signal attributable to formula
mean:    mean of m/z values in dataset
median:    median of  m/z values in dataset

m/z      sigm  diff  %max     mean    median    formula   comment
 18.033 |  27 |  45 |         |               H3NH+ |  805.206 |  735.570 | no isotope peak
 21.021 |  30 |  51 |         |             H218OH+ |   99.779 |  100.030 | No C or N
 27.023 |  28 |  -2 |         |              C2H2H+ |    4.481 |    2.533 | no isotope peak
 28.019 |  41 | -29 | >100% |               CHNH+ |    4.105 |    2.001 |
 29.013 |  21 |  14 |         |                N2H+ |  789.203 |  833.382 | no isotope peak
 29.021 |  79 |  18 |  <10% |             13CHNH+ |  275.677 |  252.651 | DISMISS: <10%
 29.037 |  46 |  54 | >100% |              C2H4H+ |   15.543 |   11.306 |
 29.997 |  17 |  14 | >100% |                 ON+ |  153.878 |  163.033 |
 30.040 |  36 |  64 |   57% |            13CCH4H+ |    0.584 |    0.387 |
 31.018 |  19 |  -5 |         |              CH2OH+ |   31.096 |   25.882 | no isotope peak
 31.022 |   0 |-134 |         |              CH2OH+ |   22.155 |   18.825 | no isotope peak
 31.988 |  18 |  40 |         |                 O2+ |  391.270 |  425.372 | No C or N
 33.032 |  20 |  45 | >100% |              CH4OH+ |   12.053 |    9.396 |
 33.993 |  11 |  15 |         |               18OO+ |    2.338 |    2.487 | No C or N
 34.036 |  26 |  24 |   52% |            13CH4OH+ |    0.272 |    0.247 |
 39.032 |  14 |  16 |         |            H418OOH+ |   21.329 |   21.143 | No C or N
 40.028 |  38 | -42 |         |           13CC2H2H+ |    0.045 |   -0.018 | DISMISS: no parent peak
 41.038 |  18 |  14 |         |              C3H4H+ |    9.695 |    5.556 | no isotope peak
 42.033 |  16 |  19 |         |             C2H3NH+ |    5.187 |    1.583 | no isotope peak
```

```
126.052 |  15 |-144 | >100% |              C9H3NH+ |    0.184 |    0.067 |
126.052 |  15 |-111 |   26% |         13CC4H4O2N2H+ |    0.184 |    0.067 |
126.052 |  15 | -97 | >100% |            C2H7O5NH+ |    0.184 |    0.067 |
126.052 |  15 | -79 |   52% |            13CC9H4H+ |    0.184 |    0.067 |
126.052 |  15 | -33 |   24% |          13CC2H8O5H+ |    0.184 |    0.067 |
126.052 |  15 |  23 | >100% |            C6H7O2NH+ |    0.184 |    0.067 |
126.052 |  15 |  55 |  <10% |          13CCH8O4N2H+ |    0.184 |    0.067 | DISMISS: <10%
126.052 |  15 |  87 |   57% |          13CC6H8O2H+ |    0.184 |    0.067 |
127.037 |  32 |-118 |  <10% |          13CC4H3O3NH+ |    1.112 |    0.424 | DISMISS: <10%
127.037 |  32 |-104 | >100% |             C2H6O6H+ |    1.112 |    0.424 |
127.037 |  32 | -62 | >100% |             C8H2N2H+ |    1.112 |    0.424 |
127.037 |  32 | -16 | >100% |           CH6O5N2H+ |    1.112 |    0.424 |
127.037 |  32 |   1 |  <10% |           13CC8H3NH+ |    1.112 |    0.424 | DISMISS: <10%
127.037 |  32 |  15 | >100% |             C6H6O3H+ |    1.112 |    0.424 |
127.037 |  32 |  47 |  <10% |          13CCH7O5NH+ |    1.112 |    0.424 | DISMISS: <10%
127.037 |  32 | 103 | >100% |           C5H6O2N2H+ |    1.112 |    0.424 |
127.037 |  32 | 135 |         |              C10H6H+ |    1.112 |    0.424 | no isotope peak
127.073 |   5 |-147 |         |              C10H6H+ |    1.774 |    0.174 | no isotope peak
127.073 |   5 |-115 |  <10% |         13CC5H7O2NH+ |    1.774 |    0.174 | DISMISS: <10%
127.073 |   5 |  18 | >100% |            C7H10O2H+ |    1.774 |    0.174 |
127.073 |   5 | 106 | >100% |          C6H10ON2H+ |    1.774 |    0.174 |
127.109 |  14 |-112 |         |         13CC6H11ONH+ |    0.910 |    0.156 | DISMISS: no parent peak
127.109 |  14 |  21 |         |             C8H14OH+ |    0.910 |    0.156 | no isotope peak
127.109 |  14 | 109 |         |            C7H14N2H+ |    0.910 |    0.156 | no isotope peak
128.032 |  30 |-147 | >100% |              C8HONH+ |    0.101 |    0.051 |
128.032 |  30 |-115 |         |        13CC3H2O3N2H+ |    0.101 |    0.051 | DISMISS: no parent peak
```

Figure 10. The *Attribute formulas tool* window and fractions of the report produced from a three day period during the SOAS campaign. The full report is also saved in a report file.



SOASjune18th20 | max averaging interv 420

INDEX_M (measurements):

Select Engineering [	Selection	OPs & CONE	Value	indM value	default
	46	=	1	100	300
	65	=	1	200	
	72	=	1	301	
	71	=	1	302	
	70	=	1	303	
	60,57	AND NO	1	308	
	64,57	AND NO	1	108	
	65,57	AND NO	1	208	

Hardcopy Section: Specify INDEX value onto which VALUES are added pointwize

INDEX	VALUES
100	1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
200	1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,

INDEX_S (sampling):

Select Engineering [	Selection	OPs & CONE	Value	indS value	ass indM
	47,48,48,49,49,4	=	5	10	100
	47,48,48,49,49,4	=	3	20	
	52,53,53,54,54,5	=	5	10	200
	52,53,53,54,54,5	=	3	20	
	61	=	1	10	300
	78,61,	subtract	0.5	20	

Figure B1. The *New index window* with the values that were used to categorise data from the SOAS campaign (as shown in Fig. 9). The two text fields in the first row allow to specify an index name and the maximum length (in seconds) of data-junks that are averaged. The *get current values* button reads the values of the selected index file (label 6 in Fig. 7a) into the respective fields. The three main sections are marked by red brackets: (1) up to 10 conditions can be used to categorise the PTR-TOF-MS data, (2) point wise addition of the provided values to specified index values, and (3) 6 conditions to define up to three sampling indices (two conditions, i.e. sample and blank, per sampling index). The different types of entry columns are marked by blue letters: (A) droplists that allow selecting engineering data and adding the corresponding index to column B; (B) engineering data indices or ion masses can be provided here and will be processed according the condition selected in column C; (C) droplists that allow selecting different rules to compare the selected data (B) with the provided target value (D); in column (E) index values are provided that are attributed if the condition is met; (F) specifies the index value to which sampling indices are associated or to which the values provided in column (G) are added.