Journal of Chromatography A, 1529 (2017) 81–92

Contents lists available at ScienceDirect

Journal of Chromatography A

journal homepage: www.elsevier.com/locate/chroma

Full length article

Automated single-ion peak fitting as an efficient approach for analyzing complex chromatographic data

Gabriel Isaacman-VanWertz[a,*], Donna T. Sueper[b], Kenneth C. Aikin[c,d], Brian M. Lerner[b,c,d], Jessica B. Gilman[c,d], Joost A. de Gouw[c,d], Douglas R. Worsnop[b], Allen H. Goldstein[e,f]

[a] Dept. of Civil and Environmental Engineering, Virginia Tech, Blacksburg, VA, USA
[b] Aerodyne Research, Inc., Billerica, MA, USA
[c] Chemical Sciences Division, Earth System Research Laboratory, National Oceanic and Atmospheric Administration, Boulder, CO, USA
[d] Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO, USA
[e] Dept. of Environmental Science, Policy, and Management, University of California, Berkeley, CA, USA
[f] Dept. of Civil and Environmental Engineering, University of California, Berkeley, CA, USA

ARTICLE INFO

Article history:
Received 22 June 2017
Received in revised form 1 November 2017
Accepted 2 November 2017
Available online 7 November 2017

Keywords:
Gas chromatography
Peak integration
Peak fitting
Data reduction
Environmental chemistry
Chemometrics

ABSTRACT

Chromatography provides important detail on the composition of environmental samples and their chemical processing. However, the complexity of these samples and their tendency to contain many structurally and chemically similar compounds frequently results in convoluted or poorly resolved data. Data reduction from raw chromatograms of complex environmental data into integrated peak areas consequently often requires substantial operator interaction. This difficulty has led to a bottleneck in analysis that increases analysis time, decreases data quality, and will worsen as advances in field-based instrumentation multiply the quantity and informational density of data produced. In this work, we develop and validate an automated approach to fitting chromatographic data within a target retention time window with a combination of multiple idealized peaks (Gaussian peaks either with or without an exponential decay component). We compare this single-ion peak fitting approach to drawn baseline integration methods of more than 70,000 peaks collected by field-based chromatographs spanning across a wide range of volatilities and functionalities. Accuracy of peak fitting under real-world conditions is found to be within 10%. The quantitative parameters describing the fit (e.g. coefficients, fit residuals, etc.) are found to provide valuable information to increase the efficiency of quality control and provide constraints to accurately integrate peaks that are significantly convoluted with neighboring peaks. Implementation of the peak fitting method is shown to yield accurate integration of peaks otherwise too poorly resolved to separate into individual compounds and improved quantitative metrics to determine the fidelity of the data reduction process, while substantially decreasing the time spent by operators on data reduction.

1. Introduction

Over the past several decades, one of the major applications of gas and liquid chromatography (GC, LC) has become the analysis of trace constituents in environmental and atmospheric samples [1]. These mixtures can be comprised of thousands to tens of thousands of compounds [2] and frequently contain many isomers with similar retention times for which significant peak separation cannot be reasonably achieved. The use of a mass spectrometer (MS) as a detector, which is commonplace in modern environmental GC applications, can improve identification and quantification of many analytes but typically cannot fully resolve all isomers of interest [3–5] unless highly selective methods are utilized. Comprehensive data reduction of a single chromatogram therefore requires quantification of several hundred known compounds of interest amongst hundreds to thousands of chromatographic peaks, many of which overlap. Measurement of difficult-to-resolve compounds or compound classes can in some cases be achieved through targeted highly selective methods, or through introducing selective chemistry to the analysis but at the expense of comprehensive analysis and/or ease-of-operation. Decades of research have consequently been dedicated to the improvement of chromato-

* Corresponding author at: 419 Durham Hall, 1145 Perry St (0246), Blacksburg, VA, 24061, USA.
E-mail address: ivw@vt.edu (G. Isaacman-VanWertz).

https://doi.org/10.1016/j.chroma.2017.11.005
0021-9673/

graphic resolution and quality of chromatographic data achieved by comprehensive techniques [6–8], but the density of peaks in environmental samples and their sample-to-sample variability remain a unique analysis and data reduction challenge.

Reduction of chromatographic data from its raw form (detector response versus retention time) to its processed form (integrated peak areas, or "zeroth moments", of target analytes versus sample collection time) often requires substantial manual interaction with the data. Most commercially available software used in the analysis of environmental GC data (e.g. Agilent© Chemstation and MassHunter, Thermo Fisher™ Xcalibur™, Waters® Empower® 3) draws baselines under chromatographic peaks to numerically calculate the zeroth moment ("drawn baseline" or "baseline" integration). Baseline integration has been automated and applied successfully to environmental data (e.g. [9–11]). However, trace analytes in a complex mixture often yield co-elutions of multiple peaks, sometimes with large ratios of relative peak sizes, which can introduce large errors in quantification by baseline integration [12]. High chromatographic complexity thus often requires substantial user interaction for error checking and correction, a process that is both time consuming and subjective. Effective application of automated baseline integration is therefore challenged by comprehensive chromatography of mixtures that contain thousands (or tens of thousands) of compounds, such as atmospheric samples [2]. Furthermore, the drawn baseline approach is sensitive to the selected endpoints of the baseline [13], which can depend on proprietary algorithms and/or user interpretation of the data. Increased interest in comprehensive chromatography of complex environmental samples therefore requires new approaches for the fast and reproducible integration of chromatographic peaks and the means to quickly assess and correct the quality of processed data.

The complex chromatograms and overlapping peaks often encountered in environmental data can be tackled through advanced algorithmic approaches to peak integration. Co-eluting spectra can be separated through targeted factor analysis to simultaneously reach a mathematical solution in both spectral and chromatographic space. This approach yields resolved chromatographic peaks, each with a deconvolved spectra – or, rather, an optimized mathematical solution approximating this condition. Several such algorithms have been applied to liquid chromatography data with spectroscopic detectors [14–16], but their relative complexity has limited their wide-spread application. Furthermore, while helpful for deconvolving spectroscopic data, the detail and computational expense of these approaches is higher than necessary for mass spectrometric methods, which have higher specificity and relatively unique spectra. Instead of complex factor-deconvolution methods, experimental chromatographic data can be reasonably described as mathematically idealized peaks [17]. A large number of peak shapes have been proposed and explored [18]; while chromatographic peaks are roughly Gaussian, an assumption of Gaussian shape can lead to considerable error, and a better description is that of a Gaussian peak convoluted with an exponential decay ("exponentially modified Gaussian" or "EMG") [19,20]. Substantial previous work (see review [18] for upwards of 50 references) has demonstrated that otherwise-unresolved isobaric species can be theoretically quantified by fitting chromatographic data with a combination of idealized Gaussian and EMG peaks. This peak fitting approach allows for decreased stringency in chromatographic resolution and improved analytical efficiency, with the additional benefit that it yields higher statistical moments that quantitatively describe the data (peak center, width, skew, etc.). However, adoption of mathematical data reduction techniques has been limited by a lack of automated peak fitting approaches that has impeded its application to large datasets. Instead, current implementations (e.g. OriginLab Peak Fitting Module™) have primarily focused on single chromatograms and have required substantial user input (e.g. [21,22]). While peak-fitting approaches have therefore been widely validated on synthetic data and chromatographic analysis of known standards, there has been little, if any, rigorous comparison between simplified peak fitting approaches and drawn baseline integration methods under real-world conditions and on real complex data. The work presented here addresses these gaps by developing and implementing an automated approach to simplified peak fitting, validating it through analysis of known compounds, and extensively comparing it to baseline integration for the reduction of highly varied environmental data.

In this work the application of field-based GCs to the study of organic compounds in the atmosphere serves as an example of the needs and challenges of the environmental analysis community. Several widely used instruments have been developed over the past decades to facilitate in-field atmospheric analysis, including instruments designed for analysis of particle composition [23–25] and a wide range compounds in the gas phase [9,26–34]. While differing in their construction and target analytes, all of these instruments share semi-hourly or hourly time sample collection with 30–300 compounds quantified per chromatogram. Operational periods are typically weeks to months, so a dataset consists of on the order of 10^5 peak integrations. These field instruments often generate data quickly and frequently with little time between samples for analysis, manual data reduction, or error checking. Constantly changing environmental conditions, sample matrices, and laboratory environments can result in some cases in less controlled and less stable chromatography such as drifts in retention times or peak widths of target analytes. Integrated chromatographic peaks from these instruments have been robustly shown to yield accurate quantification of trace environmental constituents in both laboratory settings and varied ambient conditions [25,27,35–38]. However, this increased run-to-run variability poses a particular challenge for integrating chromatographic peaks of highly variable concentrations of trace constituents in a complex mixture. Due to the complexity and variability of collected samples, most field-deployable gas chromatographs (including all of the referenced instruments) suffer the same bottleneck in data reduction: relying on user-intensive processes for integration and/or quality control. The difficult and time-intensive process of producing peak integrations from their raw chromatographic data has stymied the widespread adoption of field-deployable chromatographs. Furthermore, rapid data analysis is particularly important for a field-based instrument because lags in data processing can result in missed opportunities to identify and fix problems, leading to data gaps or reductions in data quality. The rate of data generation will significantly increase with the next generation of field-based GCs incorporating recent developments in "fast GC" [39] to improve time-resolution and enable new applications of in-situ GCs (e.g. airborne measurements, [34,40,41]). These new tools will produce a chromatogram every few minutes, an increase of an order of magnitude in data quantity over previous instruments. Improved automated integration of raw chromatographic is necessary to facilitate the continued and future development of field-based instruments that produce complex, variable, time-resolved datasets (thousands of chromatograms per month).

Effective application of chromatographic tools to varied environmental samples requires robust solutions to several key challenges: (a) deconvolution of co-eluting peaks due to the complexity of samples, (b) accounting for variance in chromatographic parameters due to in-field operation, (c) fast processing to facilitate in-field data analysis, (d) quantitative metrics to improve the efficiency of quality control, and (e) sufficiently flexible implementation to facilitate application to a varied suite of instrumentation. While these are general chromatographic issues, the work pre-

sented here optimizes solutions with an eye toward maintaining speed and efficiency in data reduction. For instance, managing co-elutions cannot occur at the expense of fast processing, and fast processing cannot preclude effective quality control. We describe here an automated single-ion peak fitting approach that yields a mathematical description of an eluted chromatographic peak without high computational expense. In this approach, both well-resolved and co-eluting chromatographic data are described by an empirical solution of Gaussian peak(s), either ideal or with an exponentially decaying fronting or tailing component (EMG); these peak shapes are fully described by relatively few (3–4) coefficients, which minimizes degrees of freedom and thus risk of poorly constrained fits [42], but there are no inherent limitations to peak shape in our approach. We investigate the utility of this approach as a means to accurately reduce large chromatographic datasets. We then compare this method to drawn baseline integration for four datasets collected by two instruments. The statistical moments of these fits and other quantitative parameters are explored as a means to evaluate, understand, and improve the data reduction process.

2. Methods

2.1. Peak fitting algorithm

2.1.1. Implementation

Peak fitting was implemented in Igor Pro 6/7 (Wavemetrics), using its internal multi-peak fitting functionality. Within the Igor Pro programming environment, a complete software package to import and analyze chromatographic data was developed and used to explore the efficacy of peak fitting as a data reduction method. This tool is made available online (https://sites.google.com/site/terninigor/) as "TERN" (the "TAG ExploreR and iNtegration package", definition of "TAG" below), which incorporates correction for drifts in retention time, tools to identify target peaks of interest (e.g. searching NIST mass spectral libraries [43]), and a complete implementation of peak fitting as described below. In this package, fitting is performed for one target peak across several chromatograms so that users can observe fluctuations in co-elutions and retention times, minimizing the likelihood of integration errors. Parameters from the peak fit (e.g. statistical moments, fit residual, etc.) are provided to the user as a means to facilitate quality control and understand the goodness-of-fits; as explored in this work, the availability of quantitative fit parameters is a primary benefit of peak fitting as a data reduction technique. Screenshots and some discussion of this package are provided in the Supplementary Information. This package and its source code is made freely available and editable by users to provide maximum flexibility, allow the analysis of data across multiple instrument platforms, and remove barriers to analyzing and sharing data. Toward this goal, current data import supports the open data formats of NetCDF (https://www.unidata.ucar.edu/software/netcdf/), a widely used format for sharing scientific data, and HDF5 (https://support.hdfgroup.org/HDF5/), an efficient format for storing large datasets such as those produced by time-of-flight mass spectrometers; users can add support for additional and/or proprietary formats as needed and feasible.

A critical analysis step prior to peak fitting is correction for variability in retention times, which for field-based instrumentation can include shifts of several seconds from run to run. Retention time correction is incorporated in this implementation on a run-to-run basis by manually identifying a small subset of peaks distributed throughout the chromatogram (at least two, but with no inherent upper bound) and calculating polynomial retention time correction coefficients to bring these peaks into agreement with expected retention times. While not fully automated, performing this correction once for each sample in a dataset demands only nominal operator effort (less than one hour per hundreds of samples) and substantially improves performance of automated peak fitting.

2.1.2. Description

An automated approach for fitting chromatographic data with a combination of mathematically described peaks is developed and described here. Peak detection and fitting is performed on response of a single mass spectrometric ion (e.g. m/z 57) within a specified retention time window. Operating on a single ion chromatogram (SIC) instead of the full spectral data minimizes computational expense by avoiding multi-dimensional fitting; as demonstrated by the results presented here, single-ion peak fitting without the additional complexity and demands of a full deconvolution of spectra is sufficient for deconvolution in nearly all cases. This simplified, single-dimensional approach has the additional benefit that it can be easily extended to any one-dimensional detector response, such as the output of a flame ionization detector (FID), or the total ion chromatogram (TIC) of an MS, and can be applied to high-resolution ions as well as the unit-mass resolution data used in this work. Complete mass spectrometric data is used as a quality control and error checking device (i.e. comparison between found and expected spectrum).

Automatic peak fitting is implemented through several steps (illustrated in Fig. 1): (1) optional background subtraction, (2) peak detection, (3) peak fitting, which includes multiple steps and fitting attempts with feedback to determine the optimal fit, and (4) selection of the target peak from the multiple peaks fit. Typically, the baseline of the peak is fit during peak fitting, either as a linear or constant baseline function (user-specified), but in some cases, peak fits are improved by prior background subtraction of the single ion chromatogram. If background subtraction is applied by the user, the points with the lowest signal within the target retention time window are taken as the chromatographic background and subtracted from the signal prior to peak detection and peak fitting. In most cases, differing approaches to baselines (constant fit, linear fit, background subtraction) result in only minor variation in the final integrated peak areas, though sloping baselines require a linear or otherwise retention-time-dependent description of the baseline. The ideal approach for a given dataset is dependent on instrument-specific considerations (noise, baseline drift, etc.).

Peak fitting is initiated by detection of peaks in the single ion chromatogram within a window around a specified retention time. These windows are assumed to be centered on a target analyte of interest, identified by the user, the idealized peak description of which will be returned to the user after fitting. While a "target" retention time window and ions are specified by the user, no additional *a priori* knowledge is assumed regarding the shapes, locations, or number of peaks within the window. These target peaks may therefore represent any known or unknown chromatographic peak of interest. Within the target window, Gaussian peak locations and shapes are estimated using the second derivative of the detector signal to estimate the centers (minima in second derivative) and edges (maxima in second derivative) of chromatographic peaks. Some empirically determined smoothing and filtering is useful at this step to account for noise in chromatographic signal; estimation of these parameters can be automated (e.g. low-pass filtering) or manually defined, with the best approach dependent on instrument operating parameters. If no peaks are located in the initial peak detection, increasing smoothing and noise estimates are attempted until either candidate peaks are detected or parameters are beyond acceptable limits (no peak found).

Initial estimates of peaks are fit to determine final output peak coefficients through a multi-step process. First, all detected Gaus-

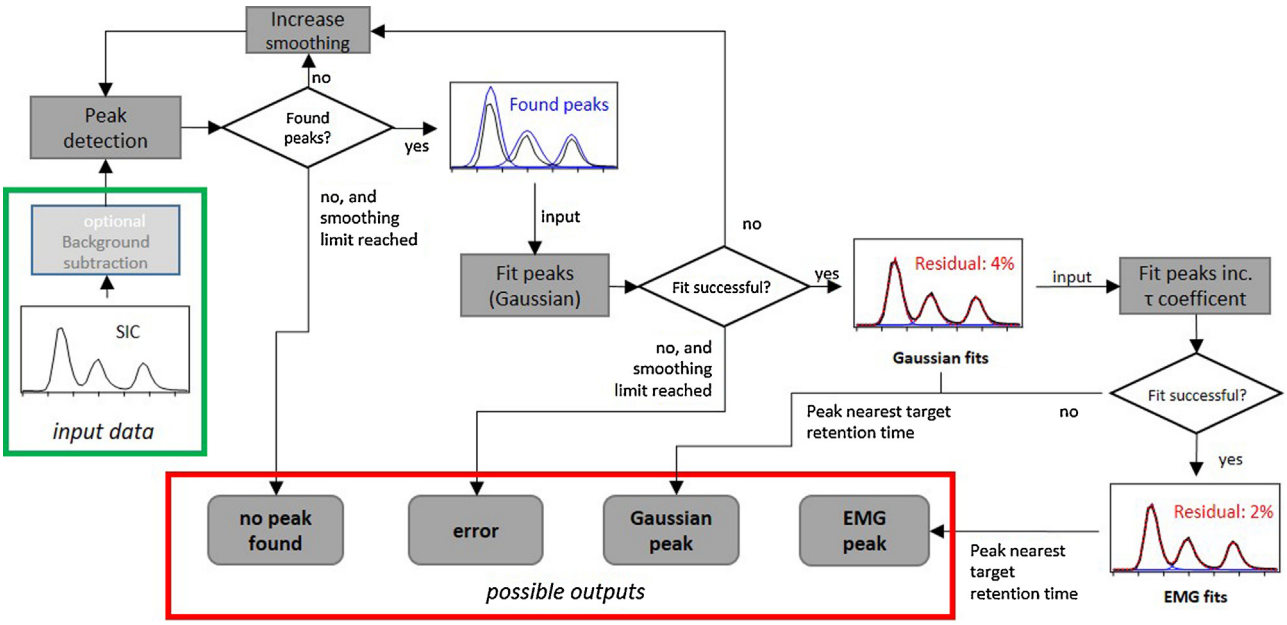

Fig. 1. Flowchart of peak detection and fitting approach developed and implemented in this work. Operations are gray squares, control points are white diamonds, sample intermediate data and results are illustrated as plots, and possible output solutions describing target peak are gray rounded squares.

sian peaks are fit simultaneously to the raw SIC data by optimizing the coefficients (center, amplitude, width) of all peaks to minimize the residual of the fit. The output of this process is a set of Gaussian peaks that describe the SIC within the target retention time window. Gaussian peaks are a reasonable approximation of chromatographic peaks, but exponentially modified Gaussian ("EMG") peaks yield lower error and are preferred. The output Gaussian peaks are therefore fit again to the raw SIC data, but with the inclusion of an additional coefficient describing the exponential decay, τ, of each peak (no constraint is made on whether the exponential convolution yields a fronting or tailing peak). This fit yields a set of EMG peaks that describe the SIC data; in cases in which this fit fails to converge to a solution, the Gaussian fit solution is taken as the best available description of the data. Similarly, a co-elution of several EMG peaks cannot be well constrained when a peak is bounded by many overlapping peaks, so the Gaussian fit solution is recommended for the description of complex co-eluting peaks (e.g. more than 4 peaks in the target retention time window). This flowchart (Fig. 1) has been optimized for the instruments employed here, so some flexibility in specifying fit parameters (e.g. retention time windows, smoothing parameters, allowable decay coefficients, etc.) is necessary to universally apply this method. For other chromatographic applications, there is no barrier to implementing non-Gaussian peak shapes and alternate optimization schemes. The results and conclusions discussed in this work are general to the application of mathematical fitting of target peaks as a method to reduce complex and variable chromatographic data.

After fitting the chromatographic data, the peak representing the target analyte must be selected from amongst the multiple peaks in the fit solution. Provided some correction for retention time drifts, the implementation described here assumes the candidate peak closest to the target retention time is considered to be the peak of interest to the user. This simple decision mechanism is found to be sufficient for the tested data, though additional information about the found peak can also be compared to "expected" information (e.g. mass spectra in the case of known peaks) to confirm the peak selection decision. There are no inherent limitations on how the "correct" peak is selected from the multi-peak fit; additional data, such as the mass spectral information, could be considered in other implementations for peak selection.

2.2. Datasets

Data from two instruments are used here to develop and evaluate the peak fitting method. Each instrument targets a different suite of organic atmospheric constituents. Gas-phase compounds more volatile than decane are measured by concentrating volatile organic compounds (VOCs) on a cryogenic absorbent trap prior to thermal desorption and GC analysis. Lower-volatility gases (less volatile than tridecane), as well as particle-phase atmospheric organic constituents are measured using a Thermal desorption Aerosol GC (TAG) employing room temperature sample collection and high-temperature thermal desorption.

Data from four measurement periods (two for each instrument) are used to provide a substantial and varied combined dataset with which to evaluate the peak fitting method, and explore its application to large and complex environmental analysis challenges. These data are all collected as part of "field campaigns," periods during which the instrument was deployed at a designated location for in-situ collection and analysis of ambient data. For all measurements presented here, the various GC systems are located at the field site within a mobile trailer, typically sharing the space with 2–4 other similarly sized analytical instruments. Samples are collected through inlets running outside through a window or hole in the trailer wall. These data were collected under a large range of atmospheric and indoor environmental conditions that may impact chromatographic reproducibility. Though nominally climate controlled, indoor temperatures typically fluctuated diurnally due to the heat generated by the instrument payload and a lack of insulation. Indoor temperatures varied by more than 10 °C (e.g. 17–29 °C for VOC-UBWOS) and in some cases reached up to 35 °C. Due to temporal variability in atmospheric conditions, collected samples frequently varied widely in relative humidity and total sample mass as well as composition.

The instruments and datasets explored in this work are described in detail elsewhere, so discussion here is limited to brief overviews and References

2.2.1. VOC-GC: volatile organic compound gas chromatograph

The "VOC-GC" described here is field-deployable GC–MS instrument that measures volatile organic compounds in ambient atmospheric samples [28–31]. In short, this in-situ GC–MS cryogenically pre-concentrates a full suite of volatile organic compounds on two separate analysis channels, in parallel, that are then analyzed in-series via a single quadrupole mass spectrometer (Agilent 5973N). Channel 1 trapping temperatures and substrates are optimized for the collection of C_2-C_5 hydrocarbons and halocarbons while channel 2 is optimized for C_5-C_{10} hydrocarbons and VOCs containing oxygen, nitrogen, sulfur, and halogens. The sample collection period (5 min) and near-immediate analysis (25 min) are repeated automatically every 30 min. After collection, samples are transferred to their respective cryofocus units by heating the sample traps to 100 °C in approximately 6 s while flushing with helium carrier gas. Channel 1 analytes are separated on an Al_2O_3/KCl PLOT column (0.25 mm ID × 25 m, RT-Alumina BOND/KCl Restek; ramped from 55 to 115 °C in 200 s). Once the channel 1 sample has eluted from the PLOT column, a heated 4-port valve (Valco) switches position to direct the channel 2 eluent to the MSD and the cryofocus unit on channel 2 rapidly heats and injects the channel 2 sample onto a DB-624 capillary column (0.18 mm ID × 20 m, MST-624 Restek; ramped from 35 to 125 °C in 800 s). Mass spectrometry is performed using a unit-mass resolution quadrupole mass spectrometer scanning in either selective ion monitoring mode, scanning a small subset of ions at specific portions of the chromatogram, or in total ion mode, scanning for all masses ranging from 29 to 150 amu. The targeted VOCs of interest included alkanes, aromatics, highly reactive alkenes, biogenic hydrocarbons, oxygenated VOCs (OVOCs), and nitrogen containing species such as alkyl nitrates.

Data from the VOC-GC collected at two field campaigns is presented here:

Campaign 1) "VOC-PAS": California Research at the Nexus of Air Quality and Climate Change (CalNex), Pasadena, CA (34.1377°N, 118.1253°W), summer 2010

VOCs were measured during the CalNex campaign (http://www.esrl.noaa.gov/csd/projects/calnex/) at a groundsite in Pasadena, CA. Data from 22 compounds are presented here, spanning the period of May 11 to June 16, 2010, totaling approximately 1700 samples with half-hourly time resolution. Details from the deployment of this instrument during this field campaign are provided elsewhere [44].

Campaign 2) "VOC-UBWOS": Uintah Basin Winter Ozone Study, Horsepool, UT (40.14370°N, 109.46718°W), winter 2012

Data from 17 compounds measured during the UBWOS campaign (http://www.esrl.noaa.gov/csd/groups/csd7/measurements/2012ubwos/) are used in this work, spanning the period of January 24 to February 20, 2012 with half-hourly time resolution totaling approximately 1300 samples. Details from the deployment of this instrument during this field campaign are provided elsewhere [45].

2.2.2. TAG: thermal desorption aerosol gas chromatograph

The Thermal desorption Aerosol Gas Chromatograph (TAG) is a field-deployable GC-based instrument that quantifies gas- and particle-phase organic compounds in ambient atmospheric samples. Operation and development of this instrument has been described in detail previously, including several improvements to the operation and quantification of this system over the last decade [23,24,46,38,47]. Particle-phase compounds and lower-volatility gases – those able to partition into the particle phase under typical atmospheric conditions – are quantified via sample concentration followed by thermal desorption and analysis by a commercially available GC–MS system (6890/5973; Agilent). Collected sample is thermally desorbed with a temperature ramp up to 320 °C (over approximately 10 min) and transferred to the head of the GC column (0.25 mm ID x 30 m × 0.25 um phase Rxi 5ms-Sil; Restek), separated using a ramped temperature profile (from 50 °C to 320 °C over approximately 20 min), and analyzed by mass spectrometry. Mass spectrometry is performed using a unit-mass resolution quadrupole mass spectrometer scanning a mass range from 29 to 500 amu at 3 scans per second. The analytical range of this instrument in the context of atmospheric organic compounds extends to compounds with volatility between that of tridecane and pentatriacontane. Data presented here represents organic compounds that can be eluted through the chromatographic column without derivatization, such as aliphatic and aromatic hydrocarbons, carbonyls, ethers and compounds containing a single hydroxyl or carboxylic acid group; a newer generation of this instrument includes in-situ derivatization [25] and the peak fitting integration method developed in this work has been extended to the analysis of derivatized, highly oxygenated compounds [48].

Data from TAG collected at two field campaigns is presented here:

Campaign 3) "TAG-SOAR": Study of Organic Aerosol at Riverside (SOAR), Riverside, CA (33.97°N, 117.32°W), fall 2005

Measurements of 137 low-volatility gases and particle-phase compounds during the SOAR campaign are used here, spanning the period of July 29 through August 8, 2005. The TAG instrument deployed during SOAR collected data with hourly time resolution (~150 samples, sample collected in first 30 min of each hour). These measurements and deployment of the TAG instrument during this field campaign are discussed in detail by Williams and co-workers [49].

Campaign 4) "TAG-BKR": California Research at the Nexus of Air Quality and Climate Change (CalNex), Bakersfield, CA (35.35°N, 118.97°W), summer 2010

During the same CalNex field campaign as VOC-PAS above, the TAG instrument was deployed at a site in Bakersfield, CA. Data from 92 compounds are shown here, spanning the period of June 1 to June 27, 2010. Samples were collected hourly before June 9 and bihourly after June 9 (~500 samples). Details from the deployment of this instrument during this field campaign are provided elsewhere [50].

2.3. Comparison to baseline integration

Peak fitting was compared to drawn baseline integration methods for approximately 70,000 peaks, comprised of a wide variety of analytes. Compounds used in these comparisons span volatilities in the range of C_2 to C_{31} *n*-alkanes, and include most atmospherically relevant oxygenated functional groups – alcohols, aldehydes, acids, ketones, esters – as well as halocarbons and nitrates with small carbon backbones. Baseline integration was performed using Agilent Chemstation (versions A.03 and D.02). After automatic baselines were determined by the software, each peak was visually inspected and new manual baselines were drawn as necessary. Co-eluting peaks are separated by "dropped baselines", in which the boundary of the peak area integration is assigned as the retention time of the saddle between two peaks. This approach is common within gas chromatographic analyses and for most peak shapes introduces less error to peak areas than other peak-splitting approaches [12,51,52]. Comparison is focused on the area (zeroth moment) of the peak as this is the relevant parameter for quantification of analytes.

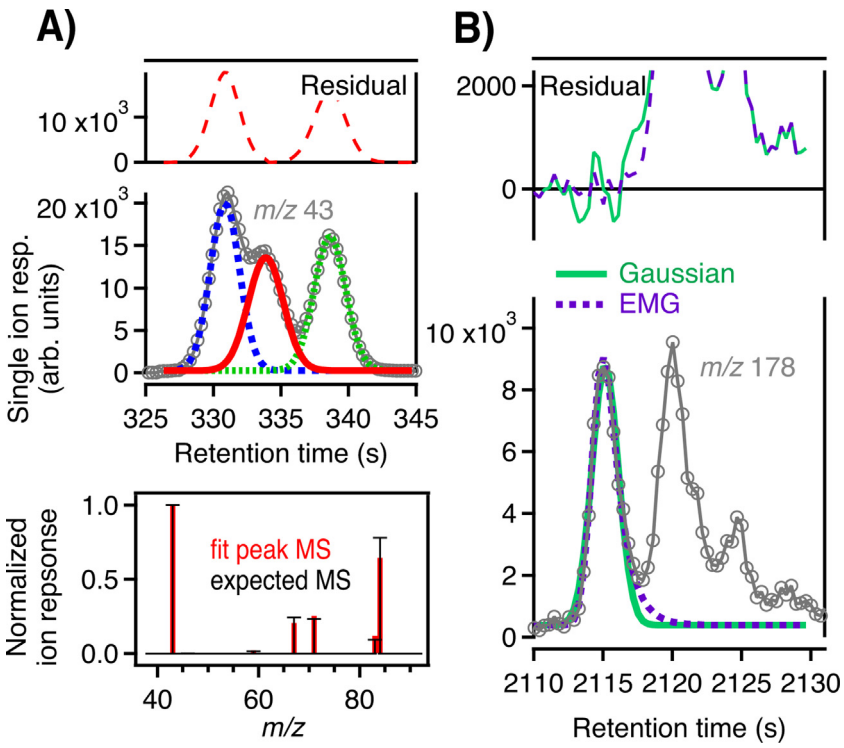

Fig. 2. Demonstration of peak fitting method on representative single ion chromatograms (SIC, gray circles). (a) Multiple Gaussian fit from the VOC-PAS dataset, fit with three peaks (blue, green, and red), with the target peak shown as a solid red line, and the residual of the sic of this fit peak shown as a dashed red line. Mass spectrum of found peak (red sticks) compared to the known spectrum (black peaks); data is only collected on some ions ("selected ion mode"). (b) Comparison of Gaussian fit (green solid line) with EMG fit (purple solid line), and the residuals of the sic of these fit peaks, using TAG-SOAR data. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3. Results

3.1. Demonstration of peak fitting and resultant data metrics

Fitting chromatographic data to multiple idealized (Gaussian, exponentially modified Gaussian, etc.) peaks has been well-characterized in previous work (see [18,19] and references therein). An example of this approach for a single ion chromatogram is shown Fig. 2. An SIC (detector response to a single mass spectrometric ion) within a retention time window is shown as a gray line with open circles, which is fit in Fig. 2a with three Gaussian peaks (colored lines). As implemented here, the fit window is centered on the expected retention time of a target peak (red line), though all detected peaks in the window are fit simultaneously. The raw chromatographic data with the target peak removed is shown as the residual (red dashed line), and the found mass spectrum of the target peak is compared to that expected for this known analyte. These tools provide quality control and error-checking metrics, which demonstrate in this case a reasonable fit and a good match between the expected and found peak. These empirical metrics can resolve cases in which multiple peak fitting solutions can reasonably describe the data. An example is shown in Fig. 2b, in which the target peak in the SIC can be fit as a Gaussian (green) or an exponentially modified Gaussian (EMG, purple) peak. The EMG fit, however, is a better fit based on its smaller and less sinusoidal residual. As discussed in the Methods, retention time is found in this work to be sufficient to identify the target peak, so these metrics (residuals, mass spectra, etc.) are provided as quality assessment tools, but not used for constraining peak fits.

Peak fitting provides robust and flexible integration of both single and co-eluting peaks described by any idealized peak shapes, with Gaussian and EMG peaks used in this work. Complex co-elution of both Gaussian and EMG peaks is demonstrated for multiple peaks of similar magnitude in Fig. 2. A variety of other common peak fitting scenarios are provided as examples in Fig. 3. Multi-peak fitting is applied in Fig. 3a to a case in which the target peak is substantially smaller than nearby or convolved peaks. Peak fitting is further extended to integrate single and co-eluting peaks that are reasonably resolved, both well above baseline noise and near signal-to-noise (Fig. 3b and c, EMG and Gaussian respectively). Peak fitting is found to reasonably apply not only to continuous SICs, but also when data is discontinuous such as in Fig. 3c, where response drops to zero at the edges of the chromatographic data (a frequent occurrence for instruments operated in "selected ion monitoring mode").

3.2. Validation of peak fitting for reduction of complex data

Peak fitting has been widely researched and implemented for the deconvolution of co-eluting peaks in individual chromatograms, based largely on the analysis of known analytes. We therefore focus our validation of peak fitting on analysis of complex environmental data collected under real-world field-deployment, which has not previously been shown. We provide only a brief analysis of known analytes (standards introduced while the instrument was operating under field-deployment conditions) to demonstrate that peak fitting as implemented here yields reasonable error; data are shown in the Supplementary Information. For injections of known varying concentrations, both well-resolved peaks and poorly resolved complex co-elutions of Gaussian and EMG peaks, are fit with approximately 10% error (mean absolute percentage error) with low (<1%) uncertainty in calibration slopes. Analysis of stable concentrations of known analytes in varied co-eluting matrices yields errors of <10% for well- and moderately-resolved peaks. Increasing error up to 30% is observed for peaks with low resolution

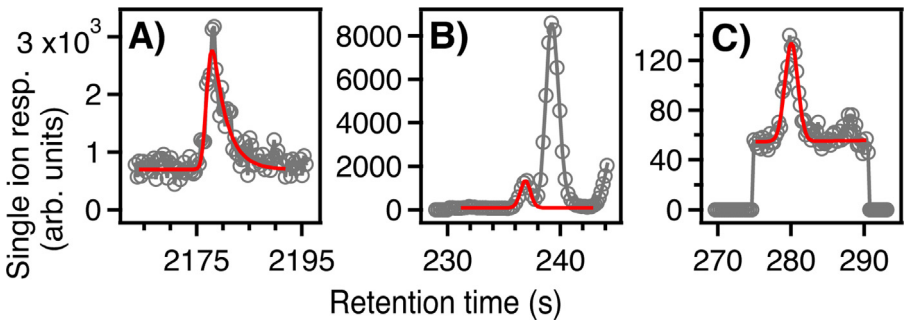

Fig. 3. Fit peaks (red line) for representative sic data (gray circles) from (a) 3-methyl-1-butane from VOC-PAS, (b) anthraquinone from TAG-SOAR, and (c) methyl-t-butyl ether from VOC-PAS. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

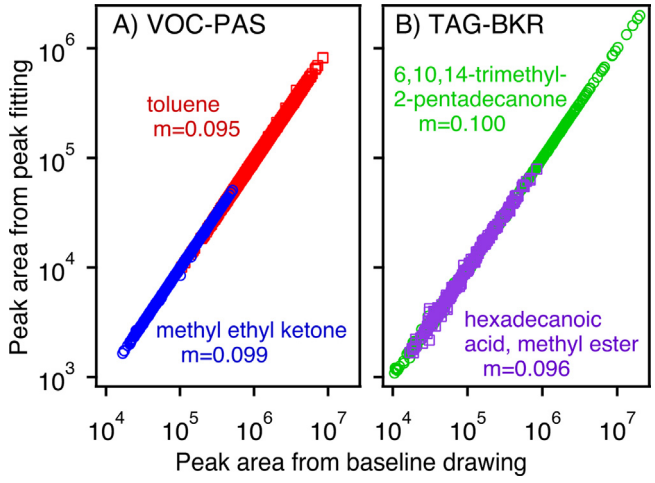

Fig. 4. Comparison of single-ion peak fitting integrated areas to drawn baseline integrated peak areas for representative compounds spanning the range in detector response observed for each instrument. Slope, m, is provided for each linear fit.

(no valley between co-eluting peaks) and high overlap (up to 35% overlap of known peak).

Though these tests provide some estimate of the trueness of the implemented peak fitting approach, the utility of peak fitting for the integration of complex and highly variable environmental data collected under varied environmental and instrument conditions cannot be reasonably assessed through injection of known standards. Instead, integration by single-ion peak fitting is validated by comparing resulting peak areas to peak areas determined by drawn baseline integration for data spanning several orders of magnitude, with variable sample and matrix composition. Representative compounds spanning the entire range of the complete dataset are shown in Fig. 4, representing a variety of volatilities and diverse functionality for both instruments tested in this work. Correlations are strongly linear, with negligible scatter. Peak fitting as implemented in this work yields peak areas that are a factor of approximately 10 times smaller than raw baseline integration values returned by Chemstation software (Agilent), but units for peak area (response × time) are arbitrary (e.g. seconds vs. milliseconds) so deviations from a slope of unity are not relevant for quantification when all data is processed by the same integration method. Calibration curves generated as part of this work are in fact found to have less than 1% uncertainty in their slopes, and to yield true calibrated masses within ~10% (Supplementary Information). Some variability in this factor is observed (i.e. differences in slopes in Fig. 4) and is thought to be due to differences in the integration calculation of each method and not due to biases in the integration, given that the factor was not found to correlate with peak shape (retention time, width, or decay coefficient). Some small intercepts are observed, but always less than 10% of the smallest peak and with no clear trends or patterns; intercepts are therefore thought to be due to random uncertainty and error, not biases between integration approaches. In real-world analyses, calibrants and analytes are integrated by the same method, so the absolute value of this slope does not impact quantification but deviations from the average slope of given compound represent real imprecision or uncertainty in integration.

To compare all compounds, peak areas are therefore scaled by multiplying each compound's average slope to explore the precision and relative differences between integration methods. Identical data reduction methods would yield identical peak areas and any deviation from unity represents real difference between the methods. Peak areas for approximately 70,000 peaks representing analytes of various chemical functionality, volatility, and chromatographic peak shape, fluctuating substantially in concentration with a variable matrix of analytes. Co-eluting peaks in these datasets frequently vary in their relative ratios by one to two orders of magnitude. Despite the complexity of these data, peak areas differ by less than 10% between integration methods in nearly all cases after scaling for differences in slope (Fig. 5); less than 15% of all data exceed this difference. Inset cumulative distribution functions reveal that large chromatographic peaks exhibit negligible variability (<5%) between methods. Differences increase with decreasing peak area (though rarely exceeding 15%) as smaller peaks have increased uncertainty in integration through both methods due to several factors. In baseline integration, defining the edges of a peak becomes more difficult at low signal-to-noise, and integrated peak area is sensitive to the selected endpoints of the drawn baseline. Similarly, parameters in peak fitting algorithms controlling baseline detection, smoothing, and background subtraction become less well defined as signal-to-noise decreases. Divergence of peak areas using these two methods can inform an estimate of the instrument level-of-detection for a given compound as the peak area at which peak integration introduces significant uncertainty to the measurement; the uncertainty in small peak areas is observed to be instrument dependent as expected. Comparison between integration techniques provides an estimate of precision that is in reasonable agreement with injection of known compounds (10%), suggesting an overall accuracy of 10–15% for most peaks, with some increasing error for poorly-resolved peaks (discussed in the next section).

3.3. *Integration of poorly resolved peaks*

The quantitative parameters that describe the fit can be used to efficiently evaluate the quality of fits and the limitations of peak fitting in general. The first and higher statistical moments of peaks are not expected to change suddenly between samples under normal operating conditions, allowing poor integrations to be quickly

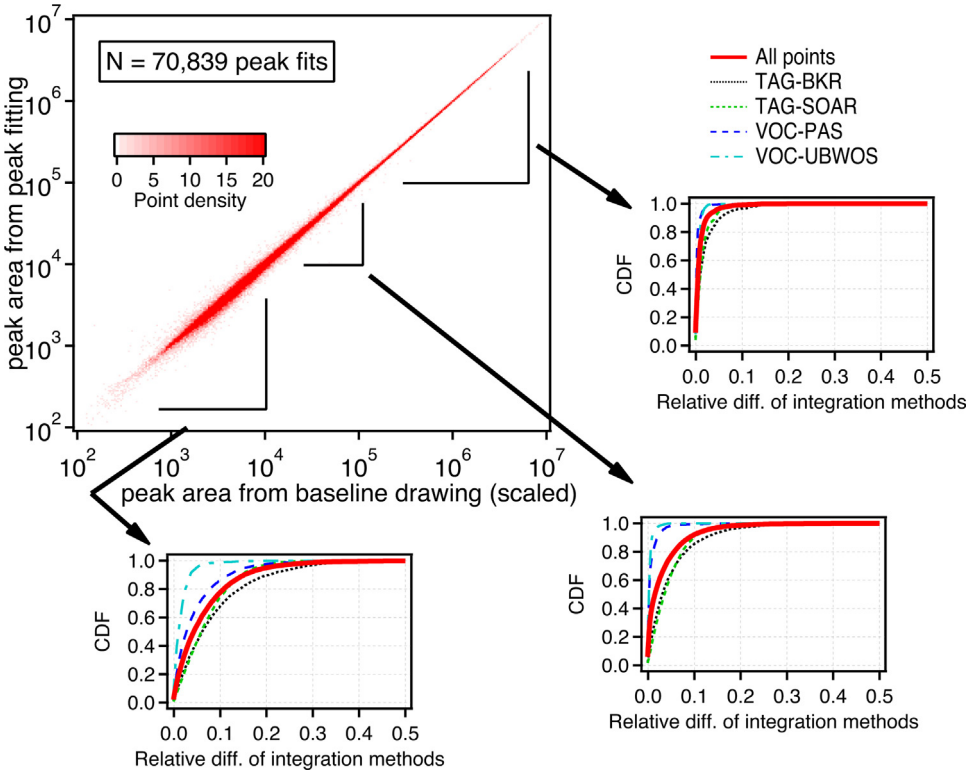

Fig. 5. Comparison of single-ion peak fit areas to drawn baseline integrated peak areas that have been scaled for differences in slopes to allow comparison to unity (i.e. removing simple multiplicative differences between software packages). Scatter plot shows all data analyzed as a plane of point density for clarity due to the high number of points. Inset plots shown cumulative distribution of relative difference between methods for high (>10^5), medium (10^4–10^5), and low (<10^4) peak areas. Total distribution shown as red solid line and distribution of each field campaign is shown in a different color and dash. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

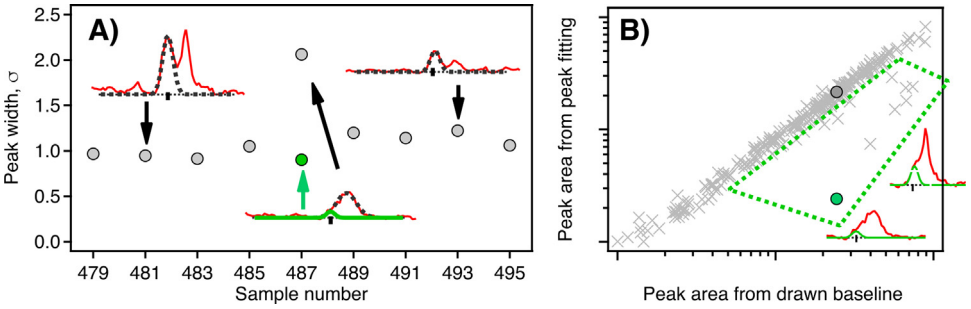

Fig. 6. Peak fit coefficients as a tool for quality control, using butanoic acid butyl ester from TAG-BKR as an example. (a) Time series of 9 consecutive automated peak fits with widths (second moment) shown (gray circles). Insets: example fits (dashed black lines) of the raw data (red lines). For sample 487, corrected fit with lower smoothing parameters shown in green. (b) Comparison of all peak areas from peak fitting to those from drawn baseline integration (gray crosses). Sample 487 peak areas shown for incorrectly integrated co-elution (gray) and corrected fit (green). All samples outlined in the green box are poorly resolved co-elutions, with incorrect drawn baseline integrations. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

identified. The second moment (peak width) of a representative co-eluting peak is shown in Fig. 6 to be relatively stable across samples despite changes in absolute and relative amplitudes. The default parameters in the peak fitting algorithm identified the black dashed peak as the target peak in each sample, with a peak width represented by gray circles. A time series of peak widths (Fig. 6a), however, exhibits a clear outlier in sample number 487 suggesting an erroneously broad fit; reducing the smoothing parameters in peak detection reveals a small co-eluting peak (green line) with the expected peak width (green circle) and retention time (hash mark on x-axis of insets) of the expected peak.

Quantitative fit metrics reduce error in integrated peak areas, as demonstrated by Fig. 6b. A comparison of peak fitting peak areas to previously reported drawn baseline peak areas reveals several samples in which peak fitting yields lower peak areas than baseline integration. Sample number 487 is one such sample, identified in this comparison by filled circles representing the co-elution peak area (gray) and the corrected deconvolved peak area (green). All points outlined in the green box are, like sample number 487, found to be poorly resolved; each of these points can be brought into agreement with reported baseline integration by erroneously fitting the co-elution with a single, wider peak. These data indicate that despite intensive user interaction with this data, baseline integration nevertheless resulted in the incorrect data reduction of many samples due to poor chromatographic resolution in some samples but not others. The well-defined empirical metrics pro-

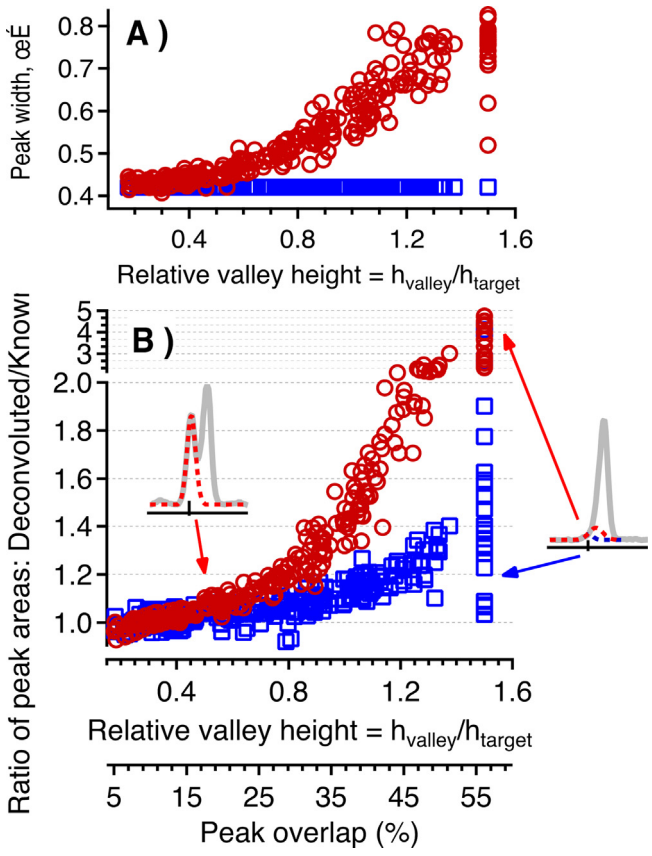

Fig. 7. Deconvolution of peaks with known peak area as a function of resolution, with (a) widths of fit peak and (b) ratio of peak area of deconvolved peak to known peak area, where a value of unity indicates no bias. Unconstrained peak fits are red circles, fit peaks with a peak width constrained to 0.42 are blue squares. Resolution is parameterized as relative valley height, with peaks exhibiting no valley or inflection points (unresolved peaks) assigned a value of 1.5 for graphical convenience. Peak overlap (percentage of target peak overlapping with neighboring peak) is closely correlated with relative valley height and provided on an additional axis. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

duced by peak fitting (e.g. statistical moments and fit residual) are therefore advantageous in assessing and improving the quality of data reduction for poorly resolved peaks.

The limitations of peak fitting for integrating poorly resolved co-eluting peaks are quantitatively determined by deconvolving peaks of known peak area. Two peaks with similar retention times but distinct mass spectra are summed, creating a synthetically combined SIC of two co-eluting peaks representing real environmental analytes; known true peak areas of the convolved peaks can be determined from their unique well-resolved peaks. Comparison of the deconvolved peak area to the known peak area then provides a measure of the trueness of peak deconvolution. An example of the relationship between this ratio and peak resolution is shown in Fig. 7, with resolution quantified as relative valley height. This parameter is calculated as the height of the valley or inflection point between the two peaks (h_{valley}) relative to the maximum height of the peak of interest (h_{target}), providing an intuitive measure of resolution; relative valley heights exceeding 1 are largely unresolved, while peaks with resolution so poor as to have no identifiable inflection point between peaks (first derivative never crosses 0) have no well-defined relative valley height so are assigned a value of 1.5 for graphical convenience. This parameter is closely related to peak overlap, the percent of the peak of interest that is contaminated by the neighboring peak. Examples of well-resolved and poorly-resolved co-elutions are shown as inset graphs. Deviation from unity of the ratio of deconvolved to known peak area represents true error in peak deconvolution and is observed to be approximately 10% for well resolved peaks and increase with decreasing resolution, concomitant with an increase in the width of the fit peak observed in Fig. 7a (red circles). Bias increases to 30% for poorly resolved peaks in cases with a clearly defined valley. These data span a larger range of resolutions and more realistic operating conditions than can be reasonably achieved by injection of known standards, but yield estimates of bias that agree with those observed by injection of poorly resolved standards (Supplementary Information). For wholly unresolved peaks, with no valley or inflection point and peak overlap of up to 50%, however, error exceeds any reasonable acceptable limits (though other integration approaches are expected to be similarly challenged under these conditions).

Peak fitting provides a means to improve these integrations by constraining parameters to known values. Erroneous broadening of the peak fit for poorly resolved peaks leads to the observed overestimation in peak area, but well-resolved peaks are observed to converge asymptotically to a given peak width (0.42 in the shown example). By constraining the width of the peak of interest to the known "correct" value, peaks can be fit with substantially lower error (Fig. 7b, blue squares). Error remains below 10% for all moderately resolved peaks, and increases to only ~30% even for unresolved peaks. Peaks that suffer up to 50% contamination with an unresolved co-eluting neighbor, which could not be integrated by drawn baselines or other approaches with no constrainable or known parameters, can thus be integrated with only moderate error. This result is achieved without initial assumptions of knowledge about the peak. Quantitative parameters provided by peak fitting can therefore be used not only for quality control, but can also be implemented in the future to improve fitting of peaks that are otherwise too poorly resolved to be accurately integrated. Furthermore, statistical moments of peaks may provide information (e.g. physicochemical characteristics) about otherwise unidentified peaks, but this potential application of fit parameters has not yet been explored deeply.

4. Discussion

Once properly processed, chromatographic data has been shown to yield reliable and accurate quantification of environmental constituents, but the necessary step of converting complex raw chromatographic data into integrated peak areas presents an analytical challenge. We implemented automated fitting of single ion chromatograms with multiple idealized peaks to efficiently reduce chromatographic data of complex environmental samples. This approach is found to provide true peak areas with ~10% error and yield peak areas that are comparable to other integration methods, but with additional advantages. Diverse quantitative parameters generated by peak fitting – from fit residuals to statistical moments of peaks – provide opportunities both to identify errors and to correct them, increasing the efficiency of quality control. Peaks that are chromatographically unresolved can be integrated with only moderate (10–30%) error by using fit parameters to correct or constrain peak fits. The accuracy and efficacy of peak fitting, coupled with the availability of quantitative metrics for error checking, is expected to speed up chromatographic data reduction and allow for many more peaks to be quantified within new and previously-collected datasets. The researchers performing the analyses presented in this work are estimated to have spent approximately a factor of 10 less time interacting with the data to produce peak fitting integrations than to produce reliable baseline integrations. This work demonstrates that simplified and automated single-ion peak fitting provides fast integration, efficient error checking, and quantitative

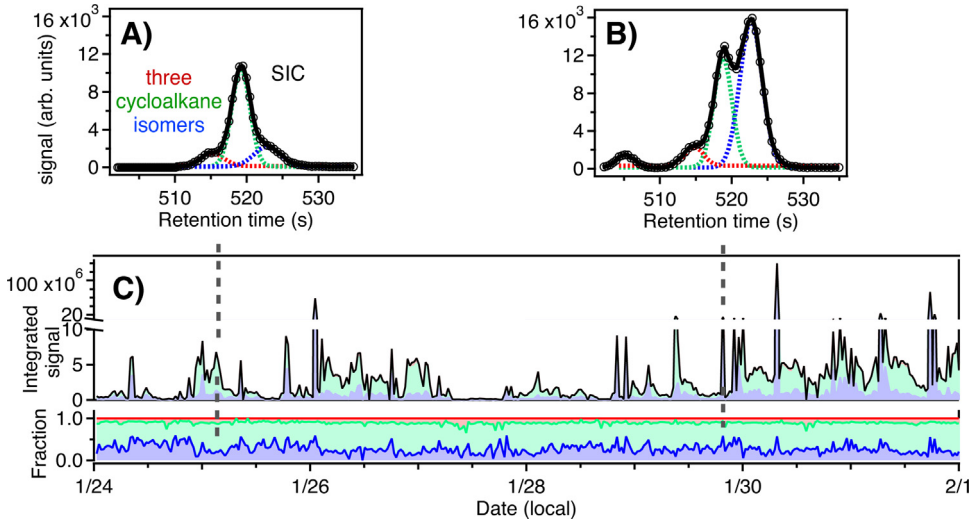

Fig. 8. Three poorly resolved, co-eluting alkylcyclohexanes from VOC-UBWOS. (a-b) Two sample SICs of co-eluting peaks of three isomers (red, green, and blue, in elution order), collected at the dashed points in the (c) timeseries of total signal. Contribution of each isomer is shown as the filled area (colored as in panels a and b) beneath the total signal (black line), and the fractional contribution of each isomer to total signal (bottom panel). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

measurements of data quality, while yielding peak areas that are comparable to drawn baseline integration.

Peak fitting provides access to new data through the deconvolution of poorly resolved peaks, allowing quantification of previously inaccessible trace compounds in environmental samples. In particular, isomers and structurally similar compounds are common in the environment, but often co-elute and share similar mass spectra. For this reason, it is not uncommon to integrate multiple compounds as a single peak (e.g. "methylbenzoic acids" [53] or "dimethyl (phenanthrenes + anthracenes)" [49]). However, in some cases, subtle differences in isomer sources and chemistry can provide insight into environmental processes (e.g. [54–56]), so improved separation and integration of these species allows novel analyses. Fig. 8 demonstrates the ability of peak fitting to access new scientific data, in this example the separation of three poorly resolved isomers emitted from gas and oil extraction. Three co-eluting peaks (alkylcyclohexane isomers) are simultaneously fit to the SIC data in this retention time window. In some cases one or more peaks are present as small, sometimes indistinguishable shoulders on a larger main peak (e.g. Fig. 8a) that would yield high error under baseline integration approaches [12], while in some cases peaks are more equal and thus better resolved (e.g Fig. 8b). The combined mass of these compounds exhibits high temporal variability, with some very high spikes in concentrations (Fig. 8c black line, top panel). However, the contribution of each isomer to total mass shown is independently variable (Fig. 8c bottom panel), so integrating all three isomers as a single peak together loses information. The two samples shown in panels (a) and (b) represent spikes in concentration dominated by two different isomers, providing insight into the different source profiles and composition of the sampled air that is not accessible without the reliable integration of poorly resolved peaks afforded by peak fitting. Furthermore, efficient integration of peaks with low chromatographic resolution relaxes constraints on instrument design, for instance increasing time resolution of field-based GCs or increasing sensitivity by reducing the number of ions scanned by the MS. Peak fitting is therefore not only a means to alleviate the burden of data reduction and improve data quality, but also to advance new scientific inquiries.

Acknowledgements

The authors would like to acknowledge the extensive effort of many individuals who contributed to the successful collection of the sample datasets used in this work, including complex logistics at all field sites. In particular, GIVW and AHG thank Yunliang Zhao, Brent Williams, and Nathan Kreisberg for their critical contributions in collecting the TAG SOAR and CalNex datasets. JBG thanks William C. Kuster for essential help collecting the VOC CalNex dataset. Support for this work was provided by the NOAA SBIR program (WC133R17CN0082), with preliminary work by GIVW supported by the NSF Graduate Research Fellowship (#DGE 1106400). TAG-SOAR data collection was supported by the California Air Resources Board (CARB) award number 03–324. TAG-BKR data collection was supported by CARB contract 09-316, and National Oceanic and Atmospheric Administration Grant NA10OAR4310104. Development for TAG instrumentation and data analysis methods was also supported by grants to UCB from the Department of Energy SBIR/STTR program (DE-FG02-08ER86335 and DE-SC0011397)

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.chroma.2017.11.005.

References

[1] J.L. Snyder, Chapter 15 – Environmental Applications of Gas Chromatography, 2004.
[2] A.H. Goldstein, I. Galbally, Known and unexplored organic constituents in the earth's atmosphere, Environ. Sci. Technol. 41 (2007) 1514–1521.
[3] J.F. Hamilton, Using comprehensive two-dimensional gas chromatography to study the atmosphere, J. Chromatogr. Sci. 48 (2010) 274–282, http://dx.doi.org/10.1093/chromsci/48.4.274.
[4] W.F. Rogge, Molecular Tracers for Sources of Atmospheric Carbon Particles: Measurements and Model Predictions (1993).
[5] J. Schnelle-Kreis, W. Welthagen, M. Sklorz, R. Zimmermann, Application of direct thermal desorption gas chromatography and comprehensive two-dimensional gas chromatography coupled to time of flight mass spectrometry for analysis of organic compounds in ambient aerosol particles, J. Sep. Sci. 28 (2005) 1648–1657, http://dx.doi.org/10.1002/jssc.200500120.

[6] T.L. Chester, Recent developments in high-performance liquid chromatography stationary phases, Anal. Chem. 85 (2013) 579–589, http://dx.doi.org/10.1021/ac303180y.
[7] S.-T. Chin, P.J. Marriott, Multidimensional gas chromatography beyond simple volatiles separation, Chem. Commun. 50 (2014) 8819–8833, http://dx.doi.org/10.1039/C4CC02018A.
[8] G.A. Eiceman, H.H. Hill, J. Gardea-Torresdey, Gas chromatography, Anal. Chem. 72 (2000) 137–144, http://dx.doi.org/10.1021/a10000054.
[9] A.H. Goldstein, B.C. Daube, J.W. Munger, S.C. Wofsy, Automated in-situ monitoring of atmospheric non-methane hydrocarbon concentrations and gradients, J. Atmos. Chem. 21 (1995) 43–59, http://dx.doi.org/10.1007/BF00712437.
[10] S.J. Andrews, L.J. Carpenter, E.C. Apel, E. Atlas, V. Donets, J.R. Hopkins, R.S. Hornbrook, A.C. Lewis, R.T. Lidster, R. Lueb, J. Minaeian, M. Navarro, S. Punjabi, D. Riemer, S. Schauffler, A comparison of very short lived halocarbon (VSLS) and DMS aircraft measurements in the tropical west Pacific from CAST, ATTREX and CONTRAST, Atmos. Meas. Tech. 9 (2016) 5213–5225, http://dx.doi.org/10.5194/amt-9-5213-2016.
[11] D. Helmig, S. Rossabi, J. Hueber, P. Tans, S.A. Montzka, K. Masarie, K. Thoning, C. Plass-Duelmer, A. Claude, L.J. Carpenter, A.C. Lewis, S. Punjabi, S. Reimann, M.K. Vollmer, R. Steinbrecher, J.W. Hannigan, L.K. Emmons, E. Mahieu, B. Franco, D. Smale, A. Pozzer, Reversal of global atmospheric ethane and propane trends largely due to US oil and natural gas production, Nat. Geosci. 9 (2016) 490–495, http://dx.doi.org/10.1038/ngeo2721.
[12] M.K.L. Bicking, Integration errors in chromatographic analysis, part II: large peak size ratios, LCGC North Am. 24 (2006) 604–616.
[13] D.J. Anderson, R.R. Walters, Effect of baseline errors on the calculation of statistical moments of tailed chromatographic peaks, J. Chromatogr. Sci. 22 (1984) 353–359.
[14] L. Cui, Z. Ling, J. Poon, S.K. Poon, H. Chen, J. Gao, P. Kwan, K. Fan, Generalized gaussian reference curve measurement model for high-performance liquid chromatography with diode array detector separation and its solution by multi-target intermittent particle swarm optimization, J. Chemom. 29 (2015) 146–153, http://dx.doi.org/10.1002/cem.2683.
[15] M.J.P. Gerritsen, H. Tanis, B.G.M. Vandeginste, G. Kateman, Generalized rank annihilation factor analysis, iterative target transformation factor analysis, and residual bilinearization for the quantitative analysis of data from liquid chromatography with photodiode array detection, Anal. Chem. 64 (1992) 2042–2056, http://dx.doi.org/10.1021/ac00042a006.
[16] J. Ohman, P. Geladi, S. Wold, Residual Bilinearization, Part 2: application to hplc-diode array data and comparison with rank annihilation factor analysis, J. Chemom. 4 (1990) 135–146.
[17] A.H. Anderson, T.C. Gibb, A.B. Littiewood, Computer resolution of unresolved convoluted gas-chromatographic peaks, J. Chromatogr. Sci. 8 (1970) 640–646.
[18] V.B. Di Marco, G.G. Bombi, Mathematical functions for the representation of chromatographic peaks, J. Chromatogr. A. 931 (2001) 1–30, http://dx.doi.org/10.1016/S0021-9673(01)01136-0.
[19] M.S. Jeansonne, J.P. Foley, Review of the exponentially modified gaussian (EMG) functionsince 1983, J. Chromatogr. Sci. 29 (1991) 258–266, http://dx.doi.org/10.1093/chromsci/29.6.258.
[20] P.J. Naish, S. Hartwell, Exponentially Modified Gaussian functions-A good model for chromatographic peaks in isocratic HPLC? Chromatographia 26 (1988) 285–296, http://dx.doi.org/10.1007/BF02268168.
[21] J. Blaško, R. Kubinec, I. Ostrovský, E. Pavlíková, J. Krupčík, L. Soják, Chemometric deconvolution of gas chromatographic unresolved conjugated linoleic acid isomers triplet in milk samples, J. Chromatogr. A. 1216 (2009) 2757–2761, http://dx.doi.org/10.1016/j.chroma.2008.11.019.
[22] J. Mydlová-Memersheimerová, B. Tienpont, F. David, J. Krupčík, P. Sandra, Gas chromatography of 209 polychlorinated biphenyl congeners on an extremely efficient nonselective capillary column, J. Chromatogr. A. 1216 (2009) 6043–6062, http://dx.doi.org/10.1016/j.chroma.2009.06.049.
[23] B.J. Williams, A.H. Goldstein, N.M. Kreisberg, S.V. Hering, J.P.D. Abbatt, An in-situ instrument for speciated organic composition of atmospheric aerosols: thermal desorption aerosol GC/MS-FID (TAG), Aerosol Sci. Technol. 40 (2006) 627–638, http://dx.doi.org/10.1080/02786820600754631.
[24] Y. Zhao, N.M. Kreisberg, D.R. Worton, A.P. Teng, S.V. Hering, A.H. Goldstein, Development of an In situ thermal desorption gas chromatography instrument for quantifying atmospheric semi-Volatile organic compounds, Aerosol Sci. Technol. 47 (2013) 258–266, http://dx.doi.org/10.1080/02786826.2012.747673.
[25] G. Isaacman, N.M. Kreisberg, L.D. Yee, D.R. Worton, A.W.H. Chan, J.A. Moss, S.V. Hering, A.H. Goldstein, Online derivatization for hourly measurements of gas- and particle-phase semi-volatile oxygenated organic compounds by thermal desorption aerosol gas chromatography (SV-TAG), Atmos. Meas. Tech. 7 (2014) 4417–4429, http://dx.doi.org/10.5194/amt-7-4417-2014.
[26] D.R. Gentner, D.R. Worton, G. Isaacman, L.C. Davis, T.R. Dallmann, E.C. Wood, S.C. Herndon, A.H. Goldstein, R.A. Harley, Chemical composition of gas-phase organic carbon emissions from motor vehicles and implications for ozone production, Environ. Sci. Technol. 47 (2013) 11837–11848, http://dx.doi.org/10.1021/es401470e.
[27] D.B. Millet, N.M. Donahue, S.N. Pandis, A. Polidori, C.O. Stanier, B.J. Turpin, A.H. Goldstein, Atmospheric volatile organic compound measurements during the Pittsburgh air quality study: results, interpretation, and quantification of primary and secondary contributions, J. Geophys. Res. Atmos. 110 (2005) 1–17, http://dx.doi.org/10.1029/2004JD004601.
[28] J.B. Gilman, W.C. Kuster, P.D. Goldan, S.C. Herndon, M.S. Zahniser, S.C. Tucker, W.A. Brewer, B.M. Lerner, E.J. Williams, R.A. Harley, F.C. Fehsenfeld, C. Warneke, J.A. De Gouw, Measurements of volatile organic compounds duringthe 2006 TexAQS/GoMACCS campaign: industrial influences, regional characteristics, and diurnal dependencies of the OH reactivity, J. Geophys. Res. Atmos. 114 (2009) 1–17, http://dx.doi.org/10.1029/2008JD011525.
[29] J.B. Gilman, B.M. Lerner, W.C. Kuster, P.D. Goldan, C. Warneke, P.R. Veres, J.M. Roberts, J.A. de Gouw, I.R. Burling, R.J. Yokelson, Biomass burning emissions and potential air quality impacts of volatile organic compounds and other trace gases from fuels common in the US, Atmos. Chem. Phys. 15 (2015) 13915–13938, http://dx.doi.org/10.5194/acp-15-13915-2015.
[30] J.B. Gilman, J.F. Burkhart, B.M. Lerner, E.J. Williams, W.C. Kuster, P.D. Goldan, P.C. Murphy, C. Warneke, C. Fowler, S.A. Montzka, B.R. Miller, L. Miller, S.J. Oltmans, T.B. Ryerson, O.R. Cooper, A. Stohl, J.A. de Gouw, Ozone variability and halogen oxidation within the Arctic and sub-Arctic springtime boundary layer, Atmos. Chem. Phys. 10 (2010) 10223–10236, http://dx.doi.org/10.5194/acp-10-10223-2010.
[31] P.D. Goldan, W.C. Kuster, E. Williams, P.C. Murphy, F.C. Fehsenfeld, J. Meagher, Nonmethane hydrocarbon and oxy hydrocarbon measurements duringthe 2002 New England air quality study, J. Geophys. Res. D Atmos. 109 (2004) 1–14, http://dx.doi.org/10.1029/2003JD004455.
[32] B.M. Lerner, J.B. Gilman, K.C. Aikin, E.L. Atlas, P.D. Goldan, M. Graus, R. Hendershot, G.A. Isaacman-Vanwertz, A. Koss, W.C. Kuster, R.A. Lueb, R.J. Mclaughlin, J. Peischl, D. Sueper, T.B. Ryerson, T.W. Tokarek, C. Warneke, B. Yuan, J.A. De Gouw, An improved, automated whole air sampler and gas chromatography mass spectrometry analysis system for volatile organic compounds in the atmosphere, Atmos. Meas. Tech. 10 (2017) 291–313, http://dx.doi.org/10.5194/amt-10-291-2017.
[33] R.G. Prinn, R.F. Weiss, P.J. Fraser, P.G. Simmonds, D.M. Cunnold, F.N. Alyea, S. O'Doherty, P. Salameh, B.R. Miller, J. Huang, R.H.J. Wang, D.E. Hartley, C. Harth, L.P. Steele, G. Sturrock, P.M. Midgley, A. McCulloch, A history of chemically and radiatively important gases in air deduced from ALE/GAGE/AGAGE, J. Geophys. Res. Atmos. 105 (2000) 17751–17792, http://dx.doi.org/10.1029/2000JD900141.
[34] E. Andrews, P. Saxena, S. Musarra, L.M. Hildemann, P. Koutrakis, P.H. McMurry, I. Olmez, W.H. White, Concentration and composition of atmospheric aerosols fromthe 1995 SEAVS experiment and a review of the closure between chemical and gravimetric measurements, J. Air Waste Manag. Assoc. 50 (2000) 648–664, http://dx.doi.org/10.1080/10473289.2000.10464116.
[35] B.J. Williams, A.H. Goldstein, D.B. Millet, R. Holzinger, N.M. Kreisberg, S.V. Hering, A.B. White, D.R. Worsnop, J.D. Allan, J.-L. Jimenez, Chemical speciation of organic aerosol during the international consortium for atmospheric research on transport and transformation 2004: results from in situ measurements, J. Geophys. Res. Atmos. 112 (2007) 1–14, http://dx.doi.org/10.1029/2006JD007601.
[36] C. Warneke, J.M. Roberts, P. Veres, J. Gilman, W.C. Kuster, I. Burling, R. Yokelson, J.A. De Gouw, International journal of mass spectrometry VOC identification and inter-comparison from laboratory biomass burning using PTR-MS and PIT-MS, Int. J. Mass Spectrom. 303 (2011) 6–14, http://dx.doi.org/10.1016/j.ijms.2010.12.002.
[37] J.A. de Gouw, P.D. Goldan, C. Warneke, W.C. Kuster, J.M. Roberts, M. Marchewka, S.B. Bertman, A.A.P. Pszenny, W.C. Keene, Validation of proton transfer reaction-mass spectrometry (PTR-MS) measurements of gas-phase organic compounds in the atmosphere during the New England air quality study (NEAQS) in 2002, J. Geophys. Res. 108 (2003) 1–18, http://dx.doi.org/10.1029/2003JD003863.
[38] N.M. Kreisberg, S.V. Hering, B.J. Williams, D.R. Worton, A.H. Goldstein, Quantification of hourly speciated organic compounds in atmospheric aerosols, measured by an in-situ thermal desorption aerosol gas chromatograph (TAG), Aerosol Sci. Technol. 43 (2009) 38–52, http://dx.doi.org/10.1080/02786820802459583.
[39] E. Matisová, M. Dömötörová, Fast gas chromatography and its use in trace analysis, J. Chromatogr. A. 1000 (2003) 199–221, http://dx.doi.org/10.1016/S0021-9673(03)00310-8.
[40] E.C. Apel, A.J. Hills, R. Lueb, S. Zindel, S. Eisele, D.D. Riemer, A fast-GC/MS system to measure C2 to C4 carbonyls and methanol aboard aircraft, J. Geophys. Res. 108 (2003) 8794, http://dx.doi.org/10.1029/2002JD003199.
[41] E.C. Apel, L.K. Emmons, T. Karl, F. Flocke, A.J. Hills, S. Madronich, J. Lee-Taylor, A. Fried, P. Weibring, J. Walega, D. Richter, X. Tie, L. Mauldin, T. Campos, A. Weinheimer, D. Knapp, B. Sive, L. Kleinman, S. Springston, R. Zaveri, J. Ortega, P. Voss, D. Blake, A. Baker, C. Warneke, D. Welsh-Bon, J.A. de Gouw, J. Zheng, R. Zhang, J. Rudolph, W. Junkermann, D.D. Riemer, Chemical evolution of volatile organic compounds in the outflow of the Mexico City Metropolitan area, Atmos. Chem. Phys. 10 (2010) 2353–2375, http://dx.doi.org/10.5194/acp-10-2353-2010.
[42] M.L. Phillips, R.L. White, Dependence of chromatogram peak areas obtained by curve-Fitting on the choice of peak shape function, J. Chromatogr. Sci. 35 (1997) 75–81, http://dx.doi.org/10.1093/chromsci/35.2.75.
[43] NIST mass spec data center S.E. stein director, mass spectra, in: P. Linstrom, W. Mallard (Eds.), NIST Chem. WebBook, NIST Stand. Ref. Database Number 69, National Institute of Standards and Technology, Gaithersburg, MD, 2012, p. 20899 http://webbook.nist.gov.
[44] A. Borbon, J.B. Gilman, W.C. Kuster, N. Grand, S. Chevaillier, A. Colomb, C. Dolgorouky, V. Gros, M. Lopez, R. Sarda-Esteve, J. Holloway, J. Stutz, H. Petetin, S. McKeen, M. Beekmann, C. Warneke, D.D. Parrish, J.A. de Gouw, Emission ratios of anthropogenic volatile organic compounds in northern mid-latitude

megacities: observations versus emission inventories in Los Angeles and Paris, J. Geophys. Res. Atmos. 118 (2013) 2041–2057, http://dx.doi.org/10.1002/jgrd.50059.
[45] P.M. Edwards, S.S. Brown, J.M. Roberts, R. Ahmadov, R.M. Banta, J.A. deGouw, W.P. Dubé, R.A. Field, J.H. Flynn, J.B. Gilman, M. Graus, D. Helmig, A. Koss, A.O. Langford, B.L. Lefer, B.M. Lerner, R. Li, S.-M. Li, S.A. McKeen, S.M. Murphy, D.D. Parrish, C.J. Senff, J. Soltis, J. Stutz, C. Sweeney, C.R. Thompson, M.K. Trainer, C. Tsai, P.R. Veres, R.A. Washenfelder, C. Warneke, R.J. Wild, C.J. Young, B. Yuan, R. Zamora, High winter ozone pollution from carbonyl photolysis in an oil and gas basin, Nature 514 (2014) 351–354, http://dx.doi.org/10.1038/nature13767.
[46] G. Isaacman, N.M. Kreisberg, D.R. Worton, S.V. Hering, A.H. Goldstein, A versatile and reproducible automatic injection system for liquid standard introduction: application to in-situ calibration, Atmos. Meas. Tech. 4 (2011) 1937–1942, http://dx.doi.org/10.5194/amt-4-1937-2011.
[47] N.M. Kreisberg, D.R. Worton, Y. Zhao, G. Isaacman, A.H. Goldstein, S.V. Hering, Development of an automated high-temperature valveless injection system for online gas chromatography, Atmos. Meas. Tech. 7 (2014) 4431–4444, http://dx.doi.org/10.5194/amt-7-4431-2014.
[48] G. Isaacman-VanWertz, L.D. Yee, N.M. Kreisberg, R. Wernis, J.A. Moss, S.V. Hering, S.S. De Sá, S.T. Martin, M.L. Alexander, B.B. Palm, W. Hu, P. Campuzano-Jost, D.A. Day, J.-L. Jimenez, M. Riva, J.D. Surratt, J. Viegas, A. Manzi, E.S. Edgerton, K. Baumann, R. Souza, P. Artaxo, A.H. Goldstein, Ambient gas-Particle partitioning of tracers for biogenic oxidation, Environ. Sci. Technol. 50 (2016) 9952–9962, http://dx.doi.org/10.1021/acs.est.6b01674.
[49] B.J. Williams, A.H. Goldstein, N.M. Kreisberg, S.V. Hering, D.R. Worsnop, I.M. Ulbrich, K.S. Docherty, J.-L. Jimenez, Major components of atmospheric organic aerosol in southern California as determined by hourly measurements of source marker compounds, Atmos. Chem. Phys. 10 (2010) 11577–11603, http://dx.doi.org/10.5194/acp-10-11577-2010.
[50] Y. Zhao, N.M. Kreisberg, D.R. Worton, G. Isaacman, D.R. Gentner, A.W.H. Chan, R.J. Weber, S. Liu, D.A. Day, L.M. Russell, S.V. Hering, A.H. Goldstein, Sources of organic aerosol investigated using organic compounds as tracers measured during CalNex in Bakersfield, J. Geophys. Res. Atmos. 118 (2013) 11388–11398, http://dx.doi.org/10.1002/jgrd.50825.
[51] M.K.L. Bicking, Integration errors in chromatographic analysis, part I: peaks of approximately equal size, LCGC North Am. 24 (2006) 402–414.
[52] B.J. Place, M.J. Morris, M.M. Phillips, L.C. Sander, C.A. Rimmer, Evaluation of the impact of peak description on the quantitative capabilities of comprehensive two-dimensional liquid chromatography, J. Chromatogr. A. 1368 (2014) 107–115, http://dx.doi.org/10.1016/j.chroma.2014.09.060.
[53] J.J. Schauer, M.J. Kleeman, G.R. Cass, B.R.T. Simoneit, Measurement of emissions from air pollution sources. 2. c $_1$ through C $_{30}$ organic compounds from medium duty diesel trucks, Environ. Sci. Technol. 33 (1999) 1578–1587, http://dx.doi.org/10.1021/es980081n.
[54] A.W.H. Chan, G. Isaacman, K.R. Wilson, D.R. Worton, C.R. Ruehl, T. Nah, D.R. Gentner, T.R. Dallmann, T.W. Kirchstetter, R.A. Harley, J.B. Gilman, W.C. Kuster, J.A. De Gouw, J.H. Offenberg, T.E. Kleindienst, Y.H. Lin, C.L. Rubitschun, J.D. Surratt, P.L. Hayes, J.-L. Jimenez, A.H. Goldstein, Detailed chemical characterization of unresolved complex mixtures in atmospheric organics: insights into emission sources, atmospheric processing, and secondary organic aerosol formation, J. Geophys. Res. Atmos. 118 (2013) 6783–6796, http://dx.doi.org/10.1002/jgrd.50533.
[55] A.B. Fialkov, A. Gordin, A. Amirav, Hydrocarbons and fuels analyses with the supersonic gas chromatography mass spectrometry-the novel concept of isomer abundance analysis, J. Chromatogr. A. 1195 (2008) 127–135, http://dx.doi.org/10.1016/j.chroma.2008.04.074.
[56] B. Nozière, N.J.D. Gonzlez, A.K. Borg-Karlson, Y. Pei, J.P. Redeby, R. Krejci, J. Dommen, A.S.H. Prevot, T. Anthonsen, Atmospheric chemistry in stereo: a new look at secondary organic aerosols from isoprene, Geophys. Res. Lett. 38 (2011) 1–5, http://dx.doi.org/10.1029/2011GL047323.