

Mass Spectra Deconvolution of Gaseous Mixtures Containing Volatile Organic Compounds

D. Nikolić¹, S. M. Madzunkov², and M. R. Darrach³

California Institute of Technology, Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, CA, 91109

The Spacecraft Atmosphere Monitor (S.A.M.) is a highly compact gas chromatograph mass spectrometer (GCMS) that will analyze the spacecraft atmosphere for all volatile organic compounds (VOCs). In the Trace Gas Analysis (TGA) mode of operation the S.A.M will detect and quantify VOCs from 50 to 10,000 parts-per-billion. GCMS mass spectra from 20 to 360 Th are generated at a rate of twenty full mass spectra per second and GC co-elutions can yield overlapping electron-impact mass spectra. Despite the presence of these molecular isobars, we report herein on the S.A.M. deconvolution algorithms that have been developed and are capable of identifying target species based on their characteristic fragmentation patterns. We investigate the efficiency of deconvolution algorithm as a function of mass resolution with which mass spectrum is acquired. Finding the balance between deconvolution accuracy and generated data volume under time constraints and limited computing resources is the main topic of this study.

Nomenclature

GC	= Gas Chromatograph	CFM	= Competitive Fragmentation Modeling
ISS	= International Space Station	m/q	= mass-to-charge quotient
JPL	= Jet Propulsion Laboratory	Th	= Thomson (1.0364E-8 kg/C)
MCA	= Major Constituents Analysis	A_{cc}	= Fragmental Similarity Matrix
MEMS	= Micro-Electro-Mechanical System	Δm	= m/q resolution (bin width)
MS	= Mass Spectrometer	N_i	= sample size (number of all initial ions)
EII	= Electron Impact Ionization	$N_i^{(c)}$	= number of initial ions of species (c)
CITA	= Computational Ion Trap Analyzer	$N_r^{(c)}$	= number of retrieved ions of species (c)
PC	= Preconcentrator	$N^{(c)}$	= number of unique ions of species (c)
S.A.M.	= Spacecraft Atmosphere Monitor	$\epsilon^{(c)}$	= deconvolution error for species (c)
QIT	= JPL Quadrupole Ion Trap	$\vec{f}^{(c)}$	= fragment distribution for species (c)
rf	= Radio Frequency	$\alpha_m^{(c)}$	= probability density for species (c) at m/q bin
SMAC	= Spacecraft Max. Allowable Conc.	$\bar{\alpha}_m^{(c)}$	= randomly perturbed $\alpha_m^{(c)}$ at every m/q bin
TGA	= Trace Gas Analysis	$\delta_m^{(c)}$	= random perturbation level at every m/q bin
VOC	= Volatile Organic Compound		

I. Introduction

THE purpose of this study is to review the role of a sample size and the degree of detail in recorded mass spectrum as it pertains to the automated interpretation of mass spectrometer (MS) data delivered by the S.A.M.'s quadrupole ion trap mass spectrometer (QITMS)¹. The purpose of S.A.M., as a technological demonstration, is to monitor the cabin air of the International Space Station (ISS) for both the major constituents and trace volatile organic compounds (VOC). Further details on instrument architecture and performance requirements can be found in latest S.A.M. progress report² and references therein.

¹ Technologist, 389 - Microdevice And Sensor Systems, Mail Stop: 306-392.

² Senior Technologist, 389 - Microdevice And Sensor Systems, Mail Stop: 306-392.

³ Senior Technologist, 389 - Microdevice And Sensor Systems, Mail Stop: 306-431.

During the Trace gas Analysis (TGA) mode of operation, intake of cabin air is routed through a preconcentrator (PC) for approximately 10 minutes, whereby the cabin-air VOCs are adsorbed. There are 22 specific SMAC compounds, shown on Table 1, targeted by S.A.M., but any VOC present in the cabin-air will be adsorbed on the PC. Following the flash-heated release of thermally desorbed species, concentrated sample is eluted through the GC micro-column into the QITMS sensor using H₂ carrier gas. Inside the QITMS all neutral molecular species are fragmented and ionized via the electron impact ionization (EII) using a 70eV electron beam. Fragment ions, generated in the center of the trapping volume are confined with respect to their mass-to-charge ratio (m/q). This is done by driving the QITMS central ring electrode with 1MHz *rf* scan function capable of trapping ions with $1 < m/q < 361\text{Th}$.

In this simulation, initial ion clouds containing N_i number of fragment ions from all species in Table 1 are propagated in time using the Computational Ion Trap Analyzer (CITA) program⁶, and detected in time-of-flight fashion by recording ion splat times. Detection times are linearly proportional to ion's m/q value. Mass spectra of various resolutions are obtained when detection times are histogrammed into bins with widths Δm ranging from 1mTh to 1Th. Each molecular species has unique EII fragmentation pattern conveniently found in NIST3 database, but these are limited to the unit m/q resolution ($\Delta m=1\text{Th}$). If sub-unit m/q resolution is required, then the Competitive Fragmentation Modeling (CFM) method⁴ may be used to obtain most probable fragment ions with non-integer masses. The isotopic distribution within each such fragment is then generated by the Isotopic Abundances computer program⁵ in the format compatible to the TrapParticle tool from the CITA⁶ suite of programs.

Appearance of less abundant fragment ions in large sample sizes can change the similarity metrics between various molecular species, and affect the deconvolution accuracy with which we can distinguish contributions of different species in a given sample mixture. The deconvolution algorithm with unit m/q resolution has been previously described in detail elsewhere⁷. Here we discuss the role of different simulation modalities in addressing deconvolution accuracy: (i) sample size effect, (ii) sub-unit m/q resolution, and (iii) effect of inaccuracies in EII fragmentation patterns. Large sample size effects ($N_i \geq 10^6$ ions) at unit m/q resolution were previously investigated by one us⁷, and here we address smaller sample sizes and sub-unit m/q resolutions that correspond to operating the QITMS sensor at lower sensitivity and higher data rates..

Table 1: List of compounds (c) used in deconvolution of simulated spectra.

(c)	$N_i^{(c)}$			CAS#	formula	(c)	$N_i^{(c)}$			CAS#	formula
	$N_i=10^4$	$N_i=10^5$	$N_i=10^6$				$N_i=10^4$	$N_i=10^5$	$N_i=10^6$		
(0)	24	46	46	1066-40-6	C ₃ H ₁₀ OSi	(11)	36	53	56	106-42-3	C ₈ H ₁₀
(1)	28	37	37	57-55-6	C ₃ H ₈ O ₂	(12)	41	63	68	108-38-3	C ₈ H ₁₀
(2)	56	109	127	541-02-6	C ₁₀ H ₃₀ O ₅ Si ₅	(13)	35	55	59	95-47-6	C ₈ H ₁₀
(3)	51	72	73	541-05-9	C ₆ H ₁₈ O ₃ Si ₃	(14)	31	40	41	108-88-3	C ₇ H ₈
(4)	54	101	112	556-67-2	C ₈ H ₂₄ O ₄ Si ₄	(15)	26	26	26	71-43-2	C ₆ H ₆
(5)	4	5	5	7664-41-7	H ₃ N	(16)	26	33	33	67-64-1	C ₃ H ₆ O
(6)	7	8	8	50-00-0	CH ₂ O	(17)	23	34	39	71-36-3	C ₄ H ₁₀ O
(7)	9	15	16	67-56-1	CH ₄ O	(18)	22	31	31	67-63-0	C ₃ H ₈ O
(8)	14	21	21	76-19-7	C ₃ F ₈	(19)	20	23	23	64-17-5	C ₂ H ₆ O
(9)	22	27	27	108-62-3	C ₈ H ₁₆ O ₄	(20)	19	20	20	107-02-8	C ₃ H ₄ O
(10)	21	31	32	75-09-2	CH ₂ Cl ₂	(21)	19	29	32	110-54-3	C ₆ H ₁₄

II. Methodology

Based on fragmentation patterns of molecular species (c) given in Table 1, we then use the TrapParticle tool to randomly generate initial ion cloud of varying sizes having in total $N_i=10^4$ - 10^6 fragment ions, as described in detail in CITA⁶. We note that in this study we work with equipartial distribution of species in gas mixtures, meaning that each species (c) in Table 1 enters the gas mixture sample with 1/22 probability. It is expected that each species (c) contribute $N_i^{(c)}$ fragment ions, a number drawn from Poisson distribution with the mean and the variance equal to $N_i/22$, and sampled under a constraint that total number of generated ions $N_i = \sum_{(c)} N_i^{(c)}$ is fixed. Due to random nature of fragment sampling algorithm, the equipartiality of gas mixtures is guaranteed only for extremely large sample sizes ($N_i > 10^8$). Nevertheless, we note that each generated fragment ion in our simulations is uniquely described with its parent molecular species (c). Therefore, by counting how many fragment ions $N_i^{(c)}$ of species (c) are present in the reference ion cloud mix and how many are retrieved $N_r^{(c)}$ from the corresponding mass spectrum by the deconvolution algorithm, the absolute deconvolution error is $\varepsilon^{(c)} = 100\% \times |1 - N_r^{(c)}/N_i^{(c)}|$.

Because each fragment ion of mass m has different probability $\alpha_m^{(c)}$ to be created from any of species (c) during the EII process, the most probable fragments will appear first in clouds with $N_i=10^4$ ions, and least abundant fragments will start to appear only in clouds with sizes greater than 10^6 ions. For example, fragmental probability density for o-Xylene, as shown in Figure 1, dominates by two unique fragments, both illustrated with their chemical structures. Inset in Figure 1 presents some of the less probable fragments of o-Xylene. The high-resolution binning clearly shows a slight shift to the right with respect to the centers of standard unit-resolution bins. This is direct

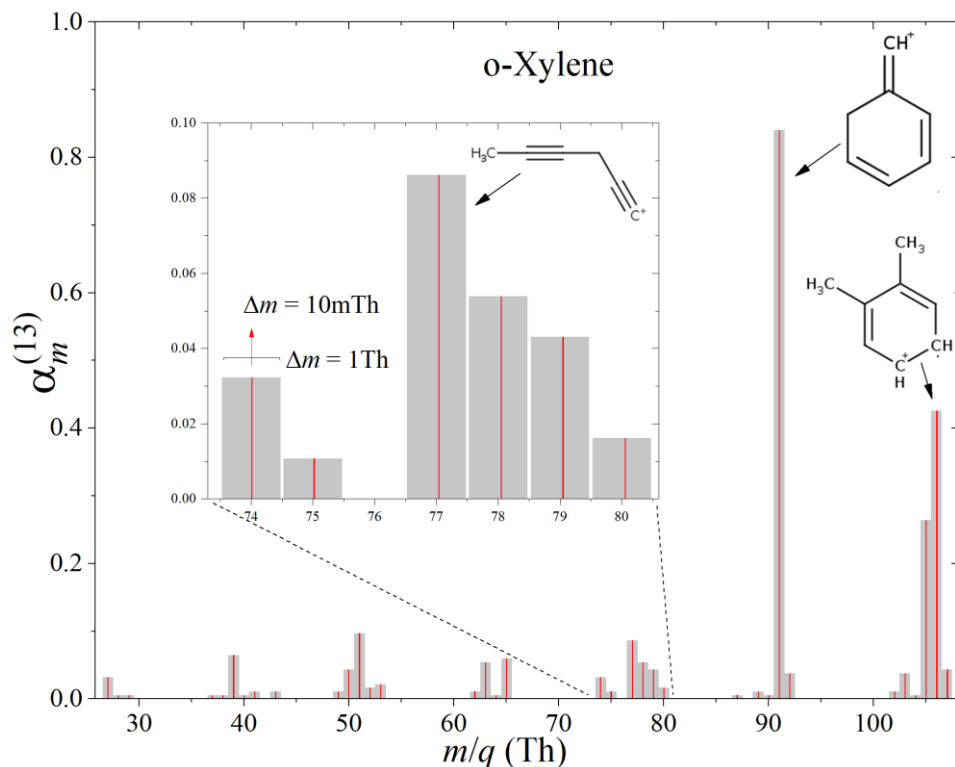


Figure 1: Fragmental probability density for o-Xylene (c=13). Initial ion cloud for equipartial mixture has size of $N_i=10^4$ ions, such that o-Xylene contributed $N_i^{(c)}=444$ fragment ions, out of which $N^{(c)}=35$ are unique. Inset enlarges the 73-81Th range, to emphasize the difference between the standard $\Delta m = 1\text{Th}$ resolution (gray bins) and sub-unit $\Delta m = 0.01\text{Th}$ resolution (red bins).

consequence of the unit-resolution binning of the mass spectrum due to the round-off errors of non-integer m/q values. It is obvious that in the unit-resolution representation any other species that might contribute the fragments that differ by at most 0.5Th will not be resolved from o-Xylene's fragments.

Table 1 also contains the number of unique fragments, $N^{(c)}$, that each molecular species (c) will contribute to initial ion clouds of varying sizes. We note that $N^{(c)}$ values weakly depend on the used m/q resolution, but are most sensitive to the total number of ions N_i present in the initial ion cloud. For example, if sample size is fixed at $N_i=10^4$ ions, the corresponding $N^{(c)}$ values reported in Table 1 remain similar for all investigated m/q resolutions ($\Delta m = 1\text{mTh}, 5\text{mTh}, 10\text{mTh}, 50\text{mTh}, 0.1\text{Th}, 0.5\text{Th},$ and 1Th). However, if sample size is increased to $N_i=10^5$ ions such that less probable fragment ions are likely to be created, then the number of unique fragments $N^{(c)}$ that each species can contribute increases as well.

III. Results and Discussion

The overview of computational algorithm as it pertains to the random walk deconvolution method can be found in detail in Nikolić et al.⁷. Convergence of iterative minimization of the 2-norm of the difference between the reference mass spectrum and the current guess relies on the subtle changes in off-diagonal elements of the fragmental similarity matrix $A_{c,ci} = \sum_m \alpha_m^{(c)} \alpha_m^{(ci)}$, as illustrated in Figure 2. The summation index m is a shorthand notation for a

bin with the width Δm and representing the m/q ratio. Thus, index m can take values from $1 \leq m/q \leq 361$ range in steps of Δm . The $\alpha_m^{(c)}$ represents the probability density that molecular species (c) contribute a fragment ion to this particular bin m . The norm of fragmental probability density, $(\sum_m \alpha_m^{(c)} \alpha_m^{(c)})^{1/2}$, is set to unit and therefore all diagonal elements satisfy $A_{c,c} = 1$ regardless of the step size Δm . Such normalization of the fragmental similarity matrix ensures that the diagonal elements for the identical species are always unit, and that species with no common fragments are represented with off-diagonal elements of value zero.

Figure 2 summarizes fragmental similarity of all species listed in Table 1 for wide range of m/q resolutions Δm and sample sizes N_i . For the fixed sample size, most changes in fragmental similarity matrix can be traced back to silanols ($c=0-4$), xylenes ($c=11-13$) and alcohols ($c=17-19$) for when m/q resolution increases from $\Delta m = 1\text{Th}$ to $\Delta m = 0.1\text{Th}$. Further increase in m/q resolution to at most $\Delta m = 1\text{mTh}$ introduces changes only for larger sample sizes and mainly through recoupling of xylenes to toluene ($c=14$) or benzene ($c=15$). Figure 2 is helpful in predicting which species are likely to interfere in convergence of deconvolution algorithm⁷ when dealing with equipartial gas mixtures where all compounds elute in similar amounts and simultaneously. This situation is often encountered in the MEMS PC/GC failure mode, where all thermally desorbed species are released at the same into the QITMS sensor. Based on similarity matrices $A_{c,c}$ an informed early decisions can be made on which species are to be introduced in the beginning of a random walk protocol and improve the overall convergence.

In order to quantify the error propagation within the random walk algorithm⁷, we perform deconvolution of synthetic ion clouds with *a priori* known composition $N_i^{(c)}$ and compare them with retrieved abundances $N_r^{(c)}$. For each species we define absolute error $\varepsilon^{(c)}$ as a relative difference between given $N_i^{(c)}$ and algorithmically retrieved $N_r^{(c)}$ abundances. Retrieval accuracy is a complex function of fragmentation patterns of all species listed in Table 1.

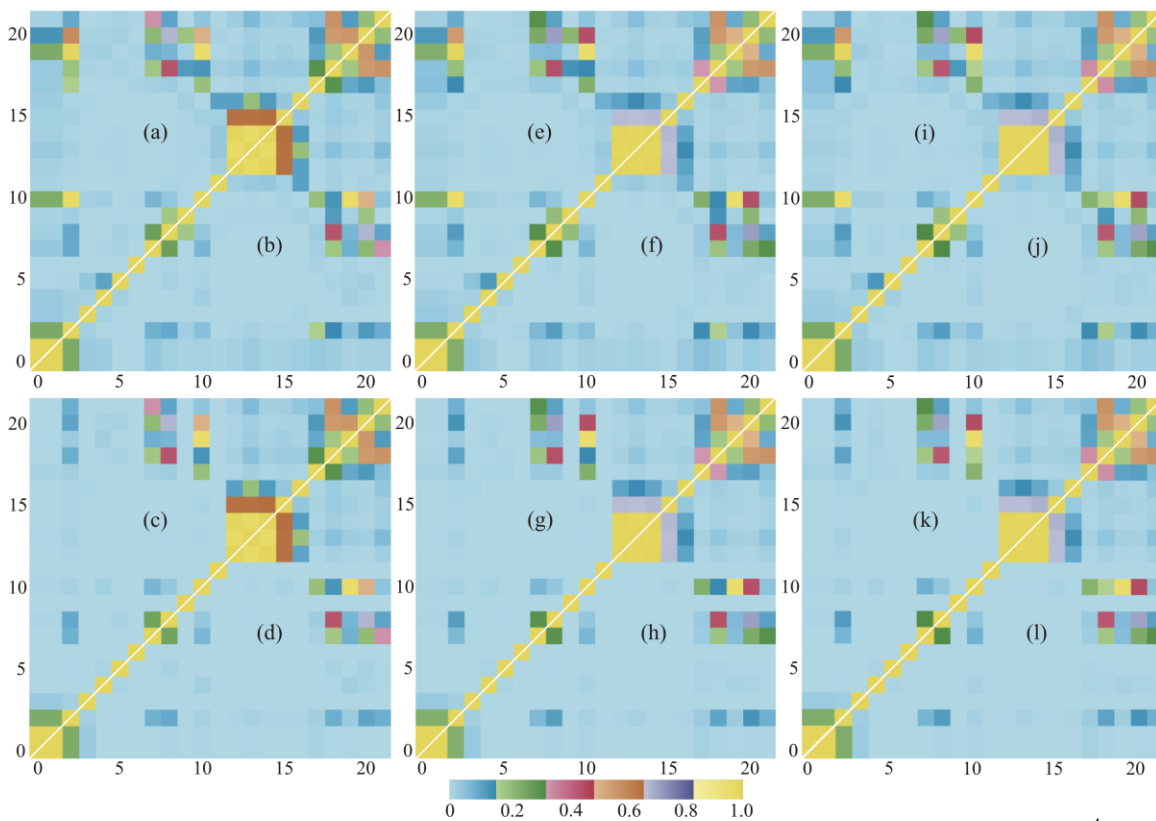


Figure 2: Fragmental similarity matrices $A_{c,c}$ of all species from Table 1. Sample sizes are: $N_i=10^4$ ions (left panels), $N_i=10^5$ ions (central panels), and $N_i=10^6$ ions (right panels). Triangular matrices correspond to different Δm values: 1Th (a,e,i); 0.1Th (b,f,j); 10mTh (c,g,k), and 1mTh (d,h,l). See text for details.

Main descriptor of single species (c) is its fragment distribution $\vec{f}^{(c)} = \sum_m \alpha_m^{(c)} \vec{e}_m$, where eigenvector \vec{e}_m uniquely defines m/q bin with width Δm , and $\alpha_m^{(c)}$ is a probability that this bin holds a fragment ion as a product of

EII fragmentation, as previously illustrated in Figure 1. In addition, our simulations allow the probabilities $\alpha_m^{(c)}$, to be randomly perturbed at every bin m by an amount of $\delta_m^{(c)}$ (at most $\pm 15\%$). These randomly-perturbed probabilities, $\bar{\alpha}_m^{(c)} = (1 + \delta_m^{(c)}/100) \times \alpha_m^{(c)}$, are then used to deconvolute the mass spectrum corresponding to the ion cloud of *a priori* known composition $N_i^{(c)}$. Introduced perturbations serve as a stress test for the robustness of the random walk algorithm⁷ and establish a tolerable level of uncertainty when defining fragmentation patterns. For example, uncertainties in fragmentation patterns can be caused by intermittent changes in electron beam energy, which directly affect the EII cross section for creating fragment ions from neutral molecular species.

Unperturbed probabilities $\alpha_m^{(c)}$ are deduced directly from the composition of the ion cloud of various sizes, by counting the number of fragment ions $N_i^{(c)}$, that originate from the given species (c). Unperturbed probabilities are then multiplied by a number randomly drawn from the 0.85 – 1.15 range for every bin m , to yield perturbed probabilities $\bar{\alpha}_m^{(c)}$. We also systematically reduce the random range to 0.99-1.01, and investigate the rate of reduction in deconvolution error as perturbed probabilities $\bar{\alpha}_m^{(c)}$ approach their unperturbed values $\alpha_m^{(c)}$.

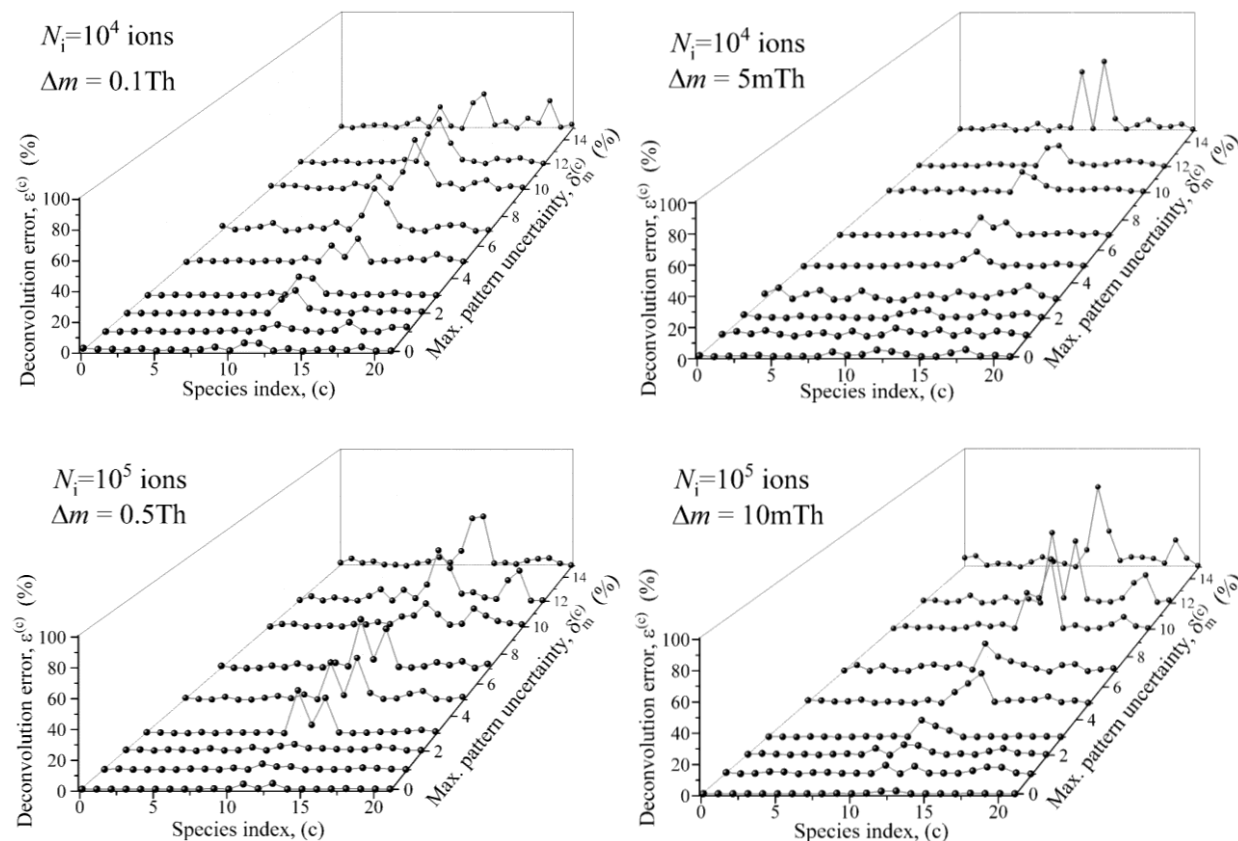


Figure 3: Effect of random perturbations of fragmentation probabilities $\alpha_m^{(c)}$ on accuracy of the mass spectrum deconvolution. Reliability of random walk algorithm for Xylene isomers ($c=11-13$) at sub-unit m/q resolutions Δm consistently improves with the accuracy of the fragmentation probabilities.

Summary of results for few representative cases is given in Figure 3. However, an extensive set of simulations followed by the random walk deconvolutions have been performed on large number of synthetic mass spectra of various m/q resolutions Δm , ranging from 1Th to 0.5mTh. In general, increasing the number of fragment ions in the simulated ion clouds will result in feature-rich mass spectra. This is because weaker mass lines start to appear due to creation of less probable fragment ions. These new weak lines can slightly change similarity matrix $A_{c,c'}$ between any of the two investigated species (c) and (c') contributing to the equipartial mixture. If similarity of $A_{c,c'}$ is reduced due to appearance of new fragments, then the random walk algorithm becomes more successful in setting apart these two species, and vice versa. However, the overall accuracy in retrieving the exact amounts of each species entering the mixture is most sensitive to the accuracy with which fragmentation patterns are *a priori* known.

This fact is illustrated in Figure 3 whenever fragmentation probabilities $\bar{\alpha}_m^{(c)}$ are randomly perturbed by $\pm 15\%$ compared to their unperturbed values. Resulting deconvolution error can be as high as 80% for xylenes, suggesting for example that the m-Xylene is falsely identified as its p- or o-Xylene isomers. Thus, it is of great importance to create a library with complete sets of fragmentation patterns of all target species. Completeness mainly refers to the availability of fragmentation patterns at several sub-unit m/q resolutions Δm that can be applied to various sample sizes. For example, if integrated count rate in the measured mass spectrum is 10^3 ions per second, then one needs to accumulate counting statistics for 10 seconds in order to use fragmentation probabilities that are calibrated on sample sizes with $N_i=10^4$ ions. Otherwise the analysed mass spectrum which has been acquired at lower counting statistics may lack few mass peaks that are exclusive to a particular species in a gas mixture.

Another example is when 1-120Th mass range is measured in 4000 mass channels, such that each mass bin has width of 30mTh. In this case it is not recommended to use fragmentation probabilities that are calibrated for resolutions $\Delta m < 30\text{mTh}$, mainly to avoid oversampling. Nevertheless the use of fragmentation probabilities that are calibrated for coarser resolutions is possible but only after rebinning the measured spectra to match the sought level of detail. We note that use of high resolution fragmentation patterns on samples of small sizes is not recommended, because random sampling of small number of ions from large population will inherently have large variance between successive samplings and thus be poorly described by a single fragmentation pattern..

IV. Conclusion

We have presented a simulation study in which we benchmark the random walk deconvolution algorithm to be used in analysis of the MS data produced by the S.A.M. instrument. Standard NIST fragmentation pattern library is compiled at unit m/q resolution and offers inadequate level of detail compared to the experimental mass spectra generated by the S.A.M. This study identifies the need for more complete library of fragmentation probabilities of target species that will correspond to actual measurement conditions, such are ion trapping and detection parameters that determine peak shapes in recorded mass spectrum as well as the data collection rates. We found that sub-unit mass resolution does improve the accuracy and efficiency with which random walk algorithm retrieves true relative abundances of Xylene isomers but only if fragmentation probabilities are calibrated with uncertainties of up to 3%. We note that the obtained results represent worst case scenario, in which employed MEMS PCGC technology fails to provide time resolved elution of studied species and instead all species simultaneously enter the S.A.M.'s QIT-MS sensor module.

Acknowledgments

This work has been carried out at the Jet Propulsion Laboratory, California Institute of Technology, under the contract with the National Aeronautic and Space Administration. © 2018. California Institute of Technology. Government sponsorship acknowledged.

References

- ¹S.M.Madzunkov et al., "Progress Report on the Spacecraft Atmosphere Monitor's Development Model", 47th International Conference on Environmental Systems, Charleston, S. Carolina; article 324 (2017).
- ²S.M.Madzunkov et al., "Development Model of Spacecraft Atmosphere Monitor", 48th International Conference on Environmental Systems, Albuquerque, New Mexico; abstract 325 (2018).
- ³NIST Standard Reference Database 1A, "NIST/EPA/NIH Mass Spectral Library with Search Program", Data Version: NIST 14, Software Version: 2.2g
- ⁴F. Allen et al., "Computational Prediction of Electron Ionization Mass Spectra to Assist in GC/MS Compound Identification", Analytical Chemistry 88(15), 7689-7697 (2016).
- ⁵D. Nikolić, "Isotopic Abundances", unpublished (2018).
- ⁶D. Nikolić et al., "Computer Modeling of an Ion Trap Mass Analyzer, Part I: Low Pressure Regime", Journal of the American Society for Mass Spectrometry 26(12): 2115-2124 (2014).
- ⁷D. Nikolić et al., "Mapping of Spacecraft Atmosphere Monitor Signal to Major Constituent Abundances", 46th International Conference on Environmental Systems, Vienna, Austria; article 285 (2016).