

Decomposition of electron ionization mass spectra for space application using a Monte-Carlo approach

Thomas Gautier, Joseph Serigano, Jérémy Bourgalais, Sarah M. Hörst, Melissa G. Trainer

▶ **To cite this version:**

Thomas Gautier, Joseph Serigano, Jérémy Bourgalais, Sarah M. Hörst, Melissa G. Trainer. Decomposition of electron ionization mass spectra for space application using a Monte-Carlo approach. Rapid Communications in Mass Spectrometry, 2020, 34 (8), pp.e8684. 10.1002/rcm.8684 . insu-02393452

HAL Id: insu-02393452
https://insu.hal.science/insu-02393452

Submitted on 6 Sep 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Decomposition of Electron Ionization Mass Spectra for Space Application using a Monte-Carlo approach

Thomas Gautier[1*], Joseph Serigano[2], Jérémy Bourgalais[1], Sarah M. Hörst[2], Melissa G. Trainer[3]

[1] LATMOS-IPSL, CNRS, Sorbonne Université, UVSQ, Guyancourt, France

[2] Johns Hopkins University, Baltimore, MD, USA

[3] NASA Goddard Space Flight Center, Greenbelt, MD, USA

*corresponding author:

Thomas Gautier
LATMOS
11 bvd d'Alembert
78280 Guyancourt, France
Thomas.gautier@latmos.ipsl.fr

KEYWORDS: Electron Ionization, Deconvolution, Space Instrumentation, Quadrupole mass spectrometer, Monte-Carlo

ABSTRACT

RATIONALE

Quadrupole mass spectrometers equipped with an Electron Ionization (EI) sources have been widely used in space exploration to investigate the composition of planetary surfaces and atmospheres. However, the complexity of the samples and the minimal calibration for the fragmentation of molecules in the ionization chambers have prevented the deconvolution of the majority of the mass spectra obtained at different targets, thus limiting determination of the exact composition of the samples analyzed. We propose a Monte-Carlo approach to solve this issue mathematically.

METHODS

We have decomposed simulated mass spectra of mixtures acquired with unit resolving power mass spectrometers and EI sources into the sum of the single components fragmentation patterns weighted by their relative concentration using interior-point least-square fitting. To fit compounds with poorly known fragmentation patterns, we use a Monte-Carlo method to vary the intensity of individual fragments. We then decompose the spectrum thousands of times to obtain a statistical distribution.

RESULTS

By performing the deconvolution on a mixture of seven different molecules with interfering fragmentation patterns (H_2O, O_2, CH_4, Ar, N_2, C_2H_4 and C_2H_6) we show that this approach retrieves the mixing ratio of the individual components more accurately than regular mass spectra decomposition methods that rely on fragmentation pattern from general databases. It also provides the probability density function for each species' mixing ratio.

CONCLUSIONS

By removing the solution degeneracy in the decomposition of mass spectra, the method described herein could significantly increase the scientific retrieval from archived space flight mass spectrometry data, where calibration of the ionization source is no longer an option.

1 INTRODUCTION

The majority of the mass spectrometers (MS) that have flown in space to date have had a resolving power close to the unit resolution, and have used an electron ionization (EI) source to ionize sample gases prior to their mass analysis. The prevalence of this approach is mainly due to the volume, mass, power, and robustness requirements of this subsystem, all of which are critical for space exploration[1,2]. To form ions through electron ionization, the incoming gas sample interacts with an electron beam at a given mean energy, typically 70 eV. It is well known that upon such ionization most, if not all, molecular compounds will fragment. This yields a mass spectrum for a molecular compound comprised not only of one peak at the nominal *m/z* but of several peaks at different *m/z*, resulting in a fragmentation pattern for each molecule analyzed. The position of the peaks in the fragmentation pattern is characteristic of the original molecule and the incident electron energy. This means that the fragmentation pattern observed in a sample can be used to trace back to the original molecule analyzed and lead to its identification (Figure 1). While this approach works extremely well for single compounds, it becomes harder to identify the analyzed material in the case of complex mixtures that could be composed of dozens of different molecular compounds. One approach for analyzing complex mixtures is to separate each individual component, for example using gas chromatography (GC) prior to the MS analysis. However, this approach is not always viable in space exploration where adding a GC to a mass spectrometer adds several kilograms and power in a finite resource environment. This explains why many of the mass spectrometers previously used for space exploration were not coupled to a chromatographic system. There have also been instances in which a GC-capable MS has been operated in MS-mode only due to sparse samples and the need to maximize sensitivity. This was the case for the COSAC (Cometary Sampling and Composition Experiment) instrument aboard the Philae lander of the Rosetta mission that landed on comet 67P/Churyumov-Gerasimenko in 2014[3] and the GCMS (Gas Chromatograph Mass

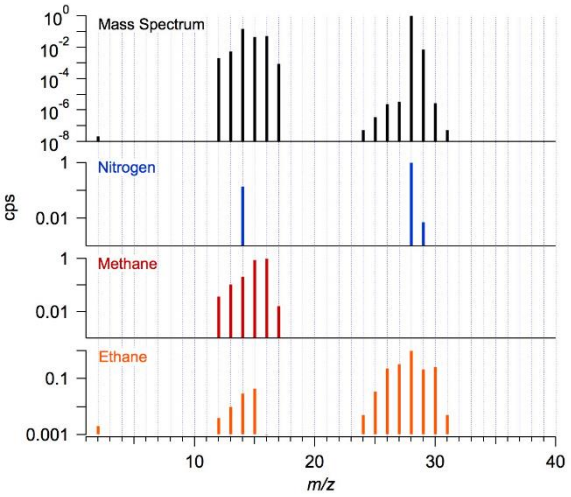


Figure 1: Example of the fragmentation of molecular compounds (nitrogen in blue, methane in red and ethane in orange) in an EI ionization source and how the measured mass spectrum (black, top) is the sum of the fragmentation patterns of all the molecules present in the sample weighted by their relative concentrations and ionization cross-sections.

Spectrometer) instrument aboard the Huygens probe that landed on Titan, the largest satellite of Saturn, in 2005[4,5].

When only a mass spectrum of the sample is available, the results tend to be an intricate mix of mass peaks corresponding to the addition of multiple fragmentation patterns of the molecules present in the sample (Figure 1). However, while the peak positions in a molecule fragmentation pattern are predictable, their relative intensities are highly dependent of the ion source geometry[6]. This explains the variability in the relative weight of fragments in the mass spectra of compounds even with laboratory grade instruments on Earth databases[7–10]. This variability is typically of 20-30% of a fragment intensity but can go much higher in extreme cases. For example, for methane, there is a factor 3.5 in the peak intensity of fragment CH_2^+ at *m/z* 14 between the value from the NIST database (20.4 % of the base peak at *m/z* 16) and the value from pre-flight calibration of the Ion and Neutral Mass Spectrometer aboard Cassini (5.7% of *m/z* 16). Thus, for space flight instruments, the ideal approach would be pre-flight calibration of all the molecules to be detected by the instrument. However, while a few compounds are usually calibrated, this approach is highly unrealistic on a large scale (1) due to the *a-priori* unknown composition of the samples to be studied, (2) the limited time available for the instrument calibration during its integration, and (3) the desire to prevent contamination of instrument surfaces that would preclude sensitive detections at the target.

The fact that the fragmentation patterns highly depend on the instrument used and that flight instruments are insufficiently calibrated introduces a strong degeneracy in the mathematical decomposition of the mass spectra, and explains why most mass spectra acquired by space instruments have not been fully deconvolved so far, with only very limited counter examples[3,11]. In this paper, we present a novel approach to deconvolution, introducing a Monte-Carlo sampling on the expected molecules fragmentation patterns, to overcome the degeneracy inherent within compositional analysis.

2 THEORY AND CALCULATION

2.1 FRAGMENTATION PATTERNS

Since the pressure in an EI ion source is kept low (typically around $1x10^{-6}$ mbar) ensuring a mean free path sufficiently large (ca. 100 m) to avoid bimolecular and collisional reactions. The only possible interaction remaining is between electrons (typically at 70 eV) and the neutral gas phase molecules from the sample. Since these interactions take place at different locations within the ion source, each resulting ion formed from a neutral molecule (M) could have a

variety of internal energies. Ions formed with low internal energy will be directly detected as the parent peak of the molecule (M^+) in the spectrum. However, if a molecule is ionized with a sufficient initial energy to reach a dissociative ionization threshold, it will decompose through different reactions depending on its internal energy to generate fragments of the original molecule. In some cases, the fragment ion will have sufficient remaining internal energy to further decompose into an even smaller ion. The decomposition processes explain the fragmentation pattern of a given compound, while the stability of the formed ions and the geometry of the interaction between the incoming neutral molecule and the electron beam explain the variation in fragment intensities depending on the measuring instrument. For further information on the fragmentation processes, we suggest reading the detailed chapters 4, 7 and 8 of *Interpretation of mass spectra* by McLafferty and Turecek[6].

2.2 MASS SPECTRA DECOMPOSITION

Deconvolving a mass spectrum to identify and quantify the different molecules present in a sample relies on being able to solve the linear system:

$$Mass\ Spectrum = Fragmentation\ Database \times Species\ concentrations \quad \textbf{(1)}$$

which equates to, in an ideal case, solving equation (1):

$$\begin{pmatrix} I_{m/z\,1} \\ I_{m/z\,2} \\ \vdots \\ I_{m/z\,j} \end{pmatrix} = \begin{pmatrix} F_{1,1} & F_{1,2} & F_{1,i} \\ F_{2,1} & F_{2,2} & F_{2,i} \\ \vdots & \ddots & \vdots \\ F_{j,1} & \cdots & F_{j,i} \end{pmatrix} \times \begin{bmatrix} N_1 \\ N_2 \\ \vdots \\ N_i \end{bmatrix} \quad \textbf{(2)}$$

where $I_{m/z\ j}$ is the ion count in the measured mass spectrum for a mass bin j, N_i is the concentration of each species i arriving on the detector of the spectrometer, and $F_{j,i}$ is the fragment intensity for species i in the mass bin j.

Several methods exist to solve such system. For example, the single value decomposition has been used to treat data from the Cassini Ion and Neutral Mass Spectrometer (INMS) instrument[11], while least-square fitting was used for the data from the COSAC instrument on the Philae lander[3]. In our case we choose to use least-squares fitting on multiple linear regressions, which can be more time consuming but allows better constraint of the solution if

boundary conditions are known on some species. Equation (1) is thus solved by finding the least squares minimum of the Euclidian norm of (2) as described in (3) and using condition (4):

$$\min_N \frac{1}{2} \|F \times N - MS\|_2^2 \quad \textbf{(3)}$$

with

$$L \leq N \leq U \quad \textbf{(4)}$$

where F is the matrix containing the fragmentation database of the molecules suspected to be present in the sample; N is the output vector containing their relative concentration at the detector of the spectrometer; MS is the measured mass spectrum; and L and U are respectively the vectors containing the lower limit and upper limit of molecules in N, both of which must be positive.

A variety of numerical recipes exist to minimize Equation (3) but we chose to use an interior-point algorithm that easily accommodates large matrices such as mass spectra[12–16]. However, mathematically decomposing a mass spectrum only works if the fragmentation patterns are well known. Failure to constrain the exact fragment ratios for all molecules present in the sample would most often result in a degeneracy of the solutions. Such degeneracy would render any relative quantification of the molecules in the sample questionable.

2.3 MONTE-CARLO APPROACH

To overcome the degeneracy in mass spectra deconvolution we introduced a Monte-Carlo approach in our algorithm. This approach has been successfully applied in planetary sciences to further refine photochemical models of planetary atmospheres such as Earth, Neptune, Saturn and Titan by randomizing the poorly known reaction rate coefficients[17–20].

In our case the randomization is applied to the nominal fragmentation pattern of each molecule in our database to create *n* different fragmentation databases *F*, typically n=10,000 for the work presented here, and then decompose the measured mass spectrum with each fragmentation database individually. Out of these 10,000 runs, the best 5% (*i.e.* with minimal residuals) are kept to retrieve a statistical solution of the problem. This allows retrieving not only the most probable concentrations for a given species, but also its probability distribution function.

The randomization process is only applied to the intensity of the fragments, not their position in the mass spectrum; *i.e.* N_2 will always have peaks at *m/z* 7, 14 and 28 (corresponding to N^{2+}, N^+ and N_2^+ ions), and only there, but their relative intensity will be varied in each Monte-Carlo

run. The relative intensity variation is made within a chosen range. Typically for molecules whose fragmentation pattern is only constrained from mass spectral databases, we allow a variation, *d*, of ±30% for each fragment's intensity around its nominal value. For every run *n*, each member $F_{j,i}$ of the fragmentation database matrices *F* can thus be expressed as :

$$F_{j_n,i_n} = F_{j_1,i_1} \times (1 + \varepsilon) \quad \textbf{(5)}$$

with

$$-d \leq \varepsilon \leq d \quad \textbf{(6)}$$

where F_{j_1,i_1} corresponds to the nominal fragmentation of the molecules and ε is a pseudorandom number comprised between -1 and + 1 and generated using a uniform pseudorandom number generator based on the Mersenne Twister method. To avoid *a priori* on the validity of the measurement, we chose to use uniformly pseudorandom distributed numbers for ε. Actually it is very likely that epsilon do not follow a uniform distribution but a normal one, centered around the experimental value determined with the same instrument. As our method intends to correct for the poor knowledge of this value when the measurement is performed with a uncalibrated instrument (and thus where epsilon is taken for example from the NIST database), we chose not to make the assumption of a normal distribution. Instead, since we cannot a priori know the bias of the database toward the actual measurement, we chose to use a uniform distribution. While this ensures that the algorithm does not favor an outcome based on false a priori, it also induces larger uncertainties on the retrieval. In cases where sufficient pre-flight calibrations are available, one would be able to determine more precisely the epsilon distribution, which would in return drastically reduce the uncertainties of our method. These pre-flight calibrations database are essential and should always be favored if available.

2.4 IONIZATION CROSS SECTIONS

Another parameter of importance to take into account when retrieving the relative concentration of molecules in a sample from an EI mass spectrum is that the intensity of a peak does not directly correlate with the concentration of a molecule, but is a product of its concentration times its ionization cross section. The molecular ionization cross sections (σ) depend on the neutral molecule ionized and on the mean energy of the electrons in the ion source (typically 70 eV). If unavailable in the literature, such cross sections can be approximated using either empirical formulae[21] or through a Binary-Encounter-Dipole (BED) or Binary-Encounter Bethe (BEB) model for electron ionization[22,23]. The origin of the electron ionization cross section

(experimental, empirical, BED or BEB) is not critical. However, since relative ionization probabilities are required to compare abundances, one must be sure to use the same origin for all the cross sections in their database to avoid introducing biases. Given the lack of availability for experimental data on most of the species of interest for planetary science, we recommend using the BED method for such applications. For each species in the sample, the concentration decomposed by the model, *N*, can thus be expressed as a function of its actual concentration in the sample, *C*, as follows:

$$\begin{pmatrix} N_1 \\ \vdots \\ N_i \end{pmatrix} = A \,.\, \begin{pmatrix} \sigma_1 \\ \vdots \\ \sigma_i \end{pmatrix} . \begin{pmatrix} C_1 \\ \vdots \\ C_i \end{pmatrix} \quad \textbf{(7)}$$

where *A* is an instrumental function, assumed constant for all of the molecules. Equation 7 can thus be transformed to express the relative concentration of two molecules *x* and *y* as:

$$\frac{C_x}{C_y} = \frac{N_x}{N_y} . \frac{\sigma_y}{\sigma_x} \quad . \quad \textbf{(8)}$$

Numerous steps are thus needed before retrieving the concentration of molecules in a sample from a direct mass spectrum measurement. The general workflow to achieve this using our method is summarized in Figure 2.

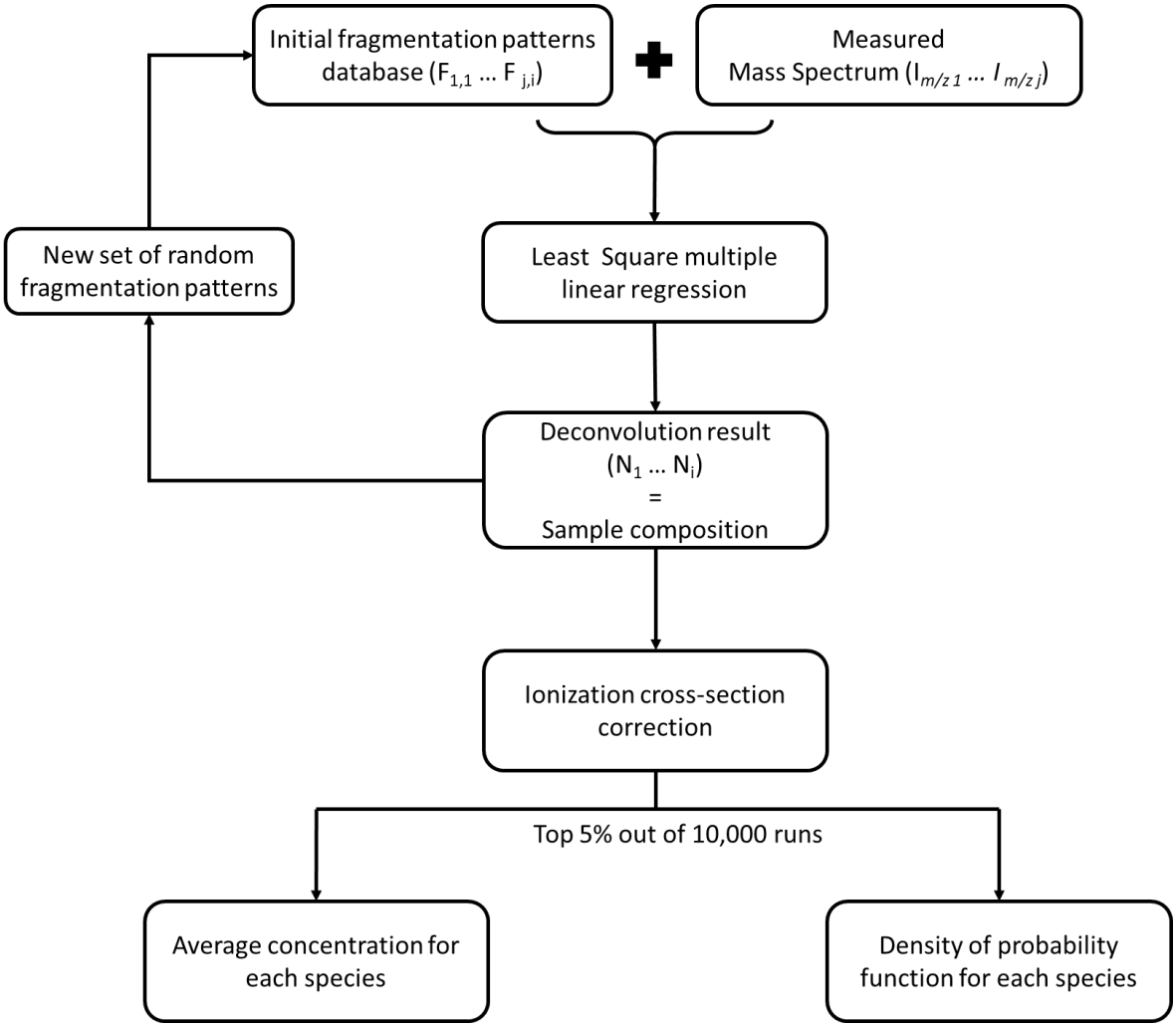


Figure 2: Schematic representation of the Monte-Carlo method used to achieve the decomposition of EI mass spectra without instrumental calibration of fragmentation patterns.

3 RESULTS AND DISCUSSION

In order to evaluate the performance of our algorithm we ran it on a synthetic mass spectrum of a mixture of 7 species – namely argon, molecular nitrogen, molecular oxygen, methane, ethane, ethylene, and water – each of which was randomly assigned a mixing ratio in the mixture, as given in Table 1, column 2. These species were chosen for testing as their fragmentation patterns particularly interfere with each other in the *m/z* 10 to 30 range. To test the limitations of our code, the sample mass spectrum was not reconstructed directly from the NIST fragmentation patterns of the 7 molecules. Instead, using the method presented in equations (5) and (6), we randomly varied their fragments with d ± 30%, and then reconstructed a mass spectrum from the newly generated fragmentation patterns. The fragmentation database

used to build the synthetic mass spectrum is given in Table SI 1 and the synthetic mass spectrum in Table SI 2. The synthetic mass spectrum used to test our approach is presented in Figure 3 (black line) with the calculated relative contribution of each species (colored bars).

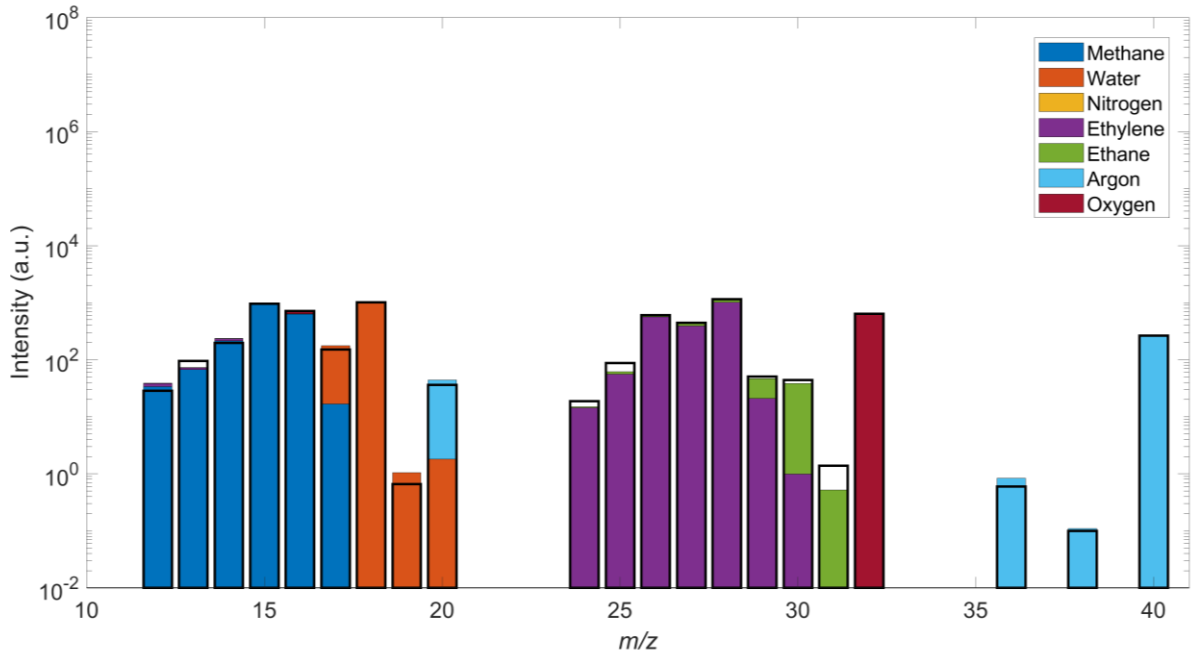


Figure 3 : Example of mass spectrum decomposition using our Monte-Carlo approach. The synthetic mass spectrum used for the decomposition is shown with the black lines. The colored bars show the calculated contribution of each of the molecules to the mixture: methane (blue), water (orange), nitrogen (yellow), ethylene (purple), ethane (green), argon (cyan) and oxygen (red).

As visible in Figure 3, our algorithm is able to precisely reconstruct the mass spectrum, with the minor exceptions of *m/z* 24, 25, 30 and 31, indicating inconsequential mismatches in the fragmentation patterns of ethylene and ethane. In order to fully evaluate the performance of our decomposition algorithm we performed three different deconvolutions using: (1) the nominal fragmentation patterns of these molecules from the constructed database given in Table SI 1, which corresponds to having a custom database on the fragmentation in an instrument's ionization chamber and should, as the most accurate method, always be favored if available; (2) using the nominal fragmentations patterns of the molecules as given by the NIST database[8] (simulating a decomposition with poorly known fragmentations); and (3) using the Monte-Carlo approach described here. A comparison of the retrieved concentration is summarized in Table 1.

	Synthetic mass spectrum	Decomposition using:			
		Known fragmentations	NIST fragmentation	Monte-Carlo (mean top 5%)	Monte-Carlo (max pdf)
Compounds	*Mixing ratio (%)*	*Mixing ratio (%)*	*Mixing ratio (%)*	*Mixing ratio (%)*	*Mixing ratio (%)* ±1σ
CH_4	23.5	23.5	24.6	21.2	21.7 ±1.9
H_2O	22.1	22.1	31	24.2	22.4 ±3.2
N_2	7.9	7.9	5.2	5.4	7.2 ±3.6
C_2H_4	21.4	21.4	25	21.6	22.5 ±2.8
C_2H_6	6.3	6.3	5.9	3.7	3.5 ±1.7
Ar	5.5	5.5	8.3	7.2	6.1 ±1.2
O_2	13.4	13.4	18.4	16.6	14.5 ±2.4

Table 1: List of compounds present in the mass spectrum with their relative concentration (columns 1 and 2) and the retrieved mixing ratios from the different decomposition methods: using exactly known fragmentation patterns (column 3); using poorly known fragmentation pattern (column 4); using a randomized distribution of poorly known fragmentation pattern using a Monte-Carlo method and calculating the average of the top 5% of 10,000 decompositions (column 5); and using the Monte-Carlo method but selecting the maximum of the probability density function and its 1σ standard deviation (column 6).

(1) Compared to the exact mixing ratio introduced in the synthetic mass spectrum, the mathematical decomposition of the mass spectrum using exactly known fragmentations yields, by far, the best output. This method allows for the exact determination of the molecular ratio for every single species. This was expected and, mathematically speaking, would remain true for complex mixtures up to *z* compounds, as long as the number of peaks with signal in the input mass spectrum is greater than z and the fragmentation patterns of the molecules are perfectly characterized with the exact same ionization chamber used to measure the samples.

(2) While the NIST database approach allows the analyst to get a fairly good idea of the sample composition, using fragmentation patterns directly from the NIST database yields the largest inaccuracies in the absolute mixing ratios. Methane and ethane are well determined, but water

is severely overestimated. This comes from the differences in the fragmentation patterns between the database and the actual mass spectrum. This is representative of what would happen when deconvolving a mass spectrum measured in space using Earth databases with no fragmentation corrections.

(3) The Monte-Carlo method we propose uses the same fragmentation pattern database as the previous one (NIST), but attempts to correct for the discrepancies by varying the fragmentation patterns a large number of times and deconvolving the spectrum with each variation of the database to run a statistical analysis on the deconvolution. It can be used in two different ways. The first one is to determine the molecular mixing ratio by taking the average output of the top 5% of simulations (Table 1, column 5), calculated on the least-square regression residue. Note that this 5% threshold is not a requirement of our approach, but mainly a cut-off we introduced to limit computational time of our method. Using this method improves the information one can have of the sample. Compared to method 2, while it slightly degrades the retrieval of ethane mixing ratio, it significantly refines the retrievals of water and ethylene, and improves the retrievals of argon and oxygen. The second use, results of which populate column 6 of Table 1, is to probe the probability density functions for each species as illustrated in Figure 4. The determination of maximum probability density functions is the method we favor for better accuracy. However, this approach is more time consuming than the top 5% method. The probability density functions reflect the likelihood of a species to be at a given mixing ratio in the sample.

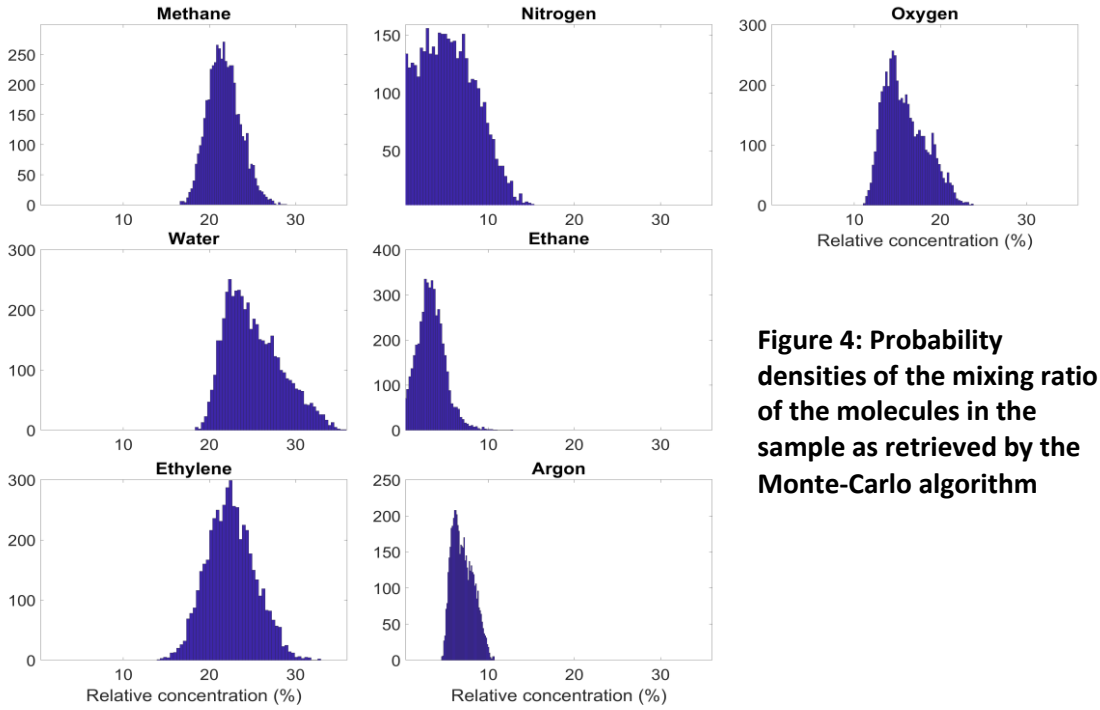


Figure 4: Probability densities of the mixing ratio of the molecules in the sample as retrieved by the Monte-Carlo algorithm

These probability densities allow for a quick determination of the parameter space occupied by each species in the sample as well as the determination of a confidence level in the retrieved mixing ratio (Table 1, column 6). Doing so demonstrates the power of this method, which can then accurately retrieve all species in Table 1 but ethane within the predicted 1σ variation. The probability density functions can also be used to probe the interdependencies between species in their mixing ratio retrievals. In our example, using a mixing ratio for ethylene in the lower side of its distribution, automatically induces a higher mixing ratio for ethane. Should ethylene mixing ratio be determined more accurately in a sample using another instrument, one could then infer the accurate mixing ratio of ethane too.

Given our observations, the proposed Monte-Carlo approach seems then to be an extremely valuable tool in mass spectral decomposition that could be applied to flight mass spectra such as the data recorded by the INMS and GCMS instruments during the Cassini-Huygens mission.

CONCLUSION

While numerous mass spectrometers have been flown in space over the last few decades, most of their data have not be deconvolved and remained under used due to the complexity of the analyzed samples and the lack of essential pre-flight calibration of the unique fragmentation patterns of molecules in the instruments.

We propose a method for the analysis of these data using Earth-based determination of the fragmentation patterns of the molecules, readily available in databases. To do so, we correct for

their divergence compared to the flight instruments' fragmentation patterns using a Monte-Carlo randomization of the fragment intensities and the generation of thousands of mass spectral decompositions. This allows for a retrieval of the probability densities for the mixing ratio of each species in the analyzed sample.

While we presented implications for space exploration mass spectrometry, the method proposed in this paper could also be applied to laboratory measurements performed with Residual Gas Analyzers, typically quadrupole mass spectrometers equipped with an electron ionization source, when direct calibration of the component is not possible.

We also believe the method presented here can significantly improve the science retrieval in archived space flight data. For such applications, we want to point out that in cases with low count rates – for example the most intense peaks in the mass spectrum recorded by the COSAC instrument at the surface of comet 67P were only at ~20 counts – the counting statistics would introduce an additional source of uncertainty. In such cases, this source of error would need to be combined with the fitting ones to retrieve proper uncertainties on the relative abundances in the sample.

Finally, we also confirm that the most accurate method, by far, to determine a sample composition analyzed by a space mass spectrometer relies on the *a priori* knowledge of the exact fragmentation pattern of every single suspected compound with the *exact* same instrument. We thus reaffirm the need for extended pre-launch calibrations of future flight mass spectrometers as practicable.

ACKNOWLEDGEMENTS

This research was supported by the National Aeronautics and Space Administration under16-CDAP16_2-0087, initially issued through the Cassini Data Analysis Program, and NASA's Planetary Science Division Internal Scientist Funding Program through the Fundamental Laboratory Research (FLaRe) work package to M.G.T.

REFERENCES

1. Palmer PT, Limero TF. Mass spectrometry in the U.S. space program: past, present, and future. *J Am Soc Mass Spectrom*. 2001;12(6):656-675. doi:10.1016/S1044-0305(01)00249-5

2. Mahaffy P. Mass Spectrometers Developed for Planetary Missions. In: ; 1999:355-376.

doi:10.1007/978-94-011-4728-6_13

3. Goesmann F, Rosenbauer H, Bredehoeft JH, et al. Organic compounds on comet 67P/Churyumov-Gerasimenko revealed by COSAC mass spectrometry. *Science (80-).* 2015. doi:10.1126/science.aab0689

4. Niemann HB, Atreya SK, Demick JE, et al. Composition of Titan's lower atmosphere and simple surface volatiles as measured by the Cassini-Huygens probe gas chromatograph mass spectrometer experiment. *J Geophys Res*. 2010;115(E12):E12006. doi:10.1029/2010JE003659

5. Niemann HB, Atreya SK, Bauer SJ, et al. The abundances of constituents of Titan's atmosphere from the GCMS instrument on the Huygens probe. *Nature*. 2005;438(7069):779-784. doi:10.1038/nature04122

6. McLafferty FW, Tureček F. *Interpretation of Mass Spectra*. University Science Books; 1993. https://books.google.fr/books?id=xQWk5WQfMQAC&redir_esc=y. Accessed January 2, 2019.

7. Ausloos P, Clifton C., Lias S., et al. The critical evaluation of a comprehensive mass spectral library. *J Am Soc Mass Spectrom*. 1999;10(4):287-299. doi:10.1016/S1044-0305(98)00159-7

8. Linstrom PG, Mallard W., Wallace E, eds. NIST Mass Spectrometry Data Center. In: *WebBook de Chimie NIST, Base de Données Standard de Référence NIST Numéro 69*. Gaityhersurbg, MD; 2019. doi:https://doi.org/10.18434/T4D303

9. Wallace WE, Ji W, Tchekhovskoi D V., Phinney KW, Stein SE. Mass Spectral Library Quality Assurance by Inter-Library Comparison. *J Am Soc Mass Spectrom*. 2017;28(4):733-738. doi:10.1007/s13361-016-1589-4

10. Valdez CA, Leif RN, Hok S, Alcaraz A. Assessing the reliability of the NIST library during routine GC-MS analyses: Structure and spectral data corroboration for 5,5-diphenyl-1,3-dioxolan-4-one during a recent OPCW proficiency test. *J Mass Spectrom*. 2018;53(5):419-422. doi:10.1002/jms.4073

11. Cui J, Yelle RV, Vuitton V, et al. Analysis of Titan's neutral upper atmosphere from Cassini Ion Neutral Mass Spectrometer measurements. *Icarus*. 2009;200(2):581-615. doi:10.1016/j.icarus.2008.12.005

12. Potra FA, Wright SJ. Interior-point methods. *J Comput Appl Math*. 2000;124(1-2):281-302. doi:10.1016/S0377-0427(00)00433-7

13. Woitke P, Kamp I, Thi W-F. Radiation thermo-chemical models of protoplanetary disks. *Astron Astrophys*. 2009;501(1):383-406. doi:10.1051/0004-6361/200911821

14. Karmarkar N. A new polynomial-time algorithm for linear programming. *Combinatorica*. 1984;4(4):373-395. doi:10.1007/BF02579150

15. Vanderbei RJ, Carpenter TJ. Symmetric indefinite systems for interior point methods. *Math Program*. 1993;58(1-3):1-32. doi:10.1007/BF01581257

16. Altman A, Gondzio J. Regularized symmetric indefinite systems in interior point methods for linear and quadratic optimization. *Optim Methods Softw*. 1999;11(1-4):275-302. doi:10.1080/10556789908805754

17. Stolarski RS, Butler DM, Rundel RD. Uncertainty propagation in a stratospheric model 2. Monte Carlo analysis of imprecisions due to reaction rates. *J Geophys Res*. 1978;83(C6):3074. doi:10.1029/JC083iC06p03074

18. Dobrijevic M, Parisot JP. Effect of chemical kinetics uncertainties on hydrocarbon production in the stratosphere of neptune. *Planet Space Sci*. 1998;46(5):491-505. doi:10.1016/S0032-0633(97)00176-1

19. Dobrijevic M, Ollivier JL, Billebaud F, Brillet J, Parisot JP. Effect of chemical kinetic uncertainties on photochemical modeling results: Application to Saturn's atmosphere. *Astron Astrophys*. 2003;398(1):335-344. doi:10.1051/0004-6361:20021659

20. Hébrard E, Dobrijevic M, Bénilan Y, Raulin F. Photochemical kinetics uncertainties in modeling Titan's atmosphere: First consequences. *Planet Space Sci*. 2007;55(10):1470-1489. doi:10.1016/J.PSS.2007.04.006

21. Fitch WL, Sauter AD. Calculation of relative electron impact total ionization cross sections for organic molecules. *Anal Chem*. 1983;55(6):832-835. doi:10.1021/ac00257a006

22. Kim Y-K, Rudd ME. Binary-encounter-dipole model for electron-impact ionization. *Phys Rev A*. 1994;50(5):3954-3967. doi:10.1103/PhysRevA.50.3954

23. Hwang W, Kim Y -K., Rudd ME. New model for electron-impact ionization cross

sections of molecules. *J Chem Phys*. 1996;104(8):2956-2966. doi:10.1063/1.471116

SUPPLEMENTARY INFORMATION

Table SI 1: Fragmentation database used to construct the synthetic mass spectrum

	m/z																					
	12	***13***	***14***	***15***	***16***	***17***	***18***	***19***	***20***	***24***	***25***	***26***	***27***	***28***	***29***	***30***	***31***	***32***	***36***	***37***	***38***	***39***	***40***
Methane	2.6	9.1	15.3	100.0	64.6	1.4																	
Water					0.6	13.7	100.0	0.1	0.2														
Nitrogen			8.7											100.0	0.5								
Ethylene	0.5	1.1	2.8	0.4						2.7	11.7	80.0	55.7	100.0	2.0	0.1							
Ethane	0.5	1.6	3.3	5.3	0.2					0.6	5.6	43.1	42.7	100.0	21.8	26.2	0.8						
Argon									12.9										0.2		0.0		100.0
Oxygene					15.0													100.0					

Table SI 2: Synthetic mass spectrum created to test the algorithm

m/z	Intensity (a.u.)	*m/z*	Intensity (a.u.)
1	0	21	0
2	0	22	0
3	0	23	0
4	0	24	18.79
5	0	25	87.16
6	0	26	603.7
7	0.01	27	441.25
8	0	28	1156.17
9	0	29	50.57
10	0	30	43.81
11	0	31	1.38
12	28.66	32	635.05
13	95.35	33	0
14	196.37	34	0
15	954.51	35	0
16	711.12	36	0.6
17	150.87	37	0
18	1007.72	38	0.1
19	0.66	39	0
20	35.99	40	264.81