



Cite this: *Analyst*, 2024, **149**, 4388

Thermogravimetry-synchronized, reference-free quantitative mass spectrometry for accurate compositional analysis of polymer systems without prior knowledge of constituents†

Yusuke Hibi, * Shiho Uesaka and Masanobu Naito *

Compositional analysis (CA)—identification and quantification of the system constituents—is the most fundamental and decisive approach for investigating the system of interest. Pyrolysis mass spectrometry (MS) with a high resolution of over 10 000 is very effective for chemical identification and is directly applicable to polymer materials regardless of their solubilities. However, it is less helpful for quantification, especially when the references, *i.e.*, pure constituents, are unknown, non-isolable and thus cannot be prepared. To compensate for this weakness, herein we propose reference-free quantitative mass spectrometry (RQMS) with enhanced quantification accuracy assisted by synchronized thermogravimetry (TG). The key to success lies in correlating the instantaneous weight loss from TG with the MS signal, enabling the quantitative evaluation of the distinct ionization efficiency for each fragment individually. The determined ionization efficiencies allow the conversion of MS signal intensities of pyrolyzed fragments into weight abundances. In a benchmark test using ternary polymer systems, this new framework named TG-RQMS demonstrates accurate CA within ± 1.3 wt% errors without using any prior knowledge or spectra of the references. This simple yet accurate and versatile CA method would be an invaluable tool to investigate polymer materials whose composition is hardly accessible *via* other analytical methods.

Received 29th April 2024,
Accepted 21st June 2024

DOI: 10.1039/d4an00624k

rsc.li/analyst

Introduction

Compositional analysis (CA) is the most fundamental approach for investigating the system of interest. “Composition” usually means chemical composition; however, recent studies showed that the CA framework can access geometric/physical information as well, *e.g.*, the monomer sequence along the main chain¹ and thermal properties,² *via* carefully selecting the system constituents, which further enhances the importance of CA. The feasibility of CA *via* some spectroscopy depends on the availability of the references, *i.e.*, pure system constituents. When a comprehensive library including all possible reference spectra is available, CA is simply reduced to spectral deconvolution under the assumption of a linear mixing model.³ A representative example is elemental CA using techniques such as electron energy loss

spectroscopy (EELS)⁴ and energy dispersive X-ray (EDX).⁵ Since the number of natural elements is limited to 92, a comprehensive library can be prepared and is usually installed by default on the instrument. In contrast, such libraries for molecular-level CAs are not feasible to be prepared, since the size of all possible molecular structures is on the order of 10^{60} .⁶ Therefore, reference spectra must be prepared for each individual case, which is feasible only when all system components are known and isolable. In mass spectrometry (MS), this approach is well-known as label-free quantitative MS and is frequently used in proteomics.⁷ This strategy was also exploited in pyrolysis-MS for quantifying microplastics in soil.⁸ However, material scientists often face situations where the system components are unknown or non-isolable and thus the reference spectra are not feasible to be prepared. Even in such cases, we recently demonstrated that CA was executable *via* reference-free quantitative MS (RQMS) based on pyrolysis-MS;¹ there, unmeasurable reference spectra were inferred from a spectral dataset of mixtures. This data-driven method inspired by unsupervised learning techniques for spectral imaging^{9,10} is effective especially when the analyte compositions are unknown but can be somehow modulated by changing experimental conditions. Indeed, by inferring reference spectra of sequence-defined copolymers based on a dataset of random

Data-driven Polymer Design Group, Research Center for Macromolecules and Biomaterials, National Institute for Materials Science (NIMS), 1-2-1, Sengen, Tsukuba, Ibaraki 305-0047, Japan. E-mail: hibi.yusuke@nims.go.jp, naito.masanobu@nims.go.jp

† Electronic supplementary information (ESI) available: Materials and methods, supplementary figures (Fig. S1 and S2) and spectral dataset with sample information (Data S1–S3). See DOI: <https://doi.org/10.1039/d4an00624k>



copolymers synthesized from various monomer feed ratios, RQMS solved a long-standing question in polymer science—how to analyze sequence distribution in copolymers.¹ With the versatility of pyrolysis-MS, which is directly applicable to any polymer/composite materials, regardless of their solubilities in solvents, RQMS has great potential to be a decisive characterization method in polymer science. However, the reported accuracy in benchmark CA for ternary polymer systems was insufficient with ± 5 wt% composition errors, which might cause non-negligible inconsistency between RQMS and conventional nuclear magnetic resonance (NMR) in sequence distribution analysis.¹ To become a more reliable and quantitative characterization method, drastic improvements in accuracy are needed for RQMS.

To this end, herein, we propose thermogravimetry-assisted RQMS (TG-RQMS). The concurrent measurement of TG and MS (TG-MS) was first proposed in the 1960s, originally designed to identify the chemical structures of pyrolyzed gases.¹¹ Later, TG-MS was utilized for quantifying small molecules such as H_2O , CO and CO_2 , based on the finding that the MS peak intensities and weight losses were identically proportional, regardless of their chemical structures.¹² However, larger fragments generated from polymer pyrolysis have individually different ionization efficiencies and also undergo

various ionization fragmentations. Therefore, the proportional constants between MS peak intensities and weight losses are peak-wisely different, obscuring how to quantitatively and complementarily analyze TG-MS. Nevertheless, MS peak intensities and derivative TG (dTG) (Fig. 1) are at least correlated, even if not proportional, which would allow converting the MS peak intensities into the weight abundances of pyrolyzed fragments. This conversion would put more focus on the fragments with lower ionization coefficients during the analysis; they are less dominant in MS spectra but may occupy a relatively large fraction by weight, as compared to those with higher ionization coefficients. To intuitively understand this key concept, assume a trace amount of contaminant with a much higher ionization coefficient as compared to major polymeric components; then, the observed MS spectra would be significantly distorted by the contaminant. Nevertheless, its importance can be properly assessed through the conversion into the weight fraction. To this end, the principal contribution of this paper is to first provide a quantitative method for determining the distinct ionization efficiency of each fragment individually. Ionization efficiency for each fragment is herein defined as the ratio of MS peak intensity to weight. Although TG data allow for the capture of instantaneous weight loss, this represents the aggregate weight loss from

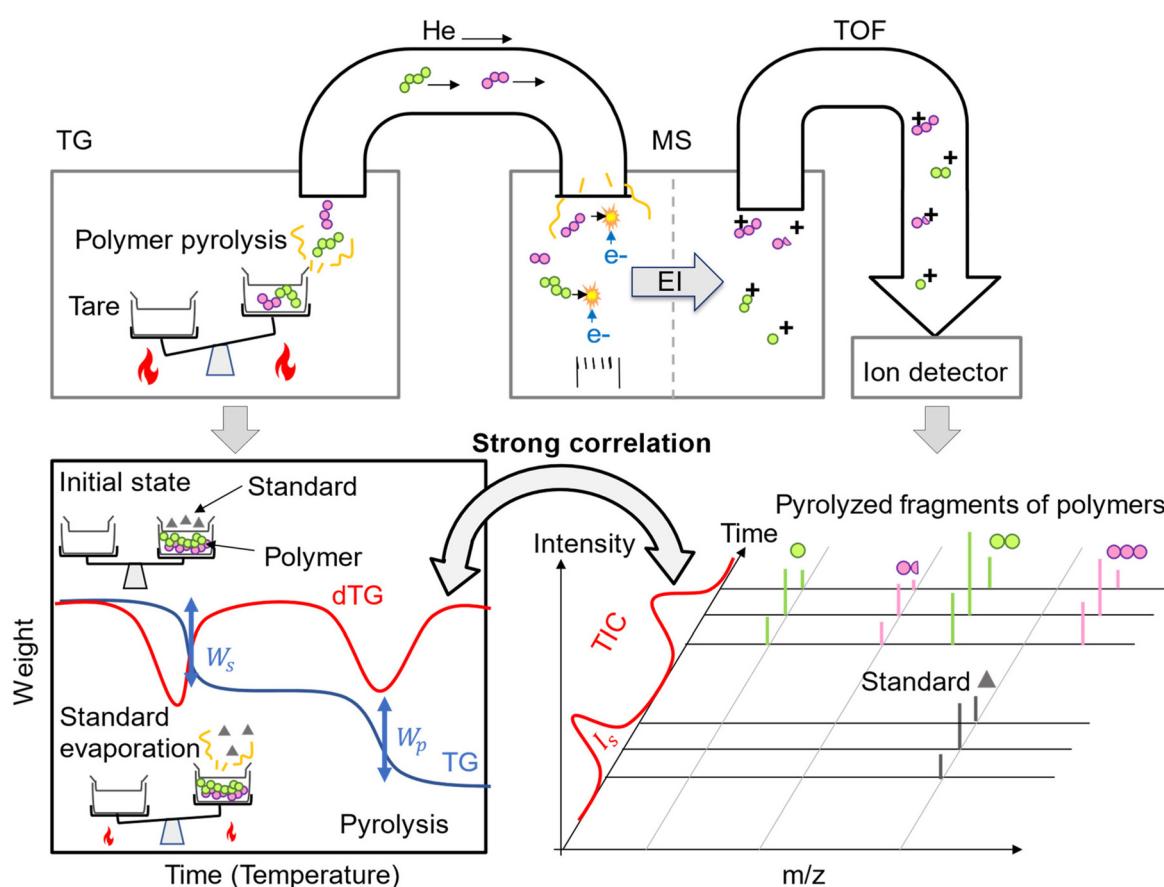


Fig. 1 TG-TOFMS setup for TG-RQMS. The weight and MS spectra are recorded every second. The synchronous analysis of instantaneous weight loss (dTG) and MS spectra allow the estimation of ionization efficiencies of pyrolyzed fragments for improving the RQMS accuracy.

hundreds of peaks that emerge simultaneously in the corresponding mass spectra. Consequently, determining ionization efficiency for each peak by solving the simultaneous equations between TG and MS peaks becomes impractical due to the excessive number of variables relative to the number of equations. Therefore, we initially consolidate the peaks into approximately ten basis spectra using the unsupervised learning algorithm of soft-orthogonality constrained non-negative matrix factorization with automatic relevance determination (ARD-SO-NMF),⁹ the linear combination of which best represents the original MS spectra. Subsequently, we determine the ionization efficiency for each of these basic spectra. Indeed, as described below, this simple strategy significantly improves the RQMS accuracy within estimated composition errors of ± 1.3 wt% in a benchmark test. TG-RQMS also allows the quantification of 1000 ppm level of additives. Furthermore, the inferred reference spectra retain the high resolution of over 10 000 at around 100 m/z provided by time-of-flight MS (TOFMS), allowing easy chemical identification as well. In the following section, we describe detailed experimental and mathematical procedures.

Experimental section

TG-TOFMS measurements were conducted at NIMS with a connected system of TG-DTA8122 (Rigaku) and JMS-T2000GC AccuTOF (JEOL) with an electron impact (EI) ion source (see the Methods section in the ESI† for detailed measurement conditions). TG-quadrupole MS measurements were conducted at Rigaku with a connected system of TG-DTA8122 and JMS-Q1500GC (JEOL) with an EI ion source. Unless otherwise noted, the ionization energy was set to 70 eV. For the benchmark CA test, binary or ternary samples of polyethylmethacrylate (E; $M_w = 515\,000$), polymethylmethacrylate (M; $M_w = 66\,700$) and polystyrene (S; $M_w = 18\,100$) were prepared as follows. First, the blended polymers at a certain mixing ratio were dissolved in 1,4-dioxane (total polymeric concentration, 3.3 wt%) and about 30 μL of the solution was dropcast on an aluminum pan (polymer sample weight, about 1 mg). After solvent evaporation under vacuum for 12 hours, an aluminum lid was put on the pan on top of which an internal standard of 4,4'-di-*tert*-butylbiphenyl (dtBbph, about 0.5 mg) was placed so that it would not contact the polymeric sample. Since dtBbph rapidly evaporates at a significantly lower temperature of around 150 °C, which is much lower than the polymer decomposition temperature of over 300 °C, there were two instances of weight loss. The first and second weight losses corresponded to dtBbph weight (W_s) and polymer weight (W_p), respectively (Fig. 1). As a result, precise measurements of exact weights were not necessary during sample preparation. The double-bottom pan was loaded into the combustion chamber of TG and pyrolyzed at elevated temperature from 50 to 600 °C at a heating rate of 25 °C min⁻¹. The pyrolyzed gases were passed through a transfer tube kept at 350 °C into the EI ion source of TOFMS (Fig. 1). The pyrolysis gases generated at

temperatures above 350 °C may condense on the transfer tube. However, the interior of the tube is under reduced pressure due to the MS vacuum chamber, creating conditions less conducive to condensation. Furthermore, since pyrolysis finely divides the polymer into small molecules, the vapor pressure of the pyrolysis gas generated at high temperatures is not necessarily low. Nonetheless, constant vigilance is required to prevent contamination from the transfer tube and chambers. We have confirmed that the total ion currents of the background mass spectra before and after dataset creation were more than four orders of magnitude lower than those measured during the experiments, confirming that contamination is negligible. The delay time for transporting and spectral intensity of MS were corrected using the internal standard as described later. The acquisition intervals of the MS spectra and weight losses were one second. The raw profile spectra were converted to centroid spectra using msAxel (JEOL) with ± 2 mDa tolerance. The RQMS analysis were conducted using the same algorithm presented in the previous reports (see the original developments^{9,10} and our implementation¹), except that the spectral fragment abundances (FA) were converted to the weight-base fragment abundances (wFA) using a dTG curve as described later.

Results and discussion

MS intensity normalization

Prior to analysis with (TG-)RQMS, the signal intensity of the MS spectra was normalized for each sample using the peak intensity of the internal standard (I_s ; see Fig. 1). Given that the signal intensity of MS in TG-MS can significantly vary from day to day based on the equipment's condition, this normalization is crucial for maintaining long-term dataset stability and for transferring the dataset to different equipment. To see the normalization effect, PS samples with various weights were measured. As shown in Fig. 2, the linearity of the MS intensities and sample amounts of PS were insufficient before the normalization; however, multiplication by the factor of W_s/I_s well established the linearity, suggesting the feasibility of quantitative analysis. The linearity was maintained even when different MS setups at different institutes were used (red: measured at NIMS with TG-TOFMS; blue: measured at Rigaku with TG-quadrupole MS). This strongly suggested long-term dataset stability and data-sharing feasibility in contrast to the reported RQMS based on environment-specific direct analysis in real-time MS (DART-MS) under open air.¹³ Since we are only interested in weight fractions, not the absolute abundances, hereafter the MS intensity was further divided by the sample weight W_p . The delay of MS spectra from TG caused by gas transportation (typically 10–20 seconds) was corrected using the standard peaks of dTG and TIC curves so that they would match on the time axis. After the intensity normalization and delay corrections, 840 MS spectra acquired at 250–600 °C (8–22 min), corresponding to the polymer decomposition stage, were extracted and integrated into N_T spectra (typically



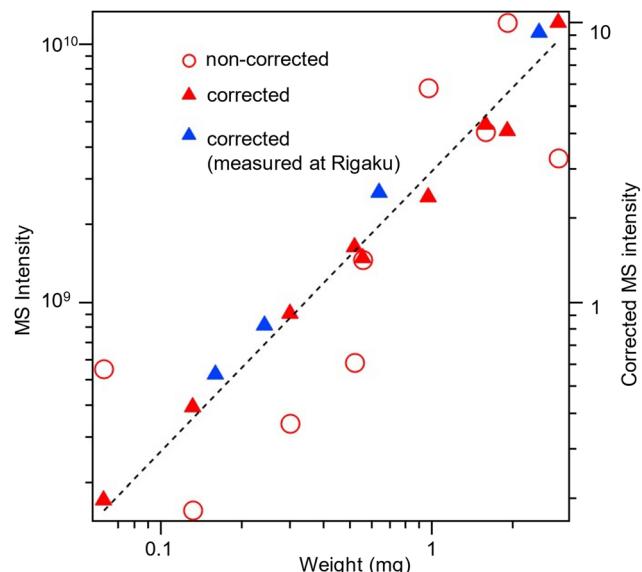


Fig. 2 MS intensity normalization using internal standards. The circles and triangles represent the raw and normalized intensities, respectively. The red datapoints represent measurements acquired at NIMS using TG-TOFMS, and the blue datapoints represent measurements acquired at Rigaku using TG-quadrupole MS. The dotted line represents the best fit line.

$N_T = 10$). The spectral dataset composed of $N_T N$ spectra (N : the number of mixed samples) was then subjected to (TG-) RQMS analysis to output the inferred reference spectra and their fraction in each sample.

Benchmark CA for ternary polymer systems

The purpose of this benchmark CA test is to verify the accuracy of (TG-) RQMS. We therefore prepared a ternary polymer system with known compositions so that RQMS accuracy can be evaluated by the difference between the inferred and known ground-truth compositions. During the composition inference, neither the known compositions of any mixture samples nor the reference spectra of the pure constituents were used. This limitation can create the same situation as CA of unknown polymer systems where only mixture spectra with unknown compositions are given. In this paper, we do not conduct CA on really unknown polymer systems, since the inferred composition cannot be readily verified. Such a practical application of RQMS, which is beyond the scope of this paper, can be found in sequence distribution analysis.¹ It should be noted that in such a practical situation, the number of system constituents is not necessarily limited to three and CAs for systems with five, nine, and 13 constituents were demonstrated in the previous report.¹ However, since it has been shown that increasing the number of constituents does not increase the difficulty of the problem,⁷ only ternary and quaternary systems were used for the verification of RQMS accuracy in this paper.

Leaving the mathematical discussion for later, we first describe the results of benchmark CA for a ternary system of polyethylmethacrylate (E), polymethylmethacrylate (M) and

polystyrene (S). The used dataset consisted of 32 samples of binary or ternary mixtures, none of which exceeded 80 wt% composition for any component (therefore, reference-free). The fractions estimated by the previously reported RQMS without utilizing TG curves significantly deviated from the ground-truth (Fig. 3A). This could be attributable to the shared substructure of methacrylic backbones of E and M, which caused significant peak-overlapping among the constituents. Indeed, the reference spectra of E and M (Fig. 3D) were very similar. Despite this inherent difficulty of polymer CA, concurrent TG analysis drastically improved the situation, even when the different constituents generated multiple identical fragments (Fig. 3B; analyzed by TG-RQMS; also see the numerical data in Table S1†). The root-mean-squared error (RMSE) for estimated fractions was reduced from 0.043 to 0.013 (from ± 4.3 wt% error down to ± 1.3 wt% error; see how to calculate the error in the “Revisiting RQMS” section). This improvement was attributable to the conversion of MS signal intensities of pyrolyzed fragments into weight abundances *via* TG data, which placed more emphasis on the fragments with lower ionization. It should be noted that various pyrolyzed fragments with different ionization efficiencies are simultaneously generated, distorting the proportional relationship between MS signal intensities (TIC) and instantaneous weight loss (dTG). Nevertheless, by introducing the concept of “inverse ionization efficiency” inherent to each fragment, TIC and dTG can be quantitatively correlated as discussed later. The concurrent analysis of TG also allowed the quantification of trace additives (Fig. 3C). These high purity samples (M fraction > 0.99) were not used for leaning the reference spectra but just projected onto the learned space spanned by the inferred references, suggesting an accurate inference. Also, the inferred reference spectra showed good consistency with the observed reference spectra (Fig. 3D). Since the inferred reference spectra preserved the high resolution (over 10 000 at around 100 m/z), the identification of polymeric species was easily executable based on the monomeric mass. To validate the transferability of the dataset from TOFMS to ubiquitous quadrupole MS, the reference spectra inferred from the TOFMS dataset were rounded into integer spectra, which were used for deconvoluting the integer spectrum acquired by quadrupole MS at Rigaku, outputting the accurate fraction (Fig. 3C). This ensured dataset transferability between different instruments and institutes.

To emphasize its distinct advantages over other analytical methods, we conducted a TG-RQMS analysis on a quaternary system of E/M/R/S, where a random copolymer (R) of methyl methacrylate and styrene with a 1:1 monomer composition was introduced into the aforementioned ternary system of E/M/S. Quantifying random copolymers within homopolymer blends is an essential technique for ensuring the quality of, for example, polymer alloys and recycled materials.¹⁴ However, it poses a significant challenge in NMR, where the peaks of each component tend to overlap almost completely. To address this, we created a new dataset consisting of 58 samples of mixtures in the quaternary system of E/M/R/S. It is noteworthy that the dataset size explosion has been mitigated

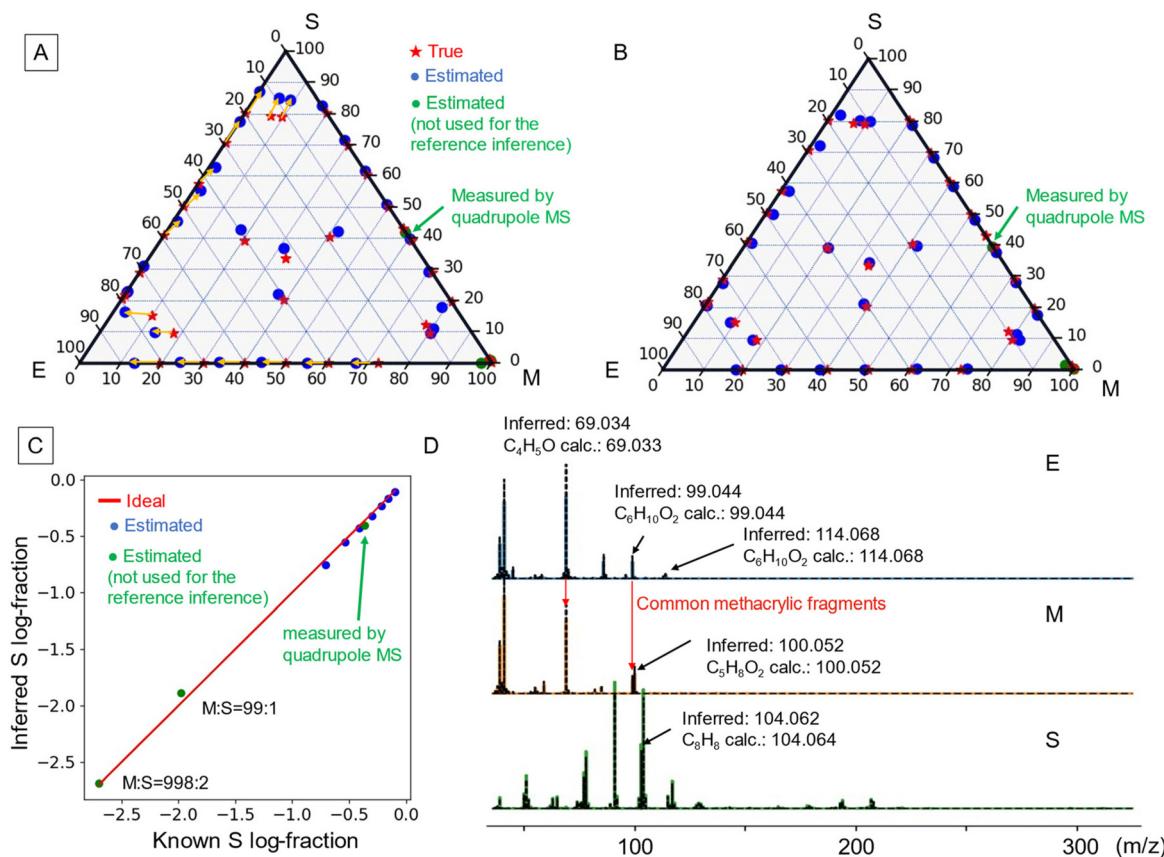


Fig. 3 Benchmark CA test for ternary polymer films. (A) RQMS-estimated composition and (B) TG-RQMS-estimated composition for the reference-free system (the maximum fraction was 80 wt%). Red stars and blue dots represent the ground-truth and inferred compositions, with parings indicated by orange arrows. Green dots represent high purity samples or samples measured by quadrupole MS, which were not used for inferring the reference spectra, but just projected on the learned simplex. (C) A log-scale plot along the M-S edge of (B). (D) Reference spectra inferred by TG-RQMS. The overlapped black dotted curves represent the observed spectra.

compared to the ternary system of E/M/S. The experimental and algorithm parameters were largely consistent with those of the ternary system (see Table S3†); however, to achieve relatively softer ionization, the ionization energy was adjusted to 20 eV. The CA result is depicted in Fig. 4, which shows the gap between the inferred and ground-truth fractions of the quaternary polymers. The RMSE was 0.029 (± 2.9 wt% inference error), which was surprisingly accurate considering the inclusion of random copolymers along with the homopolymers (also see the numerical reports in Table S2†). The unique ability of TG-RQMS to accurately estimate compositions without references, even in mixtures containing random copolymers and polymers with slightly different side chains, underscored its utility.

Revisiting RQMS

We first briefly review the previously reported RQMS algorithm¹ without synchronous analysis of TG, which output the inferred composition, as shown in Fig. 3A. In this section, a MS spectrum is represented by a D -dimensional row vector preserving the non-negative signal intensities in D -channels. A pyrolysis-MS spectrum with N_T temperature bands is a two-

dimensional spectrum represented by a (N_T, D) -matrix. However, for simplicity, for a while, we assume a one-dimensional spectrum by summing up the N_T -spectra along the temperature axis. A spectral dataset \mathbf{X} composed of N -samples is then represented by a non-negative (N, D) -matrix $\mathbf{X} \in \mathbb{R}_+^{N \times D}$ storing the n th sample spectrum at the n th row ($\mathbf{X}_{n:}$). The mission of RQMS is the simultaneous identification of K -reference spectra $\mathbf{P} \in \mathbb{R}_+^{K \times D}$ ($K \ll N$) and their fraction in each of the N -samples $\mathbf{C} \in \mathbb{R}_+^{N \times K}$. In this paper, we do not consider chemical reactions/interactions between the components—see Fig. S1† for how the final CA accuracy is adversely affected when the constituent polymers are reactive/interactive with each other. Under the linear mixing assumption, any observed spectrum $\mathbf{X}_{n:}$ would be represented by a linear combination of the K -reference spectra with individually different mixing fractions, *i.e.*,

$$\mathbf{X}_{n:} \approx \sum_{k=1}^K C_{nk} \mathbf{P}_{k:}, \quad \text{s.t. } \sum_{k=1}^K C_{nk} = 1, (n = 1, \dots, N) \quad (1)$$

where C_{nk} is the weight fraction of the k th component in the n th sample. Since $\mathbf{C}_{n:}$ represents the fraction of the n th sample, it should satisfy the sum-to-one condition. This can



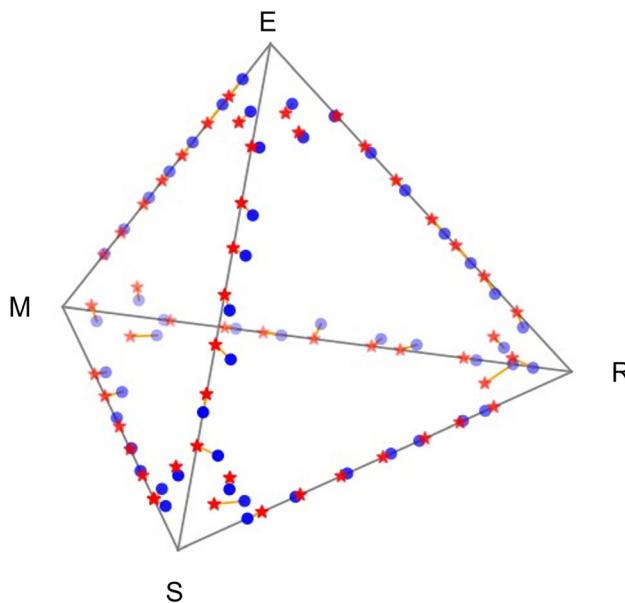


Fig. 4 Fractions of E/M/R/S quaternary polymers inferred by the TG-RQMS algorithm without using any reference spectra. Blue dots and red stars represent the inferred and ground-truth fractions, respectively, with orange lines indicating the gaps. The RMSE of the entire dataset was 0.029.

be simply written as $\mathbf{X} \approx \mathbf{CP}$ in a matrix form called non-negative matrix factorization (NMF),¹⁵ which is a mathematical representation of CA. We want to find the most suitable \mathbf{C} and \mathbf{P} best representing the observed spectra \mathbf{X} . The long history of NMF developments tackling difficult questions,¹⁶ *e.g.*, how to find a better solution⁹ and how to ensure its uniqueness,¹⁰ can be found in the references. Besides such mathematical discussions, we herein focus on only one practical issue specific to pyrolysis-MS; the direct factorization of \mathbf{X} into the reference spectra \mathbf{P} and their fraction \mathbf{C} would not be successful, since pyrolysis-MS does not observe polymers themselves, but measures their thermally and/or ionically decomposed fragments. We therefore proposed two-step NMFs formulated as $\mathbf{X} \approx \mathbf{AS} \approx (\mathbf{CB})\mathbf{S}$.¹ First, the observed spectra \mathbf{X} would be factorized into M -fragment spectra $\mathbf{S} \in \mathbb{R}_+^{M \times D}$ and their abundances in each of the N -samples $\mathbf{A} \in \mathbb{R}_+^{N \times M}$ (fragment abundances, FA). Subsequently, \mathbf{A} would be factorized into FA of the K -pure constituents $\mathbf{B} \in \mathbb{R}_+^{K \times M}$ and their fractions in each of the N -samples $\mathbf{C} \in \mathbb{R}_+^{N \times K}$. From these definitions, $\mathbf{P} = \mathbf{BS}$ apparently holds, deriving \mathbf{C} and \mathbf{P} *via* $\mathbf{X} \approx \mathbf{C}(\mathbf{BS}) = \mathbf{CP}$. To incorporate TG data into RQMS, in the following section we consider the temperature axis; therefore, the first NMF was conducted on $\tilde{\mathbf{X}} \in \mathbb{R}_+^{N_{TN} \times D}$ rather than $\mathbf{X} \in \mathbb{R}_+^{N \times D}$, outputting spectrum-wise FA $\tilde{\mathbf{A}} \in \mathbb{R}_+^{N_{TN} \times M}$. The (TG-) RQMS accuracy was evaluated by the root mean squared error (RMSE) of \mathbf{C} from the ground-truth fraction $\tilde{\mathbf{C}}$ only known in the benchmark CA test, *i.e.*, RMSE = $\sqrt{\frac{1}{N} \sum_k^K \sum_n^N (\mathbf{C} - \tilde{\mathbf{C}})_{kn}^2}$.

Incorporation of TG curves into RQMS

Here, we describe the key contribution of this paper, *i.e.*, TG-synchronized RQMS for more accurate compositional analysis (Fig. 3B). After the first NMF $\tilde{\mathbf{X}} \approx \tilde{\mathbf{a}}\mathbf{S}$, we have FAs for each of the N_{TN} -spectra ($\tilde{\mathbf{a}}$), which are mixing coefficients of the L2-normalized fragment spectra (\mathbf{S}). The abundances are, however, strongly biased by their individually different ionization efficiencies and, therefore, do not directly indicate their weight fractions in the analyte. To convert the spectral FA into weight-base FA (wFA), we exploited TG curves which should hold:

$$\tilde{\mathbf{a}}_i z = w_i (i = 1, \dots, N_{TN}) \quad (2)$$

where $z \in \mathbb{R}_+^M$ represents the inverse ionization coefficients of M -fragments and w_i is the sample weight change during the i th temperature band. The vector $\mathbf{w} \equiv (w_1, \dots, w_{NT})^T$ can be calculated as $-\Delta T$ from TG; therefore, eqn (2) mathematically connects TG and MS. The unknown z can be easily obtained by solving eqn (2) *via* non-negative least-squares algorithms,³ allowing the conversion of the spectral FA $\tilde{\mathbf{a}}$ into wFA: $\tilde{\mathbf{a}}^{(w)} = \tilde{\mathbf{a}}\mathbf{Z}$, where $\mathbf{Z} = \text{diag}(z)$. Since the temperature distribution of wFA is not helpful for CA, $\tilde{\mathbf{A}}^{(w)} \in \mathbb{R}_+^{N_{TN} \times M}$ was converted into $\mathbf{A}^{(w)} \in \mathbb{R}_+^{N \times M}$ *via* sample-wise integration along the temperature axis and subsequently input into the second NMF, *i.e.*, $\mathbf{A}^{(w)} \approx \mathbf{CB}$, using the reported algorithm (see the original development¹⁰ and our implementation¹). The modified part of the entire algorithm is illustrated in Fig. S2.† This simple update drastically improved the CA accuracy for the benchmark test as presented in Fig. 3.

Conclusion

This paper first demonstrated how to incorporate TG curves into quantitative MS analysis. The update toward TG-RQMS was trivial: just converting spectral fragment abundances into weight-base fragment abundances *via* synchronized TG data. Nevertheless, this TG reinforcement drastically improved the analytical performances: decreasing the compositional estimation errors from ± 4.3 wt% down to ± 1.3 wt% and allowing the contaminant quantification on order of 1000 ppm. This technique, capable of detecting extremely small impurities with high precision, is expected to be essential for applications such as monitoring the leaching of weathering agents present in practical polymer materials at very low levels and for quality control in recycled materials. TG-RQMS is performed under the linear mixing assumption, yet even for a non-linear mixing system with reactive components, “good enough” CA accuracy was demonstrated (Fig. S1†). To achieve more accurate CA for strongly interacted systems, a non-linear CA framework explicitly considering interactions/reactions between components should be developed, which could be achievable based on a bi-linear mixing model¹⁷—extended version of the linear mixing model. Overall, TG-RQMS allows accurate CA without using any prior knowledge about the system, which would solve long-standing questions in polymer science, *e.g.*, how to analyze sequence distribution.



Author contributions

Y. H. conceived the research, conducted the experiments, developed the software, analyzed the data, and wrote the manuscript. S. U. helped with the experiments. Y. H. and M. N. supervised the research.

Conflicts of interest

Y. H. and M. N. are owners of patent applications on TG-RQMS.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number JP24K08520 (to Y. H.) and the Core Research for Evolutional Science and Technology Program of the Japan Science and Technology Agency under Grant JPMJCR19J3 (to M. N.). The authors thank Yoshinobu Hosoi (Rigaku Corporation) for his help with TG-quadrupole MS measurements.

References

- Y. Hibi, S. Uesaka and M. Naito, A Data-Driven Sequencer That Unveils Latent “Codons” in Synthetic Copolymers, *Chem. Sci.*, 2023, **14**, 5619–5626.
- Y. Hibi, Y. Tsuyuki, S. Ishii, E. Ide and M. Naito, Decoding Thermal Properties in Polymer-Inorganic Heat Dissipators: A Data-Driven Approach Using Pyrolysis Mass Spectrometry, *Sci. Technol. Adv. Mater.*, 2024, **25**(1), DOI: [10.1080/14686996.2024.2362125](https://doi.org/10.1080/14686996.2024.2362125).
- D. C. Heinz and C.-I. Chang, Fully Constrained Least Squares Linear Spectral Mixture Analysis Method for Material Quantification in Hyperspectral Imagery, *IEEE Trans. Geosci. Remote Sens.*, 2001, **39**, 529–545.
- F. Hofer, W. Grogger, G. Kothleitner and P. Warbichler, Quantitative Analysis of EFTEM Elemental Distribution Images, *Ultramicroscopy*, 1997, **67**(1–4), 83–103.
- J. J. Friel and C. E. Lyman, Tutorial Review: X-Ray Mapping in Electron-Beam Instruments, *Microsc. Microanal.*, 2006, **12**(1), 2–25.
- P. Kirkpatrick and C. Ellis, Chemical space, *Nature*, 2004, **432**, 823.
- W. Zhu, J. W. Smith and C. M. Huang, Mass Spectrometry-Based Label-Free Quantitative Proteomics, *J. Biomed. Biotechnol.*, 2010, DOI: [10.1155/2010/840518](https://doi.org/10.1155/2010/840518).
- J. David, Z. Steinmetz, J. Kučerík and G. E. Schaumann, Quantitative Analysis of Poly(Ethylene Terephthalate) Microplastics in Soil via Thermogravimetry–Mass Spectrometry, *Anal. Chem.*, 2018, **90**(15), 8793–8799.
- M. Shiga, K. Tatsumi, S. Muto, K. Tsuda, Y. Yamamoto, T. Mori and T. Tanji, Sparse Modeling of EELS and EDX Spectral Imaging Data by Nonnegative Matrix Factorization, *Ultramicroscopy*, 2016, **170**, 43–59.
- X. Fu, K. Huang, B. Yang, W. K. Ma and N. D. Sidiropoulos, Robust Volume Minimization-Based Matrix Factorization for Remote Sensing and Document Clustering, *IEEE Trans. Signal Process.*, 2016, **64**(23), 6254–6268.
- F. Zitomer, Thermogravimetric–Mass Spectrometric Analysis, *Anal. Chem.*, 1968, **40**(7), 1091–1095.
- G. Hotová and V. Slovák, Determination of the Surface Oxidation Degree of the Carbonaceous Materials by Quantitative TG-MS Analysis, *Anal. Chem.*, 2017, **89**(3), 1710–1715.
- J. H. Gross, Direct Analysis in Real Time—a Critical Review on DART-MS, *Anal. Bioanal. Chem.*, 2014, **406**, 63–80.
- D. Rigby, J. L. Lin and R. J. Roe, Compatibilizing Effect of Random or Block Copolymer Added to Binary Mixture of Homopolymers, *Macromolecules*, 1985, **18**(11), 2269–2273.
- D. D. Lee and H. S. Seung, Learning the Parts of Objects by Non-Negative Matrix Factorization, *Nature*, 1999, **401**, 788–791.
- X. Fu, K. Huang, N. D. Sidiropoulos and W. K. Ma, Nonnegative Matrix Factorization for Signal and Data Analytics: Identifiability, Algorithms, and Applications, *IEEE Signal Process. Mag.*, 2019, **36**(2), 59–80.
- N. Dobigeon, J. Y. Tourneret, C. Richard, J. C. M. Bermudez, S. McLaughlin and A. O. Hero, Nonlinear Unmixing of Hyperspectral Images: Models and Algorithms, *IEEE Signal Process. Mag.*, 2014, **31**(1), 82–94.

